Institut für Organische Chemie und Chemische Biologie

The Dodecin of *Mycobacterium tuberculosis*: Biological Function and Biotechnical Applications

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Florian Alexander Bourdeaux

aus Frankfurt am Main

Frankfurt am Main (2020)

(D30)

Vom Fachbereich Biochemie, Chemie und Pharmazie der
Johann Wolfgang von Goethe-Universität als Dissertation angenommen

Dekan:	Prof. Dr. Clemens Glaubitz
Gutachter:	Prof. Dr. Martin Grininger
	Prof. Dr. Stefan Knapp

Datum der Disputation: 5.05.2021

“Indeed. The pursuit of knowledge is hopeless and eternal.”

“Hooray!”

Professor Hubert J. Farnsworth

Futurama Season 6 Episode 26; 20th Century Studios, Inc

Table of Contents

1 Zusammenfassung

In dieser Dissertation werden bakterielle Dodecine aus zwei Gesichtspunkten heraus diskutiert, erstens die biologische Rolle von bakteriellen Dodecinen als Homöostasefaktoren für Flavine und zweitens die Nutzung von bakteriellen Dodecinen als Trägermaterial von biologisch aktivem Material, wie z.B. Antigene.

Zur Biologischen Funktion von Dodecinen

Überleben und Wachstum von Zellen basiert auf einem fein ausbalancierten Zusammenspiel von Substraten, Kofaktoren und Enzymen, welche alle kontinuierlich neu aufgenommen, aufgebaut und abgebaut werden müssen. Hier haben reaktive Kofaktoren einen besonderen Stellenwert, da sie in ausreichender Menge für verschiedene Enzyme zur Verfügung stehen müssen, aber ein Überschuss ungewünschte Reaktionen in der Zelle katalysieren kann. Zu dieser Gruppe von Kofaktoren gehört die Stoffklasse der Flavine, welche nahezu in allen Bereichen der Zelle von Enzymen genutzt werden, um verschiedenste Reaktionen zu katalysieren und somit unverzichtbar für das zelluläre Überleben sind.

Flavine sind reaktive Verbindungen, die mit ihrem aromatischen Ringsystem ein oder zwei Elektronen übertragen können und somit an diversen Redox-Reaktionen als Kofaktoren in der Zelle beteiligt sind. Neben einfachen Redox-Reaktionen sind sie aber auch in Prozesse wie DNA-Reparatur, Apoptose und Proteinfaltung involviert. Die hohe Reaktivität und Vielfältigkeit von Flavinen basiert auf ihrem heteroaromatischen Ringsystem, genannt Isoalloxazin. Ihre hohe Reaktivität kann jedoch dafür sorgen, dass reaktive Sauerstoff- oder Stickstoffspezies im Zytosol entstehen, was bedingt, dass die Menge von freien Flavinen und nicht mehr gebrauchten Flavoenzymen niedrig gehalten werden muss. Jedoch muss anderseits die Menge an freien Flavinen groß genug sein, um den Bedarf der Flavoenzyme zu befriedigen. Die biologisch relevanten Flavine sind Riboflavin (RbF), Flavinmononukleotid (FMN) und Flavin-Adenin-Dinukleotid (FAD), wobei die beiden letzteren als Kofaktor fungieren und RbF als deren Synthesebaustein dient.

Menschen und andere Tiere können RbF nicht synthetisieren und müssen dieses über die Nahrung aufnehmen und spezielle RbF-Bindeproteine sorgen dafür das RbF gespeichert und transportiert werden kann. Im Gegensatz dazu können viele Bakterien und Archaeen RbF synthetisieren. Flavinarme Lebensräume sind für Bakterien und Archaeen daher nicht problematisch, aber dafür muss von diesen Organsimen die mehrstufige Synthese von RbF

kontrolliert werden, um die Versorgung zu sichern, aber gleichzeitig keinen Überschuss zu erzeugen. Ein Schlüsselelement zur Kontrolle der RbF-Synthese ist die Konzentration von freien FMN, welches durch Binden an einen Riboswitch, ein regulatorisches RNA-Element, die Translation von Proteinen der RbF-Synthese stoppt. Obwohl Bakterien und Archaeen nicht auf externes Flavin angewiesen sind, besitzen manche Spezies Methoden Flavine aus der Umgebung aufzunehmen, was zeigt, dass es für Organsimen günstig ist, wenn möglich, die RbF-Synthese zu umgehen, da es ihnen erlaubt ihre Ressourcen anderweitig zu nutzen. Alternative Flavinquellen können daher als förderlich angesehen werden, besonders wenn Ressourcen knapp sind. Eine Mögliche alternative Quelle wären Flavinspeicherproteine, welche bei überschüssigen Ressourcenaufkommen beladen werden, um unter Mangel oder hohen Bedarf diese wieder abzugeben.

Das Dodecin des Archaeon *Halobacterium salinarum* (*Hs*Dod) wird als ein solcher RbF-Speicher betrachtet. *Hs*Dod war das erste untersuchte Dodecin und damit der Namensvater der Dodecin-Proteinfamilie. Zwölf *Hs*Dod Monomere bilden einen kugelförmigen Komplex, fortan Dodecamer (Zwölfmer) genannt, welcher in der Lage ist zwölf Flavinmoleküle zu binden. Der Name Dodecin setzt sich aus „Dodec“ von Dodecamer und „in“ von Flavin zusammen. Besonders hierbei ist, dass Dodecine nur etwa 70 Aminosäuren lang sind, was sie zu den kleinesten bekannten Flavoproteinen macht und dadurch, dass praktisch ein Flavinmolekül pro Monomer (Zwölf pro Dodecamer) gebunden wird, zählen sie auch zu den effizientesten Flavinbindern. Die hohe Flavinbindeeffizienz wird durch die einzigartige Bindetasche von Dodecinen ermöglicht, in welcher zwei Flavinmoleküle zwischen zwei Tryptophanseitenketten einklemmt werden. Stabilisiert wird diese Tryptophan-Flavin Tetrade durch π-Stapeleffekte zwischen den Tryptophanseitenketten und den Isoalloxazinringsystemen der Flavine. Aufgrund dieses Effektes bindet *Hs*Dod RbF mit hoher Affinität, bzw. die Dissoziationskonstante liegt im zweistelligen nanomolaren Bereich. Der Bindemodus von Flavinen in Dodecin verhindert auch, dass Flavine ungewollte Reaktionen eingehen, da die reaktiven Positionen des Isoalloxazinringsystems vom Zytosol abgeschirmt sind – gebundene Flavine sind damit unschädlich gemacht. Zusätzlich verhindert die Bindung von Flavinen zu Dodecin die photoinduzierte Degradierung oder andere photoinduzierte Reaktionen der Flavine, da durch Dodecine der angeregte Zustand schnell wieder in den Grundzustand überführt wird. *Hs*Dod ist damit ein idealer RbF-Speicher, es besitzt hohe Affinität zu RbF und sorgt dafür, dass das gebundene RbF keine Reaktionen eingehen kann

Obwohl alle untersuchten bakteriellen Dodecine bis auf eines hohe Ähnlichkeit zu *Hs*Dod haben, sie formen Dodecamere und binden Flavindimere, zeigten genauere Untersuchungen, dass die bakteriellen Dodecine generell niedrigere Affinitäten zu Flavinen haben und FMN anstatt RbF als Ligand präferieren. Damit können bakterielle Dodecine nicht als RbF-Speicher fungieren und die insgesamt niedrigeren Affinitäten zu Flavinen stellen die Rolle als Flavinspeicher überhaupt in Frage. Dazu, anders als bei Archaeen, bei denen Dodecine fast ausschließlich in der Klasse der Halobacteria vorkommen, sind bakterielle Dodecine weit verbreitet und Dodecin-kodierende Gene liegen auch in Pathogenen, wie z.B. *Mycobacterium tuberculosis* und *Pseudomonas aeruginosa*, vor. Folglich stellt sich die Frage, was bakterielle Dodecine für eine biologische Rolle haben und ob das Verständnis dieser genutzt werden könnte, um Infektionen von z.B. *M. tuberculosis* zu bekämpfen.

In dieser Arbeit wurden die Dodecine von *M. tuberculosis* (*Mt*Dod), *Streptomyces coelicolor* (*Sc*Dod) und *Streptomyces davaonensis* (*Sd*Dod) untersucht, um der Aufklärung der biologischen Rolle von bakteriellen Dodecinen ein Stück näher zu kommen. Kinetische Messungen der Flavinbindung von *Mt*Dod ergaben, dass die Dodecinbindetasche in zwei verschiedenen Schritten gefüllt wird, für die dann ein kinetisches Modell erstellt und durch experimentelle Daten verifiziert wurde. Die Analyse mit dem zweistufigen Modell zeigte, dass die einzigartige Bindungstasche von Dodecinen es ihnen ermöglicht, überschüssige Mengen an Flavinen zu binden, während bei niedrigen Flavinkonzentrationen Flavin freigesetzt und nur schwach gebunden wird. Diese Funktion der Flavinpufferung verhindert die Anreicherung von frei oxidierten Flavinen und trägt daher dazu bei, das Redoxgleichgewicht der Zelle aufrechtzuerhalten, und verhindert mögliche Zellschäden, die durch zu hohe Konzentrationen von freien Flavine verursacht würden. Um weitere Einblicke in die Rolle von bakteriellen Dodecinen zu erhalten, wurde die Auswirkung des Ausschaltens des Dodecin-kodierenden Gens in *S. davaonensis* analysiert. Der Knockout-Stamm zeigte erhöhte Konzentrationen verschiedener stressbedingter Metaboliten, was darauf hinweist, dass ohne Dodecin das Zellgleichgewicht gestört ist, was die Rolle von Dodecinen als Flavin-Homöostasefaktor unterstützt.

Mit einer selbst entwickelten Affinitätsmessmethode, die auf der temperaturabhängigen Dissoziation des Dodecin:Flavin-Komplexes basiert und ein paralleles Screening mehrerer Bedingungen ermöglicht, konnte gezeigt werden, dass *Mt*Dod, *Sc*Dod und *Sd*Dod unter sauren Bedingungen eine viel höhere Affinität zu FMN und FAD aufweisen. Unter diesen Bedingungen können die drei Dodecine als FMN-Speicher fungieren. *M. tuberculosis* trifft

während seines Infektionszyklus beim Menschen auf mehrere saure Umgebungen und kann zum Schutz einen Ruhezustand annehmen, welcher auch als Dormanzphase bezeichnet wird. Während dem Erwachen aus dem Ruhezustand könnte ein Flavinspeicher vorteilhaft sein, da er es erlaubt, dass Ressourcen von der Flavinsynthese auf andere Prozesse umverteilt werden. Für einige *Streptomyces*-Arten wurde berichtet, dass die gebildeten Sporen leicht sauer sind und daher *Sc*Dod und *Sd*Dod als Flavinspeicher für die Sporen fungieren könnten. Weitere Details zum Flavinbindungsmechanismus von *Mt*Dod wurden durch eine Mutagenesestudie enthüllt, in der die Bedeutung des Histidinrests an der vierten Position der Proteinsequenz für die Flavinbindung identifiziert wurde. Entgegen den Erwartungen scheint dieser Rest jedoch nur teilweise an der pH-bezogenen Affinitätsveränderung beteiligt zu sein.

Die in dieser Arbeit aufgeführten Daten zeigen, dass bakterielle Dodecine wahrscheinlich als Flavin-Homöostasefaktoren fungieren, die insgesamt höhere Mengen an Flavin in der Zelle ermöglichen, ohne das zelluläre Gleichgewicht zu stören. Ferner legt der berichtete säureabhängige Anstieg der Bindungsaffinität nahe, dass bakterielle Dodecine unter bestimmten Bedingungen auch als Flavinspeichersystem fungieren können.

Anwendung des Dodecins von *M. tuberculosis*

Auf dem Gebiet der angewandten Biotechnologie sind Träger-/Gerüstproteine von zunehmendem Interesse, da mit diesen die Eigenschaften gebundener Enzyme oder anderer bioaktiver Verbindungen, wie z.B. von Antigenen, verbessert werden können. Mögliche Anwendungen sind die Schaffung künstlicher Multienzym-Anordnungen, die darauf abzielen, den Substrat-/Zwischenproduktfluss der katalysierten Reaktionskaskaden zu verändern, oder das Design neuer Impfstoffe durch Verwendung von Trägern zur Präsentation von Antigenen und zur Verbesserung ihrer Immunität. Abhängig von der Anwendung ist es wichtig, dass das verwendete Träger-/Gerüstprotein stabile und definierte Nanopartikel bildet, wie z.B. in der Rolle als Antigen-Träger. Das Dodecamer von Dodecinen stellt ein solches Teilchen dar und könnte daher ein vielversprechender Träger-/Gerüstproteinkandidat sein, da frühere Berichte eine hohe thermische Stabilität von *Mt*Dod und des Dodecins von *Thermus thermophilus* (*Tt*Dod) zeigten.

In dieser Arbeit wurde die Stabilität von *Mt*Dod, *Sc*Dod *Sd*Dod und *Hs*Dod analysiert, um ein geeignetes Dodecin für die Verwendung als Träger-/Gerüstprotein zu finden. Daher wurde eine Methode zur einfachen Messung der Stabilität von Dodecinen entwickelt, die die Rückbindung von Flavinen, welche das intakte Dodecamer vorrausetzt, nach einer Erhitzungsphase mit schrittweise ansteigenden Temperaturen misst. Die Verwendung dieses Assays und das Testen der Stabilität gegen Detergenzien durch SDS-PAGE zeigten, dass das Dodecamer von *Mt*Dod eine ausgezeichnete Stabilität gegen eine Vielzahl von Bedingungen wie Temperaturen über 95 °C, niedrigen pH-Wert und etwa 2% SDS besitzt. Durch Lösen der Kristallstruktur von *Sc*Dod und *Sd*Dod, wobei letzteres ein weniger stabiles Dodecamer bildet, in Kombination mit einer Mutagenesestudie, wurde die Bedeutung einer spezifischen Salzbrücke für die Dodecamer-Stabilität deutlich und könnte dafür hilfreich sein weitere hochstabile Dodecine zu finden.

Zusätzlich zu der intrinsisch hohen Stabilität des *Mt*Dod-Dodecamers wurde auch die Robustheit der Faltung getestet, indem verschiedene *Mt*Dod-Fusionskonstrukte geplant und in *Escherichia coli* hergestellt wurden. Hier wurde gezeigt, dass *Mt*Dod die Anlagerung von Proteinen bis zum 4-fachen seiner eigenen Größe toleriert und dass beide Termini modifiziert werden können, ohne das Dodecamer merklich zu beeinflussen. Ferner wurde gezeigt, dass *Mt*Dod und viele *Mt*Dod-Fusionskonstrukte in hohen Ausbeuten über ein einfaches Protokoll gereinigt werden können, welches auf der Entfernung von *E. coli*-Proteinen durch Hitzedenaturierung und anschließender Zentrifugation basiert. In einer Fallstudie wurde durch Anhängen verschiedener Antigene an *Mt*Dod und Verwendung dieser *Mt*Dod-Antigen-Konstrukte zur Herstellung von Antikörpern in Kaninchen gezeigt, dass *Mt*Dod immunogen ist und die gebundenen Antigene dem Immunsystem präsentiert.

Die hier beschriebenen Eigenschaften von *Mt*Dod und in geringerem Maße von anderen bakteriellen Dodecinen zeigen, dass bakterielle Dodecine eine wertvolle Ergänzung für die Gruppe von Gerüst- und Trägerproteinen darstellen und ein großes Potenzial als Antigenträger haben.

2 Background

The research area of dodecin proteins is relatively young with the first dodecin described in 2003 by Bieger *et al.*[1] The first member of the dodecin family was isolated from the archaeon *Halobacterium salinarum* (*H. salinarum*) and identified as a flavin binding protein. The late discovery of this protein family would suggest that dodecins are restricted to a small group of organisms, like the class of Halobacteria, but surprisingly dodecins are rather common in the whole bacterial domain. Hence, the reason for its late discovery can rather be attributed to its small size of only about 70 amino acids, which makes the recognition as an actual gene difficult even when the sequence of the genome is available. For example, in *Mycobacterium tuberculosis* (*M. tuberculosis*) the dodecin gene was noticed four years after the whole genome was annotated.[2,3] Furthermore, its purpose in bacteria is still elusive up to date.

The archaeal dodecins are in principle similar to the avian riboflavin (RbF) binding protein, in the sense that both bind RbF,[4,5] but dodecins differ significantly in protein structure and in the mode of riboflavin binding, making the dodecin protein family in this sense unique.[1,6] Similar to the avian riboflavin binding protein, which functions as a RbF storage in eggs for the evolving embryo,[4] halobacterial dodecins are thought to bind and store RbF during phases of slow growth, which is then released under conditions that allow fast growth again.[7] This hypothesis is based on the sudden blooms of *H. salinarum* observed when specific conditions in salt lakes are met, while at unfavourable conditions, which can last years, *H. salinarum* has slow growth.[8,9] While plausible, it is still not clear how halobacterial dodecins actually switch between storage and release, since the dodecin of *H. salinarum* (*Hs*Dod) is constantly present in the cell,[7] whereas the avian RbF binding protein is degraded after a certain time during the embryonal development in the egg.[10–12]

The dodecin protein family is not limited to archaea and dodecin encoding genes are ubiquitous spread among the different Phyla of bacteria. Despite this, so far, only the bacterial dodecins of *Halorhodospira halophila* (*H. halophila*; *Hh*Dod)[5,7], *Thermus thermophilus* (*T. thermophilus*; *Tt*Dod)[13] and *M. tuberculosis* (*Mt*Dod)[14] were characterized rudimentarily with a focus on structural aspects and basic ligand binding. While bacterial dodecins overall can be considered similar to *Hs*Dod, different amino acid residues are involved in flavin binding, leading to an altered arrangement of the bound flavins.[13,14] These variances compared to *Hs*Dod cause bacterial dodecins to have overall lower affinities towards flavins and also to preferably bind flavin mononucleotide (FMN), which is synthesized from RbF and functions

as a coenzyme. Because of these dissimilarities it is unlikely, that bacterial and archaeal dodecins fulfil the same biological role. Here, the biological role or biological function refers to the benefit for the organism gained by the production of the protein, while the function or molecular function of the protein itself refers to its action, for example the binding of flavins.

For bacterial dodecins, which are not restricted to a single phylogenetic class, a similar correlation between biological function and an event in the life cycle, like for *Hs*Dod, cannot be made, simply because there is no obvious event shared by the entire spectrum of these species. Hence, the broad occurrence of the dodecin gene in bacteria suggests a more general biological role, which would be applicable to all dodecin-encoding organisms. An alternative explanation would be that each bacterial dodecin is utilized for a specific purpose and thereby indicating that dodecins have highly diverse biological roles. Due to the relative high sequence similarity of all bacterial dodecins, especially for the amino acid involved in flavin binding,[13] the second hypothesis sounds less probable.

Both of these hypotheses represent rather extreme cases and in reality, it is more likely that there is a set of subgroups of dodecins each fulfilling a distinct biological role different to the other subgroups. Subgroup refers here to dodecins, which have different purposes and thereby to some degree different molecular functions, for example due to different affinities for the specific flavins. The archaeal dodecins can be seen as such a subgroup, since *Hs*Dod has higher affinity for smaller flavins, like RbF, compared to the so far studied bacterial dodecins. That archaeal dodecins are a subgroup is also supported by their evolutionary origin, as they were acquired by the ancestors of the Halobacteria class via horizontal gene transfer from bacteria,[15] which would explain their nearly exclusive presence in Halobacteria and distinct differences from the bacterial dodecins.

Since bacterial dodecins are not restricted to extremophiles, their biological role must go beyond the assistance to overcome environmental changes, like suggested for *Hs*Dod.[7] Further, the presence of bacterial dodecins in pathogens, like *Pseudomonas aeruginosa*, *M. tuberculosis* and *Ralstonia solanacearum* (plant pathogen), indicates the importance of dodecins for their hosts. Disclosing the biological function of bacterial dodecins is of general interest to better understand microbial life, and might specifically help to combat diseases caused by pathogens encoding dodecin.

2.1 Flavins and their Role in Cells

The biological role of bacterial dodecins is clearly related to flavins and therefore, the importance of bacterial dodecins can only be as important as flavins are.

2.1.1 Flavins

Flavins are heteroaromatic molecules that function as cofactors, or more precisely coenzymes, for a vast number of enzymes. The name flavin is based on the Latin word *flavus* meaning "yellow" and was chosen because watery solutions of flavins have a yellow colour. Their yellow colour originates from the 7,8-dimethylisoalloxazin ring system, which is the conserved moiety of all flavins. Different flavin species only vary in the modification of the N-10 position of the heterocyclic ring system and the chemical name of the respective attached moiety is part of the specific name of the flavin.[16] Biological relevant flavins are riboflavin (RbF, also called vitamin B2), FMN, also called riboflavin 5'-monophosphate, and flavin adenine dinucleotide (FAD) (**Figure 1**).[16]

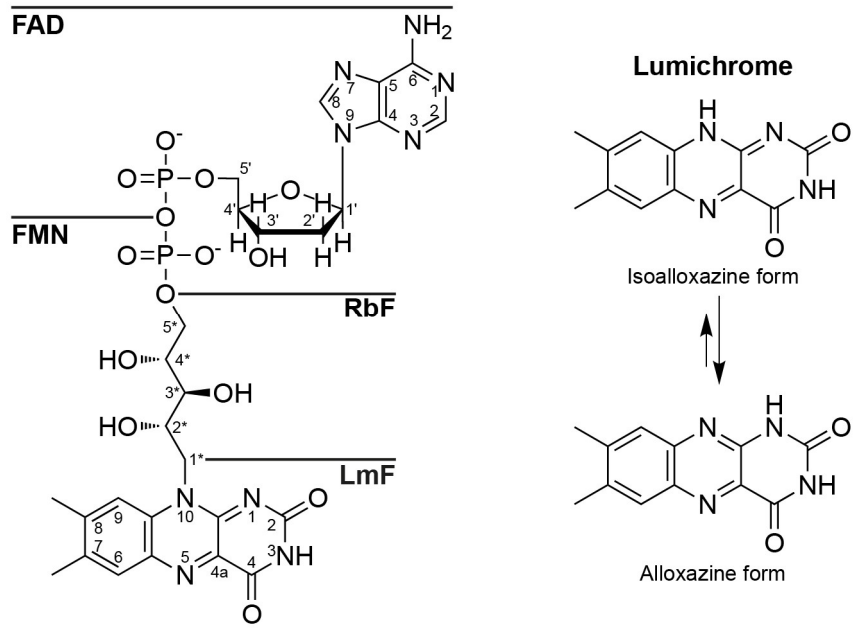


Figure 1: Structure of lumiflavin (LmF), RbF, FMN, FAD and lumichrome.

While all flavins are similar in structure, mainly FMN and FAD are utilized as cofactors and RbF typically acts as their biosynthetic precursor.[17–19] LmF and other flavins with a small moiety at the N-10 position are only relevant as degradation products of the other flavins, for example induced by exposure to light (photodegradation).[20] Lumichrome, which is basically

the 7,8-dimethylisoalloxazin ring system alone, does not belong to the class of flavins, since the hydrogen at position N-10 allows tautomerization to the alloxazine form, which has different properties compared to the isoalloxazine form of "real" flavins (see **Figure 1**).[16]

The isoalloxazine ring system can adopt two different redox and also protonation states, which enable catalysis of one or two electron transfer reactions.[18,21] During this process the isoalloxazine ring system can switch between different redox states, the oxidised form, the partly reduced form (semiquinone) and the fully reduced form (hydroquinone) (**Figure 2**).[18,21]

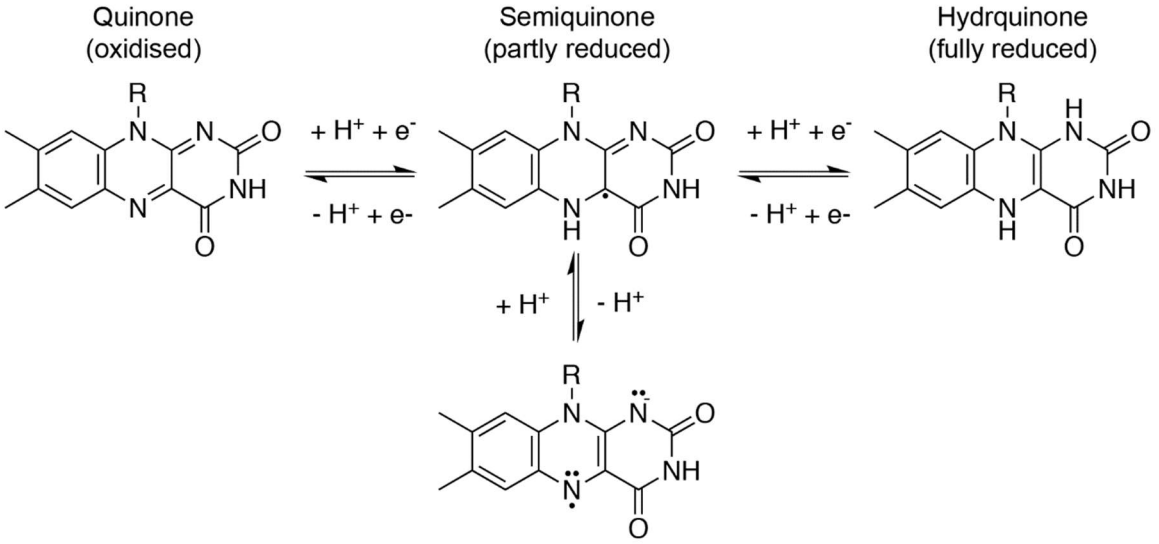


Figure 2: Different redox states of the isoalloxazine ring system of flavins.

Oxidised flavins have two distinct absorption maxima at about 375 nm and about 450 nm.[22,23] Upon reduction the absorption maxima are diminished and thereby flavin solutions lose their yellow colour (a pale yellow colour remains).[22,23] Flavins (oxidised) are also fluorescent and emit light at about 520 nm after excitation,[22,23] which is part of the cellular autofluorescence.[24]

In general, FMN and FAD are the most prominent cellular flavins, of which the main amount is bound to proteins, while only a small fraction remains free in the cytosol.[25–29] The actual cellular amount and the relative share between FAD, FMN and RbF seem to be dependent on the organism and the growing conditions, which hinders general assumptions about the cellular flavin levels.[26,27,29]

Other compounds closely related to flavins are roseoflavin, which is 8-dimethylamino-8-demethyl-riboflavin, and coenzyme F420, which is a derivate of 7,8-didemethyl-8-hydroxy-5-deazariboflavin (**Figure 3**).[30,31]

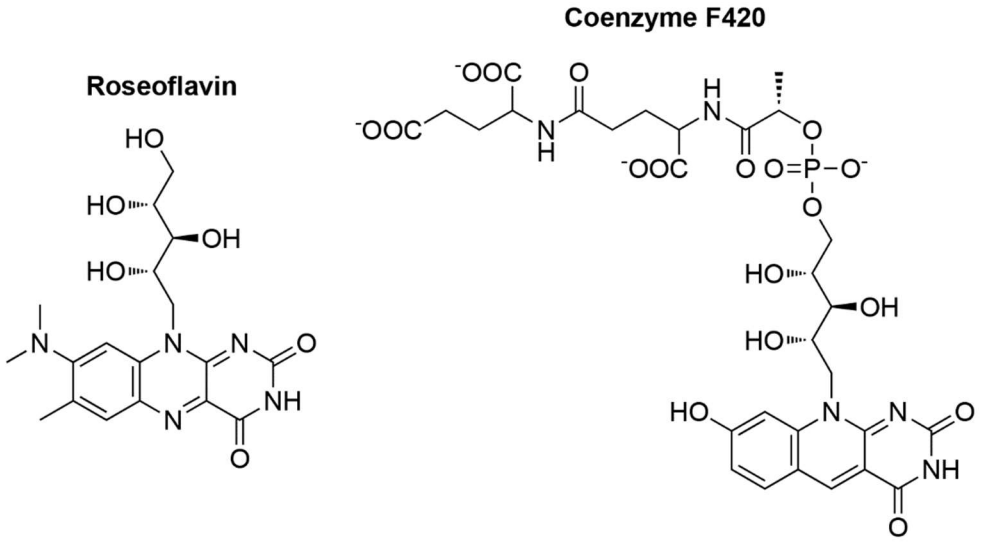


Figure 3: Structure of roseoflavin and coenzyme F420.

Because of their altered aromatic systems, these compounds have different redox potentials compared to flavins,[32,33] and coenzyme F420, for example, is utilized in the methanogenesis due to its lower redox potential.[33] Roseoflavin does not function as a cofactor and has antimicrobial properties, as it interferes with the flavin biosynthesis and the function of flavoenzymes by replacing the cofactors FMN and FAD.[32,34]

2.1.2 Biosynthesis

Plants and many bacteria can produce RbF based on guanosine triphosphate (GTP) and ribulose-5-phosphate, while most animals (like humans) require to obtain enough RbF through their diet.[35,36] FMN and FAD can be synthesized based on RbF and ATP through the riboflavin kinase (produces FMN) and the FAD synthase.[35,37] The biosynthesis pathway of RbF, FMN and FAD was elucidated for *E. coli* and *B. subtilis*.[35,37] Via homology searches, the flavin biosynthesis can also be modelled in other organisms, and the respective flavin biosynthesis pathway, can be accessed via the online database BioCyc.[38]

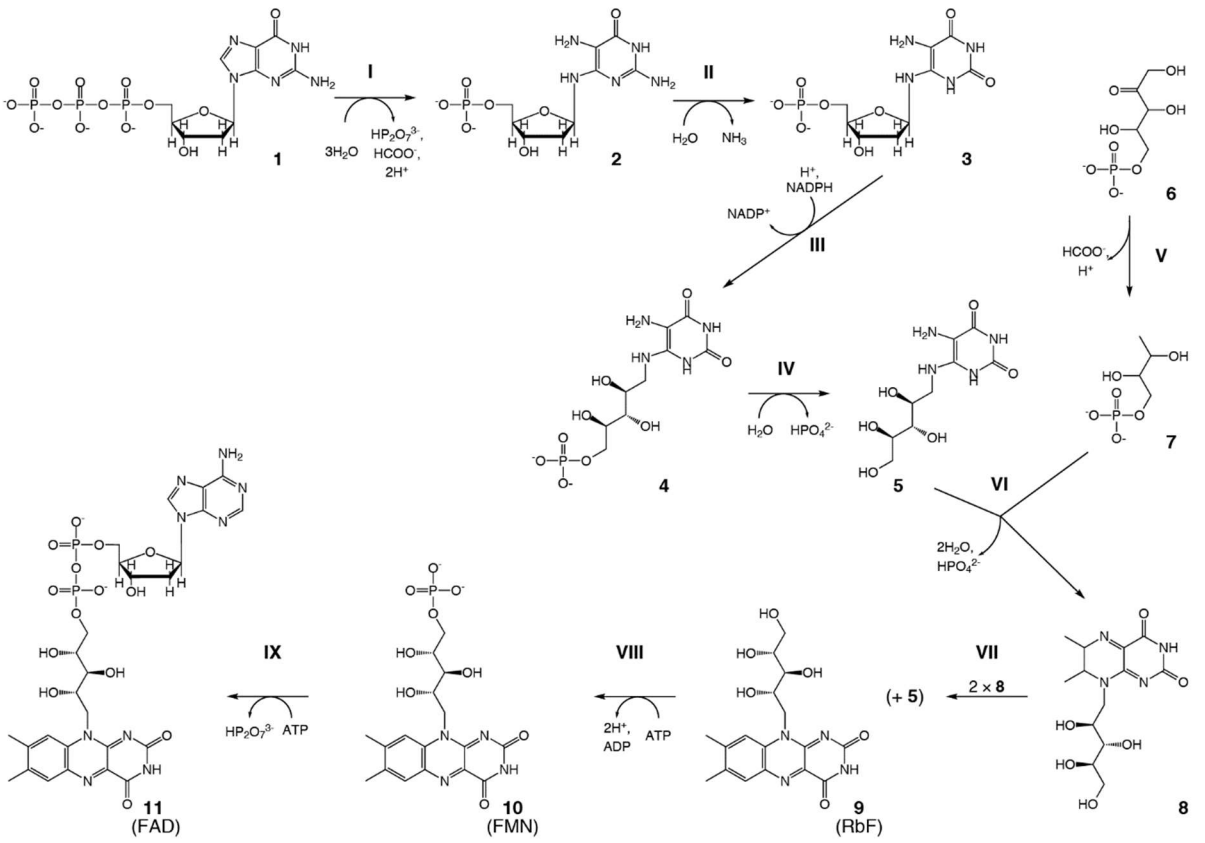


Figure 4: Predicted flavin synthase pathway for *M. tuberculosis* based on gene homology. Molecules in the synthesis pathway: **1:** GTP ; **2:** 2,5-diamino-6-(5-phospho-D-ribosylamino)pyrimidin-4(3H)-one; **3:** 5-amino-6-(5'-phosphoribosylamino)uracil; **4:** 5-amino-6-(5-phospho-D-ribosylamino)uracil; **5:** 5-amino-6-(D-ribitylamino)uracil; **6:** D-ribulose-5-phosphate; **7:** 1-deoxy-L-glycero-tetrulose 4-phosphate; **8:** 6,7-dimethyl-8-(1-D-ribityl)lumazine; **9:** RbF; **10:** FMN; **11:** FAD. Enzymes involved in the reactions: **I:** Potential GTP-cyclohydrolase II (Gene-ID: *RV1415*); **II:** Potential bifunctional diaminohydroxyphosphoribosylaminopyrimidin deaminase (Gene-ID: *RV1409*), potential bifunctional diaminohydroxyphosphoribosylaminopyrimidin deaminase (Gene-ID: *RV2671*); **III:** Potential bifunctional 5-amino-6-(5-phosphoribosylamino)uracil reductase (Gene-ID: *RV1409*), potential bifunctional 5-amino-6-(5-phosphoribosylamino)uracil reductase (Gene-ID: *RV2671*) **IV:** Unknown protein; **V:** Unknown 3,4-Dihydroxy-2-butanone-4-phosphat synthase ; **VI:** 6,7-dimethyl-8-ribityllumazin synthase; **VII:** RbF synthase: α chain (Gene-ID: *RV1412*) and β chain (Gene-ID: *RV1416*); two molecules **8** react to molecule **5** and molecule **9**. **VIII:** Potential bifunctional RbF kinase (Gene-ID: *RV2786C*) **IX:** Potential bifunctional FAD synthetase (Gene-ID: *RV2786C*).

To ensure survival of the cells, the cellular flavin demand needs to be satisfied, while overproduction should be minimized to avoid wasting metabolic energy and accumulation of flavins, which can cause the formation of reactive oxygen species.[39,40] Therefore complex strategies evolved to control the cellular flavin levels.[40] Flavin biosynthesis is dominantly regulated by a FMN riboswitch (also called RFN-element), which inhibits the translation of the RbF biosynthesis related enzymes upon FMN binding, but also other proteins might be involved.[40–42] Since the flavin biosynthesis is controlled by the cellular FMN level, the inhibition of FMN and FAD production causes an overproduction of RbF.[41] In addition to its biosynthesis, some bacteria contain RbF transporter to allow the intake of extracellular RbF.[40,42]

2.1.3 Flavoenzymes and flavoproteins

Flavoenzymes have a plethora of biological functions ranging from simple oxidation and reduction reactions to functions like DNA repair, light emission, oxygen activation, apoptosis and detoxification.[17–19,43] A reason for this versatility is that the redox potential of flavins can differ depending on the surrounding amino acids in the catalytic site. While free flavins have a redox potential of about −200 mV, the potential of bound flavins can vary from −400 mV to 60 mV.[43,44] As the isoalloxazine ring system, especially the N5-position, is the active part of the flavin cofactor, it needs to be accessible for substrates in flavoenzymes.[43] Depending on how the flavin is bound either the *Re*-side or the *Si*-side (faces) of the isoalloxazine ring participates in the reaction, which can be used to classify flavoenzymes, but it is not an indicator for the specific reaction catalysed by the respective flavoenzyme.[43]

Since the isoalloxazine ring system needs to be accessible and the redox potential optimised for catalysis, interactions between the N10-moiety and flavoenzymes are important for the binding of FMN and FAD, which allow more binding interactions than RbF.[18,19] On the example of *Anabaena* flavodoxin, it was shown that the ribityl chain itself of FMN has the lowest enthalpic contribution to cofactor-binding (about −1 kcal/mol), while the phosphate group (about −7 kcal/mol) as well as the isoalloxazine ring system (about −6 kcal/mol) are the major contributors.[45] The strong impact of the group attached to the ribityl chain was also shown for flavodoxin of *Azotobacter vinelandii* by analysing binding enthalpies, here FMN (about −28 kcal/mol) is the preferred ligand compared to FAD (about −17 kcal/mol) and 8-carboxy-8-demethylriboflavin (about −14 kcal/mol).[46] Non-catalytic flavin binding proteins like the RbF-binding protein of chicken and reptiles found in their eggs, bind flavins by sandwiching them between two tryptophans utilizing π-system interactions.[6,47,48] The chicken RbF-binding protein is able to bind lumichrome (K_D = ~93 nM), lumiflavin (K_D = ~47 nM) and RbF (K_D = ~1 nM), indicating that here the major contributor to binding are the interactions with the isoalloxazine ring system and the attached moiety at the N-10 position is used to distinguish between the different flavins.[47] The lower affinities measured for FMN (K_D = ~1 µM) and FAD (K_D > 14 µM) are likely caused by repulsion of the negatively charged phosphate groups of those flavins.[47] A similar way of discriminating between FMN and FAD was found for *Hs*Dod, where the flavins are bound via π-system interactions and FMN as a ligand is unfavoured by negative repulsion of the phosphate group (see Chapter 2.2.1).[1,5]

These factors indicate, why RbF is not utilized as cofactor for flavoenzymes. Without the phosphate group of FMN and the adenine dinucleotide moiety of FAD a high enough binding

strength might not be possible without tight binding of the isoalloxazine ring system, which would block catalytic reactions. Also since typical cofactors stay bound during the catalytic cycle, binding modes relying highly on the aromatic ring system of flavins, could cause the loss of the reduced flavins, which is not the case for most flavoenzymes.[49] Because of the tight binding of flavins to flavoenzymes they are also sometimes referred to as prosthetic groups.[17,49,50]

2.2 Flavin Binding and Structure of Dodecins

Since dodecins have no catalytic activity, or at least so far none is known, their biological role relies on the formation of the dodecin-flavin complex itself. Therefore, different to enzymes where the generated product is the main benefit for the organism, dodecins molecular function itself causes the benefit.

2.2.1 The Dodecin Protein Family

In general, dodecins are small, about 70 amino acids long, dodecameric complex forming flavoproteins that have no known catalytic function.[5,7,13] The first studied protein of the dodecin protein family, designating the name for the protein family, was the dodecin of *Halobacterium salinarum* (*Hs*Dod).[1] The name dodecin ("dodec" + "in") was given since *Hs*Dod forms dodecameric complexes that bind flavins (preferably RbF).[1] *Hs*Dod is only 68 amino acids long, well representing the average length of dodecins (**Figure 5**), making them the smallest known flavoproteins (binds one flavin per monomer, 12 per dodecamer).[1] Unique to dodecins is the binding mode of the flavins, where two flavins are bound between two tryptophans forming an aromatic tetrad in the binding pocket.[1,13,51] The crystal structure of *Hs*Dod revealed that the dodecin monomers have a $\beta_1\alpha_1\beta_2\beta_3$ topology resulting in single α-helix partly enwrapped by a three stranded antiparallel β-sheets.[1] So far, all studied dodecins share this structure (see **Figure 5**).[1,13,14,52,53,7] Dodecin encoding genes can be found in several bacteria (3269; UniProt hits for dodecin (PF07311) in NCBI RefSeq; April 2020) and in some archaea (194; UniProt hits for dodecin (PF07311) in NCBI RefSeq; April 2020).[54] While the dodecin gene is found in many bacteria classes, in archaea it is nearly exclusive to the class of Halobacteria (of the 194 entries: 189 Halobacteria, 4 Methanomicrobia and 1 Nitrososphaeria). A dodecin encoding gene seems to be absent in eukaryotes (the occasional hits in UniProt[54]

are likely database errors and are ignored here). In the following, the focus is on *Mt*Dod and bacterial dodecins in general.

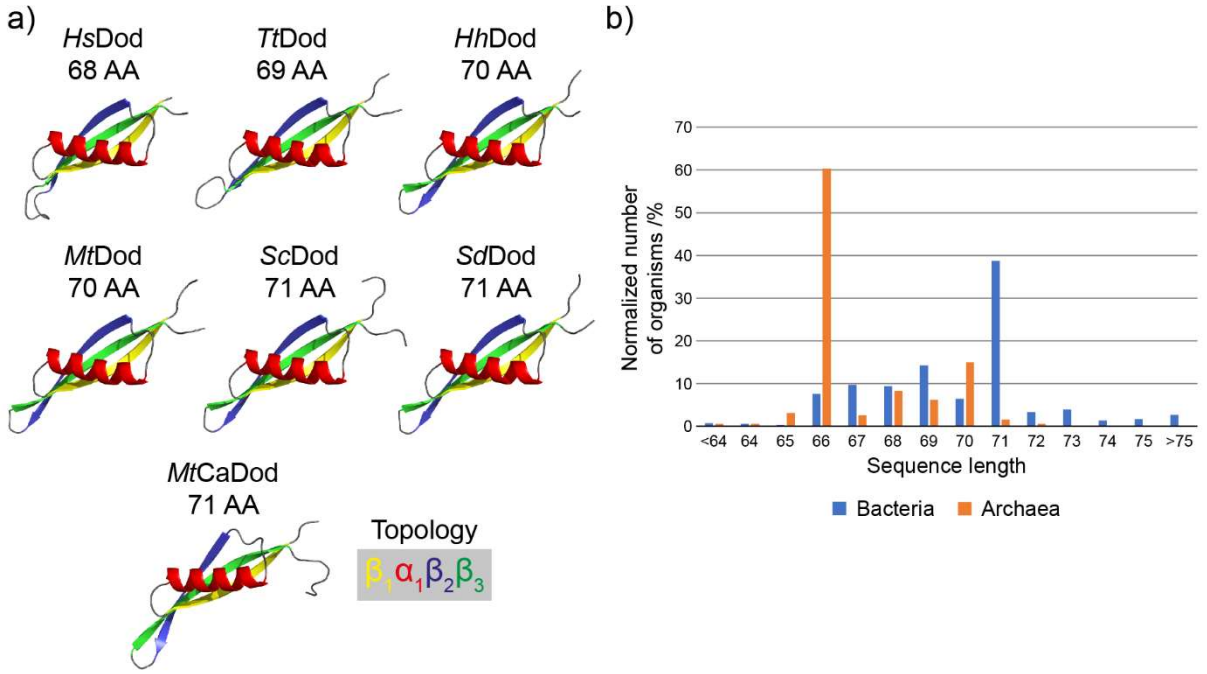


Figure 5: Overview of known dodecin structures and length distribution found in UniProt. a): Monomers of all solved dodecins structures. *Hs*Dod: The dodecin of *H. salinarum* dodecin (PDB ID: 2CCB)[5]. *Tt*Dod: The dodecin of *T. thermophilus* (PDB ID: 2V21)[13]. *Hh*Dod: The dodecin of *H. halophila* (PDB ID: 2VXA)[7]. *Mt*Dod: The dodecin of *M. tuberculosis* (PDB ID: 2YIZ)[51]. *Sc*Dod: The dodecin of *Streptomyces coelicolor* (PDB ID: 6R1E)[53]. *Sd*Dod: The dodecin of *Streptomyces davaonensis* (PDB ID: 6RI3)[53]. *Mt*CaDod: The calcium-dodecin of *M. tuberculosis* (PDB ID: 3ONR)[52]. b): Relative distribution of the dodecin sequence lengths of archaea (194 sequences) and bacteria (3269 sequences).

While dodecins in general are flavin binding proteins, concluded from sequence analysis, there seems to be a small number of dodecins that are exceptions. *M. tuberculosis* has another gene that encodes a protein that is counted to the dodecin protein family, but binds Ca^{2+}-ions instead of flavins; the so called calcium-dodecin (*Mt*CaDod).[52] *Mt*CaDod has a similar structure (see **Figure 5**) and forms a dodecamer like other dodecins,[52] although its sequence similarity to *Mt*Dod is only about 20% (30% to *Tt*Dod). Because only one of the non-flavin binding dodecins is studied up to date, not much is known about this subgroup of dodecins. In addition to Ca^{2+}-ions and flavins, it is also known that some dodecins can bind CoA (shown for *Mt*Dod, *Tt*Dod and *Sc*Dod), but no quantitative data is available.[13,51,53]

2.2.2 The Dodecin Dodecamer

All dodecins that were structurally elucidated so far, form a hollow spherical dodecamer making the overall protein structure likely the only feature that all members of the dodecin protein family share, including the calcium dodecins, but further data is needed to fully validate this.[1,13,14,52,53,7] Twelve dodecin monomers assemble to the highly symmetrical dodecamer with 23-cubic point group symmetry.[1,13,14,52,53,7] This dodecameric assembly is extraordinary stable and it was shown for the dodecamers of *Mt*Dod and *Tt*Dod that high temperatures up to 95 °C and above are tolerated (*Mt*Dod: 123 °C[14]).[13,14] In the dodecamer, there are six flavin binding pockets (each can bind two flavins), four trimer interfaces called channel 1 and four trimer interfaces called channel 2 (**Figure 6**), where channel 2 is also the CoA binding site for CoA binding dodecins.[13,51]

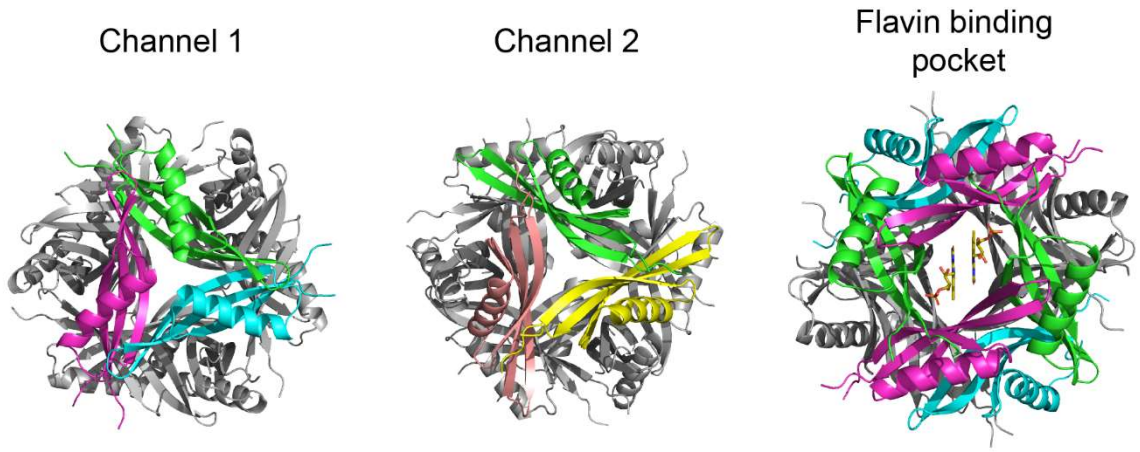


Figure 6: Assembly motifs of the dodecin dodecamer shown for *Mt*Dod dodecin (PDB ID: 2YIZ)[51]. Monomers involved in the assembly motifs are coloured, while others are grey.

The centres of channel 1 and channel 2 seem to be potential ion binding sites in which the bound ions further stabilize the trimeric interface by additional ionic interactions (structures with ions bound, PDB IDs: 1MOG[1], 2YIZ[51], 2DEH, 2CCB[5], 3ONR[52] and 3OQT[14]).[1] Although not in all dodecin crystal structures reveal ions bound in the channels, the potential ion-binding amino acid residues seem to be conserved (for alignment see ref. [13] or ref. [55], for *Mt*Dod the respective potential positions are aspartic acid 20 in channel 1 and arginine 7 and lysine 62 in channel 2). In *Mt*CaDod, channel 1 is the binding site for calcium and here glutamic acid 18 (respective position to *Mt*Dod aspartic acid 20) binds the calcium ion in the middle of the channel.[52] In addition to potential binding of negatively charged ions, lysine 62 can build a salt bridge with glutamic acid 10 (also highly conserved, for alignment see ref. [13] or ref. [55]) of another monomer in channel 2.[14] The importance of salt bridges for the

dodecamer stability and assembly was shown by mutagenesis of *Mt*Dod and disrupting the salt bridges Glu10-Lys62 and Glu41-Arg46 caused only the production of insoluble monomer.[14] In addition to salt bridges and ionic bonds at the centre of the channels, each monomer is in direct contact to five other monomers forming extended antiparallel β-sheet interactions. This causes the 12 monomers to form one antiparallel β-sheet that spans over the whole dodecamer (**Figure 7**). Altogether, the extended antiparallel β-sheet, the ionic bonds and the salt bridges in the channels could well explain why the dodecin dodecamer has such a high thermostability.

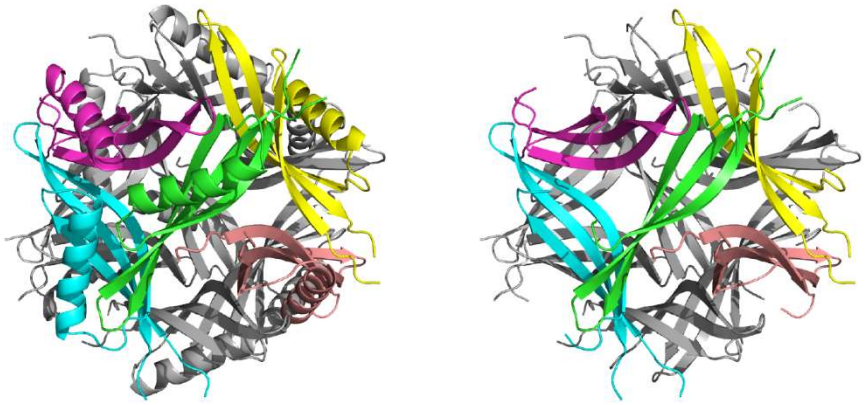


Figure 7: Direct contacting monomers in the dodecamer complex of *Mt*Dod (PDB ID: 2YIZ)[51]. The monomer in the centre (green) is part of channel 1 (green, magenta and cyan) and of channel 2 (green, yellow and salmon). Monomers not in direct contact to the green one are coloured in grey. The dodecamer on the left side is shown with helices and the one on the right side without to highlight the extended antiparallel β-sheet.

The flavin binding pocket is formed by the contact of four monomers and is only present in the fully assembled dodecamer, since four monomers alone build not a stable complex. The reason for this can be illustrated by using the trimer as the stable subunit of dodecamers and gradually assembling the dodecamer with it (**Figure 8**). While a binding pocket is formed by the interaction of two trimers (hexamer), the hexamer is not a preferred oligomeric species, because it can oligomerise to the dodecamer by the same contacts stabilizing the hexamer. The high number of stabilising interactions generated by dodecamer formation (inducing cubic 23 symmetry) leads to the preference of the dodecameric state (see **Figure 8**).

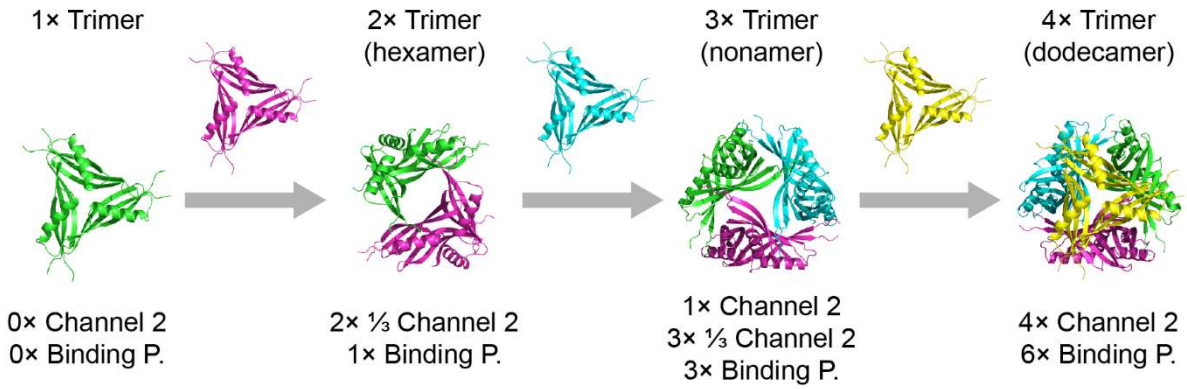


Figure 8: Assembly motifs formed by different oligomeric states based on channel 1 timers. Channel 2 is divided into full trimeric assemblies and partial assemblies (1/3 channel 2). Partial formed channels (2 monomers) are likely to have very low effect on stabilization since only 1/3 of the protein-protein interactions are formed and the ion binding should be limited (only two of the three ion binding amino acids are present). Binding P.: Flavin binding pocket.

2.2.3 Flavin Binding Site

In all dodecins (calcium-dodecins will be ignored here), the binding pocket is built up by four monomers and can bind up to two flavins. The four involved monomers are part of two different trimers, which will be called trimer 1 (T(1)) and trimer 2 (T(2)) in the following (**Figure 9**). For the unique flavin binding mode of dodecins, two amino acid residues are crucial, the first is tryptophan interacting via π-stacking with the bound flavin and the second is glutamine building two hydrogen bonds with the bound flavin (**Figure 9**).[1,13,14,53,7] The monomer that interacts via the tryptophan with the bound flavin is called monomer A (M(A)) here and the other monomer of the same trimer is called monomer B (M(B)) (**Figure 9**).

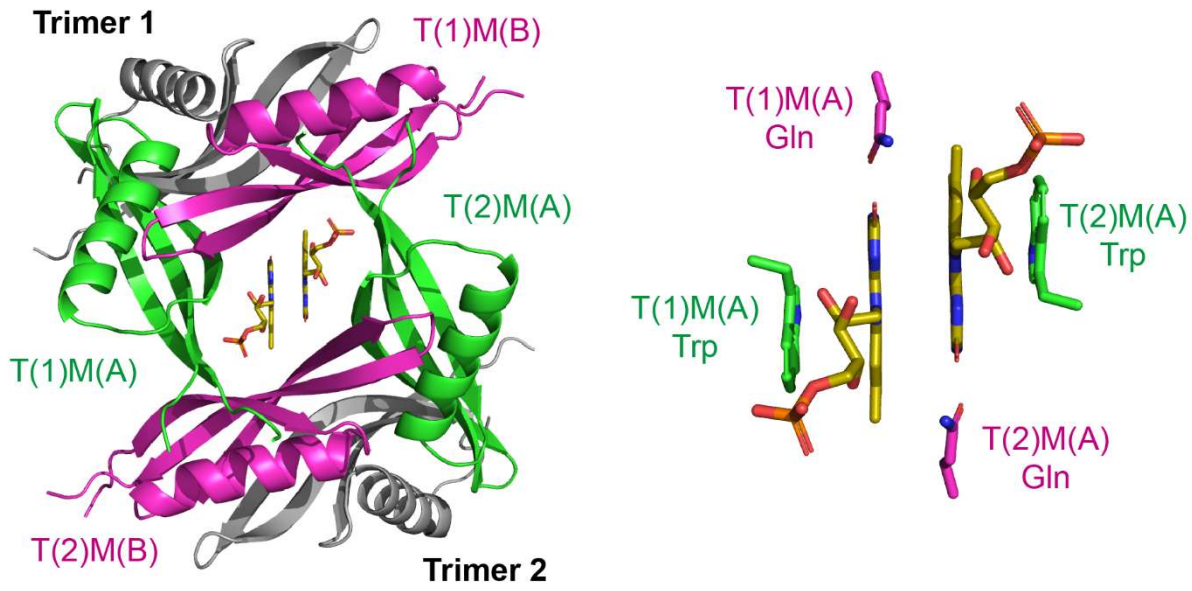


Figure 9: Flavin binding pocket (view along the C2 axis of the binding pocket) and core residues for the flavin binding. a) The four monomers of the two trimers generating the binding pocket are coloured (Trimer 1: red, yellow. Trimer 2: blue and green) and the remaining one's grey. b) Core residues for the flavin binding. FMN coloured green and amino acid residues coloured based on the colour of their monomer. Based on the *Mt*Dod structure (PDB ID: 2YIZ)[51].

While in all studied dodecins these two core interactions with the isoalloxazine ring system of the flavin are similar, the relative orientation of the bound flavins can differ. In the bacterial dodecins, these small relative orientational differences affect the aromatic tetrad (Trp-flavin-flavin-Trp) by altering the π-system overlap and the distances within it.[13,14,7] So far, all studied bacterial dodecins bind flavins in the *Si-Si* orientation (isoalloxazine rings contact each other via their *Si*-faces, see **Figure 9**), while in *Hs*Dod the bound flavin is flipped so that the binding glutamine becomes part of T(2) and not T(1).[1] The flipped binding orientation in *Hs*Dod is called *Re-Re*.[1,13] Since the aromatic plane is also a mirror plane, it can be assumed that the electronic distribution is also mirror-symmetric and the difference between the *Re-Re* and the *Si-Si* orientation should not have any impact on binding affinities, but so far no experimental data verifies this. In general, it seems that bacterial dodecins have smaller π-system overlaps and thereby a less energetic beneficial aromatic tetrad than *Hs*Dod,[13,51] which correlates with lower affinity for lumiflavin of *Tt*Dod (K_d=141 ± 11 nM)[13] compared to *Hs*Dod (K_d=18 ± 4 nM)[5] This lower affinity of bacterial dodecins for the isoalloxazine ring system could be an important factor for dodecins to distinguish between RbF, FMN and FAD, since it allows to counter select against the smaller flavins by making the flavin binding more reliant on the interactions with the N-10 attached moiety, like the phosphate group of FMN.

In contrast to the two core residues of the flavin binding pocket (see **Figure 9** b), other flavin-binding residues are less conserved and indicate that different dodecins are optimized for the flavin metabolism/content of their specific organism.[7,13] For example, in bacterial dodecins, a highly conserved arginine residue (*Mt*Dod arginine 46, **Figure 10**) is involved in the binding of the isoalloxazine ring system, which is missing in archaeal dodecins.[1,13,51]

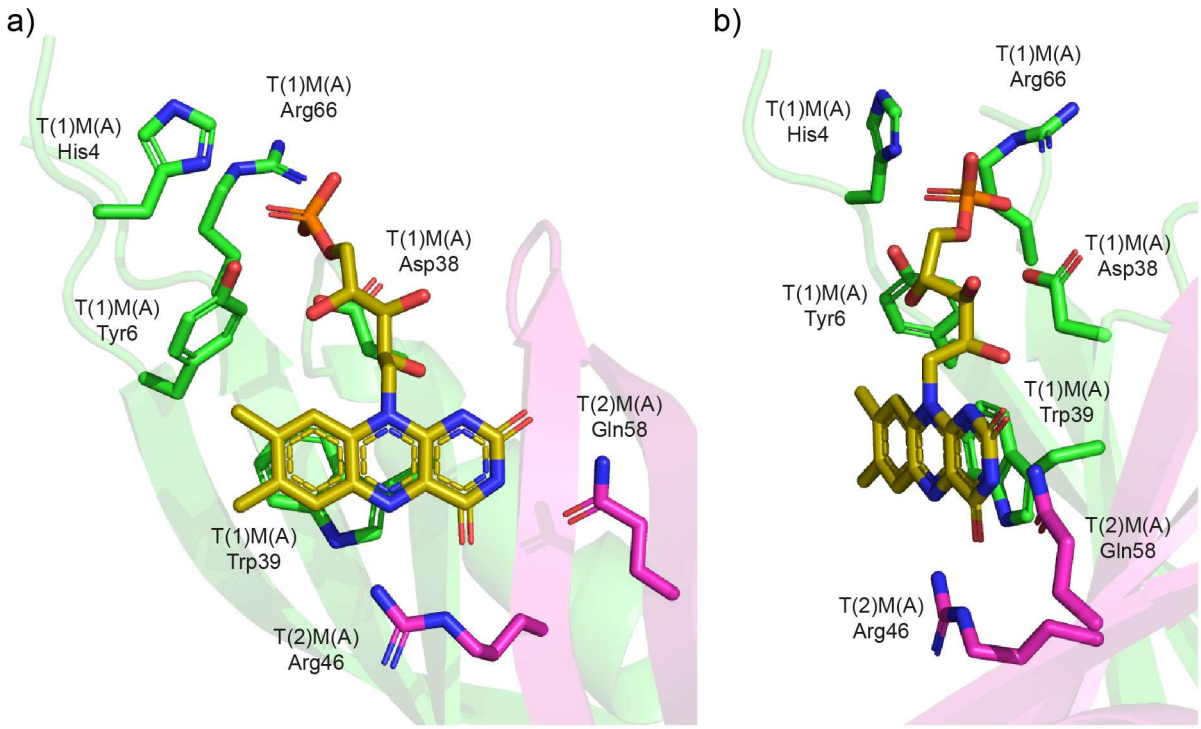


Figure 10: A single FMN bound in the *Mt*Dod binding pocket, with the interacting amino acids highlighted (PDB ID: 2YIZ)[51]. a): Front view of the isoalloxazine ring system. b): Side view (slightly tilted) of the isoalloxazine ring system.

The most relevant residues for flavin binding in bacterial dodecins are shown in **Figure 10** (histidine 4 added, based on conservation, see alignment ref. [55]).[13] Arginine 66 is of special interest, since it interacts with the phosphate group of FMN and in *Hs*Dod, there is glutamic acid 63 at the corresponding position, which makes the binding of FMN unfavourable.[7,13] While dodecins have different affinities for different types of flavins, the binding mode is in principle the same.[5,13,14] A special case is FAD, which can be bound in two states, either as the typical flavin dimer (FAD:*Tt*Dod complex PDB ID: 2CZ8) or as a monomer, where the adenine moiety bends back in the binding pocket and forms an aromatic tetrad with the isoalloxazine ring system.[5]

2.3 Life of *M. tuberculosis*: Where does *Mt*Dod fit in?

M. tuberculosis, the main causative agent of tuberculosis (TB), is one of the most severe human pathogens causing millions of deaths every year and evolved to nearly exclusively proliferate inside humans. Making its dodecin interesting as a potential drug target and raising the question, what does dodecin do that such an optimized human pathogen didn't lose the gene. To figure out the biological role of *Mt*Dod, the highly complex life cycle of *M. tuberculosis* needs to be understood in detail.

2.3.1 Mycobacteria in general and *Mycobacterium tuberculosis*

Like all mycobacteria, *M. tuberculosis* is an aerobe, acid-fast gram-positive bacterium with mycolic acid esters containing cell walls (the characteristic of all mycobacteria).[56–59] There are several ways to classify or group mycobacteria due to historical reasons based on, for example, growth speed or pathogenicity.[56,57,60,61] *M. tuberculosis* is counted to the slow growing mycobacteria (these require more than 7 days to form visible colonies) and is part of the *Mycobacterium tuberculosis* complex (MTC: mycobacteria that cause TB or similar diseases in humans and animals).[2,56] While these classifications of mycobacteria are still valid and used, the availability of sequenced genomes and bioinformatic tools allows a more defined and systematic classification. Based on a genome analysis of 150 genomes of mycobacteria by Gupta *et al.* in 2018, the genus *Mycobacterium* was split into the five genera *Mycobacterium*, *Mycolicibacterium*, *Mycolicibacter*, *Mycolicibacillus*, and *Mycobacteroides*.[59] In accordance to the revised genera descriptions of Gupta *et al.*, the term mycobacteria will be used to refer to all members of the five genera, while *Mycobacterium* (here *M.* is only used for *Mycobacterium*) is used to only describe members of the "*Tuberculosis-Simiae*" clade.[59]

The newly defined genus *Mycobacterium* contains 100 defined species (based of genome availability at NCBI: https://www.ncbi.nlm.nih.gov/ for genus *Mycobacterium*; access 29.03.2020) and several unclassified mycobacteria. *M. tuberculosis* is the selected type species (6619 genome assemblies at NCBI: https://www.ncbi.nlm.nih.gov/ for *M. tuberculosis*; access 29.03.2020) and the strain *M. tuberculosis* H37Rv the reference strain of the species.[2] All major human and animal pathogens of the mycobacteria are part of the genus *Mycobacterium* (e.g. *M. tuberculosis*, *M. leprae*, *M. ulcerans* and *M. bovis*), while the other four genera contain mostly non-pathogenic environmental species, although several species seem to be opportunistic pathogens (can infect humans under some conditions, but are considered not

severe life-threatening).[58,59,62] The comparison of the genomes of slow and rapid growing mycobacteria indicates that the slow growing mycobacteria (genera *Mycobacterium*, *Mycolicibacter*, and *Mycolicibacillus*)[59] evolved from ancestral rapid growing species.[63,64] While slow growth rates bear the dangers of being overgrown by fast organisms, they allow species to adapt easier to hostile environments, like e.g. the presence of antibacterial agents.[65,66] This is in agreement with the fact that most relevant pathogenic mycobacteria are slow growing, since they have to survive the hosts defend mechanisms, while most fast growing mycobacteria tend to be soil or water bacteria.[57,62,65] In addition to slow growth rates, the complex cell wall of mycobacteria, containing an outer waxy layer of mycolic acid esters, acts as a strong permeation barrier that protects mycobacteria form many hostile conditions.[67] The importance of this waxy layer was shown by inhibiting the production of full length mycolic acids in *M. tuberculosis* – the studied mutants lost their acid-fast character and failed to cause active TB in the mice.[68] Whereas slow growth and the presence of a waxy cell wall are common features of all mycobacteria, giving a reason why many mycobacteria are opportunistic pathogens and/or parasites, these two features alone are not sufficient to explain the high pathogenicity of some mycobacteria. Pathogenic mycobacteria evolved highly complex systems to cope with the host defence mechanisms, which will be shown on the example of TB.

2.3.2 The Burden of Tuberculosis

TB is the 9th (including HIV/AIDS positive cases) or 10th (excluding HIV/AIDS positive cases) leading cause of human death worldwide, estimated to have caused 1.0-1.5 million (0.5-1.2 excluding HIV/AIDS related cases) deaths in 2018.[69] The estimated global epidemiological burden of TB in 2018 is 9.0-11.1 million incidents, of which about 50% alone came from India, China, Indonesia and the Philippines.[69] TB can be separated in two types, the pulmonary (limited to the lung) and the extrapulmonary type, where the latter is much rarer and often related to weakened immune systems.[70] Only the pulmonary type will be discussed further. In general, TB is treatable and the typical treatment with a daily drug application (isoniazid, rifampicin, ethambutol and pyrazinamide) takes about 6 months, which costs about US$ 40 per person and has a success rate of at least 85%.[69] In cases of drug resistant species, alternative drugs are required and the treatment duration can take up to 20 months, which raises the costs to more than US$ 1000 per person and lowers the treatment success rate to only about 56%.[69] For untreated TB, the average duration of illness is about 3 years (death or cure) with varying

mortality rates depending on the presence of either smear-positive TB (sputum contains active/acid-fast bacteria), which has a mortality rate of about 70%, or smear-negative TB, which has a mortality rate slightly above 20%.[71] In addition to active TB (with symptoms), there is also a latent variant of TB (LTBI: latent TB infection). Here, patients are infected with *M. tuberculosis*, but no symptoms are observed – estimates suggest that over two billion people have LTBI.[72]

Many species of the MTC (List of MTC species: *M. tuberculosis*, *M. africanum*, *M. bovis* (and the *Bacillus Calmette–Guérin* strain (BCG)), *M. microti*, *M. canetti*, *M. caprae*, *M. pinnipedii*, *M. orygis*, *M. suricattae*, *M. mungi*, and the dassie bacillus)[73–76] can cause tuberculosis or similar diseases in humans, but the most relevant species are *M. tuberculosis*, *M. africanum*, and *M. canetti*.[65] Species of the MTC complex seem to only proliferate inside a host and have no environmental habitat and *M. tuberculosis*, *M. africanum*, and *M. canetti* are nearly exclusively restricted to human hosts.[65,77] While the proliferation is limited to a suitable host, many pathogenic mycobacteria can survive in soil and water for an extended period of time (months to years), although the epidemiology impact is debatable.[65,78–80] Since *M. tuberculosis* is the most relevant cause of TB only this species will be discussed further.

2.3.3 Infection: *The Immune Response Determines the Cause of the Disease*

Two key features make *M. tuberculosis* an exceptional pathogen, its ability to hijack alveolar macrophages and to adopt a dormant or latent state.[81–83] Because of the complex interplay of *M. tuberculosis* and the immune system, the relevant steps of the infection and disease will be fundamentally discussed without giving details on the cellular mechanisms. **Figure 11** shows the critical infection steps.

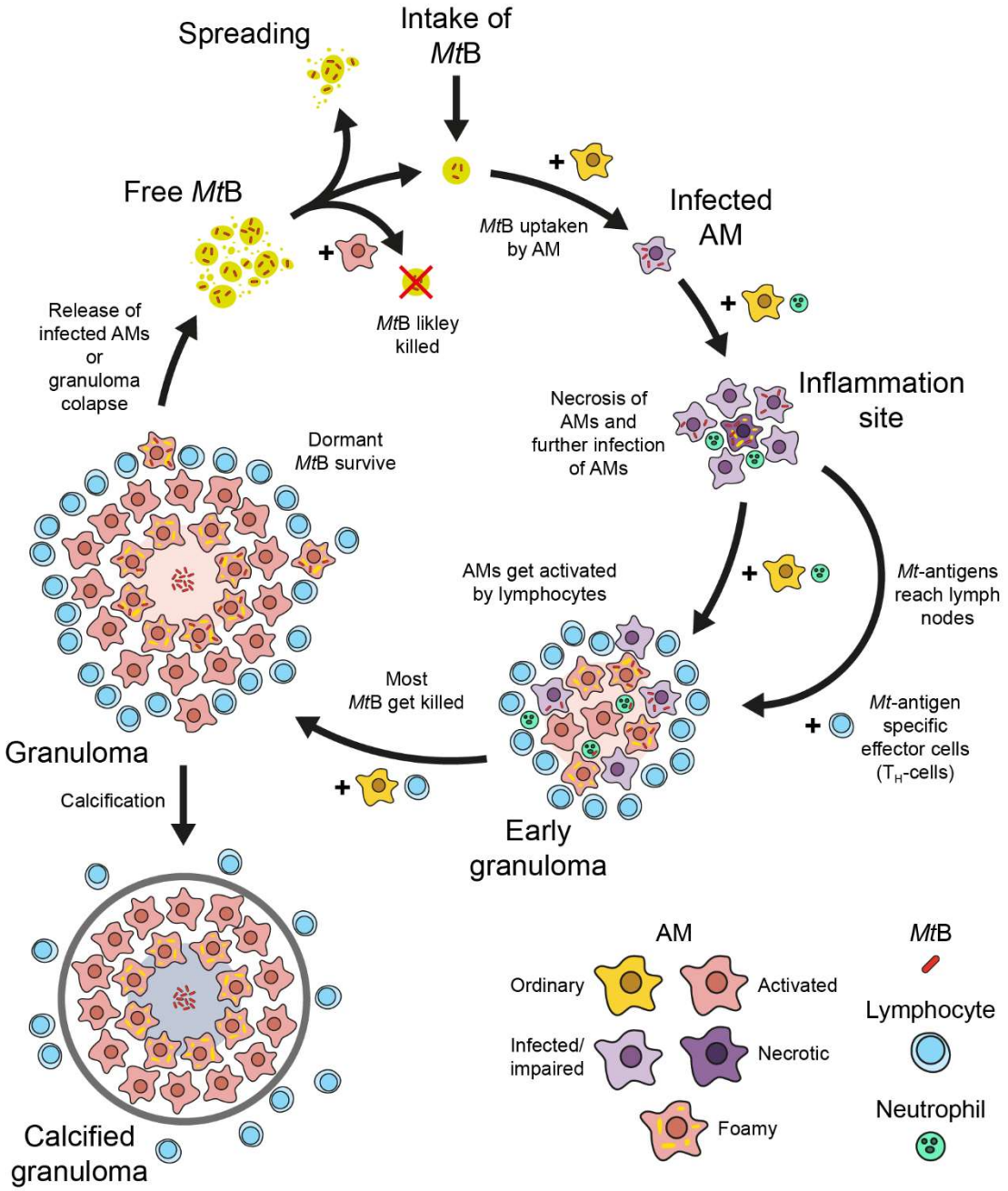


Figure 11: Critical steps during the TB infection. *Mt*B: *M. tuberculosis* bacillus/bacilli. AM: Alveolar macrophage. T_H-cell: Helper T-cell. The role of calcified granuloma is so far debated and shown here as an exit out of the circle.[84–86] The figure is mainly based on the ref. [84] and the ref. [87] while also including points of the ref. [85]

Alveolar macrophages are part of the innate immune system of the lung, that ingest foreign material entering the lung, while tightly controlling inflammation (attracting other cells of the immune system).[88,89] The tight control of inflammation is important to limit tissue damage in the lung and alveolar macrophages can even induce tissue repair.[88] When a *M. tuberculosis* bacillus or rather a cluster of bacilli enters the lung via droplets, it is recognized by alveolar macrophages and engulfed.[85,90] Inside the alveolar macrophages, *M. tuberculosis*

compromises the defence mechanism, ensures that it can proliferate and prevents/alters the communication with surrounding cells by a multitude of actions.[81,85,86,91] Impaired defence mechanisms of alveolar macrophages are: phagosome acidification (acidification stops at about pH 6.4), phagolysosomal fusion (phagosome does not fuse with the lysosome and is unable to degrade captured bacteria), and the production of reactive oxygen and nitrogen species (used by macrophages to kill engulfed bacteria). To ensure proliferation, *M. tuberculosis* prevents apoptosis and autophagy of infected cells (to allow to further multiply inside macrophages), causing the accumulation of lipids and cholesterol as nourishment (formation of foamy macrophages) and induces necrosis of infected cells (dissemination and release of nutrients). It further prevents the presentation of antigens (to prevent T-cell recognition of infected macrophages) by the infected macrophages and induces apoptosis of bystander cells (to inhibit T-cell interactions) to inhibit the activation of the intercellular defence systems. Overall, all these processes can render alveolar macrophages inactive and *M. tuberculosis* can multiply inside them and spread by necrosis of the infected alveolar macrophages. While a single *M. tuberculosis* bacillus might be able to overcome alveolar macrophages, it is assumed that *M. tuberculosis* clusters, agglomeration of bacilli, are the more relevant cause of infection.[85,92] At some point of the infection, inflammation is triggered, which can be caused by *M. tuberculosis* by crossing the basal membrane of the lung (process is not clear) or simply by the increasing amount of bacilli and the necrosis of alveolar macrophages.[84,85,87] The inflammatory response allows the influx of monocytes, more alveolar macrophages and neutrophils (other phagocytic cells).[85,87] While some of the *M. tuberculosis* bacteria are killed by the neutrophils, the enzymes released during this process and toxic oxygen radicals cause necrosis of the lung tissue forming a lesion – neutrophils are in general short lived and possible partly turned necrotic by *M. tuberculosis*.[85] At some point after the infection and inflammation, T-cells (more specifically, effector T-cells, part of the adaptive immune system) will reach the infection sites.[84,87] The T-cells start to activate infected alveolar macrophages and the activated macrophages are able to degrade most *M. tuberculosis* bacilli, although it was also shown that *M. tuberculosis* can survive even in activated macrophages.[93] During these stages granulomas start to form (beginning as a lesion), which are very typical for TB.[84,87] These basically consist of infected alveolar macrophages, free *M. tuberculosis* bacilli, T-cells and cell debris (caused by necrosis) surrounded by other macrophages (often foamy macrophages are close to the centre, while normal macrophages build an outer ring), neutrophils and T-cells.[84,87] *M. tuberculosis* bacilli released from necrotic alveolar macrophages seem to form necrosis-associated extracellular clusters (NEC).[85,90] NECs have

the characteristics of biofilms (higher resistant's to hostile agents) and lost their acid-fastness (indicating changes in the cell wall structure).[85,90] The actual composition of the granuloma and its further development depend on the immune system and genetic predisposition of the patient.[84,87] In the case of an impaired immune system, the T-cell response is weak, while a high number of macrophages is observed, which leads to the formation of large granulomas.[87] In these large macrophage rich granulomas, *M. tuberculosis* proliferates easily and causes the necrosis of macrophages and other cells.[84,87] This leads to a collapse of the granulomas and high amounts of *M. tuberculosis* are released into the lung.[84,87] On the other hand, an exaggerated immune response (excessive inflammatory response) causes the recruitment of high amounts of neutrophils (turned necrotic by *M. tuberculosis*) towards the infection site, which causes severe necrosis of surrounding cells forming growing lesions.[84] Around the original lesions, further new lesions start to form, which at some point fuse with each other (lesion coalescence), forming active TB lesions (in contrast to the initial infection lesions) and massive tissue destruction (creating cavities).[84,85] During a "balanced" immune response, small stable granuloma start to form and infected alveolar macrophages are tightly surrounded by activated alveolar macrophages and then T-cells (forming a sort of cellular wall).[84,87] The low amount of neutrophils keep the tissue damage under control and the amount of *M. tuberculosis* bacilli in the granuloma is decreasing or at least remains the same.[84,87] Over time the granuloma gets encapsulated and calcified stopping the spreading of *M. tuberculosis*.[84,87]

2.3.4 Infection: *M. tuberculosis* – the Sleepy Parasite

Inside the granuloma and active alveolar macrophages, *M. tuberculosis* bacilli encounter hostile environments (hypoxia, low pH (about pH 5.0))[94] and high concentrations of radical oxygen or nitrogen species (mainly inside macrophages).[82,86,94] To survive under these conditions *M. tuberculosis* can adapt a dormant state.[82,86,94] During dormancy, *M. tuberculosis* downregulates it cellular functions to a minimum and stops replication, while upregulating stress related genes.[82,85,86] Dormant *M. tuberculosis* bacilli (at least in some cases) lose their acid-fast character as described for NECs.[85,94–96] This dormancy is or was often related to the latent form of TB, called latent TB infection (LTBI).[97] In some models of LTBI, *M. tuberculosis* stays dormant over long periods of time (here dormancy is more seen as a complete halt of the bacillus) trapped in the granuloma or in alveolar macrophages, "waiting" for the right time to reactivate causing active TB.[82,86] This idea originates from findings made in TB infected mice in the 1950s and is often referred to as "Cornell model".[98–100] The

connection between dormant cells and LTBI is highly debated, since it cannot explain many phenomena, like how treatment with drugs that only affect replicating *M. tuberculosis* bacilli (e.g. isoniazid) cures LBTI or how the reactivation that requires resuscitation factors of active bacilli is working.[85,101] In more recent models, it is thought that LBTI is more related to *M. tuberculosis* surviving/growing outside of the granuloma and alveolar macrophages. In these models the fully encapsulated calcified granulomas are just parts of the primary infection that were not coughed out.[85,101,102] Here calcification separates the uninfected lung tissue from necrotic cells in the granuloma and "pushes" *M. tuberculosis* towards the edge of the lesion into the inner surface of the alveola (towards air lung interface).[85] This happens because the proceeding calcification compresses the NEC containing liquefied cells (by necrosis).[85] When the *M. tuberculosis* NECs reach the surface, alveolar macrophages ingest them and likely degrade them, because effector T-cells are already present (keeping the alveolar macrophages active). But if this fails, the infection starts anew and NECs can spread via the respiration to other parts of the lung.[85] Alternatively *M. tuberculosis* can grow in the formed cavity (as the granuloma is cleared or "coughed out") in a biofilm like manner (pellicle) causing low inflammation (no TB symptoms), while constantly spreading clusters and bacilli.[85,102] As NECs and *M. tuberculosis* growing in pellicle form are not acid-fast, it would explain why LBTI shows low counts of acid-fast bacteria.[85,102] In this model, dormancy is rather a slowdown than a complete halt, which enables *M. tuberculosis* to prolong its survival in the decaying granuloma waiting for its release, but not a mechanism to survive long time encapsulation in granulomas.[85,101] While the precise role of dormancy in LTBI is still unclear, it is overall accepted to be a mechanism of downregulation to overcome hostile environments.[82,85,86,94] Starvation might trigger dormancy, but since *M. tuberculosis* is able to process foreign lipids and cholesterol, which should be present in high amounts through the foamy macrophage,[81,103] this cause of dormancy might not be highly relevant during infection.[85,104]

2.3.5 Flavoproteins and Flavin Biosynthesis in *M. tuberculosis*

While flavins are important for all organisms, the unusual high amount of potential flavoproteins encoding genes in the genome of *M. tuberculosis* highlight the special importance of flavins for this species.[105] The occurrence of the high number of flavoprotein encoding genes is thought to be related to the need to adapt to different carbon and nitrogen nutrients.[105] Surprisingly, *M. tuberculosis* seems not to possess a flavin transporter and must hence rely on

its own biosynthesis.[42] Accordingly, the predicted genes for flavin biosynthesis seem to be in general essential for *M. tuberculosis* growth, but their seems to be some discrepancies between ref. [106] and ref. [107]. But how the RbF biosynthesis is actually regulated is not clear, since *M. tuberculosis* seems to not utilize a FMN riboswitch.[108,109]

2.3.6 Important Nutrition and their Relation to Flavins

The high number of flavoproteins encoded in the genome of *M. tuberculosis* is an indicator that flavins play a more important role in the life of *M. tuberculosis* compared to other organisms.[105] Two examples for the critical role of flavins and flavoproteins for *M. tuberculosis* to survive inside humans are the cholesterol and lipid metabolisms and its iron homeostasis.

The ability to utilize the hosts cholesterol and lipids is a key feature of *M. tuberculosis* and ensures a carbon source during infection.[81,110] In addition, the utilization of host cholesterol and lipids also seems to be involved in the processes of entering macrophages, preventing the phagolysosomal fusion, growth under acidic conditions, and persistence and reactivation.[81,110–113] The catabolic process for the breakdown of cholesterol and lipids is called β-oxidation, in which aliphatic groups/chains are degraded to acetyl-CoA (main product) and to propionyl-CoA (odd numbered fatty acids and cholesterol).[114–118] Important cofactors for the β-oxidation are flavins (cofactor of acyl-CoA dehydrogenases) and CoA (intermediate/product-carrier).[119] In mammals, flavin deficiency impairs the β-oxidation.[120,121] The β-oxidation of the aliphatic chain of cholesterol (involved enzymes ChsE1-ChsE2, ChsE3, ChsE4-ChsE5)[116,122] requires FAD and creates two equivalents of propionyl-CoA and one equivalent of acetyl-CoA.[116–118] The remaining ring system of cholesterol, the androstenedione, is further degraded to pyruvate, acetyl-CoA and succinyl-CoA, but the full pathway is not yet understood.[117,118] This breakdown of the ring system involves a flavin-dependent monooxygenase (HsaAB) that can utilize FMN or FAD.[123] Acetyl-CoA, propionyl-CoA and succinyl-CoA (only energy generation) can be used as building blocks for new fatty acids/polyketides or for energy generation (via the citric acid cycle).[111,117,118] Propionyl-CoA is either processed to methylmalonyl-CoA for the synthesis of polyketide virulence lipids (e.g. phthiocerol-dimycocerosate (PDIM)) or converted into succinyl-CoA.[111,117,118] The intracellular amount of propionyl-CoA needs to be controlled as its accumulation is toxic for *M. tuberculosis*.[117,118,124] One way of *M. tuberculosis* to avoid propionyl-CoA accumulation is by

converting it into methylmalonyl-CoA and then funnelling it into the methylmalonyl pathway.[118] This solution requires vitamin B12, which can be synthesized by *M. tuberculosis de novo* and also utilizes FMN.[118,125]

In addition to carbon sources, iron is a growth limiting factor for *M. tuberculosis*. Iron availability inside the host is limited since *M. tuberculosis* needs to compete with the hosts own cells.[126] To scavenge iron from the host, *M. tuberculosis* utilizes two siderophores (iron binders), mycobactin and carboxymycobactin.[126] Both substances have nearly the same chemical structure and only differ in the length of an acyl moiety that alters their hydrophilicity. Both mycobactins are strong iron Fe^{3+} binders and can either take up insoluble iron or iron bound to proteins.[126] Mycobactin is hydrophobic and restricted to the cell-envelope of *M. tuberculosis*, while carboxymycobactin is hydrophilic and scavenges iron from the extracellular surroundings.[126] Carboxymycobactin transfers scavenged iron to the mycobactin present in the cell envelope.[126] The mycobactin bound iron is then transported into the cell by IrtAB that utilizes a bound FAD to reduce the iron bound to mycobactin.[127] The reduced iron Fe^{2+} is released from mycobactin and available for the organism. IrtAB knockouts or IrtAB mutants unable to bind FAD are not able to replicate in low iron-media, human macrophages and mice.[127] The connection between flavins and iron reduction and uptake was also observed for other organisms.[128]

2.4 Dodecins as Carrier for Bioactive Components

While the biological function of dodecin is related to flavins, their extraordinary stability and spherical structure indicate that they are suitable to be used as carrier/scaffold protein in biotechnological applications.

2.4.1 Carrier and Scaffold Systems in Biotechnology

In regards of the here discussed aspects, the terms scaffold and carrier are not strictly defined, as they are in general used in their literal senses.[129–136] Thereby, a scaffold can be anything that adds a spatial component to the system of interest and a carrier is simply an object to which something was or will be attached to. Because of this function-based definition the term scaffold applies to a plethora of systems like cell membranes or organelles,[133] since they are used in nature to control the spatial arrangement. Here, the term is used to describe very limited

systems that are reduced in their functionality to simply add a spatial component to a small and selected group of components. Therefore, a scaffold needs specific binding sites that allow the recruitment of the selected components, which is often used as requirement to define a scaffold.[130,131] In a similar way the term carrier is here used for components with the single purpose of allowing something to be attached and have no other active function, like for example to transport.

Since a scaffold is basically also a carrier and a carrier can be used as a scaffold, the terms are distinguished based on what is attached or recruited and the purpose of the recruitment. The term scaffold is mainly used if the aim is to create a catalytic active entity (from here called assembly) with improved properties compared to the unrecruited enzymes.[129–136] In contrast, the term carrier is used, when the aim is simply to create the carrier cargo attachment, for example attaching antigens to BSA.[137]

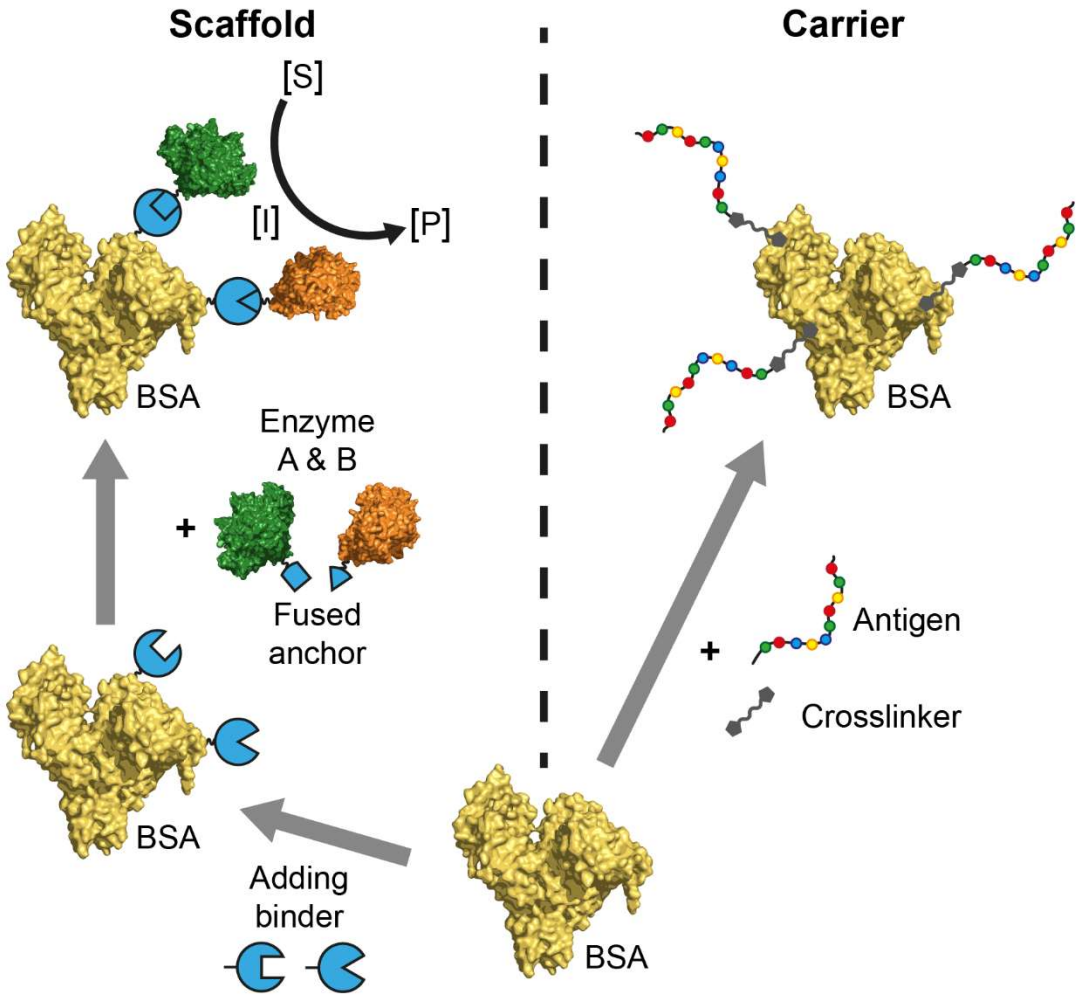


Figure 12: Difference between scaffold and carrier shown on the example of BSA (PDB ID: 4F5S)[138].

Dependent of the used scaffold, the resulting assembly can be highly defined or totally undefined in its composition and/or structure.[139,140] A defined assembly is often called complex, while the undefined variant is referred to as a cluster (sometimes also agglomerate or aggregate).[139,140] But there is not a strict separation and "clustering" is also used to describe complex formation.[140] For clarity, the terms complex or hub will only be used for assemblies with a theoretical defined composition (like a scaffold with a distinct number of binding sites), whereas the term cluster is used when the assembly composition is not predictable (**Figure 13**). The terms assembly (describing the formed structure) and assembling (describing the forming process) are used to include all previously mentioned processes. While a complex needs a scaffold or carrier that controls the assembly, clusters can also be formed without any structure or composition enforcing unit. The process of protein/enzyme clustering is also sometimes associated with immobilization (although this process often refers to adsorption of an enzyme on a matrix) and used to make enzymes more suitable for industrial applications.[141–143] The third type of assembly in addition to clusters and hubs are compartments, where the protein/enzymes are encapsulated in a more or less defined hull (see **Figure 13**).[144,145]

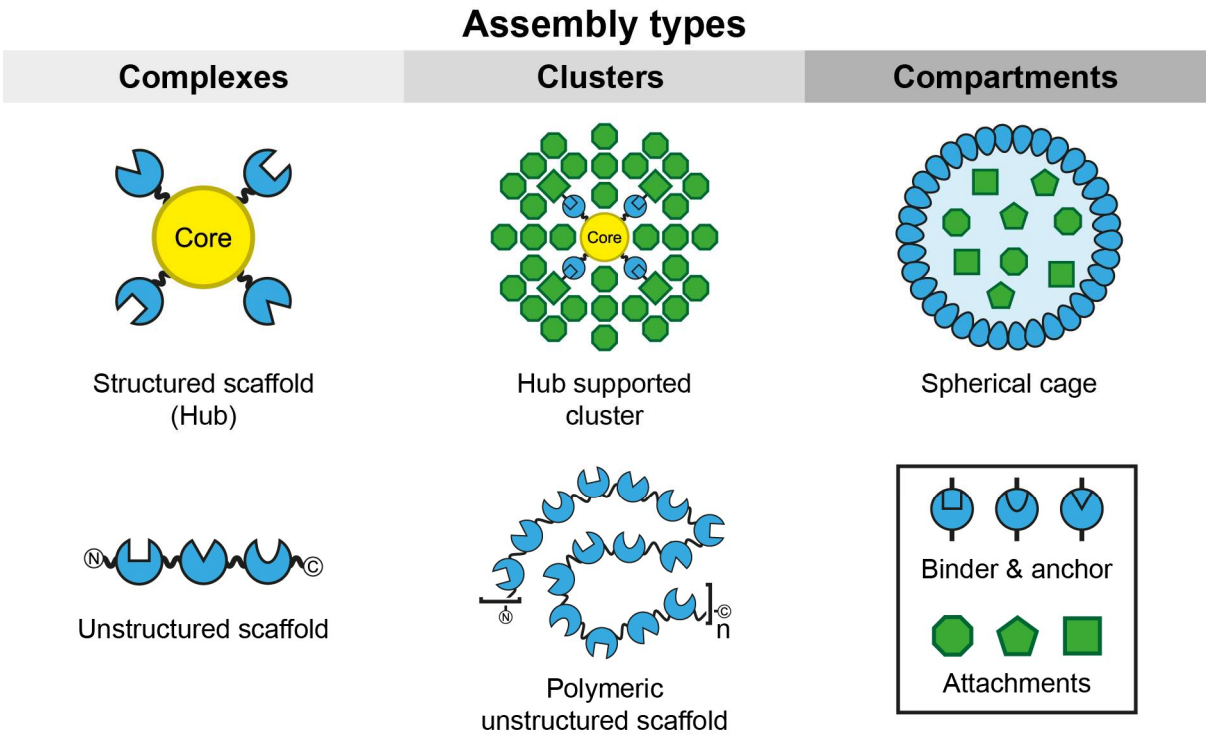


Figure 13: Assembly types. Complexes are defined assemblies, meaning they have a finite composition based on the present binders. In contrast, clusters have no clear finite composition, but can have a defined relative composition. For example, the cluster formed by a "polymeric unstructured scaffold" has the relative 1:1:1 composition, but since the polymerisation degree "n" is unknown the absolute composition is unclear. Compartments refer to all assemblies with an enclosed space formed by scaffolds.

The formed structures/assemblies formed by scaffolds are also called nanoparticles or in the case of compartment forming scaffolds nanoparticle cages (nanocages), since the terms are just based on shape and size.[146] The definition of a nanoparticle is: All dimensions of the object must be in the range of 1-100 nm (upper limit can differ: max 1 μm).[146] Therefore, technically every nanocage is a nanoparticle and therefore also sometimes referred to only as nanoparticle. The general distinction between a nanoparticle and a nanocage is based on the placement of the attachments, if the attachments are place on the surface,[147,148] it is mostly referred to as a nanoparticle and if the attachments placed inside the enclosed space,[149,150] the term nanocage is used. The terms nanoparticle and nanocage will here only be used if the formed particle is the focus of the scaffold and not its function. Therefore, the term core or core-forming carrier/scaffold refers to the centre of a complex/hub (see **Figure 13**), while on its own the core would be a nanoparticle. But this strict separation of those terms is problematic and also not done in the literature, as a functionalized nanoparticle is basically a hub and therefore the nanoparticle would now be considered to be a core.

While for all "assembly types" examples are reported in literature with potential applications, the compartment type might be the most promising one, as compartments protect the enclosed enzymes form interactions with denaturing interfaces/surfaces and can create a diffusion barrier that limits the escape of enzyme cascade intermediates (although the initial substrate intake might be inhibited).[151,152] The diffusion of intermediates is a main point for bioengineering pathways, as the intermediates can be toxic for the production host, and ensuring a low concentration of cellular intermediates is often the purpose of artificial enzyme complexes.[131,140] Unfortunately, compartments are also the most complex group, since factors like compartment size and the insertion of enzymes need to be controlled during the formation process (once formed the enzymes cannot be imported).[151] In contrast, the hub design can be saturated with cargo at any time, giving more freedom during production. Carriers and hubs can also be used to form clusters (clusters of complexes), which combines the industrial benefits of clusters with the ability to control hetero protein/enzyme composition.[139,140]

In the following chapters, we will focus on hubs and clusters since dodecins cannot be used to form compartments.

2.4.2 Why Enzyme Assembling?

The assembling of enzymes in hubs, clusters and compartments is done for two main reasons: making them more suitable for technical applications and enhancing/altering the enzyme cascade flux (simplified effects on concentration and diffusion of substrates, intermediates and products).[134,140,141] Increasing their suitability refers here to stability and recovery of the enzymes during and after applications. A higher stability can also increase the enzyme performance (single and cascades), as it allows to use the enzymes under conditions (e.g. pH, temperature or organic solvents) that are beneficial for product formation.[141]

The origin of the increased enzyme stability in assemblies is likely related to repulsive interactions (minimizing excluded volume, molecular crowding) and limited space (confinement), disfavouring unfolding since this would require a transition away from the "compact" folded native state.[153–156] While the effects of molecular crowding and confinement on protein fold stabilization are overall excepted, the magnitude of the molecular crowding effect is debatable and protein confinement can in some cases even promote denaturing.[156–158] Since clusters are tightly packed proteins or enzymes, their increased stability can be attributed to the factor of confinement. Less confined clusters (e.g. one-layer clusters formed around a support particle) can also be stabilized by encapsulation after cluster formation, which also stabilizes the whole cluster.[141,159,160] In contrast, hubs or complexes are less dense packed and therefore more flexible, which in general limits the potential stabilization affects to molecular crowding.

The other driving force of the creation of artificial enzyme assemblies is to increase the product formation or flux of enzyme cascades, based on the idea that naturally occurring assemblies can be mimicked by bringing enzymes in proximity with each other.[129,131] The overall aim here is to use such artificial assemblies to improve *in vivo* generation of specific products, which is counted to the research field of "metabolic pathway engineering". (e.g. 1-butanol[161] or itaconic acid[162] production in *E. coli*).[131,134,140] It needs to be noted that this topic is highly debated and naturally occurring assemblies are still not fully understood.[135,163] This said, one of the biggest advantages of enzyme assemblies is thought to be substrate channelling (here called intermediate channelling).[164,129,134,135,165,140,163] For reasons of clarity, channelling will be discussed only for a two-step enzyme cascade with and without a branching point (**Figure 14**).

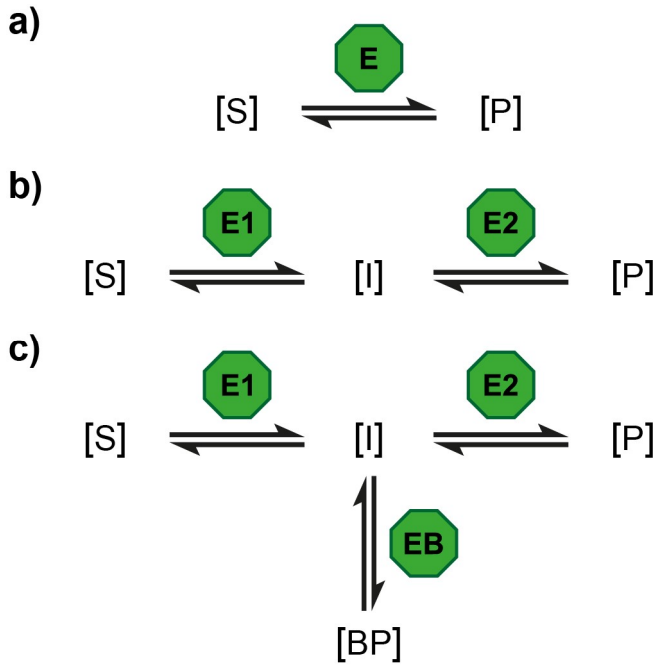


Figure 14: Schematic depictions of enzyme catalysed reactions. a) Single-step reaction. b) Two-step reaction. c) Two-step reaction with branching point. E: enzyme of a standalone reaction. E1 and E2: first and second enzyme of the enzyme cascade. EB: enzyme that catalyses the reaction from the intermediate to the product of the EB branch. [S]: substrate concentration. [I]: intermediate concentration (here product of E1 catalysed reaction). [P]: product concentration (product of the single enzyme catalysed reaction or the product of the E1-E2 enzyme cascade. [BP]: concentration of the product of the EB branch.

Intermediate channelling refers to a process, where the intermediate (product of E1) is passed towards the enzyme E2 without diffusing out of the proximity of both enzymes (entering the surrounding area, like the cytosol). In simple terms, the active sites of both enzymes seem to be connected via a channel through which the intermediate travels. Because of this channelling, the branching point is skipped and no product of the enzyme EB is formed (the reaction **c)** in **Figure 14** basically behaves like the reaction **b)** through channelling). Intermediate channelling for example, is observed in the tryptophan synthase.[166] In the di-enzyme tetrameric complex (αββα) of the tryptophan synthase a literal channel between the α-subunit and the β-subunit is formed.[166] This channel directs the diffusion of the intermediate indole towards the β-subunit, which is therefore literal channelling.[166] In addition to literal channels, there are also charged surface area patches that allow channelling between active sites, referred to as electrostatic highways (electrostatic surface channelling).[164,167,168] Examples for electrostatic highways are the dihydrofolate reductase thymidylate synthase adduct (DHFR-TS),[168,169] and the malate dehydrogenase citrate synthase adduct (MDH-CS).[170–172] The benefit or importance of channelling is that the intermediate does not or only to a limited degree diffuses away from the enzymes and thereby is kept in the reaction path through the cascade (flux).[163,165] Without channelling enzymes outside of the cascade (enzyme EB, see

Figure 14 c)) could scavenge the intermediate and lower the product formation or cause the production of toxic side products (inside a cell).[129,140,163,173]

Whether simple enzyme assemblies can create intermediate channelling is questionable, since enzyme proximity is not sufficient to cause channelling (in simple terms, proximity does not create channels).[163,174] In clusters of different enzymes (for example E1 and E2 from **Figure 14 b)**) something like intermediate channelling and therefore often called intermediate channelling can be observed.[165,163,175] Since such clusters have high enzyme densities, they increase the chance of the intermediate to interact with the second enzyme of the cascade (E2) before it leaves the cluster and thereby preventing the loss of the intermediate.[163] This form of intermediate channelling is therefore sometimes called probabilistic channelling.[163] An example for a natural occurring multi-enzyme cluster is the purinosome that contains six different core enzymes and other proteins.[176,177] While metabolic flux is mentioned as the reason for the formation, in depth studies seem not to be lacking.[176,177]

2.4.3 Examples for Natural Scaffolds and Carriers

Scaffolds play a key role in cellular signalling pathways; here spatial and also temporal control of the pathway participants is used to ensure the correct information flow.[132,133] Their roles range from simple tasks like assembly and/or localization of participants to allosteric regulation, as for example feedback loops.[132,133] For simplicity only scaffolds proteins are further described in their biological role, but also other cellular components can be utilized to control spatial arrangement.[133] A scaffold needs to fulfil two core aspects to function: domains that control the recruitment of the right components and a way to facilitate spatial arrangement.[132,133] A main benefit of scaffold proteins is that they allow the regulation of pathways without the need to affect the functional participants themselves.[133] Even with only a defined set of participants, different scaffold proteins can create different pathways, enabling the cell to have diverse signalling responses with a limited set of components.[133] An example for such a system is the mating or filamentation pathway in yeast.[130] In the presence of the scaffold protein Ste5 the mating signal is transmitted from Ste7 (activated by Ste11) towards Fus3 (**Figure 15**).[130] Without Ste5, Ste7 (again activated by Ste11) can only activate Kss1 that will induce filamentous growth (starvation induced).[130] Ste5 allows tight control of the mating pathway (requirement of Fus3 and Ste5) without requirement to alter the starvation induced pathway for filamentous growth.[130]

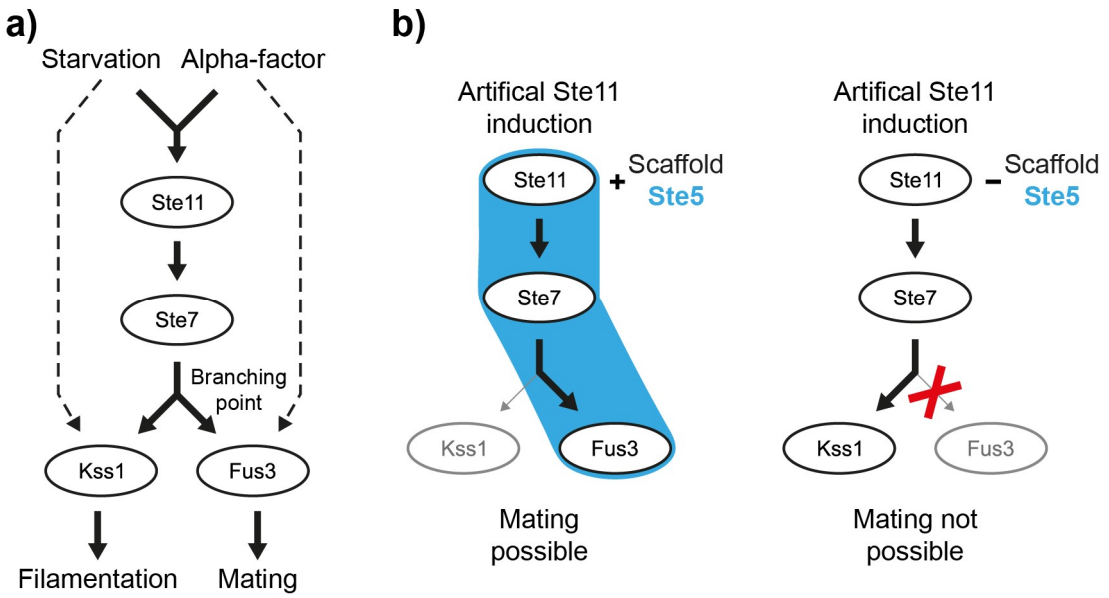


Figure 15: Effect of the scaffold Ste5 on the filamentation or mating pathway. a) Shared steps between the mating and filamentation pathway. Starvation will not trigger the Fus3 activation and thereby prevents unwanted induction into the mating pathway. This is similar to the branching described in **Figure 13 c)**. b) Artificial induction of Ste11 causes the activation of Ste7,which will activate Fus3, if the scaffold Ste5 is present, or, if the scaffold Ste5 is not present, will activate Kss1. While the scaffold Ste5 is required for the mating pathway, its absence is not a requirement for the filamentation pathway.

Different to the term scaffold protein, the term carrier protein, and more general carrier, is used ambiguously and can describe a big variety of systems that are beyond the here focussed application on information carrier. One use, likely the most common general one, of the term carrier is to define a group of membrane transporter systems/proteins that transport molecules through a membrane without forming constantly open pores (the other group being channels).[178,179] A natural example for a carrier, which is more related to herein focussed function, is the acyl carrier protein.[180,181] In simple terms, these proteins carry an acyl-moiety (substrates and intermediates of fatty acid or polyketide synthases), which is covalently attached by an enzyme, and ensure that the bound moiety interacts with the right enzymes of the synthesis cycle.[181] Technically, acyl carrier proteins do a bit more than just carrying substrates and intermediates, for example the aliphatic chain of longer intermediates is securely bound within their hydrophobic core, a process referred to as sequestration and likely protects those intermediates from side reactions.[181]

2.4.4 Examples for Biotechnological Scaffolds and Carriers

In biotechnology scaffolds and carriers are used to create artificial protein complexes or larger artificial structures based on multiple scaffold units, often, to enhance the formation of a desired product or to enhance other parameters like enzyme stability, as described above.[134,140,151,175] Examples for the latter are DNA frameworks/lattices, artificial membrane-like vesicles/tubes (often unordered, e.g. amphiphile peptides) and virus like particles.[145,151,182–185] These scaffold systems fit more in the field of compartmentalization and are not further discussed.

Since biotechnological complex or hub forming scaffolds are basically "just" devices to bring different functional proteins in proximity, they are often comprised of directly linked domains or proteins that can recruit and bind (binder) their specific counterpart (anchor).[131,136,186] Alternatively binders can be fused to oligomer forming proteins, which creates a sort of assembly core (**Figure 16**).[131,136,186]

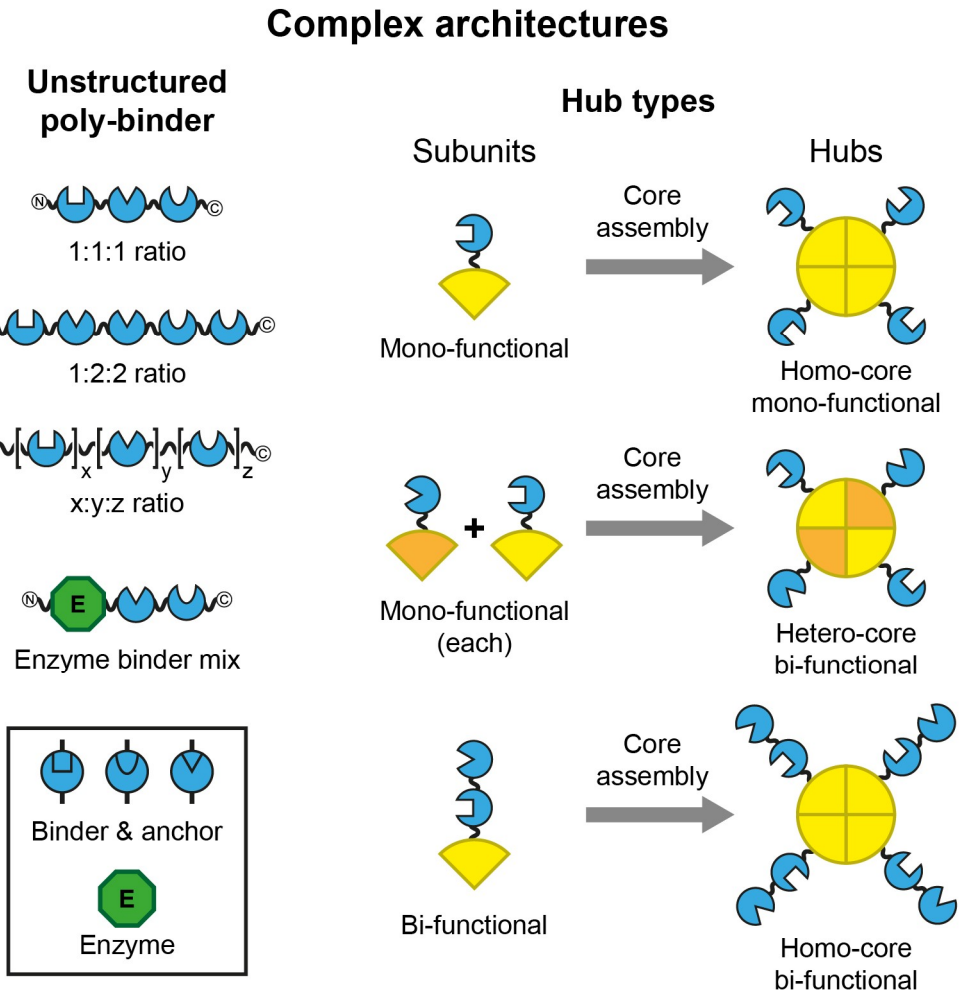


Figure 16: Different types of complex architectures.

The specific coupling of different orthogonal binding systems creates a tool to control the stoichiometry of the formed complexes and hubs.[131] Depending on the oligomeric state of the anchored functional proteins (here the focus is on enzymes, but systems are not limited to them), these hubs will form clusters theoretically defined in relative composition by the scaffold design.[134,140] Binders can also be combined with proteins that form stable oligomers to create hubs with a defined core (see **Figure 16**). Examples for such "core" forming proteins are concanavalin A (tetramer)[187], streptavidin (tetramer)[139,188], shell protein EutM (hexamer, forms larger clusters)[189], IMX313 (heptamer)[190], leucine dehydrogenase (octamer)[191] and DNA binding protein from nutrient starved cells (Dps; dodecamer).[192] These proteins form homo-oligomers (homo-core) and therefore typically homo-binder cores, as every subunit contains the same binder units.[139,187,189,191] With such systems it is very difficult to obtain defined hetero-cores (e.g. 3-times subunit with binder A and 1-time a subunit with binder B), because there is no direct way for a controlled formation of such particles. Undefined hetero-cores can be formed by using a mixture of different subunits for core formation (statistical controlled), but their purification is problematic.[188] Homo-binder scaffolds (non-core or core variants) can also be loaded with different enzymes by using a mixture of these enzymes with the same anchor during charging of the scaffold.[186,189] Further, also multiple binders could directly be fused to a functional protein (ref. [136] uses single domain antibodies) to create an enzyme scaffold hybrid (see **Figure 16**). Oligomeric enzymes could here also function as a core.

Examples for binder domains are the GTPase binding domain (GBD),[131,193] the SH3 domain,[131,193] the PSD95/DlgA/Zo-1 (PDZ) domain,[131,193] the SpyCatcher[190], the SnoopCatcher[190], SYNZIPs,[136,194] and streptavidin[139].

2.4.5 Carrier Protein Supported Antibody Production

While scaffolds are used to create artificial enzyme assemblies, carriers can also be used to fine-tune the recognition of bioactive material, for example antigen-carrier conjugates. By attaching antigens to appropriate carriers, antibodies (AB) can be produced even for non- or weak immunogenic antigens. The cellular processes important for AB production and core features of ABs are described in Chapter 8.1.

Since AB production is closely related to vaccines, a lot of different antigen and carrier systems (proteins and artificial systems)[195–199] exist and to discuss them all is beyond the scope of this thesis. Therefore, this chapter will be focussed mainly on carrier proteins and peptide antigens.

To produce ABs against a selected target (e.g. a protein or specific peptide) the used compounds need to activate antigen presenting cells (APC), facilitate helper T-cells (T_H-cell) activation (requires major histocompatibility complex class II (MHC II) epitopes) and contain the actual target or epitope (B-cell receptor (BCR) epitope) (**Figure 17**).[200,201] Different to BCR epitopes, which define what epitopes the later produced ABs recognize, MHC II epitopes have no direct impact on the specificity of the produced ABs and are required to activate T_H-cells.[202] In simple terms, MHC II proteins, also called MHC II molecules, are membrane bound proteins that get loaded with fragments of digested extracellular proteins, which were acquired through phagocytosis.[202] If the presented fragment is recognized by the T-cell receptor, it signals T_H-cells that the fragment presenting cell (here APCs and B-cells) has taken up foreign proteins.[202] These protein fragments are called MHC II epitopes or more general T-cell epitopes (includes also other types).[202]

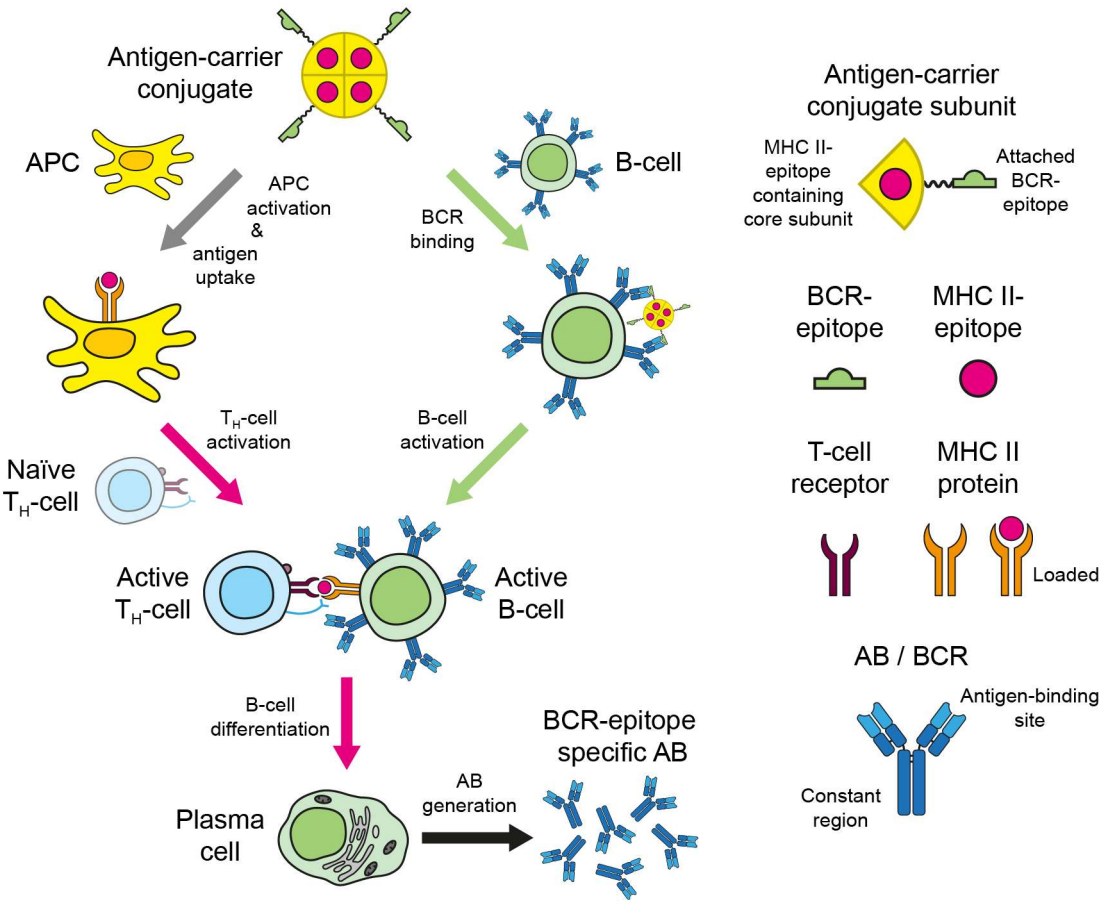


Figure 17: Key steps in AB generation. The colour of the arrows represent, which epitope is important for this step. Grey arrow: The activation of APC and the antigen uptake is not dependent on the used MHC II- or BCR-epitope and can be affected by adjuvants and the size of the antigen-carrier conjugate. Magenta arrow: These steps depend on the MHC II-epitope and are required for the T_H-cell activation and the B-cell differentiation induced by the active T_H-cells. Green arrow: Steps depend on the BCR-epitope, which represents the antigen recognized by the later produced AB. The BCR is depicted here as a membrane bound AB, which is sufficient for this level detail.

An optimal antigen-carrier protein conjugate can fulfil all three aspects, but they often simply act as multivalent BCR-epitope carrying particles that contain the MHC II-epitopes.[203] For antigen-carrier conjugates that cannot or only weakly activate APCs, adjuvants are added during the process of immunization (injection of the antigen-carrier conjugate).[204,205] In simple terms, adjuvants are compounds that cause a general immune response, which attracts and activates APCs.[204,205] A classic adjuvant is "Freund's Complete Adjuvant" (FCA), that is a mineral oil/water emulsion with heat-killed dried *M. tuberculosis* bacteria (surfactants or emulsifier are added).[204,206,207] A factor related to the APC activation is the particle size of the antigen-carrier conjugate, since too small particles are less efficiently taken up by APCs.[208] Optimal particle sizes seem to be in the range of 10 nm to 100 nm (about 40 nm being the suggested optimum), since APC uptake is facilitated and the antigen can still enter lymph vessels, which ensures an easy availability to B-cells.[208]

Since APC activation can be accomplished by adjuvants, the main attributes of carrier proteins are the ability to present multiple copies of an antigen (multivalent) and the presence of MHC II-epitopes. The most common and broadly available method to attach multiple copies of an antigen (here a synthetic peptide) to a carrier protein is the use of bifunctional crosslinkers, which create artificial covalent bonds between the antigen and the amino acids on the carrier protein's surface.[195,209–211] The typical strategy is to use hetero-bifunctional crosslinkers, like *m*-maleimidobenzoyl-*N*-hydroxysuccinimide (MBS),[212] which connects amines on the carrier protein's surface with thiols present in the synthetic peptide (added to a terminus).[211,213,214] In addition, alternative crosslinking agents exist, which also utilize other functional groups.[209,211] Typical carrier proteins for crosslinking approaches are bovine serum albumin (BSA), chicken ovalbumin (OVA) and keyhole limpet hemocyanin (KLH) and are used by many companies to produce custom ABs, e.g. *Eurogentec*, *Davids Biotechnologie* or *Thermo Fisher scientific*.[209] KLH is possibly the most commonly used carrier protein, because of its high immunogenicity (contains diverse immunogenic epitopes and forms up to about 8000 kDa complexes, subunits are about 400 kDa).[215,216] A disadvantage of KLH is its tendency to aggregate, which makes the handling and storage problematic.[201,210]

An alternative to the crosslinking of synthetic peptides with carrier proteins is fusing the antigen to a carrier protein, for example a core-forming protein, on gene level and thereby directly expressing the antigen-carrier conjugate as a fusion protein.[203,217] Here, the carrier protein or rather complex are self-assembling peptides or proteins, which assemble to particles that present multiple copies of the fused antigens, examples are VLPs or self-assembled peptide nanoparticles (SAPN).[218] This method is rather new and in general more intended for vaccine design, but basically fulfils the same aspect of linking MHC II-epitopes with BCR-epitopes like crosslinking approaches.[218,219] Further, carrier proteins can be fused to binder domains, like the SpyCatcher system, to allow the attachment of whole antigens (proteins), which was used to couple antigens to VLP and IMX313.[190,220]

Similar to the APC activation by adjuvants, T_H-cell activation can be accomplished by covalently adding known MHC II-epitopes to carriers that do not contain them, like e.g. artificial non-protein carriers,[195,221–223] but normally MHC II-epitopes are an inherent part of the general used carrier proteins.[216,224–227] The simplest artificial carrier system consists of multiple BCR- and MHC II-epitopes covalently linked to a small molecule, e.g. branched poly-lysines, called multiple antigen peptides (MAP).[195,221,222]

Regardless of the chosen method to produce the antigen-carrier conjugate, an important factor is the density of antigens on the carrier's surface, typically higher densities produce more anti BCR-epitope ABs (facilitates BCR clustering, B-cell activation)[202,228–230], while lower densities create a less diverse AB populations.[231–234]

3 Aim of the Thesis

The aim of this thesis can be separated into two topics, the understanding of the biological role of dodecins and the utilization of dodecins as carriers.

Biological function of dodecins

Although dodecin encoding genes are present in a vast spectrum of bacteria, only the dodecin of the archaeon *H. salinarum* is in depth characterized so far. *Hs*Dod binds RbF with high affinity and is supposed to function as a RbF storage. The role of *Hs*Dod as a RbF storage is supported by the periodic appearing blooms of *H. salinarum* in salt lakes after phases of low salinity. During these phases *H. salinarum* is more or less inactive and *Hs*Dod binds unneeded RbF in the cytosol to prevent the degradation of it. When the salinity reaches again a high enough concentration, *H. salinarum* regains its activity and can utilize the stored RbF to obtain an advantage during its early growth phases.

The few available studies of bacterial dodecins showed, that bacterial dodecins have a slightly altered flavin binding pocket, which cause them to bind preferably FMN instead of RbF. Further, bacterial dodecins seem to be less affine towards flavins in general, which limits the ability of bacterial dodecins to tightly bind flavins and thereby to protect them from degradation, which would make the storage role unlikely. These finding indicated that bacterial dodecins have another biological role than the archaeal dodecins, which is so far not elucidated.

Aim of the thesis was to gain further insight into the biological role of bacterial dodecins by characterizing the flavin binding process in detail. For this, the bacterial dodecins of *M. tuberculosis*, *S. coelicolor* and *S. davaonensis* were studied. The first goal was to establish a binding model that can describe the unique binding pocket of dodecins in more detail than the so far used single-step model. For this, the flavin binding of *Mt*Dod was in depth analysed and numerical functions created. Further the aim was to develop methods to quantify the flavin binding without the need of an accurate binding model to allow parallel screening of various conditions. Initial experiments showed that *Sc*Dod and *Sd*Dod have substantially lower affinities to flavins, when compared to other bacterial dodecins, and *Sd*Dod also displayed unusually low thermal stability for a bacterial dodecin. Therefore, to understand these findings the crystal structure of *Sc*Dod and *Sd*Dod was solved. Lastly the findings were used to propose

the biological role of bacterial dodecins as flavin buffer and discussed with the outlook of dodecins as a potential drug target.

Dodecins as carrier/scaffold proteins

While flavin binding might be the key for their natural function, dodecins have also evolved unique protein properties, which make them interesting for biotechnology. They are exceptionally stable, well-expressible in bacteria, highly symmetric and all termini are accessible on the protein's surface. The combination of all these features makes bacterial dodecins valuable for biotechnological applications as it can be used as a protein-based nanoparticle. Possible applications range from antigen carriers in vaccines to localization hubs of enzymes in metabolic engineering.

As so far, no dodecin was used as a carrier/scaffold in this regard, the aim of this thesis was to evaluate, if *Mt*Dod can actually be used as a carrier for antigens and how its potential as a scaffold for more complex attachments holds up.

4 Results

This chapter includes published data regarding stability, flavin affinity and biological function of *Mt*Dod, *Sc*Dod and *Sd*Dod. Further the published material includes the use of *Hs*Dod as a diffusion probe and the biotechnological application of *Mt*Dod as scaffold and carrier protein. Publications sorted by their focus.

Focus biological function:

- **Bourdeaux, F.**; Hammer, C. A.; Vogt, S.; Schweighöfer, F.; Nöll, G.; Wachtveitl, J.; Grininger, M. Flavin Storage and Sequestration by *Mycobacterium tuberculosis* Dodecin. *ACS Infect. Dis.* **2018**, *4* (7), 1082–1092. https://doi.org/10.1021/acsinfecdis.7b00237.
  Author contribution: F. Bourdeaux designed and performed most experiments, except measurements regarding the spectroscopic and spectroelectrochemical properties of *Mt*Dod. F. Bourdeaux analysed most data and together with M. Grininger wrote the manuscript.

- Ludwig, P.; Sévin, D. C.; Busche, T.; Kalinowski, J.; **Bourdeaux, F.**; Grininger, M.; Mack, M. Characterization of the Small Flavin-Binding Dodecin in the Roseoflavin Producer *Streptomyces davawensis*. *Microbiology* **2018**, *164* (6), 908–919. https://doi.org/10.1099/mic.0.000662.
  Author contribution: F. Bourdeaux assisted in data analysis and preparation of the manuscript.

Focus structure and stability:

- **Bourdeaux, F.**; Ludwig, P.; Paithankar, K.; Sander, B.; Essen, L.-O.; Grininger, M.; Mack, M. Comparative Biochemical and Structural Analysis of the Flavin-Binding Dodecins from *Streptomyces davaonensis* and *Streptomyces coelicolor* Reveals Striking Differences with Regard to Multimerization. *Microbiology* **2019**. https://doi.org/10.1099/mic.0.000835.
  Author contribution: F. Bourdeaux together with P. Ludwig designed and performed all experiments, except crystallization of *Sc*Dod. F. Bourdeaux analysed most data and together with M. Mack wrote the manuscript.

Focus biotechnological application:

- Nöll, T.; Wenderhold-Reeb, S.; **Bourdeaux, F.**; Paululat, T.; Nöll, G. Diffusion-Ordered NMR Spectroscopy of Guest Molecules in DNA Hydrogels and Related Matrices. *ChemistrySelect* **2018**, *3* (37), 10287–10297. https://doi.org/10.1002/slct.201802364.
  Author contribution: F. Bourdeaux designed and performed *Hs*Dod stability measurements and wrote the respective paragraphs in the publication.

- **Bourdeaux, F.**; Kopp, Y.; Lautenschläger, J.; Gößner, I.; Besir, H.; Vabulas, R. M.; Grininger, M. Dodecin as Carrier Protein for Immunizations and Bioengineering Applications. *Scientific Reports* **2020**, *10* (1), 13297. https://doi.org/10.1038/s41598-020-69990-0.
  Author contribution: F. Bourdeaux designed and performed most experiments in the publication, exceptions are western blots and expressions of some constructs performed by J. Lautenschläger. and I. Gößner. F. Bourdeaux analysed data and together M. Grininger wrote the manuscript.

ACS Infectious Diseases

Article

Cite This: ACS Infect. Dis. 2018, 4, 1082−1092

pubs.acs.org/journal/aidcbc

Flavin Storage and Sequestration by *Mycobacterium tuberculosis* Dodecin

Florian Bourdeaux,[†] Christopher A. Hammer,[‡] Stephan Vogt,[§] Felix Schweighöfer,[‡] Gilbert Nöll,[§] Josef Wachtveitl,[‡] and Martin Grininger*[,†]

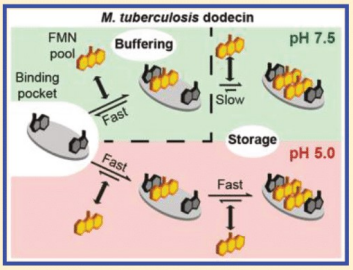
[†]Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Cluster of Excellence for Macromolecular Complexes, Goethe University Frankfurt, Max-von-Laue-Str. 15, D-60438 Frankfurt am Main, Germany

[‡]Institute of Physical and Theoretical Chemistry, Cluster of Excellence for Macromolecular Complexes, Goethe University Frankfurt, Max-von-Laue-Str. 7, D-60438 Frankfurt am Main, Germany

[§]Nöll Junior Research Group, Organic Chemistry, University of Siegen, Adolf-Reichwein-Str. 2, D-57068 Siegen, Germany

Supporting Information

ABSTRACT: Dodecins are small flavin binding proteins occurring in archaea and bacteria. They are remarkable for binding dimers of flavins with their functional relevant aromatic isoalloxazine rings deeply covered. Bacterial dodecins are widely spread and found in a large variety of pathogens, among them *Pseudomonas aeruginosa*, *Streptococcus pneumonia*, *Ralstonia solanacearum*, and *Mycobacterium tuberculosis* (*M. tuberculosis*). In this work, we seek to understand the function of dodecins from *M. tuberculosis* dodecin. We describe flavin binding in thermodynamic and kinetic properties and achieve mechanistic insight in dodecin function by applying spectroscopic and electrochemical methods. Intriguingly, we reveal a significant pH dependence in the affinity and specificity of flavin binding. Our data give insight in *M. tuberculosis* dodecin function and advance the current understanding of dodecins as flavin storage and sequestering proteins. We suggest that the dodecin in *M. tuberculosis* may specifically be important for flavin homeostasis during the elaborate lifestyle of this organism, which calls for the evaluation of this protein as drug target.



KEYWORDS: *flavoproteins, flavoenzymes, homeostasis, storage protein, tuberculosis, pH control, antibiotics*

Flavin binding proteins (flavoproteins) occur ubiquitously in living entities. Mediated by the versatile chemistry of the isoalloxazine moiety, flavoproteins catalyze a wide range of redox or light-induced reactions. The human proteome contains about 90 flavoproteins, which are involved in primary metabolic pathways such as the citric acid cycle and β-oxidation or biosynthetic pathways of other cofactors.[1] The biological relevant flavins are riboflavin (RbF), flavin mononucleotide (FMN), and flavin adenine dinucleotide (FAD) (Figure 1).

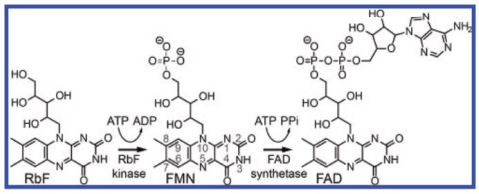


Figure 1. Biosynthesis of flavins. The physiological important flavins, riboflavin, FMN, and FAD, differ in their substituents at the ribityl O5′ position. Flavins are oriented with the view onto the *re*-faces of the isoalloxazine rings.

FMN and FAD are the commonly used flavins in enzymes. The use of FMN and FAD, but not of RbF, as active cofactors may be based on the phosphate (FMN) or adenosine diphosphate (FAD) moieties, which allow tight coordination of flavins to the proteins.[2,3]

RbF binding proteins that store and sequester RbF have been identified in eukaryotes.[4,5] About 15 years ago, the flavoprotein dodecin was discovered in archaea.[6] Dodecins have further been identified as being broadly distributed in prokaryotes. Dodecins are about 70 amino acids large and form homododecameric hollow spherical structures. They carry six binding pockets for flavin dimers, giving an overall 1:1 protomer/flavin stoichiometry. The high flavin load of dodecins is realized by sandwiching two isoalloxazine moieties of the flavin dimer between two tryptophans. Flavin binding upon formation of an aromatic tetrad seems to be common in dodecins, although the relative orientation of the flavins within the tetrad differs.[7] In archaeal dodecins, analyzed on the example of *Halobacterium salinarum* dodecin (*Hs*Dod), flavins are stacked via their *re*-faces (*re-re* arrangement), while flavins

Received: November 17, 2017
Published: April 2, 2018

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

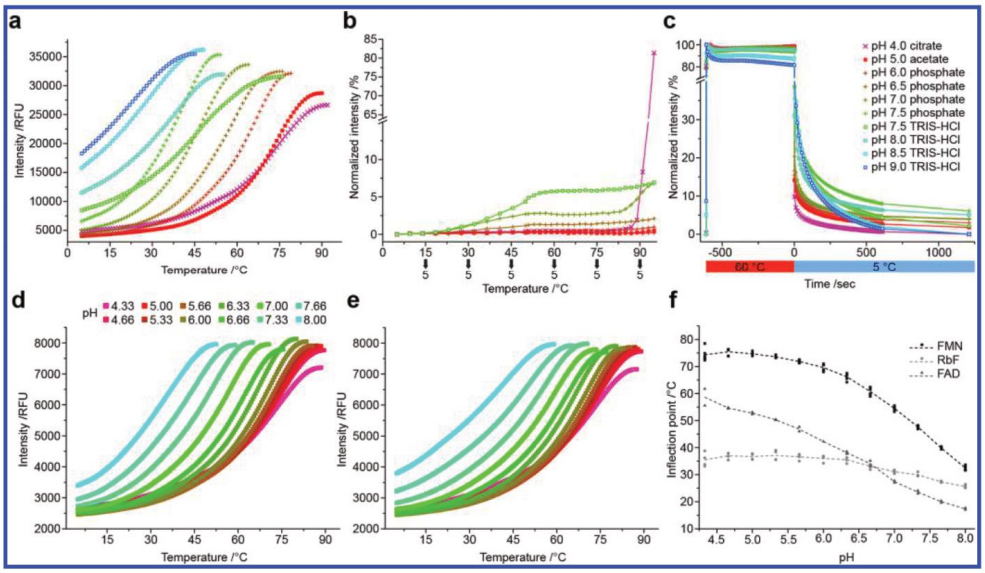


Figure 2. Temperature-dependent dissociation of FMN:*Mt*Dod. Curves are color coded as indicated in figure inset. (a) Thermal fluorescence assay of FMN:*Mt*Dod at conditions of 300 mM NaCl, 5 mM $MgCl_2$, and 20 mM of the respective buffer and recorded with a stepwise temperature increase of 1.0 °C/60 s. Initial low fluorescence is based on fluorescence quenching of *Mt*Dod bound flavins. For clarity, curves are shown until a fluorescence maximum was reached. (b) Thermocyclic fluorescence assay using thermal cycles, performed with FMN:*Mt*Dod at selected conditions (fluorescence intensity vs heating step temperature). Fluorescence intensity was generally taken after cooling to 5 °C. Up to 50 °C, the temperature was increased in steps of 4.5 °C. When 50 °C was reached, steps were reduced to 2 °C. (c) Fluorescence quenching profile during the alternative thermocyclic fluorescence assay protocol. Data are shown for a single selected cooling step from 60 to 5 °C. Rebinding is faster under acidic conditions. (d) Thermal fluorescence assay of FMN:*Mt*Dod at conditions of 300 mM NaCl, 5 mM $MgCl_2$, 20 mM acetic acid, 20 mM ADA, and 20 mM bicine (universal buffer system AAB), recorded with a stepwise temperature increase of 0.5 °C/30 s. (e) Similar experiment as (d) but with the universal buffer system MMT: 20 mM malic acid, 20 mM MES, and 20 mM Tris. (f) Inflection points of thermal fluorescence assay of FMN:*Mt*Dod (6 technical replicates), holo complex of *Mt*Dod and RbF (RbF:*Mt*Dod, 4 replicates), and holo complex of *Mt*Dod and FAD (FAD:*Mt*Dod, 2 replicates) in the universal buffer system AAB. The dashed lines connect mean values; no significant difference between two different protein preparations was observed. The universal buffer system MMT showed similar inflection points but with a higher error for RbF:*Mt*Dod (data not shown).

are *si-si* arranged in bacterial dodecins.[6,8−10] Current data also suggest different flavin binding specificities of archaeal and bacterial dodecins. FMN is the native flavin of bacterial dodecins (*si-si* arranged), whereas RbF is the native ligand in archaeal dodecins (*re-re*).[8,9,11] Archaeal dodecins have been proposed to serve as RbF storage devices involved in flavin homeostasis.[10,12] Bacterial dodecins have been shown to bind FMN and coenzyme A (CoA), and a function in cofactor storage was suggested.[9,13]

We recently started to characterize key properties of *Mycobacterium tuberculosis* (*M. tuberculosis*) dodecin (*Mt*Dod).[8,13] *M. tuberculosis* is the main causative agent of tuberculosis (TB) and encodes for a wealth of flavoproteins, suggesting a high relevance of the flavin metabolism in its life cycle.[14,15] In this study, we report biophysical properties of *Mt*Dod as flavin affinity, flavin binding kinetics, and protein stability and disclose its redox and spectroscopic properties. We show that *Mt*Dod is a FMN storage and sequestering device, which can respond in its binding mode to varying concentrations of free flavin and to pH conditions. An increase in FMN concentration favors double occupied binding pockets, which are kinetically stabilized, and further acidification leads to increased affinity by the improved stability of the holo complex of *Mt*Dod and FMN (FMN:*Mt*Dod). These properties allow *Mt*Dod to respond to changing conditions, which occur during the complex lifestyle of *M. tuberculosis*. For example, considering the ability of *M. tuberculosis* to enter a dormant state[16−19] as well as *M. tuberculosis* being exposed to acidic conditions in the phagolysosome or the granuloma,[16,20] a pH-controlled flavin storage system may be beneficial for strain survival and contribute to *M. tuberculosis* pathogenicity.

■ RESULTS AND DISCUSSION

Thermal Shift Assay with FMN Bound to *Mt*Dod. A common characteristic of bacterial dodecins is their high stability under various conditions.[8,9] To disclose properties that promote or compromise dodecin stability, we screened conditions by a thermal shift assay (TSA). Specifically, we were working with flavin saturated *Mt*Dod (flavin:*Mt*Dod) and monitored the release of the flavin by increasing fluorescence.[21] We term this assay thermal fluorescence assay. As recorded by the temperature that is necessary to dissociate the holo-complex, we observed a stabilized FMN:*Mt*Dod complex at low pH values (Figure 2a–f).

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

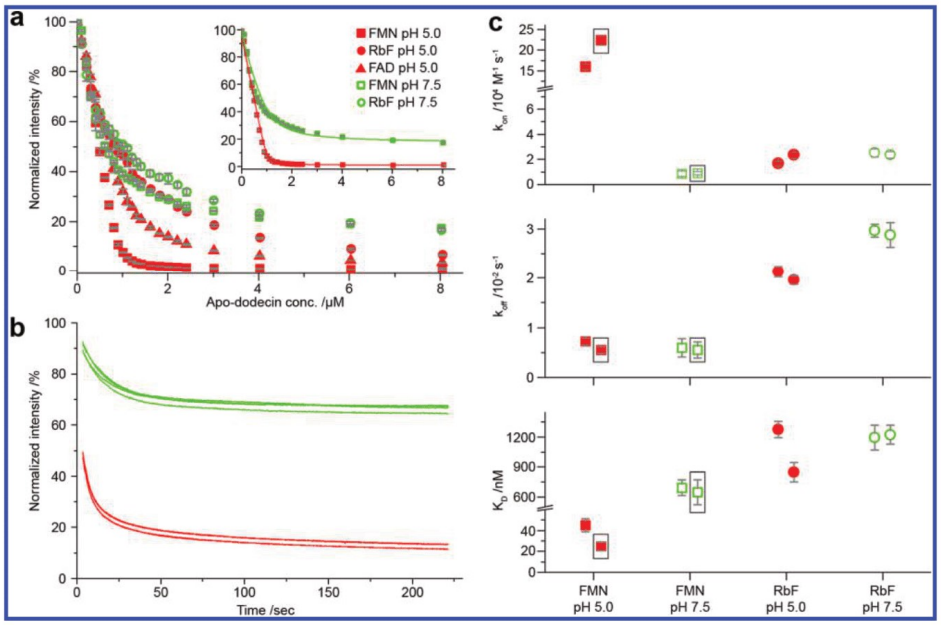


Figure 3. Equilibrium binding data and kinetic binding data. Legend applies to all graphs. (a) Equilibrium data of FMN (□), RbF (○), and FAD (▲) binding at pH 5.0 (red) and pH 7.5 (green). Each curve is based on 24 data points, which were each collected in triplicate (three data points at the highest concentration) and quadruplicate. The inset shows fits of FMN binding curves at both pH values. A simple 1:1 binding model was used for fitting data (same parameter constraints used for pH 7.5 and pH 5.0). The fit does not entirely describe binding data at pH 7.5, indicating deviations from the 1:1 binding model. (b) Kinetic data on FMN binding at pH 5.0 (red) and pH 7.5 (green). Data were collected in quadruplicate at *Mt*Dod to FMN ratio of 1:1. (c) Overview of constants (K_D calculated from determined k_{on} and k_{off}) obtained from kinetic measurements (average and standard deviation from data in quadruplicate). Two independent protein preparations were analyzed, shown next to each other in the graph. Black boxes highlight data shown in (b). A single step binding model was used for fitting (for more information, see Figure S3 and Table S1).

An alternative TSA protocol, which we term thermocyclic fluorescence assay, was additionally set up, which is based on a cyclic temperature profile of heating and cooling, while overall increasing the temperature in heating phases. Under the assumption that *Mt*Dod is just able to efficiently bind flavin in its fully assembled dodecameric state, which can be clearly deduced from the structural properties of dodecins,[13] the ability to rebind flavins after a heating phase was taken as a measure for the structural integrity of the *Mt*Dod dodecamer. Likewise, lost ability of rebinding, indicated by nondecreasing fluorescence during cooling, was interpreted by protein disassembly or denaturation. Compared to the thermal fluorescence assay, this protocol allows one to monitor protein decomposition and gives a better indication of flavoprotein stability. The rebinding ability of *Mt*Dod was recorded from 5 to 95 °C. Just below pH 5.0 and in pure water (data not shown), a loss of rebinding ability was observed (at about 80 and 90 °C, respectively), while at all other conditions, no significant increase of fluorescence occurred (see Figure 2b,c). The two TSA protocols provide valuable information; *Mt*Dod preserves its dodecameric integrity over a wide range of conditions, and the pH dependent FMN release, as observed in the thermal fluorescence assay, occurs while the protein integrity essentially remains intact. Affinity for FMN is highest under acidic conditions (see Figure 2a), where also rebinding of FMN is fastest (see Figure 2c).

Fluorescence Based Binding Assays. The flavin binding behavior of *Mt*Dod was investigated by fluorescence titration experiments, similarly as reported for other dodecins before.[9,11] In addition to equilibrium titration experiments, also kinetic measurements were performed and analyzed with a single step and a two-step model (Figure S1). The apo-*Mt*Dod was obtained via a denaturing/refolding protocol and further selected for the dodecameric oligomer by size exclusion chromatography (SEC). For recording dissociation constants, an increasing amount of apo-*Mt*Dod (0 to 8 μM final concentration) was added to a constant amount of the respective flavin (final concentration 1 μM). After incubation of 1 h at room temperature, fluorescence intensity of the samples was measured by recording the emission at 520 nm after excitation at 450 nm.

At pH 7.5, RbF and FMN showed comparable affinities (dissociation constant, K_D, for RbF of 393 nM (pH 7.5) and for FMN of 118 nM; Figure 3a), while FAD did not show binding to *Mt*Dod (Figure S2), as also reported by Liu et al. before (Table S1).[8] At pH 5.0, FMN and FAD bound with significantly higher affinities. FMN is the favored ligand with a K_D of 18 nM compared to 157 nM for FAD (both pH 5.0). RbF affinity remained essentially unchanged at pH 5.0 (K_D

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

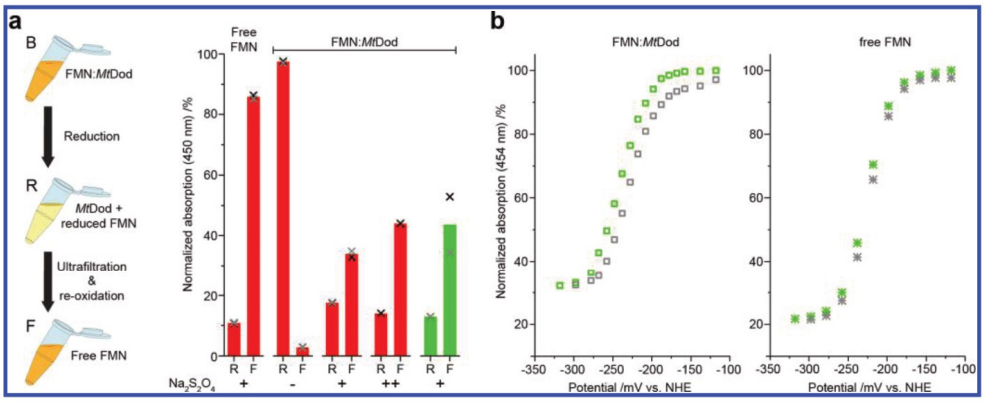


Figure 4. Redox and flavin binding properties of *Mt*Dod. (a) Chemical reduction of FMN:*Mt*Dod with sodium dithionite at pH 5.0 (red) and pH 7.5 (green), as in Figures 2 and 3. Labeling of cartoon: B, before reduction (corresponds to 100%), initial absorption used for normalization; R, after reduction; F, flow-through after filtration (30 kDa cutoff). Bars show average of two measurements (black ×, gray ×). Symbols -, + , and ++ refer to 0, 20, and 40 mM $Na_2S_2O_4$ final concentration, respectively. (b) Spectroelectrochemical investigations of FMN:*Mt*Dod with reduction shown in green and reoxidation in gray. Data of FMN:*Mt*Dod (□) and of free FMN (* with an ×) are given. The spectral changes, proportional to the transition from oxidized to reduced FMN and back, were determined from the absorption at 454 nm.

values of 598 nM; all values refer to the average of two protein preparations; see Figure 3a and Table S1). It should be noted that, throughout all measurements, fluorescence intensity curves at pH 5.0 were reaching lower final values than at pH 7.5. The fitting model suggests nearly complete quenching of fluorescence at pH 5.0 but remaining intensities of above 10% at pH 7.5 (fluorescence of free FMN set to 100%). Dissociations constants were calculated with a single step binding model, as done before.[9,11] However, as visible in Figure 3a, this simple binding model does not seem to fully describe the binding mode (Supporting Note 1).

To determine kinetic properties, flavin binding was also recorded over time. Data were collected as a function of flavin concentrations (FMN and RbF) at pH 5.0 and 7.5 and fitted as a single step binding event. To reveal any effect that may arise from such simple treatment of binding properties, kinetic data were collected at varying *Mt*Dod/flavin ratios (0.25:1, 0.5:1, 0.75:1, 1.0:1, 1.25:1, 1.5:1, 2.0:1, 2.5:1, and 3.75:1; see Supporting Note 1 and Supporting Data "Kinetic_single-step" for all ratios). Data of the 1.0:1 ratio were fitted to k_{on} of 0.8 × 10^4 M^{-1} s^{-1}, k_{off} of 5.7 × 10^{-3} s^{-1}, and K_D of 669 nM for FMN under mild basic pH conditions (pH 7.5) and 19.2 × 10^4 M^{-1} s^{-1}, 6.4 × 10^{-3} s^{-1}, and 35 nM under acidic pH of 5.0 (all values refer to the average of two protein preparations; see Figure 3c and Table S1). Data for RbF binding, fitted with 1.0:1 ratio data, gave k_{on} of 2.4 × 10^4 M^{-1} s^{-1}, k_{off} of 29.2 × 10^{-3} s^{-1}, and K_D of 1.21 μM for RbF under mild basic pH conditions (pH 7.5) and 2.0 × 10^3 M^{-1} s^{-1}, 20.4 × 10^{-3} s^{-1}, and 1.06 μM under acidic pH of 5.0 (again as the average of two protein preparations). Data reveal several aspects of FMN binding (see also Supporting Note 1): (i) k_{on} rate constants are 4 to 5 orders of magnitude lower than diffusion controlled binding events (1 × 10^9 M^{-1} s^{-1}).[22] (ii) k_{on} rate constants are about 24-fold increased for FMN under pH 5.0. Accordingly, faster binding under acidic condition causes the increased affinity for FMN. (iii) k_{off} rate constants are increased for RbF under the tested pH conditions (up to a factor of about 5). The half-life of the flavin:*Mt*Dod complexes (when half of the initially present holo-complexes have dissociated) is within the range of a few seconds. (iv) K_D of kinetic measurements (kin., average of both protein preparations, 1.0:1 ratio) tend to deliver higher K_D than equilibrium measurements (equ., average of both protein preparations) as follows: pH 7.5, FMN 118 nM (equ.) vs 669 nM (kin.); RbF 393 nM vs 1.21 μM; pH 5.0, FMN 18 nM vs 35 nM; RbF 598 nM vs 524 nM. For a comparison of flavin binding data reported in the literature and for data on kinetic binding data of *Hs*Dod, see Table S1, Supporting Note 2, and Supporting Data "Kinetic_single-step".

Kinetic data for RbF and FMN binding were additionally fitted by a two-step binding model. Here, we performed global fitting, i.e., the fitting of all concentrations per "system" (FMN/pH 7.5, FMN/pH 5.0, RbF/pH 7.5, and RbF/pH 5.0), simultaneously. The fit protocol was able to describe binding kinetics in sufficient quality, suggesting two-step sequential binding mode with fast binding and unbinding of FMN in single occupied pockets. We note that, for stable curve fitting, the value for the initial fluorescence, describing the fluorescence of unbound flavin, was released, and the parameter space of rate constants was constrained by defining an upper limit for k_{off_1}. For more details on the fitting procedure, please see Supporting Materials and Methods, Supporting Note 1, and Supporting Data "Kinetic_single-step" for all ratios. With the two step model, we received values for FMN at pH 7.5 of k_{on_1} of 1.7 × 10^4 M^{-1} s^{-1}, k_{off_1} of 134 × 10^{-3} s^{-1}, k_{on_2} of 6.7 × 10^4 M^{-1} s^{-1}, and k_{off_2} of 4.7 × 10^{-3} s^{-1} and for FMN at pH 5.0 of k_{on_1} of 92 × 10^4 M^{-1} s^{-1}, k_{off_1} of 300 × 10^{-3} s^{-1}, k_{on_2} of 27.5 × 10^4 M^{-1} s^{-1}, and k_{off_2} of 5.2 × 10^{-3} s^{-1}. Further, we receive values for RbF at pH 7.5 of k_{on_1} of 8.4 × 10^4 M^{-1} s^{-1}, k_{off_1} of 241 × 10^{-3} s^{-1}, k_{on_2} of 6.9 × 10^4 M^{-1} s^{-1}, and k_{off_2} of 23 × 10^{-3} s^{-1} and for RbF at pH 5.0 of k_{on_1} of 4.8 × 10^4 M^{-1} s^{-1}, k_{off_1} of 41 × 10^{-3} s^{-1}, k_{on_2} of 0.9 × 10^4 M^{-1} s^{-1}, and k_{off_2} of 7.6 × 10^{-3} s^{-1} (Figure S4; see also Supporting Note 1 and Supporting Data "Kinetic_two-step"). All values are the means of measurement performed with the two independent protein

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

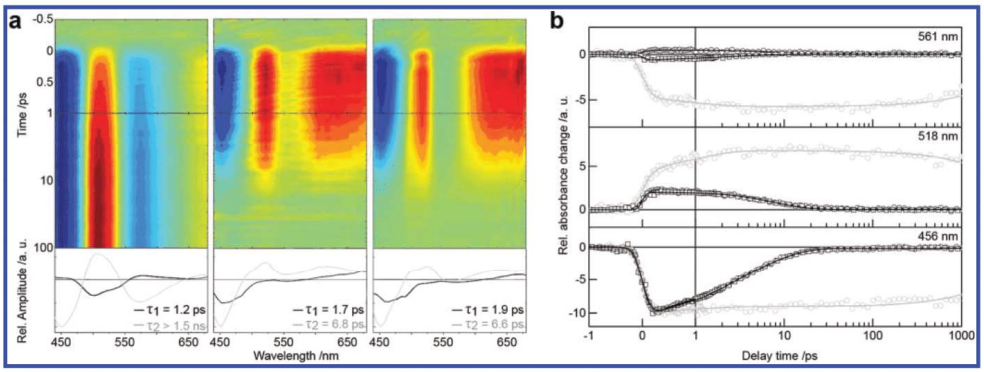


Figure 5. Spectroscopic characterization of free and dodecin-bound FMN. (a) Transient absorption spectra of free FMN in buffer (left) and FMN:*Mt*Dod at pH 7.5 (middle) and at pH 5.0 (right) after photoexcitation at 388 nm. Red indicates positive and blue negative absorbance changes. (b) Transient absorbance changes at three distinct probe wavelengths selected as key representatives of the spectral properties of free and *Mt*Dod-bound FMN. Data is shown as points: Free FMN (gray circles) and FMN:*Mt*Dod at pH 7.5 (black circles) and at pH 5.0 (black squares). The corresponding fits of the global lifetime analysis are depicted as lines.

preparations. Fitting of FMN binding at pH 5.0 reached the upper limit of k_{off_1} of 300×10^{-3} s^{-1}, and given values for FMN binding at pH 5.0 do, therefore, not claim any absolute accuracy, while nevertheless the comparison of parameters can be used for judging the dynamics of binding processes.

Impaired Binding of Reduced FMN ($FMNH_2$). As part of this study, we aimed to analyze whether *Mt*Dod is also able to tightly bind reduced FMN. After chemical reduction with 20 mM sodium dithionite in 300 mM NaCl, 5 mM $MgCl_2$, and 20 mM acetic acid (pH 5.0) or 20 mM Tris-HCl (pH 7.5) buffer, *Mt*Dod is depleted from FMN both at pH 5.0 and pH 7.5 (Figure 4a). To study the redox chemistry of FMN:*Mt*Dod in more detail, spectroelectrochemical measurements were carried out at pH 7.5. Since the different flavin redox states can be distinguished by their absorption spectra, the redox behavior of a distinct flavoprotein can be elucidated by spectroelectrochemical measurements.[23,24] The absorption spectra of FMN:*Mt*Dod were recorded at fixed potentials between −118 and −318 mV vs standard hydrogen electrode (NHE) after reaching electrochemical equilibrium. Thereafter, in the oxidative half cycle, the potential was stepwise switched back to −118 mV vs NHE leading to reoxidation of the reduced FMN (Figure 4b). For free FMN, a redox potential of −223 mV vs NHE was received during reduction and reoxidation, which is in agreement with the pH adjusted published values of −220 mV vs NHE and −226 mV vs NHE (original values: −205 mV vs NHE[25] and −211 mV vs NHE[26] measured at pH 7.0 were adjusted by −15 mV based on the Nernst equation as done in ref 24). The reduction of *Mt*Dod bound FMN proceeds at potential values, which are only slightly more negative than those for free FMN (an inflection point determined at potentials to −240 mV vs. NHE), whereas during the reoxidation the titration curve and the corresponding redox potential could not be distinguished from that of free FMN (see Figure 4b).

The kinetic equilibrium between protein bound and free FMN is fast with respect to the time scale of a spectroelectrochemical measurement with a delay (equilibration time) of a few minutes at each potential step (see Figure 2c). Hence, experimental data reflect either the reduction of FMN:*Mt*Dod followed by dissociation or the reduction of free FMN (while FMN is residing outside the binding pocket in the dynamic equilibrium). In either case, as indicated by the redox-potential during reoxidation, the free reduced FMN, which we expect to be $FMNH^-$,[21] is reoxidized during the subsequent reoxidation. Spectroelectrochemical data agrees well with data from chemical reduction (see Figure 2a). The stability of $FMNH^-$:*Mt*Dod seems to be significantly weakened and therefore is a compelling reason against the capability of *Mt*Dod to store reduced FMN (see also Supporting Note 3).[27,28]

Transient Absorption Spectroscopy. Transient spectroscopic properties of FMN:*Mt*Dod were observed in UV/vis-pump/vis-probe experiments. Photodynamics of dodecins have been analyzed before, where an ultrafast deactivation of excited states of the bound flavins takes place.[12,29] We have performed transient absorption spectroscopy, as we were interested in analyzing the pH dependent FMN binding to *Mt*Dod. Specifically, we aimed to understand whether the 10-fold increased affinity of FMN under pH 5.0 is achieved by a significantly altered binding mode, which we sought to detect by spectroscopic analysis. Transient absorption spectroscopy is an excellent reporter for the electronic environment of chromophoric systems, and a pH dependent repositioning of the isoalloxazine moiety should be visible. Data were collected on free FMN in buffer, FMN:*Mt*Dod at pH 5.0, and FMN:*Mt*Dod at pH 7.5 after photoexcitation at 388 nm. Transient absorption spectra showed that lifetime signals of FMN bound to *Mt*Dod are dramatically shortened compared to free FMN. This is in line with the previously recorded data on *Hs*Dod and on *Halorhodospira halophila* dodecin (*Hh*Dod),[12] which have also shown strong lifetime quenching of RbF excited states, where two quenching pathways for archaeal and bacterial dodecins were found. The slower quenching mechanism (here described with τ_2) is the dominating one for bacterial dodecins, while the quenching mechanism via an electron transfer (τ_1) is the dominating one for archaeal dodecins. At first sight, the data do not show different properties in FMN deactivation at pH 7.5 ($\tau_1 = 1.70 \pm 0.10$ ps, $\tau_2 = 6.80 \pm 0.18$ ps) and pH 5.0 ($\tau_1 = 1.87 \pm 0.17$ ps, $\tau_2 = 6.6 \pm$

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

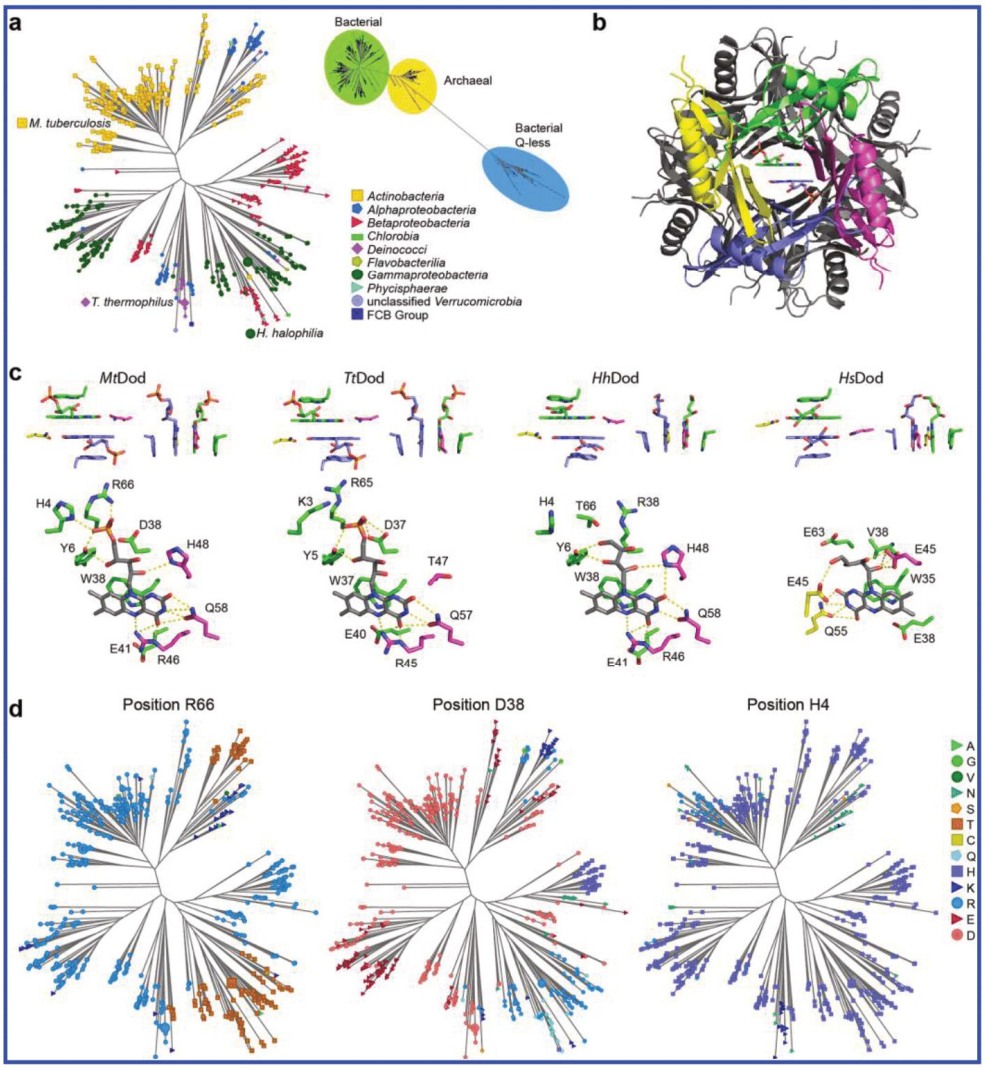


Figure 6. The dodecin protein family. (a) Phylogenetic tree of dodecins. Inset shows distribution of dodecins in three classes, of which archaeal and bacterial Q-less dodecins reflect smaller nodes. Q-less-dodecins do not show the isoalloxazine holding active site glutamine and may not be capable of flavin binding. The large bacterial (flavin binding) dodecin node (of 619 sequences; sequence identity higher than 46%) is shown in more detail. The three bacterial dodecins, solved in structure and all showing a conserved dodecameric fold, are highlighted. Data underlying this phylogenetic tree are available as Supporting Data "Metadata_phylogenetic-tree". (b) Cartoon illustration of *Mt*Dod with view one of the six binding pockets. Four of the 12 chains, which comprise a binding pocket, are colored. (c) Binding site architecture of *Mt*Dod (PDB 2YIZ[13]) and *Thermus thermophilus* dodecin (*Tt*Dod) (PDB 2V21[9]) with bound FMN and *Hh*Dod (PDB 2VXA[10]) and *Hs*Dod (PDB 2CCB[11]) with bound RbF. For all structures, the PDB_REDO optimized version was used.[32] Upper figures show bound flavins and key residues in flavin binding. In the bottom figures, binding sites are reduced to a C2-related part. Residues are in the color code of the chains shown in (b), except that in the bottom figures flavins are in gray for clarity. Except for *Hs*Dod V35, which interacts with its backbone, only side chains are shown. (d) Large bacterial node, as shown in (a), for representation of amino acid distribution in key positions. Positions are indicated for *Mt*Dod sequence; phylogenetic trees show, e.g., the complementary distribution of R66 and R38.

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

0.3 ps) indicating that the electronic environment of FMN is essentially unchanged under mildly basic and acidic conditions (Figure 5a,b), since the repopulation of the ground state is dominated by τ_2 and therefore rate determining. This essentially implies similar positioning of the chromophoric isoalloxazine moiety in FMN:*Mt*Dod complexes under mildly basic and acidic conditions. We note that we have recorded higher fluorescence quenching levels of flavins in binding assays at pH 5.0 (see Figure 3a) and by static spectroscopy (Figure S5).

Specificity of Bacterial Dodecins for FMN. The dodecin encoding gene is broadly spread in bacteria, including pathogens like *Pseudomonas aeruginosa*, *Streptococcus pneumonia*, and *M. tuberculosis*, all of which are representing major threats to human health. The bacterial dodecins are divided into several classes (Figure 6a; for sequences, see Supporting Data "Alignment"). Judged on the basis of the occurrence of a prominent tryptophan, clamping the isoalloxazine ring of flavins, most of these dodecins seem to have flavin binding properties.[7] Several flavin binding dodecins have been solved in the structure showing an overall conserved fold (Figure 6b). Bacterial dodecins align isoalloxazine rings in *si-si* orientation in an overall less compact arrangement of the aromatic tetrad than found in archaeal dodecins (on the example of *Hs*Dod) (Figure 6b,c). In most bacterial flavin binding dodecins, an arginine or a lysine residue can be identified at the binding pocket entry (R66 in *Mt*Dod), that may generally serve in FMN phosphate coordination (Figure 6c,d). The remaining flavin binding dodecins show threonine in this position but have arginine or lysine located in proximity (position homologous to D38 of *Mt*Dod; e.g., R38 in *Hh*Dod; see Figure 6d), which may take over FMN phosphate binding, implying specificity of bacterial dodecins for FMN. Additional residues interacting with the FMN phosphate group in *Mt*Dod and highly conserved in other flavin binding bacterial dodecins are H4 and Y6 (see Figure 6d). The specificity for RbF binding of archaeal dodecins may have evolved upon adaption of the protein to high-salt conditions (see Supporting Note 4).

Molecular Mechanism of Improved FMN Binding under Acidic Conditions. As an intriguing finding of this study, *Mt*Dod showed significantly improved binding behavior of FMN under acidic conditions (pH 5.0) (see Figure 3a). Kinetic analysis, disclosing the specifically promoted binding of FMN upon acidification (19- to 26-fold increase of k_{on} rate constants for the two independent protein preparations; see Supporting Data "Kinetic_single-step"), implies a steering of FMN into binding sites under these conditions. The rate of the dissociation of FMN:*Mt*Dod was essentially not affected by acidification (k_{off} rate constants are just slightly reduced), indicating the overall unchanged positioning of FMN in the binding pocket. As spectroscopic properties are sensitively reporting the electronic environment of a chromophoric system, any conformational change of FMN:*Mt*Dod, rearranging the binding pocket under acidic pH conditions, should be visible by transient spectroscopy. Data on FMN:*Mt*Dod at pH 5.0 did not show differences to data collected at pH 7.5, well reflecting the assumption from kinetic binding experiments that indeed the positioning of the FMN ring in the binding pocket, i.e., the chromophoric isoalloxazine, is not influenced upon acidification to pH 5.0 (see Figure 5a,b).

More detailed insight into the molecular basis of FMN binding and its pH dependence is possible when including the implications from the two-step fit and comparing FMN with RbF and FAD binding data. The overall affinity of RbF is not significantly affected by pH changes (pH 7.5 to 5; see Table S1). Just a slight slow down in binding dynamics of RbF (lower k_{on} and k_{off} rates) is observed at pH 5, and it remains to be speculated whether this effect is a direct consequence of, e.g., changed steering or whether simply a changed solubility accounts for deviations. In contrast to the minor changes for RbF binding, the pH dependence of *Mt*Dod in binding of FMN is pronounced, which identifies the phosphate group of FMN as involved in this process. With a pK_{a2} ($H-PO_4^-$–R, single protonated) value of 6.2,[30,31] the phosphate should be single deprotonated at pH 5.0 and double-deprotonated under mildly basic conditions. Two scenarios may be envisioned. First, the less negatively charged phosphate under pH 5.0 may cause less repulsive interaction between the phosphates of the bound FMN dimer, allowing a less restrained binding process. This is not supported by our data. The constrained modeling approach for the two step binding model does not deliver accurate absolute rates, and data should be just used for judging relative aspects in binding pocket dynamics. The rate constant k_{on_1} is high in relation to its "off" rate constant (k_{off_1}) at pH 5.0, as compared to the pH 7.5, indicating the higher stability of the single filled pocket under acidic conditions. The dynamics of the double filled binding pocket is similar at pH 5.0 and 7.5. This does not reflect a "scenario of a less repulsive interaction between FMN phosphates", which should rather act on rate constants in the second binding step that saturates the pocket. Further, also FAD shows an increased affinity at pH 5.0, in spite of the two phosphate groups (diphosphate) retaining protonation states between pH 7.5 and pH 5.0 (pK_{a2} value of 3.7 and pK_{a1} value of 2.1)[30] and keeping repulsion between FAD molecules unchanged. A more likely scenario does therefore account for the first binding event as being relevant for pH dependent interaction and focuses on the interactions of the FMN phosphate with the *Mt*Dod binding pocket. On the basis of the *Mt*Dod X-ray structure (PDB 2YIZ;[13] optimized from PDB_REDO[32]), two histidines are involved in FMN binding; H4 holds the phosphate in cooperation with R66, and H48 coordinates the ribityl chain (see Figure 6c). With a pK_a value of about 6 (imidazole ring), histidines change protonation states in the here studied pH range, and it is therefore plausible that protonated H4 and/or H48 can affect the binding process. To validate the impact of histidine H4 on FMN binding, we performed a thermal fluorescence assay with a set of H4-mutated *Mt*Dod (RbF and FMN saturated) at varying pH values, similarly as described above for wild-type protein (see Figure 2d–f). These data do not deliver information on the binding process itself but disclose the role of H4 in pH dependent properties. All FMN-saturated H4-mutated *Mt*Dod (H4A, H4D, H4N, and H4R) showed a lowered inflection point, which supports the role of H4 in FMN binding. In comparing inflection point vs pH profiles, the pH dependence was found most pronounced for wild-type protein (Figure 7a), indicating the contribution of H4 to the pH dependent binding properties of *Mt*Dod. Interestingly, the FMN:H4R–*Mt*Dod complex was less stable than the wild-type complex at acidic conditions (pH 4.0–6.5) but more stable above pH 7.0. This effect may be based at the preserved protonation of the arginine (of H4R) at pH 7.0–8.0 and the conservation of a stabilizing polar interaction under these conditions, while this interaction is lost for the wild-type *Mt*Dod (pK_a of 12.5 for arginine vs 6 for histidine). H4-mutated *Mt*Dod saturated with RbF did not significantly change inflection points within the tested range of

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082–1092

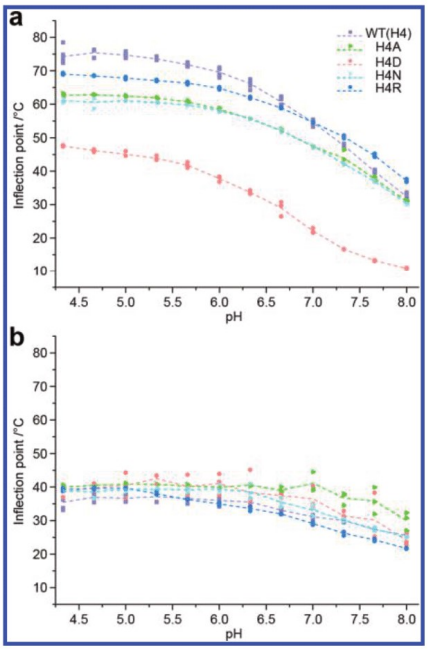


Figure 7. Thermal fluorescence assay of H4-mutated *Mt*Dod, with FMN (a) and RbF (b) as ligands. Thermal fluorescence assays were performed in the universal buffer system AAB (see also Figure 2d–f). While the H4 mutations have a small impact on the stability of the RbF holoproteins, the stability of FMN holoproteins is overall weakened, especially for the H4D mutation. Three data points are show for each condition, and the dashed lines connect mean values.

pH values, further supporting a role of the interaction of phosphate with H4 in pH dependent binding properties (Figure 7b). 529 of 619 bacterial dodecins also have histidine at the H4 equiv position (see Figure 6c,d), indicating that pH dependent binding of FMN may be a more widely distributed phenomenon in bacterial dodecins.

Biological Implications of *Mt*Dod Properties. *Mt*Dod appears as a well-tuned FMN buffering device that can take up and release FMN upon the need of the flavoproteome and in response to changing environmental conditions. Compared to flavoenzymes, where FMN and FAD (in general) behave similar to a prosthetic group and therefore rarely dissociate from enzymes,[33] the affinity of *Mt*Dod for FMN at pH 7.5 (see Figure 2a) is relatively moderate. A K_D of 118 nM at physiological pH conditions allows *Mt*Dod to still efficiently bind FMN, while the dynamics of binding provides the flavoproteome access to enough unbound FMN (see Supporting Data "Kinetics_two-step" and "Kinetic_single-step"). Kinetic analysis of the FMN binding shows that the half-life of the FMN:*Mt*Dod complex is in a range of a few seconds.

The idea of *Mt*Dod as a well-tuned FMN buffering device is further supported by kinetic data fitted to a two-step (sequential) binding mode. Here, a single filled binding pocket appears to be more dynamic (higher k_{on_1} and k_{off_1}) than a double filled binding pocket (lower k_{on_2} and k_{off_2}). The two-step binding mode enables *Mt*Dod (and likewise probably other dodecins) to respond to quickly changing FMN concentrations with the single filled binding pocket and to sequester FMN at persistently high concentrations with the double filled binding pocket. We assume that the normal *Mt*Dod mode of action is to respond quickly to fluctuations in FMN concentration by utilizing the fast dynamics of the single filled binding pocket, keeping the amount of free FMN relatively constant. Under the conditions of persistently high FMN levels, as it may for example be caused by stress induced degradation of flavoenzymes, *Mt*Dod shifts into another mode of action, in which it is sequestering FMN with its more stable double filled binding pocket. The difference of the rates of the two binding steps is significant, which may indicate a well tuned system (pH 7.5: k_{on_2} is 4-fold higher than k_{on_1} and k_{off_2} is almost 30-fold lower than k_{off_2}; average of both protein preparations; see Supporting Data "Kinetic_two-step").

Under acidic conditions, *Mt*Dod shows increased specificity and affinity for FMN. This is accomplished mainly by higher stability of the single filled binding pocket (K_{D_1} of 332 nM at pH 5.0 vs 7.9 μM at pH 7.5; calculated from rate constants; average of both protein preparations), which essentially drives formation of double filled binding pockets and turns *Mt*Dod into a FMN sequestering device.

While there is clear experimental evidence for pH dependent binding, a physiological relevance of this phenomenon is difficult to assess. Considering the lifestyle of *M. tuberculosis*, pH dependent binding properties may be beneficial when *M. tuberculosis* is encapsulated in the phagolysosome and granuloma.[16−19] However, *M. tuberculosis* is able to control its cellular pH in acidic in vitro environments (pH values of 4.5) or even in matured phagolysosomes.[34] Under conditions mimicking those in granuloma (pH 5.0, but including additional stress conditions), *M. tuberculosis* changes into a dormant state, in which the membrane/wall integrity is impaired as indicated by its lost acid-fast character.[17,35] In such a scenario, leading to lowered cellular pH, *Mt*Dod can then turn into an efficient FMN-sequestering device. The flavoproteome may then also tolerate high-affinity binding by *Mt*Dod, as *M. tuberculosis* is in a dormant state and at minimal metabolic activity. A reservoir of FMN, accumulated by *Mt*Dod, may finally be reconstituted to the flavoproteome under physiological conditions (mildly basic pH), allowing a fast return to metabolic activity.

Systems biological studies have identified *Mt*Dod as not essential[36] but are pointing toward a role of *Mt*Dod under acidification. *Mt*Dod is 2.8 times more abundant than the average protein in *M. tuberculosis* (*Mt*Dod concentration of 4.7 μM, calculated from ref 37 at exponential growth) and up-regulated under acidic conditions (1.85-fold at pH 6.5 and 1.83-fold at pH 5.5)[19,38] as well as under starvation (about 1.3-fold).[39] Interestingly, *Mt*Dod was found to be up-regulated in a multidrug resistant strain of *M. tuberculosis* compared to the strain H37Rv.[40] It should also be noted that *Mt*Dod has been found in the medium in liquid cultures.[41,42] Since a signal sequence for exporting *Mt*Dod across the membrane is absent, we speculate that these findings may rather be based on autolysis of *M. tuberculosis* under these conditions and the stability of *Mt*Dod in the culture medium.[43] It is unlikely that *Mt*Dod acts as extracellular flavin scavenger.

In summary, we suggest that *Mt*Dod can fulfill two roles: First, under beneficial growth conditions, it is a dynamic and

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

moderately affine FMN buffer that exchanges FMN with the flavoproteome.[44] Second, under stress conditions, *Mt*Dod can turn into a more tightly FMN-binding storage device. *Mt*Dod can respond to two stress conditions: An increased FMN concentration leads to kinetically more stable double filled binding pockets, and acidified cellular conditions, as putatively occurring in a dormant state, lead to a several fold increased FMN affinity. Binding FMN under stress maintains the valuable cofactor as resource within the cell and protects the cell from unwanted side reactions that may occur from free FMN.[10,45] In light of the suggested role of *Mt*Dod as an important factor in FMN distribution, we encourage the collection of further data. For example, data on cellular *Mt*Dod and free flavin concentrations is required to assess a physiological relevance of the observed different dynamics in binding properties (single filled vs double filled). We particularly encourage dodecin knockout studies in *M. tuberculosis* or other suited bacterial species. Owing to a proposed function as flavin storage and sequestering protein, dodecins may be involved in flavin homeostasis and broadly interfere in metabolic properties. As a putative factor for *M. tuberculosis* in enduring and surviving hostile conditions, *Mt*Dod may be a relevant target in combinatorial antibiotic treatment.

METHODS

Protein Preparation. The *Mt*Dod encoding gene was expressed in *Escherichia coli* under the control of a T7 promotor from pET22b plasmid construct (Figure S6). Harvested cells were broken by French Press, and *Mt*Dod was purified via a heat denaturation/$(NH_4)_2SO_4$ protocol. For apoprotein preparation, *Mt*Dod was denatured for cofactor depletion and then refolded and polished by SEC. Several hundred milligrams of *Mt*Dod were received per liter of culture. To obtain FMN:*Mt*Dod, apo-*Mt*Dod was incubated overnight at 4 °C with FMN [F6750; Sigma; 70% pure; free RbF ≤ 6.0%] in excess (above the solubility), and unbound FMN was removed by SEC. Proteins were generally analyzed in 300 mM NaCl, 5 mM $MgCl_2$, and 20 mM Tris-HCl (pH 7.5, 300N5M20T7.5) or 300 mM NaCl, 5 mM $MgCl_2$ and 20 mM acetate (pH 5.0, 300N5M20A5.0).

Protein Analysis. TSA data, with the protocols termed thermal fluorescence and thermocyclic fluorescence assay, were collected in 25 μL volumes (mixture of flavin:*Mt*Dod in the respective buffer) in 96 well plates with excitation/emission filter bandwidths of 450−490/560−580 nm. Buffers contained 300 mM NaCl, 5 mM $MgCl_2$, and 23.8 mM of the respective buffer reagent. The temperature was raised from 5 to 95 °C in 0.5 °C steps every 30 s (or 1.0 °C steps every 60 s; see Supporting Materials and Methods). The thermocyclic fluorescence assay protocol was performed by stepwise increasing temperature from 5 to 95 °C. Here, each step contained a heating phase for 6 min and a cooling to 5 °C phase for 30 min.

Chemical reduction of FMN:*Mt*Dod was performed with sodium dithionite. Reduced samples were filtered (about 1 h), and FMN concentrations were read out at 450 nm. Spectroelectrochemical data were collected with a Ag/AgCl/$KCl_{(300\ mM)}$ reference electrode with a redox potential of +262 mV vs NHE and Pt wires as counter electrodes. During the spectroelectrochemical measurements, the applied potential values were varied in the range between −100 mV vs NHE and −350 mV vs NHE. In measurements with FMN:*Mt*Dod, the equilibration time was varied for the reductive half cycle between 2 and 75 min and for the oxidative half cycle between 2 and 90 min. All measurements were conducted in 20 mM Tris-HCl buffer at pH 7.5 containing 300 mM KCl and 5 mM $MgCl_2$.

Transient absorption (TA) experiments were performed by using a Clark MXR-CPA-2110 laser system with a central wavelength of 775 nm, a repetition rate of 1 kHz, and a pulse length of 150 fs. Pump pulses at 388 nm were realized by second harmonic generation of the laser fundamental in a BBO crystal. The white light was split for referenced measurements and recorded by 63-channel photodiode arrays [Hamamatsu]. Samples were measured in quartz cuvettes of 1 mm optical path length. Analysis of data of the time-resolved measurements was carried out with the computer program OPTIMUS.[46] Steady-state fluorescence spectra were measured using a Jasco FP-8500 spectrometer. For further information concerning the experimental setup, see ref 47.

Flavin Binding Assays. Equilibrium binding data were collected in volumes of 100 μL of mixtures of flavins (RbF [R4500; Sigma], FMN [F2253; Sigma], and FAD [F6625; Sigma], used without further purification) and apo-*Mt*Dod at varying concentrations in 96 well plates. Flavins were used at concentrations of 1 μM. The fully pipetted plate was incubated for 1 h at room temperature in the dark. Flavin binding was recorded by reading out unbound flavin by excitation at 445 ± 9 nm and emission at 520 ± 20 nm.

In kinetic measurements, *Mt*Dod concentrations were varied (0.0 (only flavin), 0.6, 0.8, 1.0, 0.2, and 1.4 μM) while the flavin concentration was fixed at 1 μM. Measurements started by addition of apo-MtDod to flavin solutions (delay of the instrument was 3.66 s). Fluorescence was measured after excitation at 445 ± 9 nm and emission at 520 ± 20 nm. Data were fitted with a single step binding model (see Supporting Materials and Methods function (3−5), based on ref 48) with Origin Pro 8.6 [OriginLab corporation]. For the numerical two-step binding model fit, data of a respective ratio were averaged separately. The numerical two-step binding model fit was performed in Excel [Microsoft] using the solver function by minimizing the mean square error between calculated and measured data.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsinfecdis.7b00237.

Supporting material and methods; Figure S1: different binding models; Figure S2: FAD binding data; Figure S3: kinetic data of *Mt*Dod and *Hs*Dod flavin binding; Figure S4: kinetic data of flavin binding by *Mt*Dod fitted by a two-step binding model; Figure S5: fluorescence spectra of FMN:*Mt*Dod at pH 7.5 and pH 5.0; Figure S6: plasmid map; Table S1: dissociation constants of fluorescence based binding assays; Supporting Note 1: binding properties of *Mt*Dod; Supporting Note 2: comparison of binding of dodecin and kinetic binding data of *Hs*Dod; Supporting Note 3: redox properties of FMN:*Mt*Dod; Supporting Note 4: archaeal dodecins, restricted to phylogenetic class Haloarchaea (PDF)

Excel Sheet with (i) "Kinetic_single-step"; (ii) Kinetic_two-step"; (iii) "Alignment"; (iv) "Metadata_phylogenic-tree" (XLSX)

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

AUTHOR INFORMATION

Corresponding Author

*E-mail: grininger@chemie.uni-frankfurt.de.

ORCID

Josef Wachtveitl: 0000-0002-8496-8240

Martin Grininger: 0000-0002-7269-0667

Author Contributions

F.B. cloned protein expression constructs and prepared dodecins in apo- and holo-state. F.B. further characterized dodecin in thermal and thermocyclic fluorescence assays, flavin binding assays, and chemical reductions. F.B. further established protocols, designed experiments, and analyzed data. C.A.H. and F.S. measured and analyzed the spectroscopic properties of dodecin. S.V. measured and analyzed the spectroelectrochemical properties of dodecin. G.N. analyzed redox data and discussed binding data. J.W. analyzed spectroscopic properties. M.G. analyzed data and designed research. M.G. and F.B. wrote the paper. C.A.H., S.V., J.W., and G.N. contributed to writing the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

M.G. is grateful to Dieter Oesterhelt for continuous support and discussing the manuscript. F.B. and M.G. thank Mike Dyall-Smith and Monica Hagedorn for discussions on dodecin function and Ines Gößner for assistance in starting the project. Furthermore, F.B. and M.G. thank Ernst Stelzer for support in binding assays and TSA (use of platereader and RT·PCR machine). This work was supported by a Lichtenberg Grant of the Volkswagen Foundation to M.G. (grant number 85 701). S.V. thanks the Konrad-Adenauer-Stiftung for a Ph.D. scholarship.

ABBREVIATIONS

AAB, universal buffer system: acetic acid, ADA, and bicine; ADP, adenosine 5′-triphosphate; ATP, adenosine 5′-diphosphate; CoA, coenzyme A; equ., equilibrium measurements; FAD, flavin adenine dinucleotide; FAD:*Mt*Dod, holo complex of *Mt*Dod and FAD; FMN, flavin mononucleotide; FMN:*Mt*Dod, holo complex of *Mt*Dod and FMN; $FMNH^-$, reduced and single deprotonated flavin mononucleotide; H4X-*Mt*Dod, X: A D N R; *Mt*Dod mutation at position 4 (His); *Hh*Dod, *Halorhodospira halophila* dodecin; *Hs*Dod, *Halobacterium salinarum* dodecin; K_D, dissociation constant; K_{D_1}; k_{on_1}; k_{off_1}, constants for the first binding step, single filled pocket; K_{D_2}; k_{on_2}; k_{off_2}, constants for the second binding step, double filled pocket; kin., kinetic measurements; LmC, lumichrome; *M. tuberculosis*, *Mycobacterium tuberculosis*; MMT, universal buffer system: malic acid, MES, and Tris; *Mt*Dod, *Mycobacterium tuberculosis* dodecin; NHE, standard hydrogen electrode; RbF, riboflavin; RbF:*Mt*Dod, holo complex of *Mt*Dod and RbF; SEC, size exclusion chromatography; TB, tuberculosis; TSA, thermal shift assay; *Tt*Dod, *Thermus thermophilus* dodecin

REFERENCES

(1) Lienhart, W.-D., Gudipati, V., and Macheroux, P. (2013) The human flavoproteome. *Arch. Biochem. Biophys. 535*, 150−162.

(2) Massey, V. (2000) The Chemical and Biological Versatility of Riboflavin. *Biochem. Soc. Trans. 28*, 283−296.

(3) Merrill, A. H., Lambeth, Edmondson, D. E., and McCormick, D. B. (1981) Formation and Mode of Action of Flavoproteins. *Annu. Rev. Nutr. 1*, 281−317.

(4) White, H. B., Armstrong, J., and Whitehead, C. C. (1986) Riboflavin-binding protein. Concentration and fractional saturation in chicken eggs as a function of dietary riboflavin. *Biochem. J. 238*, 671−675.

(5) Foraker, A. B., Khantwal, C. M., and Swaan, P. W. (2003) Current perspectives on the cellular uptake and trafficking of riboflavin. *Adv. Drug Delivery Rev. 55*, 1467−1483.

(6) Bieger, B., Essen, L.-O., and Oesterhelt, D. (2003) Crystal Structure of Halophilic Dodecin: A Novel, Dodecameric Flavin Binding Protein from *Halobacterium salinarum*. *Structure 11*, 375−385.

(7) Arockiasamy, A., Aggarwal, A., Savva, C. G., Holzenburg, A., and Sacchettini, J. C. (2011) Crystal structure of calcium dodecin (Rv0379), from *Mycobacterium tuberculosis* with a unique calcium-binding site. *Protein Sci. Publ. Protein Soc. 20*, 827−833.

(8) Liu, F., Xiong, J., Kumar, S., Yang, C., Ge, S., Li, S., Xia, N., and Swaminathan, K. (2011) Structural and biophysical characterization of *Mycobacterium tuberculosis* dodecin Rv1498A. *J. Struct. Biol. 175*, 31−38.

(9) Meissner, B., Schleicher, E., Weber, S., and Essen, L. O. (2007) The dodecin from *Thermus thermophilus*, a bifunctional cofactor storage protein. *J. Biol. Chem. 282*, 33142−33154.

(10) Grininger, M., Staudt, H., Johansson, P., Wachtveitl, J., and Oesterhelt, D. (2009) Dodecin Is the Key Player in Flavin Homeostasis of Archaea. *J. Biol. Chem. 284*, 13068−13076.

(11) Grininger, M., Zeth, K., and Oesterhelt, D. (2006) Dodecins: A Family of Lumichrome Binding Proteins. *J. Mol. Biol. 357*, 842−857.

(12) Staudt, H., Oesterhelt, D., Grininger, M., and Wachtveitl, J. (2012) Ultrafast excited-state deactivation of flavins bound to dodecin. *J. Biol. Chem. 287*, 17637−44.

(13) Vinzenz, X., Grosse, W., Linne, U., Meissner, B., and Essen, L.-O. (2011) Chemical engineering of *Mycobacterium tuberculosis* dodecin hybrids. *Chem. Commun. 47*, 11071−11073.

(14) Macheroux, P., Kappes, B., and Ealick, S. E. (2011) Flavogenomics − a genomic and structural view of flavin-dependent proteins. *FEBS J. 278*, 2625−2634.

(15) Dresen, C., Lin, L. Y.-C., D'Angelo, I., Tocheva, E. I., Strynadka, N., and Eltis, L. D. (2010) A Flavin-dependent Monooxygenase from *Mycobacterium tuberculosis* Involved in Cholesterol Catabolism. *J. Biol. Chem. 285*, 22264−22275.

(16) Rohde, K., Yates, R. M., Purdy, G. E., and Russell, D. G. (2007) *Mycobacterium tuberculosis* and the environment within the phagosome. *Immunol. Rev. 219*, 37−54.

(17) Deb, C., Lee, C.-M., Dubey, V. S., Daniel, J., Abomoelak, B., Sirakova, T. D., Pawar, S., Rogers, L., and Kolattukudy, P. E. (2009) A Novel In Vitro Multiple-Stress Dormancy Model for *Mycobacterium tuberculosis* Generates a Lipid-Loaded, Drug-Tolerant, Dormant Pathogen. *PLoS One 4*, e6077.

(18) Dannenberg, A. M. (2006) *Pathogenesis of Human Pulmonary Tuberculosis: Insights from the Rabbit Model*, 1st. ed., pp 1−453, ASM Press, Washington, DC.

(19) Baker, J. J., Johnson, B. K., and Abramovitch, R. B. (2014) Slow growth of *Mycobacterium tuberculosis* at acidic pH is regulated by phoPR and host-associated carbon sources. *Mol. Microbiol. 94*, 56−69.

(20) Kiran, D., Podell, B. K., Chambers, M., and Basaraba, R. J. (2016) Host-directed therapy targeting the *Mycobacterium tuberculosis* granuloma: a review. *Semin. Immunopathol. 38*, 167−183.

(21) Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R., and Mattevi, A. (2009) ThermoFAD, a Thermofluor®-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J. 276*, 2833−2840.

(22) Alberty, R. A., and Hammes, G. G. (1958) Application of the Theory of Diffusion-controlled Reactions to Enzyme Kinetics. *J. Phys. Chem. 62*, 154−159.

(23) Nöll, G. (2008) Spectroscopic investigation of flavoproteins: Mechanistic differences between (electro)chemical and photochemical reduction and oxidation. *J. Photochem. Photobiol., A 200*, 34−38.

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

(24) Vogt, S., Schneider, M., Schäfer-Eberwein, H., and Nöll, G. (2014) Determination of the pH Dependent Redox Potential of Glucose Oxidase by Spectroelectrochemistry. *Anal. Chem.* *86*, 7530−7535.
(25) Clark, W. M., and Lowe, H. J. (1956) Studies on oxidation-reduction. XXIV. Oxidation-reduction potentials of flavin adenine dinucleotide. *J. Biol. Chem.* *221*, 983−992.
(26) Draper, R. D., and Ingraham, L. L. (1968) A potentiometric study of the flavin semiquinone equilibrium. *Arch. Biochem. Biophys.* *125*, 802−808.
(27) Nöll, G., Trawöger, S., von Sanden-Flohe, M., Dick, B., and Grininger, M. (2009) Blue-Light-Triggered Photorelease of Active Chemicals Captured by the Flavoprotein Dodecin. *ChemBioChem* *10*, 834−837.
(28) Gutiérrez Sánchez, C., Su, Q., Schönherr, H., Grininger, M., and Nöll, G. (2015) Multi-Ligand-Binding Flavoprotein Dodecin as a Key Element for Reversible Surface Modification in Nano-biotechnology. *ACS Nano* *9*, 3491−3500.
(29) Staudt, H., Hoesl, M. G., Dreuw, A., Serdjukow, S., Oesterhelt, D., Budisa, N., Wachtveitl, J., and Grininger, M. (2013) Directed Manipulation of a Flavoprotein Photocycle. *Angew. Chem., Int. Ed.* *52*, 8463−8466.
(30) Bidwell, J., Thomas, J., and Stuehr, J. (1986) Thermodynamic and kinetic study of the interactions of nickel(II) with FMN and FAD. *J. Am. Chem. Soc.* *108*, 820−825.
(31) Sigel, H., Song, B., Liang, G., Halbach, R., Felder, M., and Bastian, M. (1995) Acid-base and metal ion-binding properties of flavin mononucleotide (FMN^{2-}). Is a 'dielectric' effect responsible for the increased complex stability? *Inorg. Chim. Acta* *240*, 313−322.
(32) Joosten, R. P., Long, F., Murshudov, G. N., and Perrakis, A. (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* *1*, 213−220.
(33) Walsh, C. T., and Wencewicz, T. A. (2013) Flavoenzymes: Versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* *30*, 175−200.
(34) Vandal, O. H., Pierini, L. M., Schnappinger, D., Nathan, C. F., and Ehrt, S. (2008) A membrane protein preserves intrabacterial pH in intraphagosomal *Mycobacterium tuberculosis*. *Nat. Med.* *14*, 849−854.
(35) Velayati, A. A., Abeel, T., Shea, T., Konstantinovich Zhavnerko, G., Birren, B., Cassell, G. H., Earl, A. M., Hoffner, S., and Farnia, P. (2016) Populations of latent *Mycobacterium tuberculosis* lack a cell wall: Isolation, visualization, and whole-genome characterization. *Int. J. Mycobacteriology* *5*, 66−73.
(36) Griffin, J. E., Gawronski, J. D., Dejesus, M. A., Ioerger, T. R., Akerley, B. J., and Sassetti, C. M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* *7*, e1002251.
(37) Schubert, O. T., Ludwig, C., Kogadeeva, M., Zimmermann, M., Rosenberger, G., Gengenbacher, M., Gillet, L. C., Collins, B. C., Röst, H. L., Kaufmann, S. H. E., Sauer, U., and Aebersold, R. (2015) Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* *18*, 96−108.
(38) Rohde, K. H., Abramovitch, R. B., and Russell, D. G. (2007) *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell Host Microbe* *2*, 352−64.
(39) Albrethsen, J., Agner, J., Piersma, S. R., Højrup, P., Pham, T. V., Weldingh, K., Jimenez, C. R., Andersen, P., and Rosenkrands, I. (2013) Proteomic Profiling of *Mycobacterium tuberculosis* Identifies Nutrient-starvation-responsive Toxin−antitoxin Systems. *Mol. Cell. Proteomics* *12*, 1180−1191.
(40) Zhang, Y., Zheng, G., Wang, Y., Chen, J., Zhu, C., Liu, R., Peng, Z., Li, Q., and Xing, L. (2012) Screening and comparison of differentially expressed genes between one MDR-TB strain and the virulent *M. tuberculosis* H37Rv. *Gene* *506*, 223−229.
(41) Mattow, J., Schaible, U. E., Schmidt, F., Hagens, K., Siejak, F., Brestrich, G., Haeselbarth, G., Müller, E.-C., Jungblut, P. R., and Kaufmann, S. H. E. (2003) Comparative proteome analysis of culture supernatant proteins from virulent *Mycobacterium tuberculosis* H37Rv and attenuated *M. bovis* BCG Copenhagen. *Electrophoresis* *24*, 3405−3420.
(42) de Souza, G. A., Leversen, N. A., Malen, H., and Wiker, H. G. (2011) Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *J. Proteomics* *75*, 502−10.
(43) Tullius, M. V., Harth, G., and Horwitz, M. A. (2001) High Extracellular Levels of *Mycobacterium tuberculosis* Glutamine Synthetase and Superoxide Dismutase in Actively Growing Cultures Are Due to High Expression and Extracellular Stability Rather than to a Protein-Specific Export Mechanism. *Infect. Immun.* *69*, 6348−6363.
(44) Martínez-Limón, A., Alriquet, M., Lang, W.-H., Calloni, G., Wittig, I., and Vabulas, R. M. (2016) Recognition of enzymes lacking bound cofactor by protein quality control. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 12156−12161.
(45) Linster, C. L., Van Schaftingen, E., and Hanson, A. D. (2013) Metabolite damage and its repair or pre-emption. *Nat. Chem. Biol.* *9*, 72−80.
(46) Slavov, C., Hartmann, H., and Wachtveitl, J. (2015) Implementation and Evaluation of Data Analysis Strategies for Time-Resolved Optical Spectroscopy. *Anal. Chem.* *87*, 2328−2336.
(47) Slavov, C., Bellakbil, N., Wahl, J., Mayer, K., Rück-Braun, K., Burghardt, I., Wachtveitl, J., and Braun, M. (2015) Ultrafast coherent oscillations reveal a reactive mode in the ring-opening reaction of fulgides. *Phys. Chem. Chem. Phys.* *17*, 14045−14053.
(48) Peuker, S., Cukkemane, A., Held, M., Noé, F., Kaupp, U. B., and Seifert, R. (2013) Kinetics of Ligand-Receptor Interaction Reveals an Induced-Fit Mode of Binding in a Cyclic Nucleotide-Activated Protein. *Biophys. J.* *104*, 63−74.

DOI: 10.1021/acsinfecdis.7b00237
ACS Infect. Dis. 2018, 4, 1082−1092

Supporting Information

Flavin storage and sequestration by *Mycobacterium tuberculosis* dodecin

Authors: Florian Bourdeaux [a], Christopher A. Hammer [b], Stephan Vogt [c], Felix Schweighöfer [b], Gilbert Nöll [c], Josef Wachtveitl [b], Martin Grininger [a]*

Author affiliations:
(a) Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Cluster of Excellence for Macromolecular Complexes, Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt am Main, Germany
(b) Institute of Physical and Theoretical Chemistry, Johann Wolfgang Goethe-University, Cluster of Excellence for Macromolecular Complexes, Max-von-Laue-Str. 7, D-60438 Frankfurt, Germany
(c) Nöll Junior Research Group, Organic Chemistry, University of Siegen, Adolf-Reichwein-Str. 2, D-57068 Siegen, Germany

*Corresponding author: grininger@chemie.uni-frankfurt.de

Contents:

Supporting Material and Methods:

Determination flavin and protein concentration:

Protein and flavin concentrations were determined via absorption spectroscopy performed in quartz glass cuvettes at a nano drop spectrophotometer [NanoDrop 2000C Thermo Scientific]. An extinction coefficient for *Mt*Dod (apo-protein) of 6970 $L \times M^{-1} \times cm^{-1}$ was calculated via CLC Main Workbench Version 6.9.1. For flavins, multiple wavelengths were used to obtain the concentration (finally using the mean value). The concentrations of RbF, FMN and FAD were determined with the absorption at 375 nm and 450 nm with average extinction coefficients of 10,000 $L \times M^{-1} \times cm^{-1}$ and 12,000 $L \times M^{-1} \times cm^{-1}$.[1] For the FMN and FAD, in addition, the absorption at 473 nm and the extinction coefficient 9,200 $L \times M^{-1} \times cm^{-1}$ was used.[2] The difference of mean extinction coefficients to specific coefficients published is less than 5%.[3]

Expression of *Mt*Dod (wildtype and H4-mutants):

pET22b-*Mt*Dod (see Supporting Figure S5) was transformed into BL21 (DE3) Gold cells and plated onto LB-agar plates containing 100 ng/µL ampicillin. 20 mL LB medium with 100 ng/µL ampicillin were inoculated with 4-5 overnight grown colonies and incubated at 37 °C 180 rpm overnight. 2 L TB medium with 100 ng/µL ampicillin were inoculated with 20-35 mL overnight LB culture, and incubated at 37 °C 180 rpm until the OD600 reached about 1.0. The culture was cooled to 20-30 °C, and the expression was induced with 2 mL 1 M IPTG solution. The culture was incubated overnight at 20-30 °C 180 rpm. Cells were harvested at 4000 rcf and frozen in liquid nitrogen.

Purification of crude *Mt*Dod (wildtype and H4-mutants):

Frozen or fresh cell pellets were resuspended in 300 mM NaCl, 5 mM $MgCl_2$ and 20 mM Tris-HCl pH 7.5 (300N5M20T7.5) buffer containing DNAseI [Agilent] and phenylmethylsulfonylfluoride PMSF [Fluka] were lysed with a French pressure cell press at about 96.5 x 10^3 kPa. The cytosol was obtained by centrifugation for 1 h at 50000 rcf 4 °C. The cytosol was further incubated at 70-95 °C for 25 min, and centrifuged at 18600 rcf for 30 min 4 °C. *Mt*Dod was precipitated twice with a final $(NH_4)_2SO_4$ concentration of 40% and resolved in 300N5M20T7.5 buffer. In an alternative protocol, 50% of DMSO or 75% MeOH (for H4-mutants) was used to precipitate *Mt*Dod. The crude *Mt*Dod was stored at -80 °C as a pellet frozen in liquid nitrogen or directly purified further.

Denaturing and refolding of apo-*Mt*Dod (wildtype and H4-mutants):

Guanidinium chloride was added to a solution of crude *Mt*Dod until the final concentration of 6 M was reached. The solution was dialyzed against its 50-100-fold volume in 6 M guanidinium chloride at 20 mM Tris-HCl pH 8.0 at room temperature to 60 °C until the solution was colorless. Dialysis was performed over night and repeated twice. Denatured *Mt*Dod solution was used directly for refolding, or was stored at -80 °C. For refolding, denatured *Mt*Dod was diluted with 6 M guanidinium chloride buffer to *Mt*Dod concentrations of 1-2 mg/mL. In addition L-arginine was added to a final concentration of 1 M. For refolding, the solution was then dialyzed against a 80-100-fold volume of the respective buffer at 4 °C without stirring. Dialysis was repeated twice with more rigorous stirring. The refolded protein solution was further purified by SEC, with a Superdex 200 16/60 or 26/60 column equilibrated with the respective buffer. The pooled fractions of apo-*Mt*Dod were stored a 4 °C.

Flavin:*Mt*Dod (wildtype and H4-mutants):

To obtain FMN:*Mt*Dod, apo-*Mt*Dod was incubated over night at 4 °C with FMN [F6750; Sigma] in excess (above the solubility). The unbound FMN was removed by SEC as described above and the purified FMN:*Mt*Dod was stored at 4 °C in the dark. For flavin:H4-mutants, RbF:*Mt*Dod and FAD:*Mt*Dod, apo-*Mt*Dod (or apo-mutants) was mixed with the equal amount of the corresponding flavin and incubated at least 1.5 h at about 5 °C or on ice.

Thermal fluorescence assay:

21 µL of the respective buffer and 4 µL of a 25 µM flavin:*Mt*Dod solution were pipetted into a white 96 well plate [HSP9655; Bio-Rad or MLL9601; Bio-Rad, for the universal buffer system measurements]. Each condition was measured in triplicate, or as otherwise stated. The plate was sealed with optical tape [iCycler iQ; Bio-Rad] and centrifuged at 3000 rcf. All steps were carried out on ice or at 4 °C. The measurement was performed in a real-time PCR instrument [C1000™ Thermal Cycler and CFX96™ Real-Time System; Bio-Rad] with excitation/emission filter bandwidth of 450-490/560-580 nm. The temperature was raised from 5 °C to 95 °C in 1.0 °C steps every 60 sec or 0.5 °C steps every 30 sec (universal buffer system measurements). One data point was taken at every step. Buffers contained 300 mM NaCl, 5 mM $MgCl_2$ and 23.8 mM the respective buffer reagent. pH-Values were screened with the following buffers: pH 4.0, citrate; pH 5.0, acetate; pH 6.0-7.5 in 0.5 steps, phosphate; pH 7.5-9.0 in 0.5 steps, Tris-HCl. For the universal buffer systems, following reagents were used: AAB: acetic acid, ADA and bicine. MMT: malic acid, MES and Tris. Inflection points were determined with

assuming a Boltzmann sigmoidal curve and fitting against function (1). $F_{highest}$ is the value of the observed fluorescence maximum and F_{lowest} the minimum and T_{Inf} is the inflection point of the temperature curve.

$$F = F_{lowest} + \frac{\left(F_{highest} - F_{lowest}\right)}{1 + e^{\frac{\left(T_{Inf} - T_{step}\right)}{slope}}} \quad (1)$$

Thermocyclic fluorescence assay:

For data collection, 20 µL 100 µM apo-*Mt*Dod solution was mixed with 105 µL 19 µM FMN solution in the corresponding buffer and incubated at 4 °C over night. Buffers contained 300 mM NaCl, 5 mM $MgCl_2$ and 23.8 mM the respective buffer reagent. pH-Values were screened with the following buffers: pH 4.0, citrate; pH 5.0, acetate; pH 5.0, citrate; pH 6.0-7.0 in 0.5 steps, phosphate; pH 7.5, Tris-HCl. 25 µL of the *Mt*Dod solution was pipetted into a clear 96 well plate [MLL9601; Bio-Rad]. The plate was sealed with optical tape [iCycler iQ®; Bio-Rad] and centrifuged at 3000 rcf. All steps were carried out on ice or at 4 °C. The measurement was performed in a real-time PCR instrument [C1000™ Thermal Cycler and CFX96™ Real-Time System; Bio-Rad] with excitation/emission filter bandwidth of 450-490/560-580 nm. The temperature was raised stepwise from 5 °C to 95 °C, till 50 °C the step size was 4.5 °C and then reduced to 2.0 °C. Each cycle contained a heating phase for 6 min and a cooling (5 °C) phase for 30 min. After each phase a measurement was taken. The complete temperature protocol was applied to every sample.

Chemical reduction of FMN:*Mt*Dod:

For pH 7.5 and pH 5.0 a FMN:*Mt*Dod stock solution was diluted with the corresponding buffer to an absorption of about 0.7 (1 cm path length) at 450 nm (about 59 µM) and a 1 M sodium dithionite solutions with the respective degassed buffer was freshly prepared. The absorption at 450 nm of the FMN:*Mt*Dod was measured in cuvettes and subsequently the sodium dithionite solution was added to a final concentration of 20 mM (40 mM for one sample). Samples were incubated for 30 min at room temperature and the absorption at 450 nm was measured. Reduced samples were transferred in 30 kDa cutoff centrifugal filters [Nanosep® Pall Corporation] and centrifuged at 14 x 10^3 rcf until fully filtered (about 1 h). The absorption at 450 nm of the re-oxidized filtrate was measured, re-oxidation already took place during the filtration.

Spectroelectrochemical determination of protein redox potentials:

The spectroelectrochemical setup used in this study has been described elsewhere.[4] Prior to measurements, the gold capillary working electrode was cleaned with piranha solution (H_2O_2:H_2SO_4 at a ratio of 1:3; CAUTION: This solution is highly corrosive and should be handled with care) and incubated overnight with 1 mM mercaptopropionic acid. A Ag/AgCl/$KCl_{(300\ mM)}$ electrode with a redox potential of +262 mV *vs.* NHE in the employed buffer served as reference electrode. As counter electrodes, Pt wires were used. For data collection, the protein solution has to be transferred into the (spectro)electrochemical cell, and the protein absorption spectra are collected depending on the electrochemical potential. From the potential-dependent spectroscopic changes also the corresponding protein redox potentials can be determined. During the spectroelectrochemical measurements the applied potential values were varied in the range between -100 mV *vs.* NHE and -350 mV *vs.* NHE. Before and after the measurements of FMN:*Mt*Dod the redox potential of free FMN was determined for calibration of the setup. For free FMN a period of 2 min after each potential step was sufficient to reach electrochemical equilibrium.

During the measurements with FMN:*Mt*Dod the equilibration time was varied for the reductive half cycle between 2 and 75 min and for the oxidative half cycle between 2 and 90 min. All measurements were conducted in 20 mM Tris-HCl buffer at pH 7.5 containing 300 mM KCl and 5 mM $MgCl_2$. The titration curves were obtained by plotting the absorption at 454 nm *vs.* the corresponding potential value.

Flavin binding assay:

For the assays, 2 µM solutions of flavins (RbF [R4500; Sigma], FMN [F2253; Sigma] and FAD [F6625; Sigma], used without further purification) in the corresponding buffer were freshly prepared. Based on a 20 µM stock solution, 24 solutions ranging from apo-*Mt*Dod concentrations of 0-16 µM in the corresponding buffer were prepared. Data at pH 5.0 were collected in 300 mM NaCl, 5 mM $MgCl_2$ and 20 mM acetate pH 5.0 (300N5M20A5.0). Data at pH 7.5 were collected in buffer 300N5M20T7.5. For the measurements, 50 µL of the respective flavin solution and 50 µL of the corresponding apo-*Mt*Dod solution were pipetted into each well of a black 96 well plate [655076; Greiner bio-one, these plates have only a minor impact on the equilibrium measurement but cause problems in the kinetic measurements]. For each condition, data were collected in triplicate or quadruplicate (only three data points were taken for the four conditions at highest protein concentration). Two buffer blanks and two 20 µM apo-*Mt*Dod blanks were used for the background correction. Apo-*Mt*Dod did not show a significant change of the background. The fully pipetted plate was incubated 1 h at

room temperature in the dark. The fluorescence was measured 4 times per well and averaged in a plate reader [Infinite® 200; TECAN] with excitation at 445 ± 9 nm and emission at 520 ± 20 nm. For the calculation of the dissociation constant, the 4 replicates were averaged, corrected by the background and normalized with the 0 µM apo-*Mt*Dod measurement signal. For the fitting of the data, function (2) was used. [A_0] (overall flavin concentration) and F_{max} (fluorescence of unbound flavin) were fixed to 1 µM and 100%. Other variables are: [D_0] (overall dodecin concentration) and [F_∞] (fluorescence of bound flavin). Fitting was performed with Origin Pro 8.6 [OriginLab corporation] without any weighting of the data points.

$$F = F_{max} - \frac{F_{max} - F_\infty}{[A_0]}\left(\frac{[D_0] + [A_0] + K_d - \sqrt{([D_0] + [A_0] + K_d)^2 - 4[A_0][D_0]}}{2}\right) \qquad (2)$$

Flavin binding kinetics:

In the kinetic measurements, the same flavins as in the equilibrium measurements were used (see above). For each flavin a 4 µM solution in the corresponding buffer was freshly prepared. Based on an apo-*Mt*Dod stock solution, a 5 µM solution in the corresponding buffer was freshly prepared. Since all solutions were kept on ice, they were equilibrated to room temperature for 30 min in the dark before pipetting. The measurement was carried out in a plate reader with a two channel dispenser [CLARIOstar BMG LABTECH] at room temperature. The corresponding buffer and the apo-*Mt*Dod solution were loaded into the dispenser, while 50 µL of the corresponding flavin solution was pipetted into the wells of the 96 well plate (Non-binding, 3650; Corning) in advance. A measurement set was composed of six kinetic measurements in quadruplicates with varying dodecin concentrations (0.0 µM (only flavin, used to obtain the 100% value F_{max}), 0.25 µM, 0.50 µM, 0.75 µM, 1.00 µM, 1.25 µM, 1.50 µM, 2.00 µM, 2.50 µM and 3.75 µM) while the flavin concentration was fixed at 1 µM. In addition, four blanks, two wells with 25 µM apo-*Mt*Dod and 1 µM flavin (25 eq.) and two wells with 50 µM apo-*Mt*Dod at pH 5.0 or 100 µM apo-*Mt*Dod at pH 7.5 and 1 µM flavin (50 eq. or 100 eq., respectively) were measured without use of the dispenser (equilibrium measurements needed for fitting, used for F_∞ in the single-step fits). The kinetic measurements were started by the addition of the corresponding volumes of buffer and apo-*Mt*Dod solution (overall added volume 150 µL, final volume 200 µL), while the delay of the instrument was 3.66 sec. The data collection interval was 0.33 sec and the whole kinetic was measured for 221.46 sec (660 data points). Fluorescence was measured with excitation at 445 ± 9 nm and emission at 520 ± 20 nm. Data points were corrected by the background and normalized with the averaged corresponding free flavin measurements.

1:1 ratio data (without weighting of data points) was fitted with a one step binding model (functions (3)-(5) based on ref[5]) with Origin Pro 8.6 [OriginLab corporation]. As above [A_0] (overall flavin concentration) and F_{max} (fluorescence of unbound flavin) were fixed to 1 µM and 100%, respectively. Other variables are: [D_0] (overall dodecin concentration, fixed for each ratio), [F_∞] (fluorescence of bound flavin) and [*DA*] (concentration of the bound state).

$$F(t) = F_\infty * [DA](t) + F_{max} * \left([A_0] - [DA](t)\right) \tag{3}$$

$$[DA](t) = -\frac{a * \left(e^{(a-b)*t} - 1\right)}{k_{on} * \left(e^{(a-b)*t} - \frac{a}{b}\right)} \tag{4}$$

$$a, b = -\frac{1}{2}\left(k_{on}[D_0] + k_{on}[A_0] + k_{off} \pm \sqrt{\left(k_{on}[D_0] + k_{on}[A_0] + k_{off}\right)^2 - 4{k_{on}}^2[A_0][D_0]}\right) \tag{5}$$

For the numerical two-step binding model fit, the quadruplicates of a ratio were averaged. The numerical two-step binding model fit was performed in Excel [Microsoft] using the solver function by minimizing the absolute error between calculated and measured data (including the 25 eq. and 50 eq. or 100 eq. measurements) for all ratios together (global fit compared to fit each ration on its own). The 25 eq. and 100 eq. measurements were included by calculating the square error between the average of the respective measurement and the last point (t= 221.46 sec) of the respective calculation, both errors together were weighted with 20% (remaining 80%: absolute error between calculated and measured curve, first 10 seconds were multiplied by 10 to keep the simulation from falling into a local optima by just optimizing the ratio of rate constants). The time step (Δt) for the calculation of the flavin binding was 0.03 sec. Starting conditions for the calculations were: [A_0]= 1 µM (overall flavin concentration, fixed), [DD_0]= 0,5*[D_0] ([DD_0] overall concentration of binding pockets, fixed), [*DDA*](t=0)= 0 µM (concentration of the half filled binding pockets; F_{DDA} corresponding fluorescence factor) and [*DDAA*](t=0)= 0 µM (concentration of the filled binding pockets; F_{DDAA} corresponding fluorescence factor). F_{max} was allowed to be optimized in the two-step fitting. No simulation altered this value below 95%.

$$F(t) = F_{DDA} * [DDA](t) + F_{DDAA} * [DDAA](t) + F_{max} * [A](t) \quad (6)$$

$$[A](t) = [A_0] - [DDA](t) - 2[DDAA](t) \quad (7)$$

$$[DD](t) = [DD_0] - [DDA](t) - [DDAA](t) \quad (8)$$

$$[DDA](t + \Delta t) = [DDA](t) + [DD](t) * [A](t) * k_{on\_1} * \Delta t + [DDAA](t) * k_{off\_2} * \Delta t - [DDA](t) * k_{off\_1} * \Delta t - [DDA](t) * [A](t) * k_{on\_2} * \Delta t \quad (9)$$

$$[DDAA](t + \Delta t) = [DDAA](t) + [DDA](t) * [A](t) * k_{on\_2} * \Delta t - [DDAA](t) * k_{off\_2} * \Delta t \quad (10)$$

Absolute parameter restraints for the calculation (to reduce calculation time) were, when reached, changed and the calculation started again, except the upper limit of k_{off_1}, which was defined as 0.3 s^{-1} (based on the fits with RbF pH 7.5). This upper k_{off_1} constraint was reached during fitting of FMN binding at pH 5.0.

Steady-state spectroscopy:

Steady-state fluorescence spectra were measured using a Jasco FP-8500 spectrometer. A slit width of 2.5 nm for excitation and 2.5 nm for emission was used. PMT voltage was set to 600 V. Emission spectra were baseline corrected and the absorption of the excitation light, the reabsorption and the wavelength-dependent instrument sensitivity were taken into account.

Supporting Figures and Tables

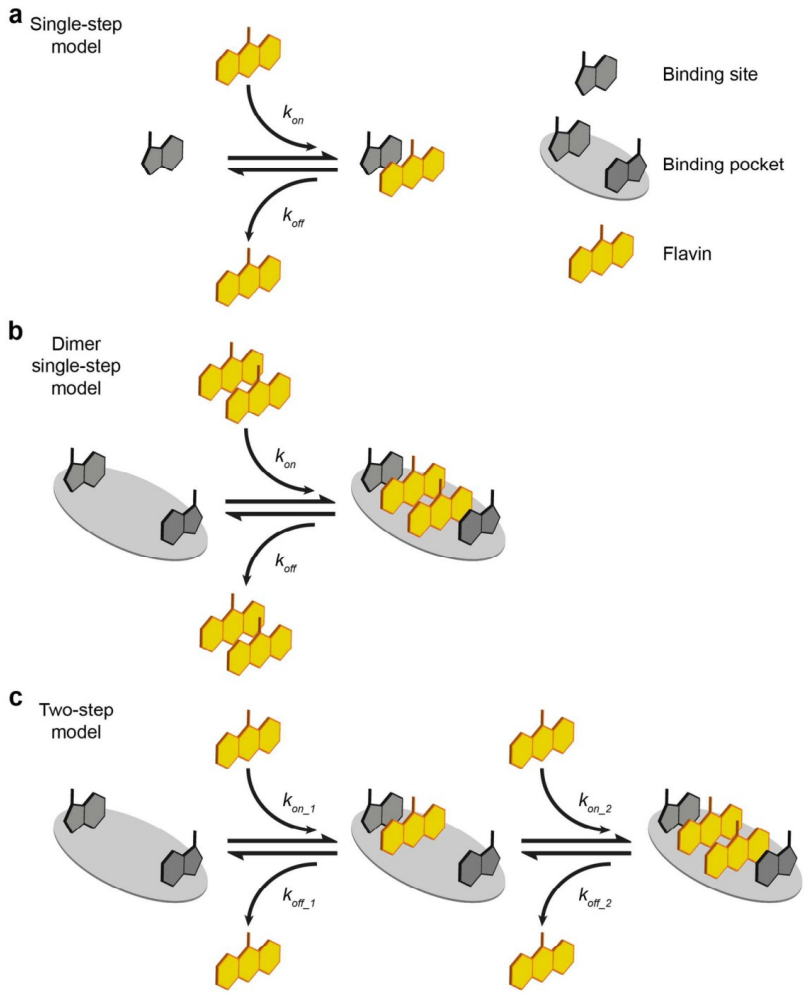


Figure S1. Different flavin binding models of dodecin. (a) Single-step binding model. Each binding site is treated as an individual binding pocket. (b) Dimer single-step model. As in (a), the binding pocket is filled in a single step, but contains two distinct binding sites. *Hs*Dod binds FAD in a similar fashion (single-step): the isoalloxazine ring system binds to one binding site, while the adenine moiety occupies the other (PDB 2CJC[1]).[1]

The isoalloxazine ring system and the adenine moiety are represented here as a dimer. For *Mt*Dod this binding mode was not reported and a crystal structure of *Tt*Dod with FAD shows two FAD molecules per binding pocket, bound via their isoalloxazine systems (PDB 2CZ8). (c) Two-step binding model. Both binding sites are occupied in distinct steps.

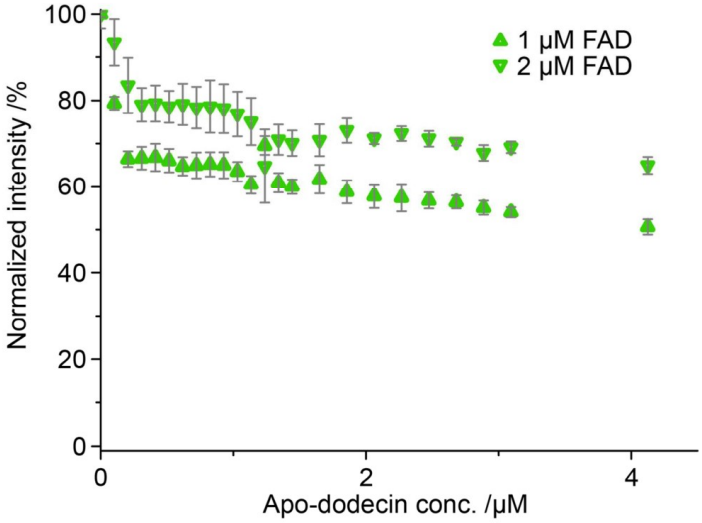
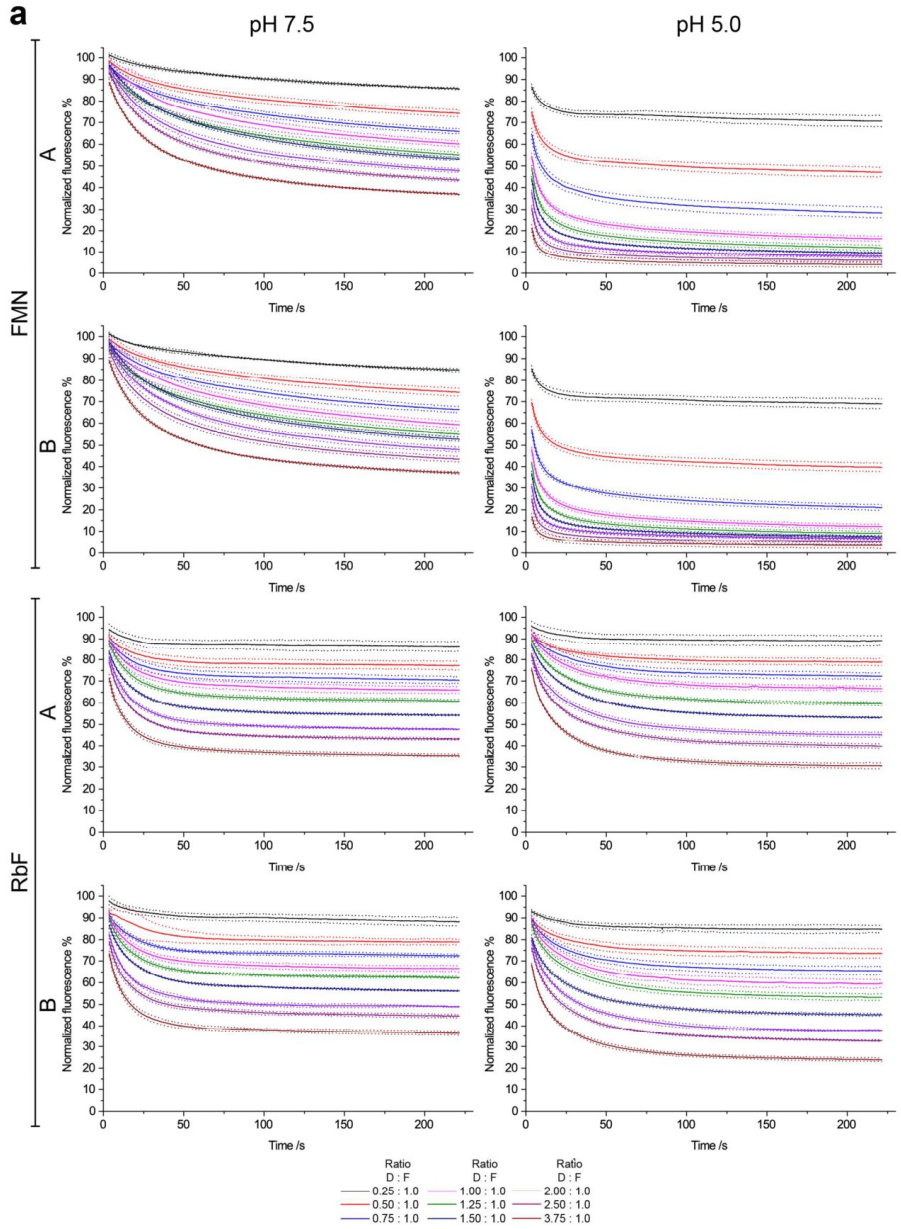


Figure S2. FAD equilibrium binding data. Titrations experiments with FAD and apo-*Mt*Dod show an initial drop of fluorescence possibly based on unspecific effects caused by the addition of protein into the solution. Such an effect may be the weakening/suppress interactions between the well surface and FAD. Since the significant drop of fluorescence stops after the addition of about 0.3 µM it cannot be described with a reasonable binding model. The other flavins did not showed this behavior in the equilibrium measurements, although minor effects may be overlain by a drop of fluorescence during specific binding.


a
pH 7.5
pH 5.0
FMN
RbF
A
B
Normalized fluorescence %
Time /s
Ratio D : F
0.25 : 1.0
0.50 : 1.0
0.75 : 1.0
1.00 : 1.0
1.25 : 1.0
1.50 : 1.0
2.00 : 1.0
2.50 : 1.0
3.75 : 1.0

S12

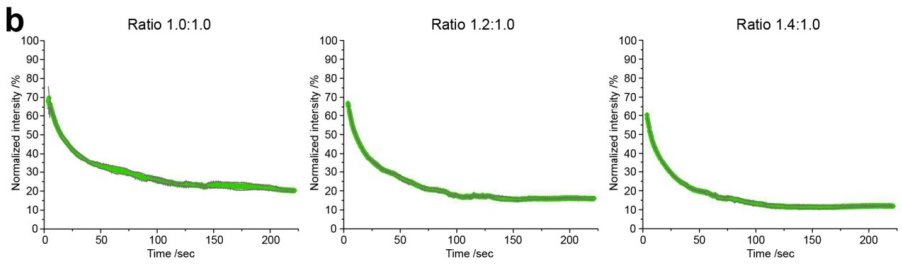
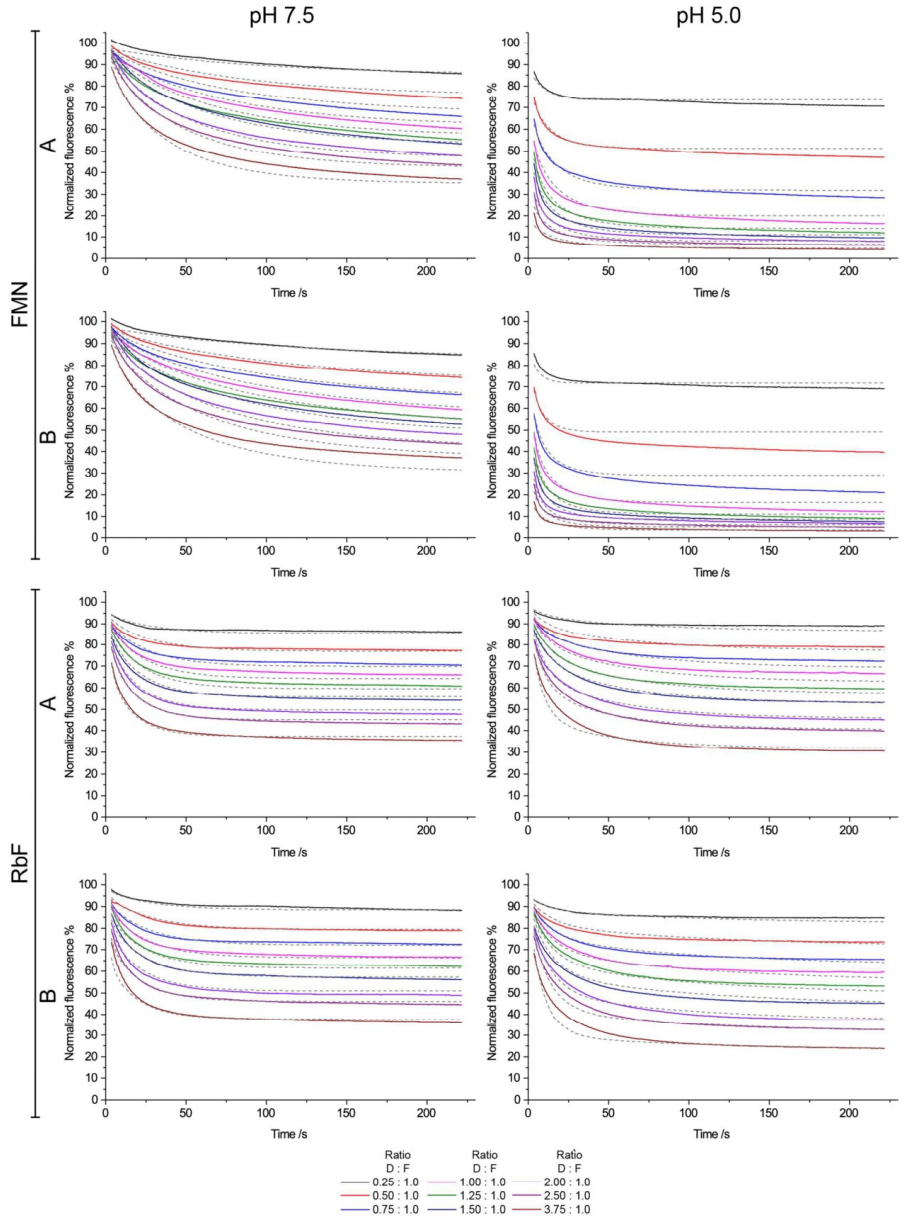


Figure S3. Kinetic data on flavin binding by *Mt*Dod. Set of kinetic measurements performed during this work. Data collected in quadruplicates were averaged and normalized. Curves at the various dodecin:flavin (D:F) ratios are displayed as solid lines, while the simulations are shown as gray dashed lines. Standard deviations are shown as dots. Protein preparations A and B, as also used for the equilibration titrations, were used in the kinetic binding measurements. (b) Kinetic measurements with *Hs*Dod (his-tagged, protein expressed and purified as described in ref[1]) were performed in addition (in medium binding well plates).


pH 7.5
pH 5.0
FMN
RbF
A
B
Normalized fluorescence %
Time /s
Ratio D : F
0.25 : 1.0
0.50 : 1.0
0.75 : 1.0
1.00 : 1.0
1.25 : 1.0
1.50 : 1.0
2.00 : 1.0
2.50 : 1.0
3.75 : 1.0

S14

Figure S4. Kinetic data of flavin binding by *Mt*Dod fitted by a two-step binding model. Data collected in quadruplicates were averaged and normalized (see Figure S3). Curves at the various dodecin:flavin (D:F) ratios are displayed as solid lines, while the simulations are shown as gray dashed lines. Standard deviations are shown as dots. Measurements with FMN show a slow fluorescence decay after an initial steep drop, which cannot be described by the two-step model. Whether this is an experimental artifact, as for example occurring from unspecific binding effects of FMN to the well plate, or a limitation of the fit model, cannot be judged at the basis of the currently available data.

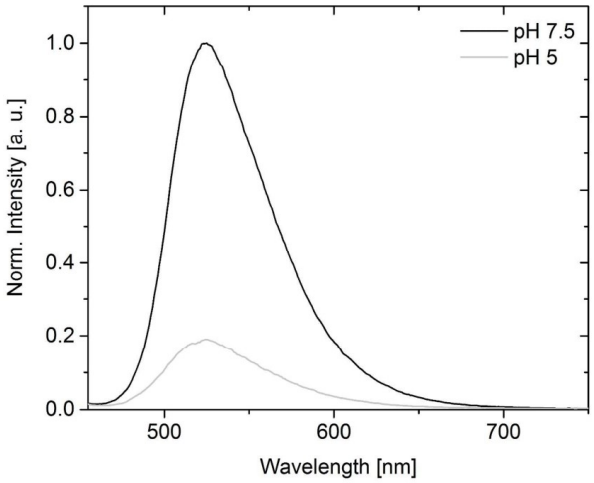


Figure S5. Fluorescence spectra of FMN:*Mt*Dod at pH 7.5 and pH 5.0. FMN fluorescence intensity of the FMN:*Mt*Dod complex at pH 5.0 and pH 7.5, normalized to the fluorescence intensity at pH 7.5. Lower residual fluorescence at pH 5.0 is pointing towards a more efficient quenching at lower pH.

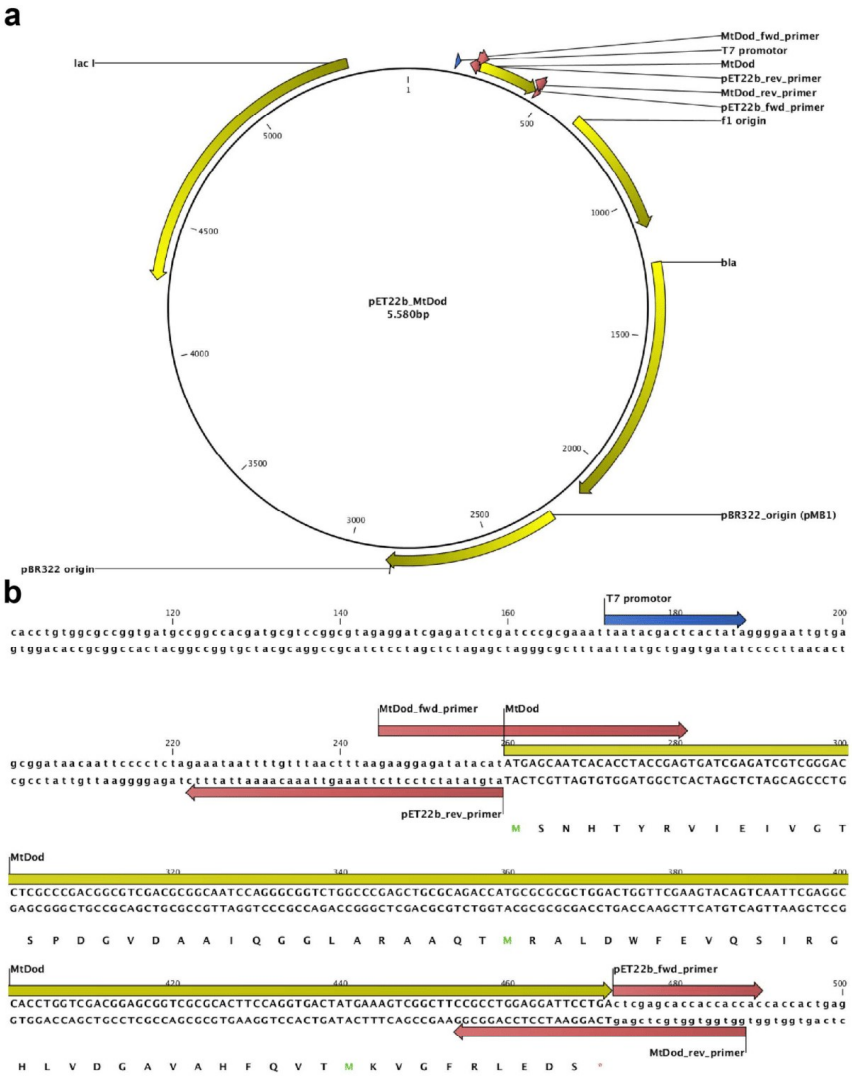


Figure S6 Plasmid map and *Mt*Dod cloning site. (a) Vector card of the used plasmid. (b) Cloning site with MtDod, primers are shown in red. pET22b_rev_primer and pET22b_fwd_primer were used to obtain the linearized vector, with primers MtDod_fwd_primer and MtDod_rev_primer the insert was amplified. The insert was cloned into the linearized vector with the In-Fusion® HD Cloning Kit (Takara Bio) using the standard protocol.

Table S1. Dissociation constants (in nM) of various dodecins (comparison of reported data with data collected in this work). Errors of the equilibration measurements are the standard error of the fits. Errors of the kinetic measurements (1:1 ratio) are the standard deviation of the quadruplicates.

Dodecin	pH; protein preparation		Ligand (K_D in nM)			
			RbF	FMN	FAD	LmC
*Mt*Dod (equilibration)	7.5	A	476 ± 39	105 ± 26	no binding	1.2 ± 0.4 x 10^3 **
		B	309 ± 39	132 ± 18	n.d.	n.d.
	5.0	A	742 ± 33	27 ± 1	163 ± 7	n.d
		B	454 ± 27	8 ± 1	151 ± 3	n.d.
*Mt*Dod (single-step kinetic)*	7.5	A	1192 ± 124	692 ± 79	n.d.	n.d.
		B	1224 ± 95	646 ± 126		
	5.0	A	1278 ± 81	45 ± 6		
		B	845 ± 98	25 ± 4		
*Mt*Dod (ref[6])	7.0		n.d.	920	no binding	n.d.
*Tt*Dod (ref[7])	8.0		233 ± 37	311 ± 18	589 ± 26	80 ± 3
*Hh*Dod (ref[1])	7.5		20 ± 4 x 10^3	6.4 ± 0.5 x 10^3	24 ± 2 x 10^3	2.5 ± 0.3 x 10^3
*Hs*Dod (single-step kinetic)	7.5		76 ± 7***	n.d.	n.d.	n.d.
*Hs*Dod (ref[1])	7.5		36 ± 4	13.7 ± 1.2 x 10^3	439 ± 48	10 ± 3

(*) We suppose that the deviation of kinetic from equilibrium data results from the limitation of the one-step model to describe the binding accurately, but it still gives a reasonable relative comparison. For relative comparison the thermocyclic fluorescence assay (see Figure 2 d-f) seems more suited, since it requires no binding model, and is in agreement with all other experiments and models (presented in this work).

(**) Based on a simplified assay to see if a higher affinity is observed, since lumichrome (LmC) showed only a relative low affinity no further measurements were concurred.

(***) Obtained by fitting the 1:1 ratio of *Hs*Dod and RbF (see Figure S3b), other constants and data of other ratios not shown.

Supporting Notes

Supporting Note 1: Binding properties of *Mt*Dod.

So far, all affinity studies with dodecins and flavins were done by using a single-step binding model (see Figure S1a), although structural data suggest a two-step binding model (see Figure S1b) and observations from Meissner *et al.* with *Tt*Dod suggest a even more complex binding model.[1,6,7] The two-step model is further supported by quartz crystal microbalance dissipation measurements with *Hs*Dod and RbF derivatives.[8] For *Tt*Dod, it was suggested that binding of a first FMN dimer to the apo-protein is kinetically hindered requiring a rearrangement of the apo-dodecin, while further binding occurs with no measurable time dependence.[7] Such a behavior was not observed with *Mt*Dod during the collection of equilibrium data.

For refining the understanding, kinetic data were collected (see Figure 3b and c). These data show that binding is indeed fast (compared to the observations of Meissner *et al.*[7]), and supporting a more complex binding mechanism than single-step binding. Kinetic data were, nevertheless, similar to equilibrium data, fitted with a single-step binding model and in addition all kinetic data was fitted with a two-step binding model. The calculations show, that the first binding step has higher dynamics than the second binding step, similar as *Hs*Dod (see Supporting Data "Kinetic_two-step" and ref[8]). Unlike the one-step model fits, the two-step model fits show no systematic deviation from the experimental data, suggesting, that the system can be described with a two-step model. On the basis of our data, we suggest a two-step binding mode for *Mt*Dod without any (or at least just insignificant) communication between six binding pockets for flavin dimers. We note fits of the kinetic measurements with RbF as ligand were in general of higher quality, and particularly fits of FMN binding data at acidic conditions showed some systematic deviation (see Figure S4).

Supporting Note 2: Comparison of binding properties of dodecin and kinetic binding data of *Hs*Dod.

Since the absolute K_D values of the different dodecins depend on the conditions of the measurement (e.g. pH value) the quotient of K_D(second highest affinity)/K_D(highest affinity) is used instead (only data of RbF, FMN and FAD considered). While *Hs*Dod shows a clear preference for RbF (the factor K_D(FAD pH 7.0)/K_D(RbF pH 7.0) is 12.2), *Tt*Dod binds RbF only slightly better than FMN with the factor K_D(FMN pH 8.0)/K_D(RbF pH 8.0) being 1.3. For *Mt*Dod, the factor K_D(RbF pH 7.5)/K_D(FMN pH 7.5) is 2.7. Comparing those factors, *Hs*Dod shows high selectivity, while *Tt*Dod and *Mt*Dod are rather unselective in flavin binding. Moreover, at neutral and slightly basic

pH (pH 7.5 and 8.0), the bacterial dodecins seem to bind their preferred ligand with less affinity compared to the archaeal dodecins. At pH 5.0, *Mt*Dod becomes more selective; K_D(FAD pH 5.0)/K_D(FMN pH 5.0) is 8.8, and (K_D(RbF pH 5.0)/K_D(FMN pH 5.0) is 33.7, and it also binds FMN with a higher affinity than *Hs*Dod binds RbF. In addition to the increase of affinity to FMN at pH 5.0, *Mt*Dod is also binding FAD. In contrast to FMN or FAD binding, affinity of RbF is only slightly affected by the change in pH; K_D(RbF pH 5.0)/K_D(RBF pH 7.5) is 1.5.

Since the affinity of RbF is only slightly affected by the pH, a change of the overall shape of the binding pocket does not seem to cause the increase of affinity for FMN and FAD at acidic pH, although the two-step kinetic calculations suggest that the RbF binding kinetic is slightly responding to pH conditions (see Supporting Data "Kinetic_two-step"). At pH 7.5, the RbF binding seems to have very similar k_{on_1} and k_{on_2} rates, while for all other measurements k_{on_1} was higher than k_{on_2}. Also for *Hs*Dod binding RbF, the first binding step has a higher k_{on_1} rate than the second binding step (k_{on_2}). Comparing the ratios of k_{on_1}/k_{on_2} (average of both biological samples) *Mt*Dod shows at pH 7.5 RbF (k_{on_1}/k_{on_2}(RbF; pH 7.5) = 1.22) a lower distinction between the two binding step rates as compared to *Hs*Dod (k_{on_1}/k_{on_2}(RbF; pH 7.5) = 3.88 (only one biological sample for *Hs*Dod, see Figure S3; data collected in medium binding plates and fitted with a non-global fit). The aromatic tetrad arrangement of *Hs*Dod is more compact than in *Mt*Dod. This may be a possible explanation for the weaker distinction of "on" rates of RbF binding to *Mt*Dod at pH 7.5, since the second binding event in *Hs*Dod may be more constraints than in *Mt*Dod.

A comparison of flavin binding affinities of archaeal and bacterial dodecins shows that FMN specificity (over RbF) is not just achieved by phosphate binding. Since RbF binds weaker to bacterial dodecins, it seems that the preferred binding of FMN was achieved by also weakening the contribution of the aromatic tetrade. Simply increasing the affinity towards the phosphate group of FMN without lowering the affinity towards the isoalloxazine ring system would otherwise be reflected in high affinities for RbF (see Table S1).

Supporting Note 3: Redox properties of FMN:MtDod.

If Apo-*Mt*Dod binds reduced FMN weaker than oxidized FMN, the association constant K_A for complex formation of *Mt*Dod and reduced FMN is lower than that for *Mt*Dod and oxidized FMN, *i.e.* $K_{A(red)} << K_{A(ox)}$. As pointed out in ref[9] differences in binding affinity result in a shift of the redox potential $E^{0'}$ of the protein bound cofactor, in comparison to that of the free cofactor in solution (without apo-protein) according to equation (11) (R is the gas constant, T the temperature, n the number of transferred electrons, and F the faradaic constant).

$$E^{0'}_{(bound)} = E^{0'}_{(free)} + 2.303\frac{RT}{nF}\log\frac{K_{A(red)}}{K_{A(ox)}} \quad (11)$$

While the value for $K_{A(ox)}$ can be calculated as $K_A = 1/K_D$ from the K_D values depicted in Table S1 (at pH 7.5 for the complex of Apo-*Mt*Dod with oxidized FMN K_D is 170 nM), a detailed value for $K_{A(red)}$ is not known. Nevertheless, a significantly lower value for $K_{A(red)}$ in comparison to that of $K_{A(ox)}$ should result in a remarkable shift in the redox potential of protein bound FMN compared to the free flavin(if for instance $K_{A(red)} \approx 1$, the redox potential of protein bound FMN should be shifted by about 200 mV to more negative values).

Supporting Note 4: Archaeal dodecins are restricted to phylogenetic class Haloarchaea.

Archaea have obtained the dodecin encoding gene *via* horizontal gene transfer from bacteria,[10] and the gene has remained narrowly distributed within the class of Haloarchaea (also called Halobacteria). Halobacteria are found under high salt or saturated salt conditions. As the accumulation of the acidic residues aspartate and glutamate as well as the reduction of the basic residues lysine an arginine at the surfaces of proteins have been accounted as one of the key strategies for stabilizing proteins under high salt conditions (the natural habitat of Haloarchaea),[11–15] the specificity of archaeal dodecins for riboflavin may have developed from such constraints. In a scenario of the evolutionary development of archaeal dodecins from the bacterial ancestor, a set of amino acid exchanges has induced RbF specificity into an initially FMN binding scaffold. The phosphate-binding arginine (R66 in *Mt*Dod), one of the main residues for FMN binding, is found exchanged by glutamate (E63 in *Hs*Dod) (see Figure 6c). Of 619 bacterial dodecins, none show an acidic residue at the respective position, indicating that indeed Haloarchaea introduced this change (see Supporting Data “Alignment”). As another prominent exchange, a highly conserved arginine in bacterial dodecins (R56 in *Mt*Dod), which is located at the inner surface of the protein and hydrogen bonding with the C4 carbonyl and the N5 group of the isoalloxazine ring, is replaced by glycine in the archaeal proteins (G43 in *Hs*Dod). This glycine is involved in an extended network of coordinated water molecules and Mg^{2+}, which is a well-known characteristic for surfaces of halophilic proteins too.[16] Further, a highly conserved glutamate occurs in the binding channel of archaeal dodecins (E45, HsDod numbering, see Figure 6c) that holds the RbF ribityl moieties. This glutamate almost exclusively occurs in archaeal dodecins (present in 4 out of 619 bacterial sequences; see Supporting Data “Alignment”). Finally, in

this scenario, in response to the adaption of the protein to high intracellular salinity, archaeal dodecins developed a different flavin-binding mode, in which flavins are bound in *re-re* orientation (*re*-sides facing), while dodecins in bacteria show *si-si* orientation. The different orientations of the flavins are accompanied with swapped arrangements of the glutamines (Q58 in *Mt*Dod, Q55 in *Hs*Dod; see Figure 6c). In the archaeal binding pocket, the aromatic systems are arranged in a more compact manner. The overlap of the aromatic systems is larger, and the orientation of the ribityl is almost parallel. Both features likely account for steering phosphates of FMN and FAD in unfavorable close proximity, finally counter-selecting against FMN and FAD binding (see Figure 6c). Owing the parallel arrangement of ribityl chains, also hydrogen bonding between hydroxyl groups was observed, additionally stabilizing the riboflavin binding.[1,7,17]

Abbreviations:

AAB	universal buffer system: acetic acid, ADA and bicine
ADP	adenosine 5′-triphosphate
ATP	adenosine 5′-diphosphate
CoA	coenzyme A
equ.	indicator for equilibrium measurements
FAD	flavin adenine dinucleotide
FAD:*Mt*Dod	holo complex of *Mt*Dod and FAD
FMN	flavin mononucleotide
FMN:*Mt*Dod	holo complex of *Mt*Dod and FMN
$FMNH^-$	reduced and single deprotonated flavin mononucleotide
H4X-*Mt*Dod	X: A D N R; *Mt*Dod mutation at position 4
*Hh*Dod	*Halorhodospira halophila* dodecin
*Hs*Dod	*Halobacterium salinarum* dodecin
K_D	Dissociation constant
K_{D_1}; k_{on_1}; k_{off_1}	constants for the first binding step, single filled pocket
K_{D_2}; k_{on_2}; k_{off_2}	constants for the second binding step, double filled pocket
kin.	indicator for kinetic measurements
LmC	lumichrome
M. tuberculosis	*Mycobacterium tuberculosis*
MMT	universal buffer system: malic acid, MES and Tris
*Mt*Dod	*Mycobacterium tuberculosis* dodecin
NHE	Standard hydrogen electrode
RbF	riboflavin
RbF:*Mt*Dod	holo complex of *Mt*Dod and RbF
SEC	size exclusion chromatography
TB	tuberculosis
TSA	thermal shift assay
*Tt*Dod	*Thermus thermophilus* dodecin

Supporting Information References:

(1) Grininger, M., Zeth, K., and Oesterhelt, D. **(2006)** Dodecins: A Family of Lumichrome Binding Proteins. *J. Mol. Biol. 357*, 842–857. DOI: 10.1016/j.jmb.2005.12.072.
(2) Aliverti, A., Curti, B., and Vanoni, M. **(1999)** Identifying and Quantitating FAD and FMN in Simple and in Iron-Sulfur-Containing Flavoproteins, in *Flavoprotein Protocols* (Chapman, S., and Reid, G., Eds.), pp 9–23. Humana Press, Totowa, NJ.
(3) Macheroux, P. **(1999)** UV-Visible Spectroscopy as a Tool to Study Flavoproteins, in *Flavoprotein Protocols* (Chapman, S., and Reid, G., Eds.), pp 1–7. Humana Press, Totowa, NJ.
(4) Vogt, S., Schneider, M., Schäfer-Eberwein, H., and Nöll, G. **(2014)** Determination of the pH Dependent Redox Potential of Glucose Oxidase by Spectroelectrochemistry. *Anal. Chem. 86*, 7530–7535. DOI: 10.1021/ac501289x.
(5) Peuker, S., Cukkemane, A., Held, M., Noé, F., Kaupp, U. B., and Seifert, R. **(2013)** Kinetics of Ligand-Receptor Interaction Reveals an Induced-Fit Mode of Binding in a Cyclic Nucleotide-Activated Protein. *Biophys. J. 104*, 63–74. DOI: 10.1016/j.bpj.2012.11.3816.
(6) Liu, F., Xiong, J., Kumar, S., Yang, C., Ge, S., Li, S., Xia, N., and Swaminathan, K. **(2011)** Structural and biophysical characterization of *Mycobacterium tuberculosis* dodecin Rv1498A. *J. Struct. Biol. 175*, 31–38. DOI: 10.1016/j.jsb.2011.04.013.
(7) Meissner, B., Schleicher, E., Weber, S., and Essen, L. O. **(2007)** The dodecin from *Thermus thermophilus*, a bifunctional cofactor storage protein. *J. Biol. Chem. 282*, 33142–33154. DOI: 10.1074/jbc.M704951200.
(8) Gutiérrez Sánchez, C., Su, Q., Wenderhold-Reeb, S., and Nöll, G. **(2016)** Nanomechanical properties of protein–DNA layers with different oligonucleotide tethers. *RSC Adv. 6*, 56467–56474. DOI: 10.1039/C6RA10090B.
(9) Pellett, J., and Stankovich, M. **(2007)** Potentiometric Measurements of Proteins, in *Encyclopedia of Electrochemistry*. Wiley-VCH Verlag GmbH & Co. KGaA.
(10) Nelson-Sathi, S., Dagan, T., Landan, G., Janssen, A., Steel, M., McInerney, J. O., Deppenmeier, U., and Martin, W. F. **(2012)** Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl. Acad. Sci. 109*, 20537–20542. DOI: 10.1073/pnas.1209119109.
(11) Madern, D., Pfister, C., and Zaccai, G. **(1995)** Mutation at a single acidic amino acid enhances the halophilic behaviour of malate dehydrogenase from *Haloarcula marismortui* in physiological salts. *Eur. J. Biochem. 230*, 1088–1095. DOI: 10.1111/j.1432-1033.1995.1088g.x.
(12) Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M., and Nishikawa, K. **(2003)** Unique Amino Acid Composition of Proteins in Halophilic Bacteria. *J. Mol. Biol. 327*, 347–357. DOI: 10.1016/S0022-2836(03)00150-5.
(13) Pieper, U., Kapadia, G., Mevarech, M., and Herzberg, O. **(1998)** Structural features of halophilicity derived from the crystal structure of dihydrofolate reductase from the Dead Sea halophilic archaeon, *Haloferax volcanii*. *Structure 6*, 75–88. DOI: 10.1016/S0969-2126(98)00009-4.
(14) Kastritis, P. L., Papandreou, N. C., and Hamodrakas, S. J. **(2007)** Haloadaptation: Insights from comparative modeling studies of halophilic archaeal DHFRs. *Int. J. Biol. Macromol. 41*, 447–453. DOI: 10.1016/j.ijbiomac.2007.06.005.
(15) Ortega, G., Diercks, T., and Millet, O. **(2015)** Halophilic Protein Adaptation Results from Synergistic Residue-Ion Interactions in the Folded and Unfolded States. *Chem. Biol. 22*, 1597–1607. DOI: 10.1016/j.chembiol.2015.10.010.
(16) Scheurer, M., Brisker-Klaiman, D., and Dreuw, A. **(2017)** Molecular Mechanism of Flavin Photoprotection by Archaeal Dodecin: Photo-Induced Electron Transfer and Mg2+-Promoted Proton Transfer. *J. Phys. Chem. B* DOI: 10.1021/acs.jpcb.7b08597.
(17) Bieger, B., Essen, L.-O., and Oesterhelt, D. **(2003)** Crystal Structure of Halophilic Dodecin: A Novel, Dodecameric Flavin Binding Protein from *Halobacterium salinarum. Structure 11*, 375–385. DOI: 10.1016/S0969-2126(03)00048-0.

MICROBIOLOGY

RESEARCH ARTICLE

Ludwig *et al.*, *Microbiology* 2018;164:908–919

DOI 10.1099/mic.0.000662

EDITOR'S CHOICE

Characterization of the small flavin-binding dodecin in the roseoflavin producer *Streptomyces davawensis*

Petra Ludwig,[1] Daniel C. Sévin,[2] Tobias Busche,[3] Jörn Kalinowski,[3] Florian Bourdeaux,[4] Martin Grininger[4] and Matthias Mack[1,*]

Abstract

Genes encoding dodecin proteins are present in almost 20 % of archaeal and in more than 50 % of bacterial genomes. Archaeal dodecins bind riboflavin (vitamin B_2), are thought to play a role in flavin homeostasis and possibly also help to protect cells from radical or oxygenic stress. Bacterial dodecins were found to bind riboflavin-5′-phosphate (also called flavin mononucleotide or FMN) and coenzyme A, but their physiological function remained unknown. In this study, we set out to investigate the relevance of dodecins for flavin metabolism and oxidative stress management in the phylogenetically related bacteria *Streptomyces coelicolor* and *Streptomyces davawensis*. Additionally, we explored the role of dodecins with regard to resistance against the antibiotic roseoflavin, a riboflavin analogue produced by *S. davawensis*. Our results show that the dodecin of *S. davawensis* predominantly binds FMN and is neither involved in roseoflavin biosynthesis nor in roseoflavin resistance. In contrast to *S. davawensis*, growth of *S. coelicolor* was not reduced in the presence of plumbagin, a compound, which induces oxidative stress. Plumbagin treatment stimulated expression of the dodecin gene in *S. davawensis* but not in *S. coelicolor*. Deletion of the dodecin gene in *S. davawensis* generated a recombinant strain which, in contrast to the wild-type, was fully resistant to plumbagin. Subsequent metabolome analyses revealed that the *S. davawensis* dodecin deletion strain exhibited a very different stress response when compared to the wild-type indicating that dodecins broadly affect cellular physiology.

INTRODUCTION

Riboflavin serves as a precursor for flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). FMN is synthesized by flavokinases (EC 2.7.1.26), whereas FAD is generated by FAD synthetases (2.7.7.2) (Fig. 1a) [1]. FMN and FAD are cofactors of flavoproteins which carry out diverse biological functions [2]. The total number and the percentage of genes encoding flavoproteins varies strongly in prokaryotic and eukaryotic organisms [3]. Several species appear to have a minimum number of flavin-dependent proteins that are required to maintain basic metabolic functions (e.g. *Thermotoga maritima* is predicted to contain 12 flavoproteins [3]), whereas others, such as the actinobacteria *M. tuberculosis* (148 flavoproteins) and *S. coelicolor* (224 flavoproteins), contain a large set of flavoproteins [3]. Most flavoproteins bind their cofactors noncovalently (ca. 90 %) and appear to predominantly utilize FAD (ca. 75 %) rather than FMN (ca. 25 %).

Dodecins (Dod) are the smallest known flavoproteins (65–73 amino acids). Dodecins were discovered in the archaeon *Halobacterium salinarum* ($HsDod^A$) during the course of an inverse structural proteomics project. The structure of archaeal $HsDod^A$ was the first dodecin structure to be published [4]. Structural analysis of bacterial dodecins followed (*Halorhodospira halophila*, $HhDod^B$ [5], *Thermus thermophilus*, $TtDod^B$ [6] and *Mycobacterium tuberculosis*, $MtDod^B$ [7, 8]) and revealed that archaeal and bacterial dodecins differ strikingly with regard to their ligand-binding spectrum and binding site architecture. All structurally characterized dodecins form dodecamers and the resulting

Received 11 January 2018; Accepted 12 April 2018

Author affiliations: [1]Institute for Technical Microbiology, Mannheim University of Applied Sciences, 68163 Mannheim, Germany; [2]Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland; [3]Center for Biotechnology, Bielefeld University, 33615 Bielefeld, Germany; [4]Institute of Organic Chemistry and Chemical Biology, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany.

***Correspondence:** Matthias Mack, m.mack@hs-mannheim.de

Keywords: *Streptomycetes*; flavins; roseoflavin; riboflavin; dodecin.

Abbreviations: Dod, dodecin; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; $HhDod^B$, dodecin from *Halorhodospira halophile*; $HsDod^A$, dodecin from *Halobacterium salinarum*; $MtDod^B$, dodecin from *Mycobacterium tuberculosis*; RoFAD, roseoflavin adenine dinucleotide; RoFMN, roseoflavin mononucleotide; $ScDod^B$, dodecin from *Streptomyces coelicolor*; $ScinDod^B$, dodecin from *Streptomyces cinnabarinus*; $SdDod^B$, dodecin from *Streptomyces davawensis*; $TtDod^B$, dodecin from *Thermus thermophilus*.

Supplementary material is available with the online version of this article.

hollow sphere-like protein complex provides six identical binding pockets for flavin (riboflavin or FMN) dimers which are sandwiched between two tryptophan residues. Consequently, 12 flavin molecules are accommodated by one dodecin dodecamer.

Most flavoproteins are redox catalysts or light sensors. These functions have been ruled out for dodecins [6] and a few reports are available which investigated the other possible physiological role(s) of these small flavin binders. One study suggests that archaeal $HsDod^{A}$ sequesters riboflavin under growth-limiting conditions and releases it during more favourable growth conditions [5]. Moreover, $HsDod^{A}$ was reported to regenerate photochemically activated flavin species [9]. Bacterial dodecins were found to be less efficient quenchers of reactive flavins indicating that bacterial and archaeal dodecins may have different functions. This is in line with the fact that archaeal dodecins incorporate riboflavin while bacterial dodecins bind FMN as well as coenzyme A [5, 10].

S. davawensis produces the toxic riboflavin analogue roseoflavin (8-demethyl-8-dimethylaminoriboflavin) (Fig. 1a) [11, 12] and contains a dodecin gene as well as the second known roseoflavin producer *Streptomyces cinnabarinus* [11]. Roseoflavin is converted to roseoflavin mononucleotide (RoFMN) and roseoflavin adenine dinucleotide (RoFAD) by flavokinases and FAD synthetases within target cells (Fig. 1a). RoFMN negatively affects FMN riboswitches which control riboflavin biosynthesis and transport [13–16]. In addition, RoFMN and RoFAD have the potential to inactivate flavoenzymes [17–21]. Flavins are highly reactive chemicals [22] – they are able to accept and donate single electrons, electron pairs as well as visible light – and at the same time are present in the cytoplasm of cells in comparably large amounts [23]. Intracellular flavins and flavoproteins were reported to be a major source for the reactive oxygen species O_2^- and H_2O_2, cytotoxic compounds which severely interfere with cellular structures, metabolism and growth [22]. Streptomycetes appear to have a 'flavin-intensive' lifestyle [3]. For example, *S. coelicolor* can adapt to various carbon and nitrogen sources and produces a large number of biologically active compounds, such as antibiotics. This organism depends on metabolic power and versatility that are certainly conferred to some degree by flavin-dependent enzymes [3]. We therefore anticipated that studying *Streptomyces* dodecins may help to better understand the function of these flavin-binding proteins in bacteria. First, we investigated a possible more specific function of dodecins in roseoflavin biosynthesis or resistance, and second, we studied a broader role of dodecins with regard to the oxidative stress response in Streptomycetes.

METHODS

Chemicals and materials

Roseoflavin was obtained from MP Biomedicals (Heidelberg, Germany). Roseoflavin mononucleotide (RoFMN) and roseoflavin adenine dinucleotide (RoFAD) were prepared enzymatically as described [24]. All other chemicals were from Sigma-Aldrich (Munich, Germany). Riboflavin and roseoflavin are light sensitive and, when not otherwise indicated, all samples and cultures containing these flavins were kept in the dark.

Database searches

All sequence comparisons and database searches (as of November 2017) were done employing the MicrobesOnline database [25].

Bacterial strains, plasmids, media and growth conditions

The strains and plasmids used in this work are listed in Table S1 (available in the online version of this article). The correctness of nucleotide sequences of all inserts of all constructs was verified by DNA sequencing. *Escherichia coli* was used for cloning and production of recombinant proteins and was grown aerobically at 180 r.p.m. in baffled Erlenmeyer flasks in lysogeny broth (LB) at 37 °C [26] in the presence of antibiotics where required. *Streptomyces* strains were aerobically cultivated at 180 r.p.m. and 30 °C. The yeast starch-medium (YS) and the mannitol soya flour-medium (MS) were prepared as described [27]. *B. subtilis* was aerobically grown at 37 °C in LB or in a minimal medium [28]. For growth experiments with Streptomycetes in the presence of plumbagin (15 or 20 µM), nitrofurantoin (100 µM) or paraquat (500 µM) 30 ml YS in 200 ml baffled Erlenmeyer flasks was inoculated with 2×10^6 spores. The cultures were incubated at 30 °C and 180 r.p.m. until an OD_{450} of 1.5±0.1 was reached. The dry weight of the cultures was determined by collection of the mycelium by centrifugation at 15 173 ***g*** (4 °C for 1 min) followed by drying the mycelium at 60 °C. For heterologous expression of the *S. davawensis* dodecin gene in *E. coli* Rosetta 2(DE3) the dod^{SD} gene (BN_159 1333) was PCR-amplified with oligonucleotides P1333NdeI-fw (5′-ATA TAT CAT ATG TCG AAC CAC ACC TAC CGG GTC ACG GAC-3′) and P1333XhoI-rv (5′-ATA TAT CTC GAG ACC GGT CTC GTC CAG GCG GAA GCC-3′) which introduced restriction sites for cloning. Genomic DNA of *S. davawensis* wild-type was used as a template and the *Nde*I/*Xho*I treated PCR product was ligated to *Nde*I/*Xho*I treated pET24a(+) (Merck KGaA, Germany). Ligation of the dodecin gene to pET24a (+) introduced codons specifying a C-terminal His_6-tag. Accordingly, for expression of *S. coelicolor* dod^{SC} (SCO0915) oligonucleotides PSCO0915NdeI-fw (5′-ATA TAT CAT ATG TCG AAC CAC ACC TAC CGG GTC ACC GAG G-3′) and PSCO0915XhoI-rv (5′-ATA TAT CTC GAG GTC GGA CTC CTC CAG GCG GAA GCC G-3′) and genomic DNA of *S. coelicolor* were employed. For expression of the *Streptomyces* dodecin genes in *Bacillus subtilis*, dod^{SD} and dod^{SC} were adapted to the codon usage of the host. The synthetic genes dod^{SD} and dod^{SC} were produced by Thermo Fisher Scientific GENEART GmbH (Regensburg, Germany) and delivered ligated to the plasmid pMA (a pUC19 derivative). To be able to verify expression of dod^{SD} and dod^{SC} additional synthetic genes (dod^{SDHis6}

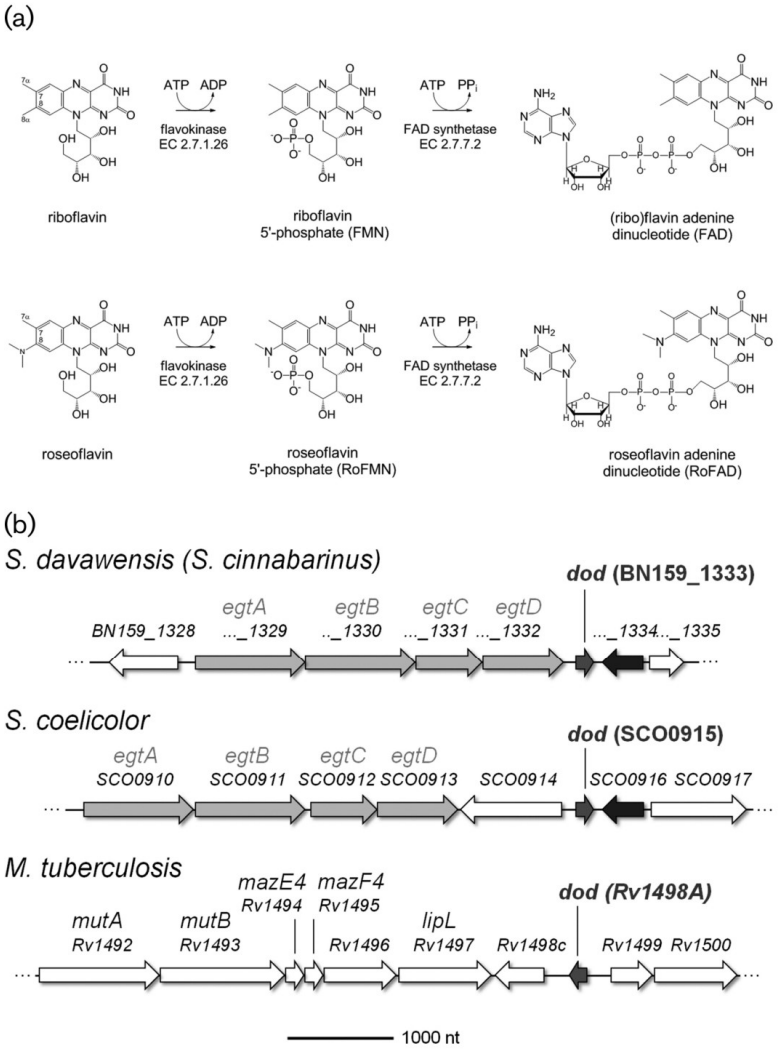


Fig. 1. Structures of flavin and the genomic context of (putative) dodecin genes in different bacteria. (a) Structures of riboflavin and roseoflavin and their intracellular derivatives are shown. Flavokinase and FAD synthetase depend on ATP and are responsible for the formation of these flavin derivatives. (b) In the closely related species *S. davawensis* [11] and *S. cinnabarinus* (genomic data unpublished) the gene order in direct vicinity to the dodecin gene (*dod*) is identical. *S. davawensis* and *S. cinnabarinus* are roseoflavin producers and the ergothioneine biosynthetic genes *egtA-D* are transcribed in the same direction as *dod*. Downstream of *egtA-D* a putative transcriptional terminator is present and within *egtD* a putative promoter is present which probably is responsible for expression of *dod*. Thus, the genes *egtA-D* and *dod* appear to not form a transcription unit. In *S. coelicolor* the dodecin gene also is located close to the *egtA-D* genes, however, SCO0914 (unknown function) seperates the *egt* genes from *dod*. In *M. tuberculosis* the *egtA-D* genes are located elsewhere in the chromosome [42] and the genes neighbouring *dod* are annotated as genes involved in lipid metabolism, virulence and detoxification. In *M. tuberculosis* the *dod* gene and the corresponding gene product has been studied in detail [7], the physiological function of this small flavoprotein, however, is still not understood for this pathogen.

and dod^{SCHis6}) were ordered which contained additional nucleotides encoding a His_6-tag. The nucleotide sequences of all synthetic dodecin genes are listed in the supplementary file. The corresponding plasmids were treated with *Bam*HI and *Sma*I and the resulting DNA fragments were ligated to the *B. subtilis* expression vector pHT01 (Mobitec, Germany).

Construction of an *S. davawensis* dod^{SD} deletion strain

Disruption of the dod^{SD} gene (BN159_1333) in *S. davawensis* essentially was carried out as described earlier [27]. The specific PCR product needed for recombination was obtained by amplification of the disruption cassette from pIJ773mod [27] with oligonucleotides P1333DC_fw (5′-ACG GTG GAG TGA CGC GTG TGC GAG AGG AGC ACC CAC ATG ATT CCG GGG ATC CGT CGA CC-3′) and P1333DC_rv (5′-AGC GCG CTC AGC CGG TCA CGC TCA GCC CAG CCG CGC TCA TGT AGG CTG GAG CTG CTT C-3′). The resulting PCR product was employed for replacement of the dodecin gene by an apramycin resistance cassette (*aacIV*) in cosmid pESAC13.4. This cosmid was obtained from a complete *S. davawensis* library and was identified by PCR screening to contain the dod^{SD} gene (BN159_1333). Disruption of the dod^{SD} gene in *S. davawensis* was verified by PCR using oligonucleotides P1333Delta_fw (5′-GCG CTG GAC CTG GCC GTC GAC TTC-3′) and P1333Delta_rv (5′-CGC ACC AAG TTG GAG ATC CTC CGC CG-3′) and genomic DNA of the disruption strains as a template.

Overproduction and purification of dodecins

Dodecins from *S. davawensis* or *S. coelicolor* were purified from recombinant *E. coli* Rosetta 2(DE3) strains and purified to apparent homogeneity (as evaluated by SDS-PAGE) by immobilized metal affinity chromatography. T7 RNA polymerase-based expression of dodecin genes was stimulated by addition of 1 mM IPTG after the culture had reached an optical density (OD_{600}) of 0.6. After 5 h of further aerobic incubation cells were harvested by centrifugation and stored at −20 °C prior to cell disruption. All chromatographic steps were performed using the ÄKTA-purifier system (GE Healthcare, Munich, Germany) at a flow rate of 2 ml min^{-1} and room temperature. Frozen cell pastes of *E. coli* cells overproducing the different dodecins were suspended in 30 ml IIisTrap-binding buffer (3 mM KH_2PO_4, 17 mM K_2HPO_4, 300 mM KCl, 10 mM imidazole; pH 8.0) containing one tablet of 'cOmplete' ('EDTA-free Protease Inhibitor Cocktail' from Roche, Mannheim, Germany). The cell suspension was passed twice through a French press at 2000 bar (2000×10^5 Pa). Two centrifugation steps (13 100 ***g***, 4 °C, 10 min followed by 108 000 ***g***, 4 °C, 30 min) removed cell debris and unbroken cells. The supernatant was loaded onto a 5 ml HisTrap column (GE-Healthcare) which was equilibrated with HisTrap-binding buffer. When the UV signal returned to baseline the column was washed with 14 column volumes (CV) of 20 % (V/V) HisTrap washing buffer (3 mM KH_2PO_4, 17 mM K_2HPO_4, 300 mM KCl, 200 mM imidazole; pH 8.0). His_6-tagged dodecins were eluted with HisTrap elution buffer (3 mM KH_2PO_4, 17 mM K_2HPO_4, 300 mM KCl, 500 mM imidazole; pH 8.0). Following elution from a HisTrap column the fractions containing His_6-tagged dodecin were applied to a HiTrap desalting column (GE-Healthcare) and chromatographed with an aqueous buffer containing 300 mM NaCl, 20 mM Tris-HCl and 5 mM $MgCl_2$ (pH 7.5).

Detemination of flavin-binding to dodecins *in vitro* and *in vivo*

For monitoring the *in vitro* binding capacity of dodecins to flavins His_6-tagged dodecins (50 µM) were incubated in the dark in the presence or absence of flavins (250 µM) at room temperature for 15 min. Subsequently, the proteins were analysed by PAGE. Documentation of the resulting gels was accomplished using visible light, UV-light or staining with Coomassie Brilliant Blue G-250. The *E. coli* strains CpXFMN and CpXFAD were employed to *in vivo* challenge dodecins with flavins (50 µM) as described earlier [17]. A scheme of the strains CpXFMN and CpXFAD is shown in Fig. S4. CpXFMN and CpXFAD are riboflavin auxotrophic, which allowed us to better control the flavin levels in our experiments. Both strains produce a riboflavin uptake system and, in contrast to wild-type *E. coli*, are able to take up flavins from the growth medium. CpXFMN contains the gene *FMN1* from *Schizosaccharomyces pombe*. The gene product FMN1 produces FMN from riboflavin and ATP and RoFMN from roseoflavin and ATP. CpXFAD contains an additional copy of *E. coli ribCF* encoding the endogenous bifunctional flavokinase/FAD synthetase which produces FMN, FAD, RoFMN and RoFAD.

Preparation of cell-free extracts of *S. davawensis* and flavin analysis

Washed mycelia were disrupted six times for 1 min at 6 m s^{-1} using a FastPrep-24T^{M} 5G Instrument (MP Biomedicals, Santa Ana, CA, USA). Cell lysates were treated with 10 % (w/V) trichloroacetic acid and filtered through a cellulose acetate membrane (pore size 0.2 µm). Riboflavin, FMN, FAD, roseoflavin, RoFMN and RoFAD levels were determined by HPLC/MS as described earlier [17]. Intracellular flavin levels were normalized to 1 mg of total protein concentration.

Monitoring gene expression in *S. davawensis* by RT-PCR

S. davawensis strains were grown in YS to the exponential growth phase (18 h, no roseoflavin production) and to the late stationary growth phase (48 h, roseoflavin production occurred). Cultures (50 ml) were inoculated with 10^8 spores. Mycelia were harvested by centrifugation (2000 ***g***; 2 min; 4 °C) and immediately frozen in liquid nitrogen. For preparation of all solutions diethyl pyrocarbonate (DEPC)-treated water was used. Mycelium pellets were suspended in 300 µl extraction buffer (10 mM sodium acetate, 150 mM saccharose; pH 4.8) and mixed with 20 µl 20 % SDS and 300 µl phenol. Cells were disrupted using the FastPrep-24T^{M} 5G

Instrument (see above). Samples were centrifuged (15 min, 16 100 ***g***, 4 °C). Supernatants were transferred without interphase to 2 ml MaXtract High Densitiy reaction tubes (Qiagen, Hilden) and carefully mixed with 500 µl chloroform/isoamyl alcohol (25 : 1, V/V in water). This procedure was repeated and supernatants were mixed with 10 % 3 M sodium acetate (pH 6.5) and 1 vol of isopropanol. DNA and RNA were precipitated overnight at −80 °C and centrifuged (1 h, 16 100 ***g***, 4 °C). DNA/RNA mixtures were washed once with 700 µl 70 % ethanol, centrifuged for 10 min, dried at room temperature and dissolved in water. The mixtures were treated with RNase-free DNase I (Qiagen, Hilden) and RNA was again precipitated as described above. Precipitated RNA was washed with 500 µl 70 % ethanol and centrifuged again. Air-dried RNA was dissolved in water. These RNA preparations were used as templates for RT-PCR using the Maxima First Strand cDNA Synthesis Kit for RTqPCR (Thermo Fisher Scientific, Darmstadt, Germany). Oligonucleotides 1333–207 bp-fw (5′-ACC ACA CCT ACC GGG TCA CGG A-3′) and 1333–207 bp-rv (5′-AAC CGG TCT CGT CCA GGC GGA A-3′) were used for detection of the dod^{SD} transcript and oligonucleotides gltA-408bp-fw (5-ACA CCG CCG CCT ATA AAT CCG C-3′) and gltA-408bp-rv (5′-ACG GGT GGC CGA TGG ACT TCT T-3′) were used to probe the transcripts of *S. davawensis gltA* (control). The oligonucleotides bind internally and their binding specificity was validated using CLC Genomics Workbench 6 (Qiagen, Hilden, Germany).

Differential metabolome analysis

Cells from *S. davawensis* or *S. coelicolor* were harvested by centrifugation (2000 ***g***; 2 min) at 4 °C. The mycelium samples (eight biological replicates for each condition) were split and an aliquot was used to determine the dry weight. The remaining mycelium was frozen in liquid nitrogen. 2 ml of a 80 °C hot extraction solution (acetonitrile/water 60 : 40 (V/V), LCMS-grade) were added to the mycelia samples. The mixtures were incubated for 3 min at 80 °C with shaking and cooled on ice for 2 min. The samples were centrifuged at 16 860 ***g*** (4 °C), the supernatant (metabolite extract) was collected and kept on ice. The extraction was repeated once and extracts were dried under vacuum and stored at −80 °C until metabolomics analysis. Evaporated extracts were re-dissolved in LCMS-grade water to 200 µl per g l^{-1} cell dry weight and analysed in three technical replicates by injecting 10 µl. The mass spectrometric analysis was performed on a platform consisting of a Hitachi L-7100 liquid chromatography pump coupled to a Gerstel MPS2 autosampler and an Agilent 6550 IonFunnel QTOF (Agilent, Santa Clara, CA) operated with published settings in negative ionization mode [29] using automated isocratic flow-injection without prior chromatographic separation. The flow rate was 150 µl min^{-1} of mobile phase consisting of isopropanol/water (60 : 40, v/v) buffered with 5 mM ammonium fluoride at pH 9. For online mass axis correction, 2-propanol (in the mobile phase), taurocholic acid, and hexakis (1 h, 1 h, 3*H*tetrafluoropropoxy)phosphazine were added to the mobile phase. Mass spectra were recorded in profile mode from m/z 50 to 1000 with a frequency of 1.4 spectra s^{-1} for 0.48 min using the highest resolving power (4 GHz HiRes). Source temperature was set to 325 °C, with 5 l min^{-1} drying gas and a nebulizer pressure of 30 psig. Spectral processing (profile alignment, peak detection, centroiding, merging) was performed as published previously [29]. All steps of data processing and analysis were performed with Matlab R2014a (The Mathworks, Natick) using functions embedded in the 'bioinformatics', 'statistics', 'database' and 'parallel computing' toolboxes. All ions were putatively annotated as metabolites within 0.001 Da deviations from their theoretical mass using the KEGG metabolite database as reference, which can lead to ambiguous annotations for compounds with identical sum formula. The mapping between m/z features and metabolites is based on accurate mass alone. This approach inherently generates putative annotations and is unable to distinguish between isomers.

RESULTS

The distribution of dodecin-encoding genes in prokaryotes and their genomic context in Streptomycetes

The present study was initiated to investigate the role(s) of dodecins in Streptomycetes. Dodecins have not been studied in Streptomycetes before. As a first step the distribution of dodecin genes in prokaryotic species in general and more specifically in *Streptomyces* species was analysed. Dodecins are apparently more widely distributed in the domain Bacteria (894 of 1752 fully sequenced genomes contain dodecin genes) than in the Archaea (18 of 94). Genes encoding dodecins are present in all *Streptomyces* species sequenced so far and the corresponding gene products share a high degree of similarity at the amino acid level (up to 71 %) and also are highly similar (66 %) to dodecin $MtDod^{B}$ of the pathogenic actinobacterium *M. tuberculosis* [7] (Fig. S1). The key tryptophan residue indicative for flavin-binding is not present in all (putative) *Streptomyces* dodecins and possibly these dodecins bind other molecules besides flavins. The well-studied archaeal $HsDod^{A}$ shows a similarity of only 29 % to the dodecins of *S. davawensis* ($SdDod^{B}$) and *S. coelicolor* ($ScDod^{B}$), proteins, which were investigated in the present work. The dodecins of the roseoflavin producers *S. davawensis* and *S. cinnabarinus* are identical in sequence. Analysis of the genomic context of dodecin genes in *S. davawensis*, *S. cinnabarinus* and *S. coelicolor* revealed that genes devoted to ergothionein-biosynthesis are located close to the dodecin genes (Fig. 1b). Ergothioneine has antioxidant properties *in vitro* and was reported to protect *S. coelicolor* from oxidative stress [30]. The chromosomal linkage of ergothionein-biosynthesis genes and the dodecin gene led us to the speculation that these genes may have related physiological functions and that dodecins and possibly flavin metabolism in general affects the oxidative stress response in Streptomycetes. Notably, in the actinobacterium *M. tuberculosis* the ergothioneine biosynthetic genes are not linked to the dodecin gene [31].

Dodecin genes in *S. davawensis* and *S. coelicolor* are expressed in the exponential as well as in the stationary growth phase

The following experiment was performed to show that the (putative) dodecin genes in *S. davawensis* and *S. coelicolor* indeed are expressed and do not represent pseudogenes. *S. davawensis* cells were grown to the stationary phase (72 h). At this stage the culture medium turned red due to synthesis of the red antibiotic roseoflavin. Cells of a separate *S. davawensis* culture were harvested earlier (in the exponential growth phase; 28 h) where no roseoflavin was present (the culture was white). Total RNA of cells from both cultures was analysed by RT-PCR using dod^{SD}- and *gltA2*-specific primers (Fig. 2). The latter primer pair was employed to monitor expression of the gene-encoding citrate synthase in *S. davawensis* (EC 2.3.3.1; $gltA2^{SD}$). Constitutive expression of the homologous gene $gltA2^{SC}$ was reported for *S. coelicolor* and thus this reaction was used as a control [32]. We found dod^{SD}- and $gltA2^{SD}$-specific PCR products using both total RNA samples as templates for RT-PCR indicating that both genes were expressed in both growth phases. Cells from *S. coelicolor* were analysed in a similar way and the results suggested as well that expression of dod^{SC} and $gltA2^{SC}$ occurred in the exponential and in the stationary growth phase. Dodecins appear to play a role in both growth phases and the following experiments were initiated to shed more light on the function of the flavin-binding dodecins in Streptomycetes.

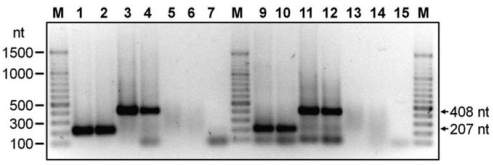


Fig. 2. (Putative) Dodecin genes are expressed in cells harvested from liquid cultures of *S. davawensis* and *S. coelicolor*. Agarose gel electrophoresis in combination with ethidium bromide staining was used to detect *dod*-specific PCR products generated by RT-PCR. Total RNA was isolated from exponentially growing cells and from cells entering the stationary growth phase. This RNA was used to generate *dod* cDNAs employing *dod*-specific oligonucleotides. In parallel expression of *gltA2* encoding citrate synthase was monitored as a control. Lanes M, nucleotide marker, 100 nucleotide pair (nt) ladder. Lane 1, RT-PCR product of the expected size (207 nt) corresponding to the *S. davawensis* dod^{SD} gene when total RNA was isolated from exponentially growing cells. Lane 2, as lane 1, however, total RNA from stationary phase cells was used as a template. Lanes 3 and 4, as lanes 1 and 2, however, oligonucleotides specific for the *S. davawensis* $gltA^{SD}$ gene were used for RT-PCR generating the expected 408 nt amplicon. Lanes 5 and 6, negative controls, RT-PCR reactions where no reverse transcriptase was added to the reaction (total RNAs from the exponential (lane 5) and stationary (lane 6) growth phases). Lane 7, negative control, RT-PCR reaction in the absence of total RNA. Lanes 9–15, same as in lanes 1–7, however, total RNA from *S. coelicolor* and oligonucleotides specific for the *S. coelicolor* dod^{SC} and $gltA^{SC}$ genes were used, respectively.

Deletion of dod^{SD} from the chromosome of *S. davawensis* leads to changes in flavin levels

To study dodecin function with regard to flavin metabolism in *S. davawensis* the corresponding gene dod^{SD} was disrupted. Successful gene deletion within the chromosomes of the resulting *S. davawensis* strains was verified by PCR and DNA sequencing revealing that dod^{SD} had been replaced by an apramycin resistance cassette (*aacIV*) (Fig. S2). In addition, two independently isolated deletion strains (*S. davawensis* Δdod.1 and *S. davawensis* Δdod.2) were analysed by RT-PCR and, as expected, a dodecin-specific cDNA could not be detected validating deletion of dod^{SD} (Fig. S3). *S. davawensis* Δdod.1 was cultivated in the standard yeast starch-medium (YS) and total cellular flavin levels (riboflavin, FMN and FAD) were compared to *S. davawensis* wild-type during the time course of growth (Fig. 3). Apparently, riboflavin present in YS (4 µM) as well as riboflavin generated intracellularly by riboflavin-prototrophic *S. davawensis* had almost completely been converted to FMN and FAD at all times (1, 2, 7, 10 and 14 days). In contrast to *E. coli* [23], the most abundant flavin within the cytoplasm was FMN (ca. 60 %) and not FAD. At the end of the growth

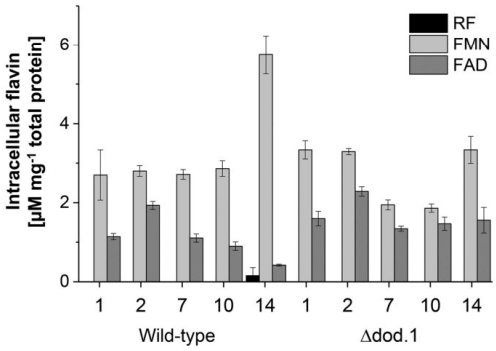


Fig. 3. Disruption of the dodecin gene (dod^{SD}) in *S. davawensis* leads to a change in flavin levels. *S. davawensis* Δdod.1 was cultivated in YS broth and total flavin levels in cell-free extracts (µM flavins normalized to 1 mg total protein of the cell-fee extract) were compared to *S. davawensis* wild-type. Mycelia of the two different strains were harvested at different times during cultivation (days 1, 2, 7, 10 and 14), disrupted and cell-free extracts were treated with trichloroacetic acid to fully denature all proteins and to release all flavins. Flavin levels in the wild-type strain are represented by the left columns whereas the corresponding data of Δdod.1 are shown in the right part of the figure. Riboflavin levels as well as FMN and FAD, which were generated in the cytoplasm of *S. davawensis* by the bifunctional flavokinase/FAD synthetase RibCF [35], were determined by LC/MS. The data represent mean values from three independent experiments with the indicated standard deviation. The total biomasses at the end of these experiments were similar in the wild-type and in Δdod.1 (not shown).

experiment (after 14 days of incubation) *S. davawensis* Δdod.1 contained less FMN, and FMN levels were also reduced at days 7 and 10 when compared to the wild-type. The total biomasses of the wild-type strain and Δdod.1 were similar in all growth experiments (also when the strains were cultivated on a minimal medium) and we conclude that $SdDod^B$ is not an essential protein under these conditions and that altered FMN levels do not affect growth.

The dodecins from *S. davawensis* and *S. coelicolor* bind flavins and flavin analogues

The following experiments were carried out to show that the dodecins from *S. davawensis* ($SdDod^B$) and *S. coelicolor* ($ScDod^B$) indeed are flavin-binding proteins. As a first step *in vivo* binding of flavins was monitored employing the specialized *E. coli* strains CpXFMN and CpXFAD (Fig. S4) [17, 33]. These *E. coli* strains were used for expression experiments since genetic manipulation and cell disruption was much easier when compared to *Streptomyces* strains. The genes dod^{SD} and dod^{SC} were overexpressed separately in CpXFMN and in CpXFAD. *E. coli* CpXFMN contains an additional flavokinase gene *FMN1* (from *Schizosaccharomyces pombe*) and especially supports formation of FMN/RoFMN within the cytoplasm of the recombinant bacterium which means that these flavins are present at higher levels when compared to the wild-type strain. *E. coli* CpXFAD contains an additional flavokinase/FAD synthetase gene (*ribCF*) and especially supports formation of FAD/RoFAD. CpXFMN and CpXFAD were transformed with expression plasmids containing dod^{SD} and dod^{SC}. At the time point of induction of $SdDod^B$ and $ScDod^B$ riboflavin or roseoflavin were added to the culture medium of the different *dod* overexpressing strains to challenge the nascent recombinant dodecins with regard to flavin-binding. The *E. coli* strains were grown further to the early stationary growth phase and disrupted. $SdDod^B$ and $ScDod^B$ contained a His_6-tag and thus could rapidly be purified to apparent homogeneity from the resulting cell-free extracts. Purified $SdDod^B$ and $ScDod^B$ preparations were denatured and their flavin content was determined by HPLC/MS. When the different strains were challenged with riboflavin (50 µM) the corresponding dodecins contained mostly FMN (Table 1) and no major differences with regard to the expression hosts CpXFMN or CpXFAD were observed. When cells were challenged with roseoflavin (50 µM) the results were different depending on the *E. coli* host. When CpXFMN was used, $ScDod^B$ contained large amounts of roseoflavin/RoFMN whereas $SdDod^B$ predominantly contained FMN. When CpXFAD was challenged with roseoflavin again $ScDod^B$ appeared to bind more roseoflavin/RoFMN when compared to $SdDod^B$. FAD or RoFAD were not released from these dodecins (Table 1) although *E. coli* CpXFAD generates more FAD/RoFAD within the cytoplasm when compared to CpXFMN [17]. These experiments were done only once, however, two different *E. coli* strains were used for this *in vivo* analysis. Our working hypothesis that $SdDod^B$ would bind more roseoflavin/RoFMN clearly was not supported by these results and our conclusion that $SdDod^B$ is not superior with regard to binding roseoflavin/RoFMN seems to be justified.

The dodecameric form of dodecin from *S. davawensis* is stabilized upon binding of roseoflavin *in vitro*

S. davawensis and *S. cinnabarinus* are the only bacteria known to synthesize roseoflavin and the primary structures of their dodecins $SdDod^B$ and $ScinDod^B$ are identical. We hypothesized that these bacteria may contain specialized dodecins with specific roles in roseoflavin metabolism. As a prerequisite to this role roseoflavin-binding to $SdDod^B$ was

Table 1. Dodecins from *S. davawensis* and *S. coelicolor* bind flavins *in vivo*

Dodecin from:	Feed[a]	RF (%)[b]	FMN (%)[b]	FAD (%)[b]	RoF (%)[b]	RoFMN (%)[b]	RoFAD (%)[b]
Test strain: CpXFMN							
S. davawensis	RF	12.4	87.6	0.0	0.0	0.0	0.0
S. coelicolor	RF	0.0	100.0	0.0	0.0	0.0	0.0
S. davawensis	RoF	0.7	84.0	0.0	0.8	13.7	0.0
S. coelicolor	RoF	1.1	6.9	0.0	50.3	41.7	0.0
Test strain: CpXFAD							
S. davawensis	RF	8.2	91.8	0.0	0.0	0.0	0.0
S. coelicolor	RF	0.0	100.0	0.0	0.0	0.0	0.0
S. davawensis	RoF	0.7	32.7	0.1	12.0	53.4	1.1
S. coelicolor	RoF	5.0	10.5	0.0	36.9	47.6	0.0

This experiment was done only once.

a 50 µM riboflavin (RF) or roseoflavin (RoF) were added to the growth medium of *E. coli* CpXFMN or CpXFAD (see also Fig. S4); these strains overproduced His_6-tagged dodecin from either *S. davawensis* or from *S. coelicolor*. The recombinant dodecins were purified to apparent homogeneity from CpXFMN or CpXFAD and denatured. The released flavins were analysed by HPLC/MS. The dodecin amounts were determined using a Bradford assay.

b The maximum flavin load (saturation) of a dodecin is 100 %; the numbers show what kinds of different flavins were bound to the different dodecins depending on the flavin feed in the culture of the test strain.

analysed *in vitro*. A His_6-tagged $SdDod^B$ recombinant protein was overproduced in *E. coli* Rosetta 2(DE3) allowing expression of genes that contain codons rarely used *in E. coli*. His_6-tagged $SdDod^B$ was purified to apparent homogeneity from *E. coli* cells which were not challenged with flavins. When purified under these conditions $SdDod^B$ did not contain bound flavins. *S. davawensis* dodecin (50 µM) was incubated in the presence of a fivefold excess of roseoflavin (250 µM) for 15 min. As a control, $SdDod^B$ was also incubated in the absence of roseoflavin. Subsequently, $SdDod^B$ was analysed by PAGE in the absence of denaturing SDS (Fig. 4). In the presence of roseoflavin an $SdDod^B$ dodecamer was found whereas in the absence of roseoflavin (control) only the monomeric form of $SdDod^B$ was present. In the presence of riboflavin this shift did not occur. The structures of riboflavin and roseoflavin are considerably different (Fig. 1a) and the additional dimethylamino group at C8 of roseoflavin appears to trigger dodecamer formation. The molecular reason(s) for this, however, remain(s) unclear and also whether this multimerization in the presence of roseoflavin occurs under physiological conditions in the *Streptomyces* cell. Still, these findings prompted us to further investigate the role of $SdDod^B$ with regard to roseoflavin-binding and with regard to a possible role in roseoflavin metabolism. Notably, $ScDod^B$ was analysed in the same way (data not shown). Under the applied conditions (and in contrast to $SdDod^B$) $ScDod^B$ was always present as a dodecamer (even in the absence of flavins). Notably, dodecameric $ScDod^B$ was found to bind roseoflavin, riboflavin or FMN when treated with these flavins prior to loading the samples to the gel.

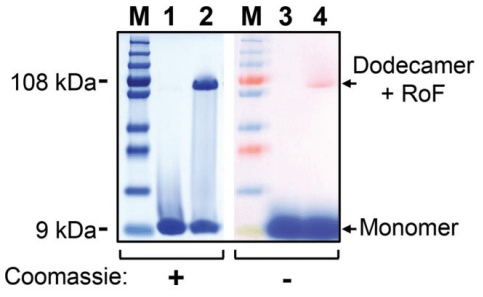


Fig. 4. Roseoflavin affects multimerization of dodecin from *S. davawensis*. Apododecin from *S. davawensis* was purified from a recombinant *E. coli* strain and incubated for 15 min in the absence (lanes 1, 3) and presence (lanes 2, 4) of 250 µM roseoflavin. Protein samples were analysed by native PAGE with a running buffer containing SDS. The left part of the gel (lanes 1 and 2) was stained with Coomassie Brilliant Blue G-250. The right part (lanes 3 and 4) was not stained, however, the running buffer contained Coomassie Brilliant Blue G-250 which is the reason for the band at 10 kDa corresponding to the monomer of dodecin. The right part was evaluated under visible light and the red band at 108 kDa shows that roseoflavin is part of the *S. davawensis* dodecin complex. A shift from 9 to 108 kDa occurred in the presence of roseoflavin, indicating formation of the dodecin dodecamer. When the protein was treated with riboflavin such a shift did not occur (not shown). Lanes M contain the molecular weight marker.

S. davawensis dodecin is not involved in roseoflavin resistance and/or roseoflavin biosynthesis

The following growth experiments were carried out to study a possible function of dodecin in roseoflavin resistance or biosynthesis in *S. davawensis*. FMN is the direct substrate for the key enzyme of roseoflavin biosynthesis (RosB) [27, 34] and we therefore hypothesized that the flavin-binding dodecins could, in a direct or indirect way, affect roseoflavin production. *S. davawensis* wild-type and the dod^{SD} deletion strains Δdod.1 and Δdod.2 (see above) were grown on solid YS which is the only growth medium known to support synthesis of roseoflavin [12]. In addition to endogenous roseoflavin (all tested strains synthesized roseoflavin) exogenous roseoflavin (200 µM) was added to the YS-plates. This amount of roseoflavin exceeds the naturally generated amount of roseoflavin by a factor of 10. *S. davawensis* wild-type and the dod^{SD} deletion strains Δdod.1 and Δdod.2 showed similar growth whereas *S. coelicolor* wild-type (which naturally does not produce roseoflavin) was not able to grow in the presence of roseoflavin (Fig. 5a). *S. davawensis* wild-type and Δdod.1 were grown in liquid YS and roseoflavin levels were monitored by HPLC (Fig. 5b). No differences in roseoflavin levels were detected even after 14 days of cultivation. In summary these results suggest that dodecin neither plays a direct role in protection from toxic roseoflavin, RoFMN or RoFAD nor in roseoflavin biosynthesis.

Heterologous expression of the dodecin gene from *S. davawensis* in *Bacillus subtilis* does not lead to roseoflavin resistance of the resulting recombinant strains

B. subtilis is especially sensitive towards roseoflavin and serves as a model system to study roseoflavin activity [14, 15, 35]. To validate the findings of the previous section which suggested that $SdDod^B$ does not protect from the toxic effects of roseoflavin, growth experiments with different recombinant *B. subtilis* strains were performed. These strains contained the dodecin genes from either *S. davawensis* or, as a control, *S. coelicolor* under control of a strong IPTG-inducible promoter. The dodecin genes were codon-adapted for efficient expression in *B. subtilis* and expressed as two different versions. One version did not contain codons specifying a His_6-tag, a second version did contain codons specifying a His_6-tag to allow immunological detection and verification of synthesis of the gene product. Successful expression of the dodecin genes in *B. subtilis* was validated upon detection of His_6-tagged dodecins in cell-free extracts of induced recombinant strains overproducing $SdDod^B$ or $ScDod^B$ (Fig. S5). Growth experiments in liquid growth media were performed in the presence/absence of riboflavin (control) or roseoflavin and the results clearly showed that the presence of the dodecin genes from *S. davawensis* (or *S. coelicolor*) did not lead to protection of

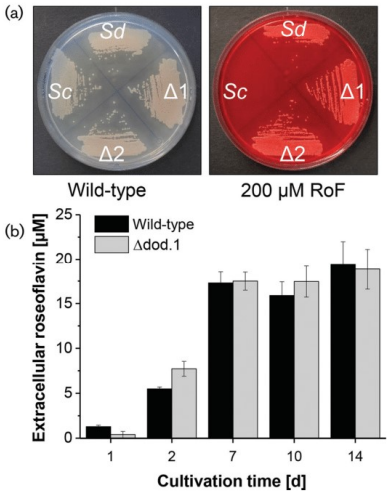


Fig. 5. Deletion of the dodecin gene in *S. davawensis* neither leads to roseoflavin sensitivity nor to a reduction of roseoflavin synthesis. (a) In total, 50 000 spores of each *S. davawensis* wild-type (Sd), two independent dodecin deletion strains (Δdod1 or Δ1; Δdod2 or Δ2) or *S. coelicolor* (Sc) were spread on YS plates and incubated for 2 days in the absence of light (to avoid degradation of roseoflavin). The left plate (control) did not contain roseoflavin whereas the right plate contained 200 µM roseoflavin (RoF). Sc could not grow in the presence of RoF whereas Sd, Δ1 and Δ2 were able to grow in the presence of this antibiotic. If $SdDod^{B}$ would contribute significantly to roseoflavin resistance no growth of the dodecin deletion strains would have been expected. (b) YS broth was inoculated with 50 000 spores either from *S. davawensis* wild-type or from Δdod1 (Δ1). Cells were grown for 14 days. At indicated times, samples were taken from the supernatants of the cultures and analysed by HPLC with regard to roseoflavin levels. The data show that deletion of the *dod* gene does not strongly affect roseoflavin synthesis under these conditions. Three independent experiments were carried out and the columns show the mean values and standard deviations, respectively.

B. subtilis with regard to the antibiotic roseoflavin (Fig. S6). A similar experiment was performed using solid growth media and an inoculum of 50 000 cells. Similar results were obtained namely that expression of dodecin genes neither led to complete roseoflavin resistance nor to an elevated tolerance of *B. subtilis* with regard to this antibiotic (Fig. S7).

Disruption of the dodecin gene leads to an increased tolerance towards plumbagin in *S. davawensis*

Previous works suggested that dodecins have a function in protection against radical or oxygenic stress and since we could rule out a specific role of dodecin in roseoflavin resistance of *S. davawensis* we anticipated that dodecins may have a more general role in the strictly aerobic Streptomycetes [4]. To test such a possible role the following experiments were carried out challenging different *Streptomyces* strains with plumbagin, nitrofurantoin and paraquat. These compounds are known to promote formation of O_2^- and are commonly used to mimic conditions of oxidative stress [36, 37]. *S. davawensis* wild-type, Δdod.1 and Δdod.2 were cultivated for 5 days at 30 °C in liquid YS. Mycelia of all strains were harvested and their dry weight was determined (Fig. 6). In contrast to all other strains, *S. davawensis* wild-type was not able to grow in the presence of plumbagin (20 µM) as only small amounts of mycelium were found at the end of the growth experiments. The deletion strains Δdod.1 and Δdod.2 could perfectly grow under these conditions and appeared to be fully resistant to this compound (Fig. 6). We were not able to complement the Δdod strains by re-introducing the wild-type gene as only one resistance cassette is known to work in *S. davawensis*. *S. coelicolor* wild-type, containing a functional dodecin gene, grew in the presence of plumbagin and thus was plumbagin-resistant (data not shown). This finding suggested that dod^{SD} had a specific function in *S. davawensis* with regard to oxidative stress.

Effects of dodecin deletion on plumbagin stressed Streptomycete metabolomes

As a first step towards elucidation of dodecin function with regard to oxidative stress metabolome analyses of *S. davawensis* wild-type, Δdod.2 and, as a control, *S. coelicolor* were carried out. The different strains were cultivated in YS in the presence/absence of sublethal amounts of plumbagin (15 µM). Polar intracellular metabolites were extracted and analysed using non-targeted flow injection time-of-flight mass spectrometry [29]. In total, 180 datasets were generated and 12 730 ions were detected, of which 1673 ions could be annotated as metabolites based on matching their accurate masses with the Kyoto Encyclopedia of Genes and Genomes (KEGG) *S. davawensis* compound database [38] (see Dataset M in the supplementary material). To obtain an overview of the main sources of variance between the samples (and to understand which metabolites contributed most to the characteristic metabolic differences in the samples) a principal component analysis (PCA) was carried out (Fig. S8a). As expected, water and medium samples separated clearly from the cellular extracts along the first principal component, which accounted for 70 % of the variance. Moreover, there was a clear separation between samples from *S. coelicolor* and *S. davawensis* along the second principal component accounting for 10 % of the variance. Effects of plumbagin treatment was found to be masked by the above described sources of variance, and for all further analysis the medium and water controls were removed. In a second step, a PCA of the cellular extracts only was carried out (Fig. S8b). The most influential factor on the remaining variance again was the taxonomic split between the two species *S. coelicolor* and *S. davawensis* (first principal component, 37.2 % of variance). The second principal component (19 % of variance) mainly explained the strong metabolome differences between plumbagin-treated *S. davawensis* Δdod.2 and all other samples, confirming the observations in the

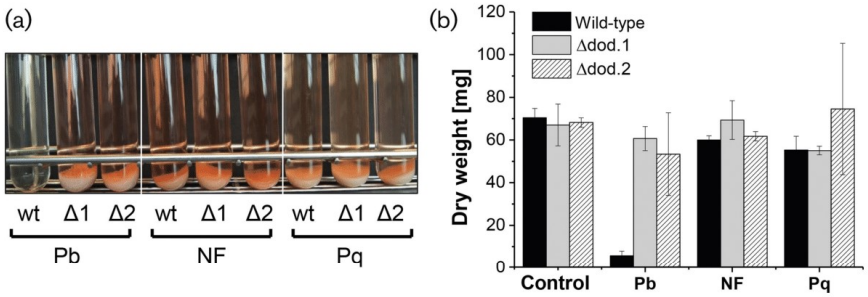


Fig. 6. Deletion of the dodecin gene in *S. davawensis* leads to an increase in resistance towards plumbagin. Two independent *S. davawensis dod*^{SD} deletion strains (Δdod1 or Δ1; Δdod2 or Δ2) and the wild-type (wt) were grown to the early stationary growth phase (5 days) in the dark in YS broth. All cultures were inoculated with 10^5 spores. Prior to inoculation, plumbagin (Pb, 20 µM), nitrofurantoin (NF, 100 µM) or paraquat (Pa, 500 µM) were added as shown. These compounds are known to induce oxidative stress in a variety of bacteria [22, 36, 37]. (a) After 5 days the cells were harvested by centrifugation. In contrast to the deletion strains, *S. davawensis* wild-type cannot grow in the presence of plumbagin. Nitrofurantoin and paraquat did not affect growth under these conditions. (b) The dry weights of the mycelia [generated as shown in (a)] were determined. Three independent growth experiments [one example is shown in (a)] were carried out and the columns in (b) show mean values and the standard deviations.

previous section, where deletion of the dodecin gene had a considerable influence on the response to plumbagin by *S. davawensis*. When focusing on *S. coelicolor*, the main source of variance was plumbagin treatment although the metabolic response was much less pronounced when compared to the *S. davawensis* strains.

As a next step in our metabolome analysis a differential analysis was carried out to identify metabolite ions of which the cellular abundance responded to plumbagin treatment. We calculated $\log_2$ fold-changes and *P*-values (Student's *t*-test, two-tailed, unpaired) for each ion between plumbagin-treated and untreated samples. Plumbagin caused complex effects on the metabolome of all strains and significant changes in the levels of mycothiol and mycothione indicated that the cells were indeed experiencing oxidative stress [30, 39]. Mycothiol is the major thiol found in Actinobacteria and has many of the functions of glutathione, which is the dominant thiol in other bacteria and eukaryotes [40]. Mycothione is the oxidized form of mycothiol. Plumbagin responses between the two *Streptomyces* species were predominantly species-specific and only a few common differential metabolites were observed (Fig. S9a, b). Interestingly, the overall number of differential metabolites between the dodecin mutant Δdod.2 and wild-type *S. davawensis* was increased in the presence of plumbagin, consistent with a specific but yet unknown physiological role of dodecins in Streptomycetes. Interestingly, the metabolomics data revealed that the observed metabolic responses to plumbagin treatment were more similar between *S. davawensis* Δdod.2 and *S. coelicolor* wild-type than between *S. davawensis* Δdod.2 and *S. davawensis* wild-type. Notably, levels of other chemicals also changed. However, we could not assign these changes to known biological processes. This is why these metabolites are not discussed here.

Metabolome analyses suggest a rather general role for *Streptomyces* dodecins

We also investigated the metabolic consequences of dodecin gene deletion in the absence of plumbagin, again by performing a differential analysis. The deletion of the dodecin gene in *S. davawensis* had profound effects on the metabolome (Fig. S9c). We observed a strong depletion of xanthine and other compounds related to nucleotide metabolism. In contrast, a strong accumulation of carotenoids, an unusual antioxidant (2-S-glutathionyl)acetyl glutathione), a large number of polysaccharides (trehalose, trehalose 6-phosphate, glycogen, tri- and tetrasaccharides) and an osmoprotectant (hydroxyectoine) occurred. Since many carbohydrates have beneficial effects on stressed cells as protein stabilizing compounds, this could explain increased plumbagin-tolerance of the dodecin deletion strain.

DISCUSSION

Our comparative study on dodecins from *S. davawensis* and *S. coelicolor* was initiated to investigate a possible specific function of these small flavin-binding proteins in roseoflavin resistance and metabolism in *S. davawensis* and *S. cinnabarinus*. The dodecins in these two bacteria are identical and *S. davawensis* and *S. cinnabarinus* are the only organisms known to synthesize toxic roseoflavin. When a His_6-tagged version of *M. tuberculosis* dodecin ($MtDod^B$) was purified from a recombinant *E. coli* strain (which had been challenged with roseoflavin) it contained roseoflavin without a significant riboflavin contamination [33]. This study led us to the speculation that dodecins may play a vital role

in roseoflavin resistance of *S. davawensis* and *S. cinnabarinus* by absorbing toxic roseoflavin. Deletion of the *S. davawensis* dodecin gene dod^{SD} did not lead to roseoflavin sensitivity of *S. davawensis* and overexpression of dod^{SD} in *B. subtilis* did not cause roseoflavin resistance in the resulting recombinant strain. When challenging $ScDod^{B}$ and $SdDod^{B}$ with roseoflavin *in vivo* both dodecins appear to bind roseoflavin and/or RoFMN. *S. coelicolor* is not a roseoflavin producer. The fact that $ScDod^{B}$ apparently is also able to bind the toxic flavins as well argues against our hypothesis that a highly specialized $SdDod^{B}$ in *S. davawensis* would absorb and thus inactivate toxic roseoflavin when cells enter the roseoflavin production phase. Recent studies on the roseoflavin biosynthetic enzymes RosA and RosB revealed that both enzymes appear to not release their toxic reaction products to the cytoplasm (roseoflavin and 8-demethyl-8-aminoriboflavin, respectively). This, in addition to a specialized FMN riboswitch [13], at present are the only known mechanisms of resistance to roseoflavin [27, 34].

We found that dodecins of *S. davawensis* and *S. coelicolor* both preferentially bind FMN over other flavins which is in line with previous reports on bacterial dodecins [6, 7]. The fact that deletion of dod^{SD} in *S. davawensis* leads to reduced overall flavin levels can be explained as follows. In *S. davawensis* riboflavin biosynthesis is regulated by an FMN-responsive riboswitch reducing expression of the riboflavin biosynthetic genes in response to elevated FMN levels [13]. In strain *S. davawensis* Δdod.1, where dodecin is not available to bind FMN, levels of free cytoplasmic FMN probably were higher when compared to the wild-type. Increased FMN levels reduced FMN-riboswitch-controlled expression of riboflavin biosynthetic genes leading to the observed reduction of cellular flavin levels.

Dodecins bind flavins with high affinity if the flavins are oxidized, whereas flavin reduction induces formation of apododecins and free flavins [41]. $SdDod^{B}$ as well as $ScDod^{B}$ appear to be abundant proteins, they preferentially bind FMN and we therefore speculated that these proteins would affect the equilibrium between FMN and $FMNH_2$ within cells. This equilibrium in turn is relevant with regard to oxidative stress management since reduced flavins were reported to be responsible for generation of highly reactive O_2^- and H_2O_2 [22]. The finding that the *S. davawensis* dod^{SD} deletion strain, in contrast to the wild-type, was fully resistant to treatment with plumbagin prompted us to carry out untargeted metabolome studies. The metabolomes of *S. davawensis* and *S. coelicolor* were found to be surprisingly different in the absence but also in the presence of plumbagin. At present we do not know why the *S. davawensis* dod^{SD} deletion strain clearly is more resistant to treatment with plumbagin than the wild-type strain – it appears that the dod^{SD} deletion strain is able to more quickly react to oxidative stress.

In summary, we think that dodecin serves as an FMN storage protein in Streptomycetes affecting flavin mediated regulation of riboflavin biosynthesis and as a result the flavin load and activity of a variety of flavoenzymes. Any perturbation of dodecin function should accordingly have a broad impact on cellular physiology, an effect which indeed was observed in our present study.

Funding information
This work was funded by the research training group NANOKAT (FKZ 0316052A) of the German Federal Ministry of Education and Research.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. **Fischer M, Bacher A.** Biosynthesis of flavocoenzymes. *Nat Prod Rep* 2005;22:324–350.
2. **Fraaije MW, Mattevi A.** Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem Sci* 2000;25:126–132.
3. **Macheroux P, Kappes B, Ealick SE.** Flavogenomics – a genomic and structural view of flavin-dependent proteins. *Febs J* 2011;278:2625–2634.
4. **Bieger B, Essen LO, Oesterhelt D.** Crystal structure of halophilic dodecin: a novel, dodecameric flavin binding protein from *Halobacterium salinarum*. *Structure* 2003;11:375–385.
5. **Grininger M, Staudt H, Johansson P, Wachtveitl J, Oesterhelt D.** Dodecin is the key player in flavin homeostasis of archaea. *J Biol Chem* 2009;284:13068–13076.
6. **Meissner B, Schleicher E, Weber S, Essen LO.** The dodecin from *Thermus thermophilus*, a bifunctional cofactor storage protein. *J Biol Chem* 2007;282:33142–33154.
7. **Liu F, Xiong J, Kumar S, Yang C, Ge S *et al.*** Structural and biophysical characterization of *Mycobacterium tuberculosis* dodecin Rv1498A. *J Struct Biol* 2011;175:31–38.
8. **Vinzenz X, Grosse W, Linne U, Meissner B, Essen LO.** Chemical engineering of *Mycobacterium tuberculosis* dodecin hybrids. *Chem Commun* 2011;47:11071–11073.
9. **Abbas CA, Sibirny AA.** Genetic control of biosynthesis and transport of riboflavin and flavin nucleotides and construction of robust biotechnological producers. *Microbiol Mol Biol Rev* 2011;75:321–360.
10. **Staudt H, Oesterhelt D, Grininger M, Wachtveitl J.** Ultrafast excited-state deactivation of flavins bound to dodecin. *J Biol Chem* 2012;287:17637–17644.
11. **Jankowitsch F, Schwarz J, Rückert C, Gust B, Szczepanowski R *et al.*** Genome sequence of the bacterium *Streptomyces davawensis* JCM 4913 and heterologous production of the unique antibiotic roseoflavin. *J Bacteriol* 2012;194:6818–6827.
12. **Otani S, Takatsu M, Nakano M, Kasai S, Miura R.** Letter: Roseoflavin, a new antimicrobial pigment from *Streptomyces*. *J Antibiot* 1974;27:88–89.
13. **Pedrolli DB, Matern A, Wang J, Ester M, Siedler K *et al.*** A highly specialized flavin mononucleotide riboswitch responds differently to similar ligands and confers roseoflavin resistance to *Streptomyces davawensis*. *Nucleic Acids Res* 2012;40:8662–8673.
14. **Lee ER, Blount KF, Breaker RR.** Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol* 2009;6:187–194.
15. **Ott E, Stolz J, Lehmann M, Mack M.** The RFN riboswitch of *Bacillus subtilis* is a target for the antibiotic roseoflavin produced by *Streptomyces davawensis*. *RNA Biol* 2009;6:276–280.
16. **Wang H, Mann PA, Xiao L, Gill C, Galgoci AM *et al.*** Dual-targeting small-molecule inhibitors of the *Staphylococcus aureus* FMN Riboswitch disrupt riboflavin homeostasis in an infectious setting. *Cell Chem Biol* 2017;24:576–588.
17. **Langer S, Hashimoto M, Hobl B, Mathes T, Mack M.** Flavoproteins are potential targets for the antibiotic roseoflavin in *Escherichia coli*. *J Bacteriol* 2013;195:4037–4045.

18. **Langer S, Nakanishi S, Mathes T, Knaus T, Binter A *et al.*** The flavoenzyme azobenzene reductase AzoR from *Escherichia coli* binds roseoflavin mononucleotide (RoFMN) with high affinity and is less active in its RoFMN form. *Biochemistry* 2013;52:4288–4295.
19. **Pedrolli D, Langer S, Hobl B, Schwarz J, Hashimoto M *et al.*** The *ribB* FMN riboswitch from *Escherichia coli* operates at the transcriptional and translational level and regulates riboflavin biosynthesis. *Febs J* 2015;282:3230–3242.
20. **Pedrolli DB, Jankowitsch F, Schwarz J, Langer S, Nakanishi S *et al.*** Natural riboflavin analogs. *Methods Mol Biol* 2014;1146: 41–63.
21. **Pedrolli DB, Mack M.** Bacterial flavin mononucleotide riboswitches as targets for flavin analogs. *Methods Mol Biol* 2014;1103:165–176.
22. **Imlay JA.** The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Rev Microbiol* 2013;11:443–454.
23. **Bennett BD, Kimball EH, Gao M, Osterhout R, van Dien SJ *et al.*** Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* 2009;5:593–599.
24. **Pedrolli DB, Nakanishi S, Barile M, Mansurova M, Carmona EC *et al.*** The antibiotics roseoflavin and 8-demethyl-8-amino-riboflavin from *Streptomyces davawensis* are metabolized by human flavokinase and human FAD synthetase. *Biochem Pharmacol* 2011; 82:1853–1859.
25. **Zhulin IB.** Databases for Microbiologists. *J Bacteriol* 2015;197: 2458–2467.
26. **Bertani G.** Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J Bacteriol* 2004;186:595–600.
27. **Schwarz J, Konjik V, Jankowitsch F, Sandhoff R, Mack M.** Identification of the key enzyme of roseoflavin biosynthesis. *Angew Chem Int Ed Engl* 2016;55:6103–6106.
28. **Coppée JY, Auger S, Turlin E, Sekowska A, Le Caer JP *et al.*** Sulfur-limitation-regulated proteins in *Bacillus subtilis*: a two-dimensional gel electrophoresis study. *Microbiology* 2001;147:1631–1640.
29. **Fuhrer T, Heer D, Begemann B, Zamboni N.** High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal Chem* 2011;83: 7074–7080.
30. **Nakajima S, Satoh Y, Yanashima K, Matsui T, Dairi T.** Ergothioneine protects *Streptomyces coelicolor* A3(2) from oxidative stresses. *J Biosci Bioeng* 2015;120:294–298.
31. **Saini V, Cumming BM, Guidry L, Lamprecht DA, Adamson JH *et al.*** Ergothioneine maintains redox and bioenergetic homeostasis essential for drug susceptibility and virulence of *Mycobacterium tuberculosis*. *Cell Rep* 2016;14:572–585.
32. **Gatewood ML, Bralley P, Weil MR, Jones GH.** RNA-Seq and RNA immunoprecipitation analyses of the transcriptome of *Streptomyces coelicolor* identify substrates for RNase III. *J Bacteriol* 2012; 194:2228–2237.
33. **Mathes T, Vogl C, Stolz J, Hegemann P.** *In vivo* generation of flavoproteins with modified cofactors. *J Mol Biol* 2009;385:1511–1518.
34. **Konjik V, Brünle S, Demmer U, Vanselow A, Sandhoff R *et al.*** The crystal structure of RosB: insights into the reaction mechanism of the first member of a family of flavodoxin-like enzymes. *Angew Chem Int Ed Engl* 2017;56:1146–1151.
35. **Grill S, Busenbender S, Pfeiffer M, Köhler U, Mack M.** The bifunctional flavokinase/flavin adenine dinucleotide synthetase from *Streptomyces davawensis* produces inactive flavin cofactors and is not involved in resistance to the antibiotic roseoflavin. *J Bacteriol* 2008;190:1546–1553.
36. **Imlay JA.** Diagnosing oxidative stress in bacteria: not as easy as you might think. *Curr Opin Microbiol* 2015;24:124–131.
37. **Farr SB, Natvig DO, Kogoma T.** Toxicity and mutagenicity of plumbagin and the induction of a possible new DNA repair pathway in *Escherichia coli*. *J Bacteriol* 1985;164:1309–1316.
38. **Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M *et al.*** Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 2014;42:D199–D205.
39. **Park JH, Roe JH.** Mycothiol regulates and is regulated by a thiol-specific antisigma factor RsrA and sigma(R) in *Streptomyces coelicolor*. *Mol Microbiol* 2008;68:861–870.
40. **Patel MP, Blanchard JS.** *Mycobacterium tuberculosis* mycothione reductase: pH dependence of the kinetic parameters and kinetic isotope effects. *Biochemistry* 2001;40:5119–5126.
41. **Gutiérrez Sánchez C, Su Q, Schönherr H, Grininger M, Nöll G.** Multi-ligand-binding flavoprotein dodecin as a key element for reversible surface modification in nano-biotechnology. *ACS Nano* 2015;9:3491–3500.
42. **Camus JC, Pryor MJ, Médigue C, Cole ST.** Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 2002;148:2967–2973.

Edited by: M. Elliot and S. V. Gordon

Characterization of the small flavin-binding dodecin in the roseoflavin producer *Streptomyces davawensis*

Running title: *Streptomyces* dodecins

Petra Ludwig[1], Daniel C Sévin[2], Tobias Busche[3], Jörn Kalinowski[3], Florian Bourdeaux[4], Martin Grininger[4] and Matthias Mack[1#]

[#]Corresponding author.

[1]Institute for Technical Microbiology, Mannheim University of Applied Sciences, 68163 Mannheim, Germany

[2]Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland

[3]Center for Biotechnology, Bielefeld University, 33615 Bielefeld, Germany

[4]Institute of Organic Chemistry and Chemical Biology, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

Correspondent footnote:

Matthias Mack

Institute for Technical Microbiology

Mannheim University of Applied Sciences

Paul-Wittsack-Str. 10

68163 Mannheim, Germany

Tel.:+49-621-292-6496

Fax:+49-621-292-6420

E-mail: m.mack@hs-mannheim.de

Overview Appendix

1. TABLE S1 Strains and plasmids used in this study

2. Dataset M, 1 Excel file with the following file name "Dataset M Metabolome profiles"

3. Dataset P, 1 Excel file with the following file name "Pathway assignment of detected metabolites"

4. Supplementary materials and methods

5. Figure legends of the appendix

6. Supplementary figures

7. References of the supplement

1. TABLE S1 Strains and plasmids used in this study

	Description	**Source or reference**
Strains		
Streptomyces davawensis	wildtype strain (JCM 4913)	Japan Collection of Microorganisms
Streptomyces coelicolor A3(2)	wildtype strain (DSMZ 40783)	German Collection of Microorganisms and Cell Cultures GmbH
*S. davawensis 1333::acc(3)IV (*Δdod.1 *and* Δdod.2*)*	Dodecin gene (dod^{SD} or BN159_1333) replaced by an apramycin resistance cassette (*aacIV*)	this work
Escherichia coli Rosetta 2(DE3)	Contains pRARE2 (Cm^R)	Merck KGaA, Darmstadt

Escherichia coli Top10	Used for cloning	Invitrogen, Carlsbad, USA
Escherichia coli DH10B pESAC13.4	Carries a 96 kbp subgenomic fragment of *S. davawensis* containing dod^{SD} and surrounding genes	this work
Escherichia coli GM2163	Methylation deficient strain used for conjugation with *S. davawensis*	Thermo Fisher Scientific, Waltham, USA
Escherichia coli DH5α pR9406	pUB307-derivative, used for triparental mating	David Figurski, Columbia University, NY
Escherichia coli CpXFMN	BL21(DE3) *ΔmanX*::*ribM*, *ΔribE*::*ribFMN1*	[1]
Escherichia coli CpXFAD	BL21(DE3) *ΔmanX*::*ribM*, *ΔribE*::*ribCF*	[1]
Plasmids		
pET24a(+)*1333*	Expression plasmid; dod^{SC} gene inserted using *Nde*I and *Xho*I	this work
pESAC13	*E. coli-Streptomyces* artificial chromosome pPAC-S1-derivative with *oriT* from RK2 replicon, (Kan^R)	[2]

36

2. Dataset M, 1 Excel file with the following file name "Dataset M Metabolome profiles"

3. Dataset P, 1 Excel file with the following file name "Pathway assignment of detected metabolites"

4. Supplementary materials and methods

Overexpression of dodecin genes in *Bacillus subtilis*. Nucleotide sequences of all synthetic (codon-adapted) *Streptomyces* dodecin genes (dod^{SD}, dod^{SC}, dod^{SDHis6} and dod^{SCHis6}) generated for efficient gene expression in *B. subtilis* are shown below. Sites for restriction endonucleases are underlined. dod^{SD}: GGA TCC ATG AGC AAT CAT ACA TAT CGC GTT ACA GAT ATT GTT GGC ACA TCA CCG GAA GGC GTT GAT CAA GCA ATT AGA AAT GGC ATT AAT AGA GCG AGC CAG ACA CTG CAT AAT CTG GAT TGG TTT GAA GTT GTT GAA GTT AGA GGC CAA CTG AAT GAT GGC CAA ATT GCA CAT TGG CAA GTT ACA ATG AAA GTC GGC TTT AGA CTG GAT GAA ACA GGC TAA CCC GGG. dod^{SC}: GGA TCC ATG AGC AAT CAT ACA TAT AGA GTT ACA GAA GTC GTT GGC ACA TCA CCG GAT GGC GTT GAT CAA GCA GTT AGA AAT GCA GTT ACA AGA GCA TCA CAG ACA CTG AGA AAA CTG GAT TGG TTT GAA GTT ACA CAA GTC AGA GGC CAA ATT GAA GAT GGC CAA GTT GCA CAT TGG CAA GTT GGC CTG AAA CTG GGC TTT AGA CTG GAA GAA TCA GAC TAA CCC GGG. dod^{SDHis6}: GGA TCC ATG AGC AAT CAT ACA TAT CGC GTT ACA GAT ATT GTT GGC ACA TCA CCG GAA GGC GTT GAT CAA GCA ATT AGA AAT GGC ATT AAT AGA GCG AGC CAG ACA CTG CAT AAT CTG GAT TGG TTT GAA GTT GTT GAA GTT AGA GGC CAA CTG AAT GAT GGC CAA ATT GCA CAT TGG CAA GTT ACA ATG AAA GTC GGC TTT AGA CTG GAT GAA ACA GGC CTG GAA CAT CAT CAT CAC CAT CAT TAA CCC GGG. $dodS^{CHis6}$: GGA TCC ATG AGC AAT CAT ACA TAT AGA GTT ACA GAA GTC GTT GGC ACA TCA

CCG GAT GGC GTT GAT CAA GCA GTT AGA AAT GCA GTT ACA AGA GCA TCA CAG ACA CTG AGA AAA CTG GAT TGG TTT GAA GTT ACA CAA GTC AGA GGC CAA ATT GAA GAT GGC CAA GTT GCA CAT TGG CAA GTT GGC CTG AAA CTG GGC TTT AGA CTG GAA GAA TCA GAT CTG GAA CAT CAT CAT CAC CAT CAT TAA CCC GGG.

5. Figure legends of the appendix

FIG S1 Dodecins of the roseoflavin producers *Streptomyces davawensis* and *Streptomyces cinnabarinus* are highly similar to other procaryotic dodecins. Primary structure alignment of dodecins from selected procaryotes (*Sdav*, *S. davawensis*, WP_015656108; *Scin*, *S. cinnabarinus* (unpublished); *Scoe*, *S. coelicolor*, NP_625213; *Save*, *Streptomyces avermitilis,* WP_037647113; *Mtub*, *Mycobacterium tuberculosis*, YP_177647; *Smel*, *Sinorhizobium meliloti*, WP_018095980; *Pstu*, *Pseudomonas stutzeri*, WP_037045112; *Bper*, *Bordetella pertussis*, JGWF01000184; *Tthe*, *Thermus thermophilus*, WP_014510626; *Ctep*, *Chlorobium tepidum*, WP_010933882; *Pput*, *Pseudomonas putida*, WP_013970282; *Paer*, *Pseudomonas aeruginosa*, OPA57256.1; *Gsul*, *Geobacter sulfurreducens*, WP_010940870; *Hsal*, *Halobacterium salinarum*, WP_012289329). *, amino acids which are conserved (green). :, amino acids with strongly similar properties (blue); ., amino acids with weakly similar properties (orange). Dodecins from *S. davawensis* and *S. cinnabarinus* are identical. The boxed tryptophane (W) is responsible for binding the aromatic isoalloxazine ring system of flavins (not experimentally validated for all species).

FIG S2 Analysis of five independent recombinant *S. davawensis* strains with regard to deletion of the dodecin gene. Agarose gel electrophoresis in combination

with ethidium bromide staining was used to detect PCR products. To generate these PCR products chromosomal DNA from different *dod* deletion strains was used as a template in combination with primers binding immediately upstream and downstream of *dod* or of the resistance cassette (in case of the *dod* disruption strains) but outside of the respective coding sequence. In lanes 1-5 chromosomal DNA from deletion strains was analyzed and a 1560 nt band corresponds to the apramycin resistance cassette (*aacIV*) replacing *dod*. In lane 6 chromosomal DNA of *S. davawensis* wildtype is shown resulting in a 683 nt DNA fragment corresponding to *dod*. In lane M the DNA standard is shown.

FIG S3 Validation of *dod* disruption in two independent recombinant *S. davawensis* strains by RT-PCR. Agarose gel electrophoresis in combination with ethidium bromide staining was used to detect specific RT-PCR products. RT-PCR products (cDNAs) were generated employing *dod* specific oligonucleotides. Expression of the constitutively expressed *gltA* gene encoding citrate synthase was monitored as a control. Lanes M, nucleotide marker, 100 nucleotide pair (nt) ladder. Lanes 1 and 2, no RT-PCR product corresponding to *S. davawensis dod* could be generated from the *dod* deletion strain (*S. davawensis* Δdod.1) neither in the exponential (lane 1) nor in the stationary growth phase (lane 2). Lanes 3 and 4, as a control an RT-PCR product corresponding to *gltA* could be produced when employing total RNA from cells harvested in the exponential (lane 3) as well as the stationary growth phase (lane 4). Lanes 5 and 6, negative controls, where no reverse transcriptase was added to the reaction ruling out the possibility that contaminating chromosomal DNA was present. Lanes 7-12, RT-PCR reactions from an independent *dod* deletion strain (*S. davawensis* Δdod.2) are shown. Lane 13, negative control, RT-PCR reaction without added RNA.

115

116 **Fig. S4 The dodecins from *S. davawensis* and *S. coelicolor* bind flavins differently**
117 Schematic view of the *Escherichia coli* strains CpXFMN and CpXFAD employed for
118 the *in vivo* generation of dodecins loaded with the FMN/FAD-cofactor analogs
119 RoFMN or RoFAD. The dodecin genes of *Streptomyces davawensis* and *Streptomyces*
120 *coelicolor* were expressed using the plasmid pET24a(+) His_6 (20). Upon induction of
121 protein synthesis, riboflavin (RF) or roseoflavin (RoF) were added to the growth
122 medium. Both flavins are taken up *via* RibM (the corresponding gene *ribM* from
123 *Corynebacterium glutamicum* replaces *manX* in the *E. coli* chromosome). *E. coli*
124 naturally does not produce a flavin transporter. The His_6-tagged recombinant dodecins
125 *in vivo* combined with FMN, FAD, RoFMN or RoFAD, were purified using affinity
126 chromatography and were analysed with regard to their cofactor content employing
127 HPLC/MS. CpXFMN contains the gene *FMN1* from *Schizosaccharomyces pombe*
128 [3]. The gene product FMN1 produces FMN from riboflavin and ATP and RoFMN
129 from roseoflavin and ATP. *FMN1* (replacing *ribE*) was inserted into the chromosome
130 to enhance intracellular synthesis of FMN analogs and to stimulate loading of FMN
131 dependent flavoproteins with FMN or RoFMN. *FMN1* is under control of the *ribE*
132 promoter P_{ribep5}. Notably, CpXFMN (as well as CpXFAD) is riboflavin auxotrophic
133 which allows control of flavin levels. Analogously, *E. coli* CpXFAD was used for the
134 analysis of FAD-dependent flavoproteins. CpXFAD contains an additional copy of *E.*
135 *coli ribCF* (replacing *ribE*) encoding the endogenous bifunctional flavokinase/FAD
136 synthetase which produces both FMN (from riboflavin and ATP) and FAD (from
137 FMN and ATP).

138

139 **Fig S5 The dodecins from *S. davawensis* ($SdDod^B$) and *S. coelicolor* ($ScDod^B$)**
140 **were overproduced in *B. subtilis*.** A, Denaturing SDS-PAGE and Coomassie

Brilliant Blue G-250 staining of cell-free extracts generated from different *B. subtilis* strains overexpressing the genes dod^{SC} and dod^{SD}. Lane M contains the molecular weigth marker. Lane 1, control, *B. subtilis* wildtype not overproducing a dodecin. Lane 2, *B. subtilis* overproducing $SdDod^{B}$. Lane 3, *B. subtilis* overproducing His_6-tagged $SdDod^{B}$. Lane 4, *B. subtilis* overproducing $ScDod^{B}$. Lane 5, *B. subtilis* overproducing His_6-tagged $ScDod^{B}$. B. His_6-tagged dodecins (lane 3, $SdDod^{B}$-His_6; lane 5, $ScDod^{B}$-His_6) could be monitored by Western-blot analysis employing anti-penta-His5 antibodies. To lanes 6 and 7 purified $SdDod^{B}$-His_6 and $ScDod^{B}$-His_6 were added (controls). Under these conditions $SdDod^{B}$-His_6 was monomeric whereas $ScDod^{B}$-His_6 was dodecameric (12-mer).

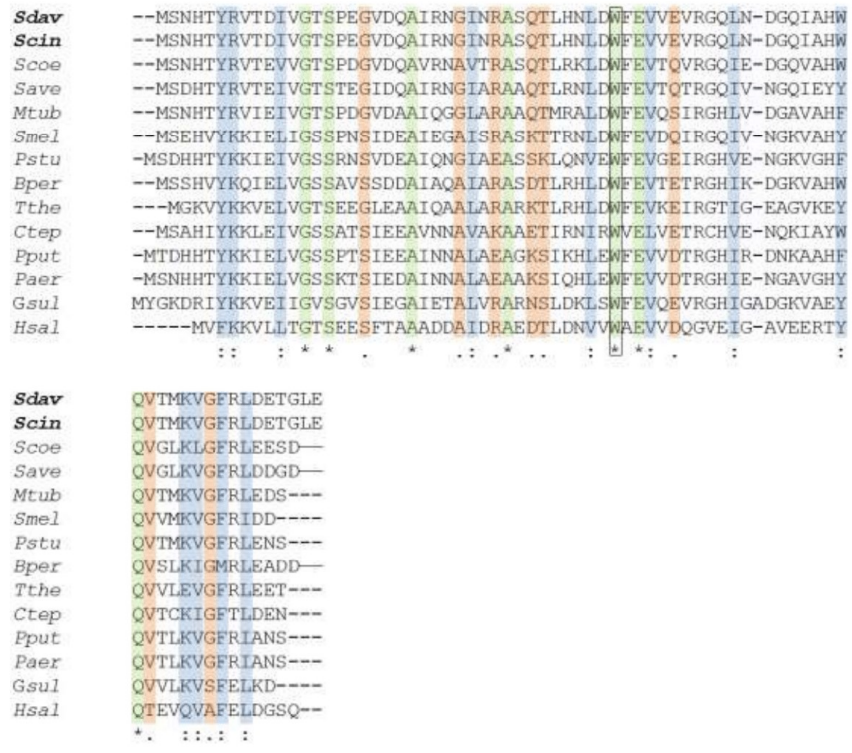
Fig S6 Heterologous expression of *Streptomyces* dodecin genes in *Bacillus subtilis* does not lead to roseoflavin resistance. For expression of *dod* genes the shuttle vector pHT01 was used. *B. subtilis* was transformed with pHT01 constructs either containing *dod* from *Streptomyces davawensis* (pHT01<1333>; dod^{SD}) or from *Streptomyces coelicolor* (pHT01<SCO0915>; dod^{SC}). The control strain contained an empty plasmid (pHT01<>). The different *B. subtilis* strains were inoculated to the same initial OD_{600} and cultivated in LB (in the absence of light) for 25 h in the presence of roseoflavin (RoF; 50 µM), riboflavin (RF; 50 µM) or in the absence of additional flavins (-). Notably, LB contains 4 µM riboflavin. At an OD_{600} of 0.1 synthesis of dodecins was induced by adding IPTG to the cultures. At indicated time points samples were taken and OD_{600} was determined. The strains expressing dod^{SD} or dod^{SC} genes did not grow better in the presence of roseoflavin when compared to the controls containing the empty plasmid. Riboflavin does not negatively affect growth. After 10 h cultivation time deregulated roseoflavin resistant strains appear which

166 contain mutations in either *ribC* or the FMN riboswitch controlling riboflavin levels
167 in *B. subtilis* [4].
168
169 **Fig S7 Expression of dodecin genes from *Streptomyces davawensis* or**
170 ***Streptomyces coelicolor* does not confer roseoflavin (RoF) resistance to**
171 **recombinant *Bacillus subtilis* strains growing on solid growth media.** Drops
172 ("circles", about 50,000 cells) and streaks of different *B. subtilis* strains
173 overexpressing *Streptomyces dod* genes employing plasmid pHT01 were applied to
174 LB plates containing the indicated amount of roseoflavin (C-E; roseoflavin was
175 dissolved in DMSO) and 1 mM IPTG (to stimulate expression of the recombinant *dod*
176 genes). The control plates (A-B) contained DMSO or not and served as controls.
177 Growth was recorded after incubation of the plates for 36 h at 30°C. Expression of
178 *dod* neither confers roseoflavin resistance to *B. subtilis* at 100 µM roseoflavin nor at
179 50 µM roseoflavin. At 10 µM roseoflavin all strains show (reduced) growth at least
180 when cell suspensions were added dropwise to the plates. 1, control, empty plasmid.
181 2, strain expressing *dod* from *S. davawensis* (pHT01<1333>). 3, strain expressing
182 dod^{His6} from *S. davawensis* (pHT01<1333-his6>). 4, strain expressing *dod* from *S.*
183 *coelicolor* (pHT01<SCO0915>). 5, strain expressing dod^{His6} from *S. coelicolor*
184 (pHT01<SCO0915-his6>). Notably, to rule out a disturbing effect of the His_6-tag,
185 strains overproducing dodecin versions not containing such a tag were also analyzed.
186
187 **Fig. S8 Principal component analysis shows effects of inter-species differences**
188 **(A) and plumbagin treatment (B) on metabolic profiles.** Principal component
189 analysis of the different samples was performed on log_{10}-transformed intensities of
190 12,730 metabolite ions measured using non-targeted metabolomics.
191

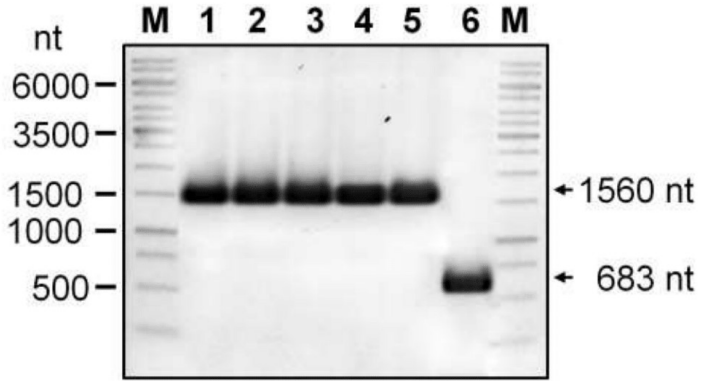
192 **Fig S9 Metabolic response *of S. coelicolor*, *S. davawensis* wildtype and *S.***
193 ***davawensis Δdod2* to plumbagin treatment.** Samples were analyzed using non-
194 targeted metabolomics. Metabolite ion levels in plumbagin-treated and untreated
195 samples were compared using differential analysis. P-values were calculated using
196 two-tailed unpaired t-tests, and adjusted for multiple hypothesis testing using Storey's
197 and Tibshirani's positive false discovery rate. Fold changes refer to fold changes of
198 means. Statistically significant differential metabolites (defined as FDR<0.1% and
199 $|\log_2$ fold-change$| > 1$) in *S. coelicolor* (panel A) included antioxidants such as
200 mycothiol, mycothioine and tocopherol, as well as nucleotides such as GDP. In *S.*
201 *davawensis* (panel B), fewer statistically significant metabolic responses to plumbagin
202 treatment were observed compared to *S. coelicolor*. In contrast, the *S. davawensis*
203 Δdod2 mutant (panel C) shows a strong metabolic response, consistent with increased
204 plumbagin tolerance in that mutant and this metabolic response contributing to that
205 resistance. The complete metabolomics dataset is included as Dataset M. Data points
206 in yellow represent metabolite ions that show a statistically significant (defined as
207 having an adjusted p-value < 0.001 in the respective contrasts) and strong (defined as
208 a |log2 fold-change| > 1 in the respective contrasts) response. Data points in blue
209 represent metabolite ions not meeting these criteria. The complete data underlying
210 these plots, including full metabolite annotations of all detected ions, is provided as
211 Dataset M in this appendix. Plots can be reconstructed from this dataset by plotting
212 the respective "log2(FC)" and -log10-transformed "adj. P-val." columns as a scatter
213 plot directly in Excel, and using the column "Metabolite name" as data point labels. A
214 pathway memberships of all detected metabolites is provided as Dataset P in this
215 appendix (please note that these assignment are ambiguous, since a metabolite can
216 participate in multiple biochemical pathways).
217

6. Supplementary figures

Fig S1



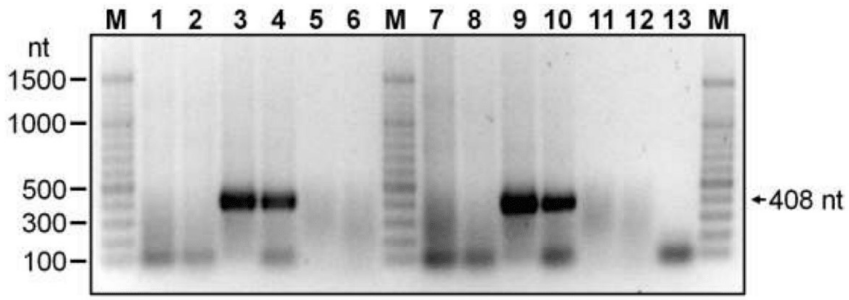
223 **Fig S2**



224

225

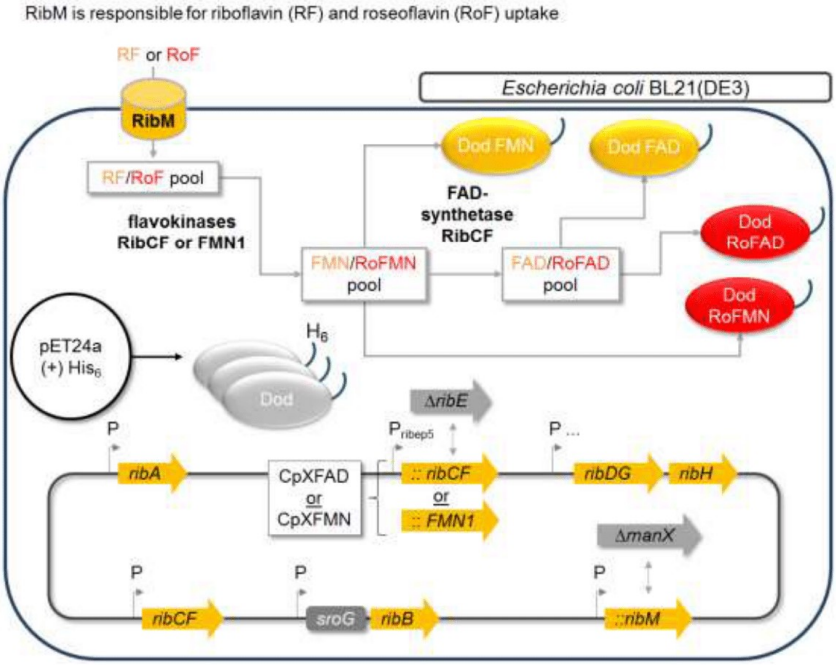
226 **Fig S3**



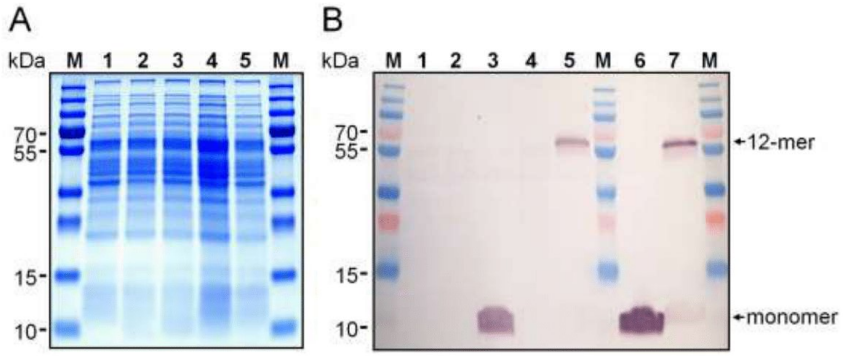
227

228

Fig S4


RibM is responsible for riboflavin (RF) and roseoflavin (RoF) uptake
RF or RoF
RibM
Escherichia coli BL21(DE3)
RF/RoF pool
flavokinases
RibCF or FMN1
Dod FMN
Dod FAD
FAD-
synthetase
RibCF
Dod
RoFAD
Dod
RoFMN
FMN/RoFMN
pool
FAD/RoFAD
pool
pET24a
(+) His6
H6
Dod
ΔribE
P
Pribep5
P ...
ribA
CpXFAD
or
CpXFMN
:: ribCF
or
:: FMN1
ribDG
ribH
ΔmanX
P
P
P
ribCF
sroG
ribB
::ribM

232 **Fig S5**

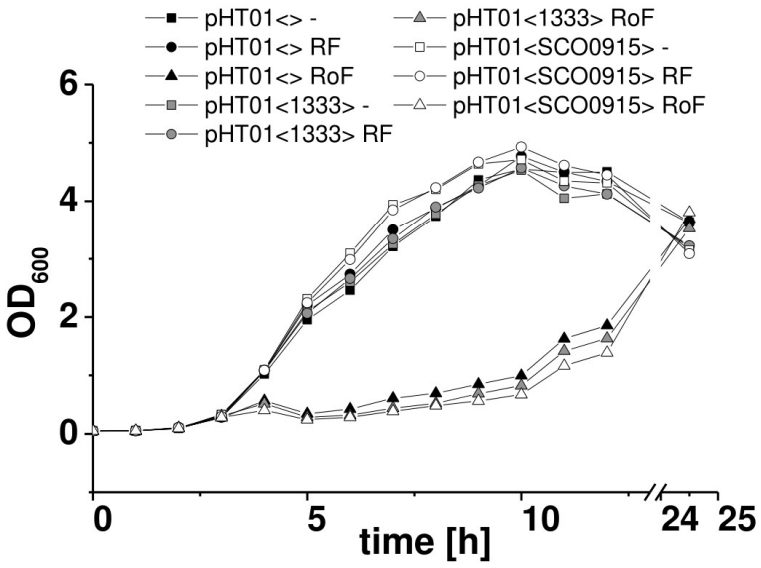


233

234

235

236 **Fig S6**

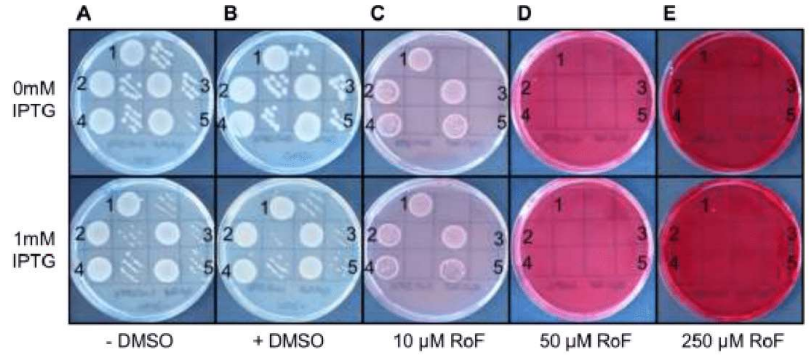


237

238

239

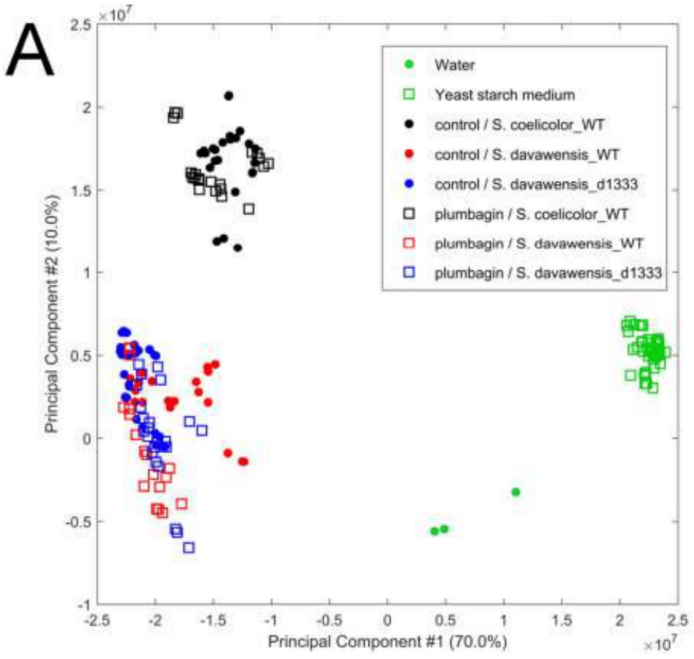
240 **Fig S7**

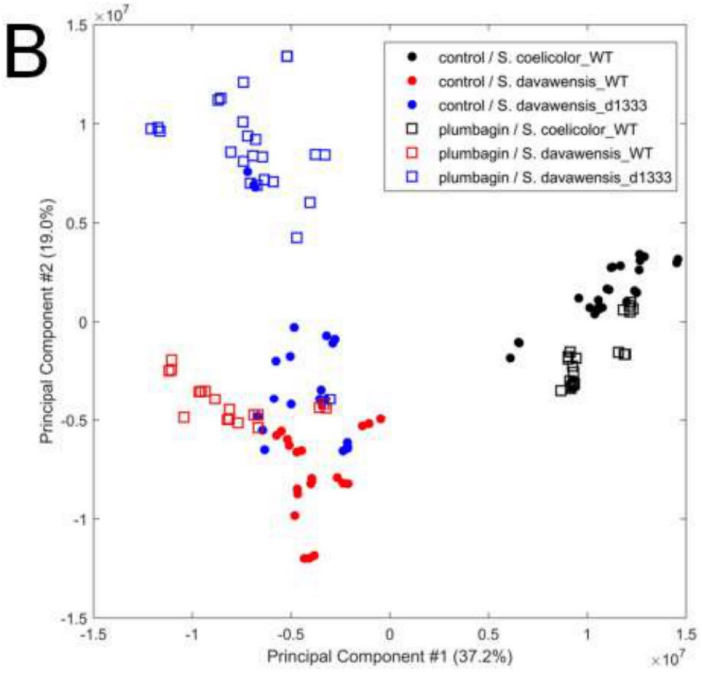


241

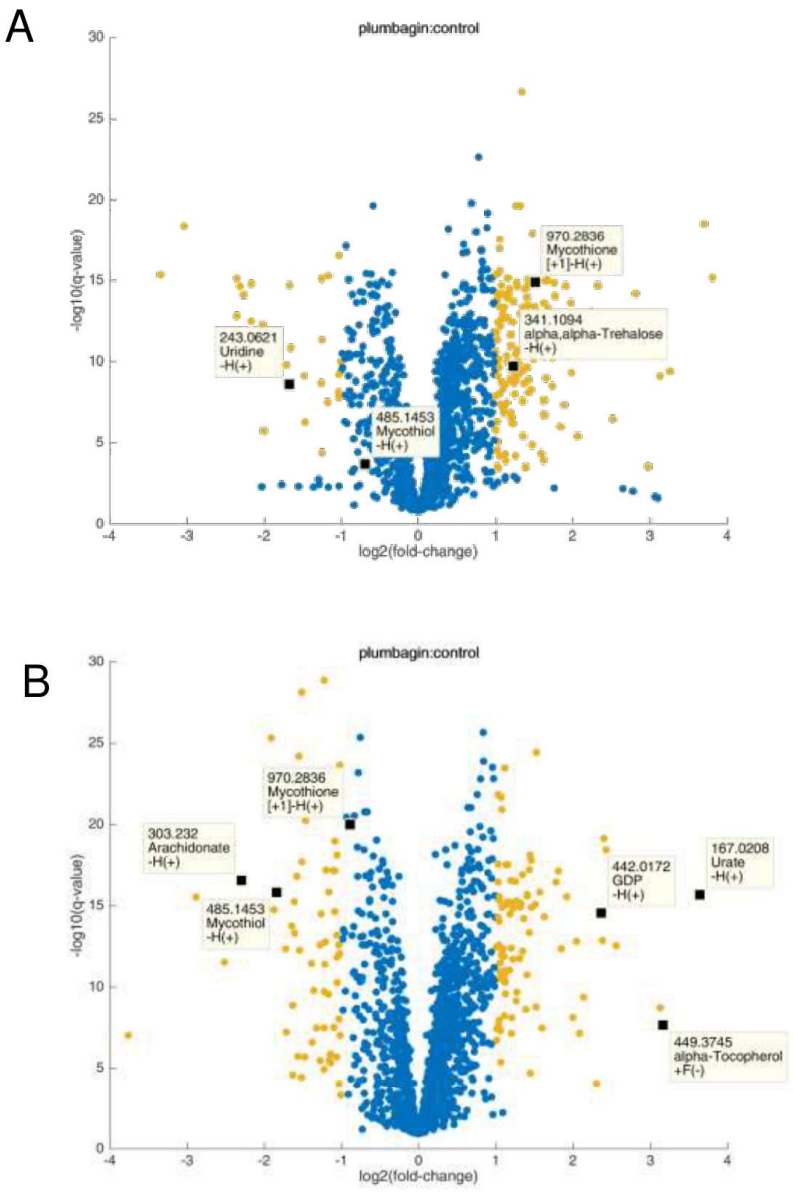
242

Fig S8


A
Water
Yeast starch medium
control / S. coelicolor_WT
control / S. davawensis_WT
control / S. davawensis_d1333
plumbagin / S. coelicolor_WT
plumbagin / S. davawensis_WT
plumbagin / S. davawensis_d1333
Principal Component #2 (10.0%)
Principal Component #1 (70.0%)
×10^7


B
control / S. coelicolor_WT
control / S. davawensis_WT
control / S. davawensis_d1333
plumbagin / S. coelicolor_WT
plumbagin / S. davawensis_WT
plumbagin / S. davawensis_d1333
Principal Component #2 (19.0%)
Principal Component #1 (37.2%)
×10^7

246 **Fig S9**

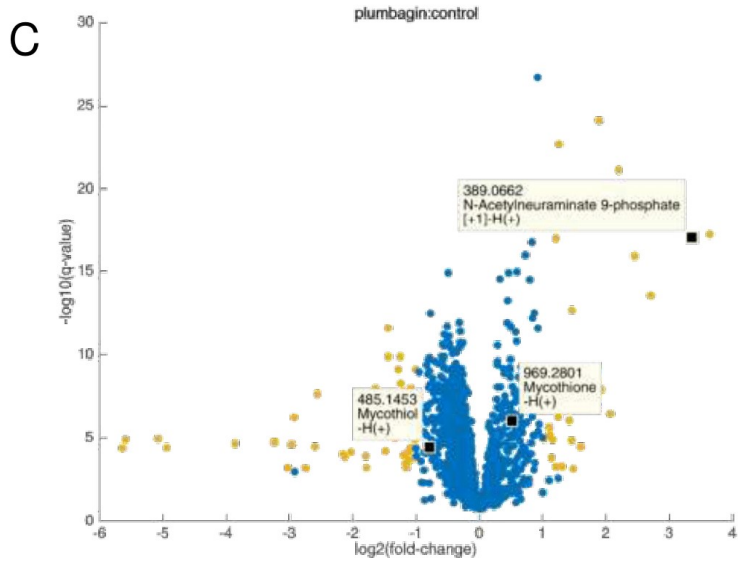


247

248

249

Fig S9 *ctd.*


C
plumbagin:control
-log10(q-value)
log2(fold-change)
389.0662
N-Acetylneuraminate 9-phosphate
[+1]-H(+)
969.2801
Mycothione
-H(+)
485.1453
Mycothiol
-H(+)

253 **7. References of the supplement**

254 1. **Langer S, Hashimoto M, Hobl B, Mathes T, Mack M**. Flavoproteins are
255 potential targets for the antibiotic roseoflavin in *Escherichia coli*. *J Bacteriol*
256 2013;195(18):4037-4045.
257 2. **Sosio M, Giusino F, Cappellano C, Bossi E, Puglia AM et al.** Artificial
258 chromosomes for antibiotic-producing actinomycetes. *Nature biotechnology*
259 2000;18(3):343–345.
260 3. **Plumbridge J**. Control of the expression of the *manXYZ* operon in
261 *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. *Mol Microbiol*
262 1998;27(2):369-380.
263 4. **Coquard D, Huecas M, Ott M, van Dijl JM, van Loon AP et al.** Molecular
264 cloning and characterisation of the *ribC* gene from *Bacillus subtilis*: a point mutation
265 in *ribC* results in riboflavin overproduction. *Mol Gen Genet* 1997;254(1):81-84.
266
267

MICROBIOLOGY

RESEARCH ARTICLE

Bourdeaux *et al.*, *Microbiology* 2019;165:1095–1106

DOI 10.1099/mic.0.000835

Comparative biochemical and structural analysis of the flavin-binding dodecins from *Streptomyces davaonensis* and *Streptomyces coelicolor* reveals striking differences with regard to multimerization

Florian Bourdeaux[1]†, Petra Ludwig[2]†, Karthik Paithankar[1], Bodo Sander[3], Lars-Oliver Essen[3], Martin Grininger[1] and Matthias Mack[2,*]

Abstract

Dodecins are small flavin-binding proteins that are widespread amongst haloarchaeal and bacterial species. Haloarchaeal dodecins predominantly bind riboflavin, while bacterial dodecins have been reported to bind riboflavin-5′-phosphate, also called flavin mononucleotide (FMN), and the FMN derivative, flavin adenine dinucleotide (FAD). Dodecins form dodecameric complexes and represent buffer systems for cytoplasmic flavins. In this study, dodecins of the bacteria *Streptomyces davaonensis* (SdDod) and *Streptomyces coelicolor* (ScDod) were investigated. Both dodecins showed an unprecedented low affinity for riboflavin, FMN and FAD when compared to other bacterial dodecins. Significant binding of FMN and FAD occurred at relatively low temperatures and under acidic conditions. X-ray diffraction analyses of SdDod and ScDod revealed that the structures of both *Streptomyces* dodecins are highly similar, which explains their similar binding properties for FMN and FAD. In contrast, SdDod and ScDod showed very different properties with regard to the stability of their dodecameric complexes. Site-directed mutagenesis experiments revealed that a specific salt bridge (D10–K62) is responsible for this difference in stability.

INTRODUCTION

Riboflavin (vitamin B_2) can be synthesized by most known prokaryotic species [1]. Riboflavin is a precursor of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) (Fig. 1). FMN and FAD are cofactors of the so-called flavoproteins, proteins that are able to carry out highly diverse biological functions [2]. Since all cells contain considerable amounts of riboflavin, FMN and FAD [3], it is likely that micro-organisms encounter flavins in their habitat as a result of cell decomposition. Since uptake of e.g. riboflavin is more economical than its biosynthesis, many bacteria and archaea contain membrane-bound transporters for riboflavin [4, 5]. Cytoplasmic riboflavin is either converted to FMN and FAD by flavokinases and FAD synthetases [6] or stored by riboflavin-binding proteins in birds [7] and mammals [8] or by dodecins in archaea [9]. Dodecins are the smallest known flavoproteins (about 70 amino acids per monomer) and appear to only be present in prokaryotic cells [9]. Dodecins form dodecameric protein complexes with six structurally identical binding pockets whereby each pocket accommodates two flavins sandwiched between two tryptophan residues. The flavin-binding dodecins can be divided into two groups, the archaeal and the bacterial dodecins. While archaeal dodecins seem to be restricted to the class *Halobacteria* (haloarchaea), bacterial dodecins are distributed among several phyla and have also been identified and characterized in the pathogens *Mycobacterium tuberculosis* (MtDod) [10–12] and *Pseudomonas aeruginosa* [9]. *Halobacterium salinarum* dodecin (HsDod) is the best-studied archaeal dodecin and has been reported to be involved in riboflavin storage and homeostasis [9, 13–15]. The primary structures of bacterial dodecins are less conserved when compared to archaeal dodecins and the

Received 07 May 2019; Accepted 26 June 2019; Published 24 July 2019

Author affiliations: [1]Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany; [2]Institute for Technical Microbiology, Faculty for Biotechnology, Mannheim University of Applied Sciences, 68163 Mannheim, Germany; [3]Unit for Structural Biology, Department of Chemistry and Biology, Philipps University Marburg, 35032 Marburg, Germany.

***Correspondence:** Matthias Mack, m.mack@hs-mannheim.de

Keywords: dodecin; riboflavin; roseoflavin; *Streptomyces davaonensis*.

Abbreviations: FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; HsDod, dodecin of *Halobacterium salinarum*; RoFAD, roseoflavin adenine dinucleotide; RoFMN, roseoflavin mononucleotide; ScDod, dodecin of *Streptomyces coelicolor*; SdDod, dodecin of *Streptomyces davaonensis*.

†These authors contributed equally to this work.

Eight supplementary figures are available with the online version of this article.

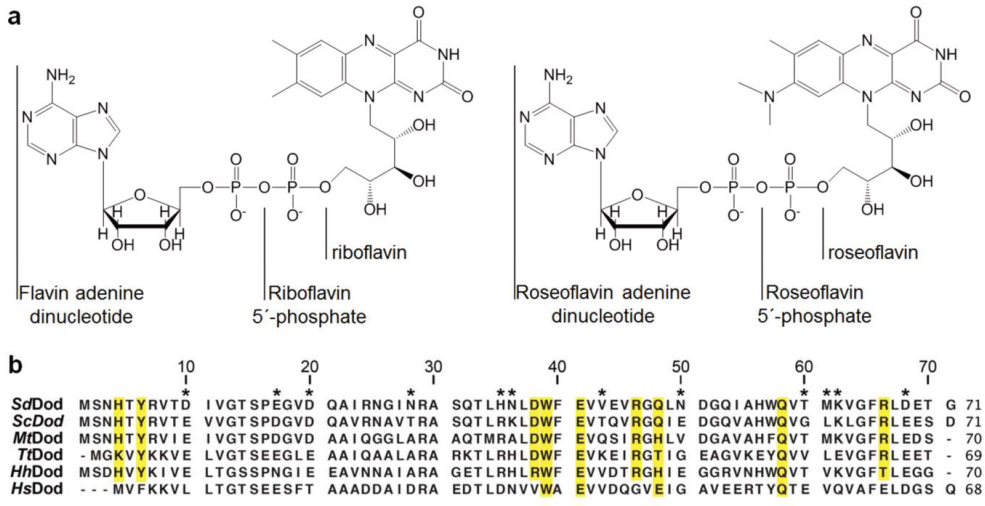


Fig. 1. Chemical structures of the flavins investigated in this study and primary structures of dodecins from different prokaryotic species. (a) Structures of riboflavin and the riboflavin-derived cofactors flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The riboflavin analogue roseoflavin is an antibiotic and is produced by *S. davaonensis* but not by *S. coelicolor*. Roseoflavin is activated by flavokinases and FAD synthetases to the cofactor analogues roseoflavin mononucleotide (RoFMN) and roseoflavin adenine dinucleotide (RoFAD). (b) Primary structures of dodecins from the actinomycetales *S. davaonensis* (SdDod), *S. coelicolor* (ScDod) and *M. tuberculosis* (MtDod) share a high degree of similarity. X-ray structures are available for the dodecins from *T. thermophilus* (TtDod) [16], *M. tuberculosis* [11, 12], *Halorhodospira halophila* (HhDod) [9] and *H. salinarum* (HsDod) [13]. Amino acids boxed in yellow are known to be relevant for flavin binding in MtDod, TtDod, HhDod and HsDod. Amino acids labelled with an asterisk were subjected to site-directed mutagenesis in the present work.

flavin-binding specificities are less pronounced, i.e. FMN, riboflavin and also FAD are accepted as ligands. While FMN seems to be the most important flavin ligand of bacterial dodecins [12, 15], dodecins were found to also bind coenzyme A [16] or calcium [17], and a function in cofactor storage has been proposed [12, 16].

The bacteria *Streptomyces davawensis* (now *Streptomyces davaonensis*) [18] and *Streptomyces cinnabarinus* are the only organisms known to produce the riboflavin analogue roseoflavin, which is activated by flavokinases and FAD synthetases to the cofactor analogues roseoflavin mononucleotide (RoFMN) and roseoflavin adenine dinucleotide (RoFAD) (Fig. 1a). RoFMN and RoFAD are inhibitors of flavoproteins [19–21] and RoFMN inhibits FMN riboswitches [22]. Since the bacterial dodecin MtDod was reported to bind roseoflavin with high affinity [23], it seemed likely that the corresponding dodecins from *S. davaonensis* (SdDod) and *S. cinnabarinus* (ScinDod) would play a role in roseoflavin resistance by tightly binding the flavin and thus inactivating this antibiotic. In-depth analysis of an *S. davaonensis* dodecin-deficient strain, however, suggested that SdDod does not contribute strongly to roseoflavin resistance [24]. The finding that an *S. davaonensis* dodecin deletion strain showed an altered metabolome and different behaviour under oxidative stress was rather in line with a flavin-buffering role [24]. As mentioned above, flavins are present at relatively high levels in prokaryotic cells (riboflavin, ~17 µM; FMN, 45 µM; FAD, 170 µM) [3] and have the potential to generate highly reactive, cytotoxic compounds in an oxic environment. Consequently, it seemed plausible that the flavin-binding dodecins shield these reactive chemicals from other cellular structures. Interestingly, MtDod was found to bind FMN more strongly at an acidic pH [10] and the hypothesis is that under these conditions MtDod helps the pathogen to survive acidic conditions encountered when entering a dormant state. Previous work has shown that the internal pH of dormant endospores of e.g. *Bacillus* species is more than 1 pH unit below that of the cytoplasm of growing cells [25]. The same was reported for spores of *Bacillus subtilis*, *Streptomyces antibioticus* and *Streptomyce kasugaensis* [26–28]. The streptomycetes *S. davaonensis* and *Streptomyces coelicolor* may have acidic spores as well and we therefore also envisaged a role for *Streptomyces* dodecins in flavin binding under acidic conditions.

The present study was initiated to shed more light on the function of the bacterial dodecins of *S. davaonensis* and *S. coelicolor* (ScDod). *S. coelicolor* is not a flavin overproducer and also does not synthesize roseoflavin. The working hypothesis was that SdDod would protect *S. davaonensis* from

roseoflavin and/or reactive flavins in general whereby ScDod was studied as a control. Our data suggest that SdDod and ScDod are involved in flavin homeostasis of the ubiquitous soil-dwelling streptomycetes.

METHODS

Chemicals

Roseoflavin was obtained from MP Biomedicals (Heidelberg, Germany). The flavins used for binding experiments and all other chemicals were obtained from Sigma-Aldrich (Munich, Germany). Flavins are light-sensitive and all samples and cultures containing flavins were kept in the dark.

Bacterial strains, plasmids, growth media and cloning

The strains and plasmids used in this work are listed in Table S1 (available in the online version of this article). The genes encoding SdDod (BN_159 1333) and ScDod (SCO0915) were amplified by PCR using specific oligonucleotide primers and chromosomal DNA as a template (for all primers see Table S2). The *Nde*I/*Xho*I-treated PCR products were ligated to *Nde*I/*Xho*I-treated pET24a(+), which introduced codons specifying a C-terminal His_6-tag. From the structural data we deduce that the His_6-tag does not interfere with oligomerization as the His_6-tag points towards the outside of the multimers. For crystallization of ScDod without a His-tag, a plasmid based on pET28a(+) was constructed. The corresponding PCR product was digested with *Pci*I/*Xho*I and ligated to *Nco*I/*Xho*I-treated pET28a(+). All inserts of all constructs were proof sequenced. *Escherichia coli* was used for cloning and was cultivated aerobically in lysogeny broth (LB) [29] at 37 °C.

Site-directed mutagenesis

Site-directed mutagenesis was carried out using the QuikChange II XL Site-Directed Mutagenesis kit according to the instructions of the manufacturer. Two complementary mismatch oligonucleotides were used to generate point mutations in the expression plasmids pET24a(+)SdDod-His_6 and pET24a(+)ScDod-His_6. The sequences of the employed oligonucleotides are listed in Table S2.

Overproduction and purification of dodecins

The genes encoding SdDod, ScDod and variants thereof were overexpressed in *E. coli* Rosetta 2(DE3), or for crystallization in *E. coli* CpXFMN [24] with 50 µM riboflavin or roseoflavin present in the growth medium. Cells were aerobically grown in LB in the presence of 50 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol to an optical density (OD_{600}) of 0.6 at 37 °C. IPTG was added to a final concentration of 1 mM and cultures were grown for another 5 h. Cells were harvested by centrifugation and stored at −20 °C prior to cell disruption. Frozen cell paste was suspended in 30 ml His Trap binding buffer (3 mM KH_2PO_4, 17 mM K_2HPO_4, 300 mM KCl, 10 mM imidazole; pH 8.0) containing one tablet of cOmplete (Roche, Mannheim, Germany). The cell suspension was passed twice through a French press at 2000 bar (200 MPa). Two centrifugation steps (13100 ***g***, 4 °C, 10 min followed by 108000 ***g***, 4 °C, 30 min) removed cell debris and unbroken cells. All affinity chromatography steps were performed using the ÄKTApurifier system at a flow rate of 2 ml min^{-1} and at room temperature (RT). The supernatant was loaded onto a 5 ml HisTrap column, which was equilibrated with His Trap binding buffer. When the UV signal returned to baseline the column was washed with 14 column volumes (CV) of 20% (v/v) His Trap elution buffer (300 mM KCl, 500 mM imidazole, 3 mM KH_2PO_4, 17 mM K_2HPO_4; pH 8.0). For elution a His Trap elution buffer gradient of 1% ml^{-1} (v/v) starting at 20% was used. Protein-containing fractions were pooled and dialyzed employing dialysis tubes (molecular weight cut-off: 8–12 kDa) in 1 L of the dodecin buffer (300 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl; pH 7.5) until no further change in conductivity was observed. Dialyzed dodecin solutions were further purified by size-exclusion chromatography (SEC) with a Superdex 200 Increase 10/300 GL column and dodecin buffer as running buffer. Dodecamer fractions were pooled and purity was analysed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Dodecin (dodecamer) samples were concentrated, shock frozen and stored at −20 °C. ScDod not containing a His_6-tag was purified by a heat denaturation, dialysis, anion exchange chromatography and SEC as described [16]. Protein solutions were concentrated with centrifugal concentrators (30.000 Da molecular weight cut-off; Amicon, Merck, Darmstadt, Germany).

Determination of protein amounts and flavin levels

Dodecins were determined by their absorbance at 280 nm (NanoDrop 2000C, Thermo Scientific, Karlsruhe, Germany) and the calculated extinction coefficient of 12490 M^{-1} cm^{-1}. Flavin levels were determined at different wavelengths (riboflavin, 375 nm and 450 nm; FMN/FAD, 375 nm, 450 nm and 473 nm; roseoflavin, 505 nm) employing the following extinction coefficients: 375 nm, 10000 M^{-1} cm^{-1}; 450 nm, 12000 M^{-1} cm^{-1}; 473 nm, 9200 M^{-1} cm^{-1}; 505 nm, 31000 M^{-1} cm^{-1}).

Analysis of proteins employing SDS-PAGE

Precast gels were used for SDS-PAGE analysis (Mini-PROTEAN TGX stain-free precast PAGE gels, BioRad, Dreieich, Germany). For standard SDS-PAGE a 2× sample buffer [126 mM Tris-HCl, pH 6.8; 20% (v/v) glycerol; 4% (w/w) SDS; 0.02% (w/w) bromophenol blue; 20 mM dithiothreitol] was added to the protein solutions and the resulting mixtures were incubated for 5 min at 95 °C. For mild denaturing or semi native SDS-PAGE a 2× SDS-free sample buffer [1 M 6-aminocaproic acid, 100 mM BisTris-HCl, 100 mM NaCl, 20% (v/v) glycerol, 0.1% (w/w) SERVA Blue G, pH 7.0] was added to the protein solutions and the resulting mixtures were loaded to the gel without heat treatment. The running buffer contained 25 mM Tris-HCl (pH 8.3), 192 mM glycine and 3.5 mM SDS. Running conditions were 15 min at 100 V followed by 45 min at 150 V. For the stability measurement of ScDod in the presence of SDS and urea, the proteins were incubated in a solution with a final concentration of

2% (w/w) SDS or a solution with 2% (w/w) SDS and 8 M urea at 95 °C. All samples were analysed by SDS-PAGE but without additional heat treatment during sample preparation. The stabilizing effect of FMN on dodecins was determined as follows. Dodecins (10 µM) were incubated at 20 °C in the dark for 15 min with a fivefold molar excess of FMN (50 µM). The samples were subsequently analysed by SDS-PAGE. To analyse the binding behaviour and effects of riboflavin and roseoflavin on SdDod (10 µM), proteins were incubated with a 2- to –20-fold molar excess of a flavin (20–200 µM), incubated in the dark at RT for 15 min and analysed by SDS-PAGE.

Determination of flavin binding to dodecins

The thermocyclic fluorescence assays and the thermofluorescence assay were essentially conducted as described recently [10]. For the present flavin binding determination a real-time PCR instrument (C1000 Thermal Cycler and CFX96 Real-Time System; Bio-Rad) with an excitation/emission filter bandwidth of 450-490/560-580 nm was used. For the thermocyclic fluorescence assay, 25 µL of a 16 µM FMN : dodecin solution was added to the 96-well plates. The basic buffer contained 300 mM NaCl, 5 mM $MgCl_2$. Depending on the pH values to be reached, the mixtures contained 23.8 mM of additional buffer reagents (for pH 4.0, citrate; for pH 5.0, acetate; for pH 5.0, citrate; for pH 6.0–7.0, phosphate; for pH 7.5, Tris-HCl). The heating-phase temperature was increased stepwise from 5 °C to 95 °C (5–50 °C, 4.5 °C steps; 50–95 °C, 2 °C steps). Following each heating phase, a 30 min 5 °C cooling phase took place, followed by fluorescence measurements. The fluorescence (cooling phase) was plotted against the heating temperature and the slope between the two steps was used to determine the start of denaturation of the dodecamer [slope above 120 relative fluorescence units (RFU) C^{-1}]. For the thermofluorescence assay 2 µL of a 50 µM flavin : dodecin solution (basic buffer solution; 300 mM NaCl, 5 mM $MgCl_2$ and 10 mM Bis-tris methane) (pH 6.17) was added to 23 µL of a basic buffer system (300 mM NaCl, 5 mM $MgCl_2$) and 23.8 mM of universal buffer system components (AAB, acetic acid, ADA and bicine; MMT, malic acid, MES and Tris-HCl). The temperatures were raised from 5 °C to 95 °C in 0.5 °C steps every 30 s and after every heating phase the fluorescence was measured. The fluorescence was plotted against the temperature and inflection points were determined by Boltzman sigmoidal curve fitting. For the titration experiments SdDod or ScDod (0–8 µM each) were incubated in the presence of 1 µM of the corresponding flavin for 1 h at 20° in the dark. For measurements at pH 5.0 a buffer solution containing 300 mM NaCl, 5 mM $MgCl_2$ and 20 mM acetic acid (pH 5.0) was used. For measurements at pH 7.5 the standard dodecin buffer solution containing 300 mM NaCl, 5 mM $MgCl_2$ and 20 mM Tris-HCl (pH 7.5) was employed. After incubation fluorescence was measured in a plate reader with excitation at 445±9 nm and emission detection at 520±20 nm. The fluorescence of a well was measured four times and averaged, and measurements were conducted in triplicates. Data were background corrected (measurement of 100 µL buffer solution), averaged and normalized with the fluorescence signal of the well with only 1 µM flavin (0 µM dodecin). Dissociation constants were determined by non linear curve fitting (Origin Pro 8.6) with function (1). Variables: $[A_0]$ (overall flavin concentration, allowed values 1±0.05 µM), F_{max} (fluorescence of unbound flavin, allowed values 100±5%), $[D_0]$ (overall dodecin concentration) and F_∞ (fluorescence of bound flavin).

$$F = F_{max} - \frac{F_{max}-F_\infty}{[A_0]}\left(\frac{[D_0]+[A_0]+K_d-\sqrt{([D_0]+[A_0]+K_d)^2-4[A_0][D_0]}}{2}\right) \quad (1)$$

Crystallization of purified dodecins

Prior to crystallization, ScDod was denatured and refolded in the presence of FMN in the dodecin buffer as reported for the dodecin of *Thermus thermophilus* (TtDod) and MtDod [10, 16] and excessive amounts of FMN were removed by SEC using a Superdex 200 16/60 column (running buffer: 100 mM NaCl, 20 mM Tris-HCl; pH 8). The SdDod used for crystallization was overproduced in *E. coli* CpXFM in the presence of 50 µM riboflavin or roseoflavin as described previously [24]. Crystallization experiments were performed by mixing 1 µL of the protein solution mixed with an equal amount of reservoir solution in a hanging droplet setup. The following conditions yielded crystals. SdDod: 6 mg ml^{-1} roseoflavin : SdDod (droplet concentration); 500 µL reservoir solution: 5% poly-γ-glutamate (PGA)-LM, 22% (w/v) PEG-3350, 0.1 M Na-cacodylate, pH 6.5; light magenta coloured crystals grew at 20 °C. ScDod: 10 mg ml^{-1} FMN : ScDod (droplet concentration); 1 ml reservoir solution: 0.2 M $(NH_4)_2SO_4$, 10% (w/v) PEG-4000, 0.1 M NaOAc, pH 4.8; yellow/orange coloured crystals grew at 18 °C. Crystals were treated with cryoprotectant solutions and cryocooled in liquid nitrogen. For SdDod crystals the reservoir solution contained 20% (v/v) ethylene glycol and for SdDod crystals the reservoir solution contained 20% (v/v) glycerol.

Data collection, processing and structure determination

SdDod crystals were exposed to single-wavelength X-radiation at the Swiss Light Source (X06DA PX3) and maintained at 100 K while data were recorded onto a detector (PILATUS 2 M-F). In the case of ScDod, data were collected with single-wavelength X-radiation from a Bruker-Nonius FR591/CuKα at 100 K. SdDod data were processed using the program *xia*2 and scaled using AIMLESS within the CCP4 suite. For both structures the phase problem was solved using the molecular replacement method with the program MOLREP, employing the structural model of MtDod (PDB accession code: 2YIZ). The structure was iteratively refined with REFMAC5 with manual model building using Coot. The side-chain positioning and model geometry improvements were performed using the PDB_REDO server and manually verified in Coot. The data collection and refinement statistics are given in Table S4. The PDB ID is 6RI3 for SdDod and 6R1E for ScDod.

RESULTS

The bacterial dodecins SdDod and ScDod are exceptional as they only bind flavins at low pH values and low temperatures

As a first step with regard to the biochemical analysis of SdDod and ScDod, the amino acid sequences of these proteins were compared to the amino acid sequences of well-studied dodecins from other microbial species (Fig. 1b). The comparison suggested that SdDod and ScDod were flavin-binding dodecins due to the presence of the conserved/characteristic amino acid residues W39 and Q58 (SdDod and ScDod numbering), amino acids known to be important for flavin binding (Fig. 1b) [13, 14, 16]. Notably, W39 builds the aromatic tetrade that is essential in sandwiching flavins. Although SdDod and ScDod were predicted to bind flavins, purification of these dodecins from recombinant *E. coli* strains yielded flavin-free preparations with no absorbance at the riboflavin-specific wavelengths of 375 nm and 450 nm. This suggested that the flavins present in the cytoplasm of *E. coli* (riboflavin, FMN and FAD; Fig. 1a) were not captured by SdDod or ScDod while being overproduced in this organism. Purified apo-SdDod was investigated with regard to the binding of flavins by titration with FMN and riboflavin (Fig. S1). SdDod showed no detectable binding of FMN and riboflavin at the physiological pH of 7.5 (Fig. S1a). Similar results were obtained when titrating ScDod (data not shown). This binding behaviour was unexpected, as eight (H4, Y6, D38, W39, E41, R46, Q58 and R66) out of nine amino acids known to interact with FMN in MtDod are present in ScDod and SdDod at a comparable position (Fig. 1b). H48, which is known to interact with FMN in MtDod [10], is the only exception, and in SdDod (and ScDod) a glutamine residue (Q48) is present at this position. At pH 5.0 binding of FMN by SdDod and also ScDod was observed (Fig. S1a) and this can be explained by the fact that protonation of the 5′-phosphate of FMN led to a more hydrophobic compound, which in turn is more likely to be bound by the tryptophan residues W39 of SdDod. Binding of riboflavin was not observed at pH 5.0 (data not shown). Measurement of tryptophan fluorescence in an FMN : dodecin titration experiment supported the idea that flavin binding does not occur at the cytoplasmic pH (Fig. S1b). Temperature-dependent dissociation of flavin : dodecin complexes is a direct quantitative measure for flavin binding and was subsequently used to characterize flavin binding by SdDod and ScDod at different pH values and temperatures in more depth (Fig. 2). At relatively low temperatures flavins were bound by dodecins and fluorescence was quenched. With increasing temperatures the flavin : dodecin complexes incrementally disassembled (dodecin does not denature above 80 °C) and the released flavins were measured using a fluorescence detector. Temperature-dependent dissociation was determined for the complexes of ScDod and SdDod with riboflavin, FMN and FAD at different pH values and temperatures(Fig. 2). The inflection points of the sigmoidal temperature–fluorescence curves were taken as a measure of complex stability, with higher values of the inflection point reflecting a more stable flavin : dodecin complex. Our data revealed that both dodecins are only able to bind FMN and FAD (but not riboflavin) weakly at physiological pH, but more strongly at a more acidic pH, supporting the results of our titration experiment. As FMN binding was observed under more acidic conditions, titration experiments with FMN and *Streptomyces* dodecins were performed at pH 5 (Fig. S2). As a result of this experiment, the K_D for FMN binding to ScDod and SdDod was estimated to be between 50 and 150 nM. This apparent binding affinity is lower than that for the bacterial

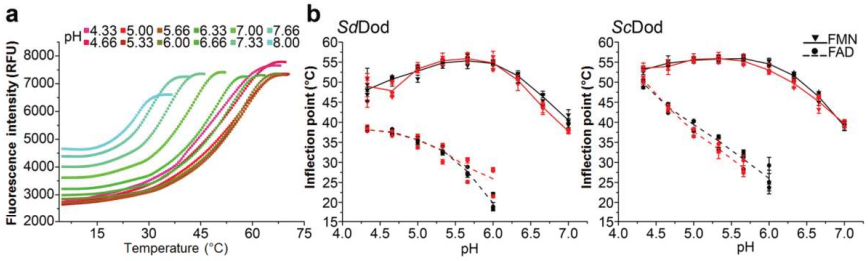


Fig. 2. Binding of the flavins FMN and FAD by ScDod and SdDod is favoured under acidic conditions. The temperature-dependent dissociation of flavin : dodecin complexes can be determined by measuring the release of fluorescent flavins. Inflection points (exemplarily shown for FMN and SdDod in panel (a) are a measure of complex stability, with higher inflection point values reflecting a more stable flavin : dodecin complex. The initial low fluorescence shown in panel (a) is caused by the quenching of bound FMN. With increasing temperature (stepwise temperature increases of 1.0 °C 60 s^{-1}), fluorescence increases as FMN is released. Experiments were carried out at different pH values and the curves are colour coded as indicated in the figure inset. (b) Inflection points of FMN : SdDod, FAD : SdDod, FMN : ScDod and FAD : ScDod complexes plotted against pH value. The average of three biological replicates is depicted as a solid (FMN) or dashed (FAD) line. Inflections points were only determined for certain pH values (FAD, pH 5.66 to 6.00; FMN, pH 7.00), because a fit with reasonable parameters was not possible above those values, although flavin binding was still observed. A lower inflection point temperature for e.g. FAD indicates that FAD : dodecin complexes are less stable and that binding is less probable.

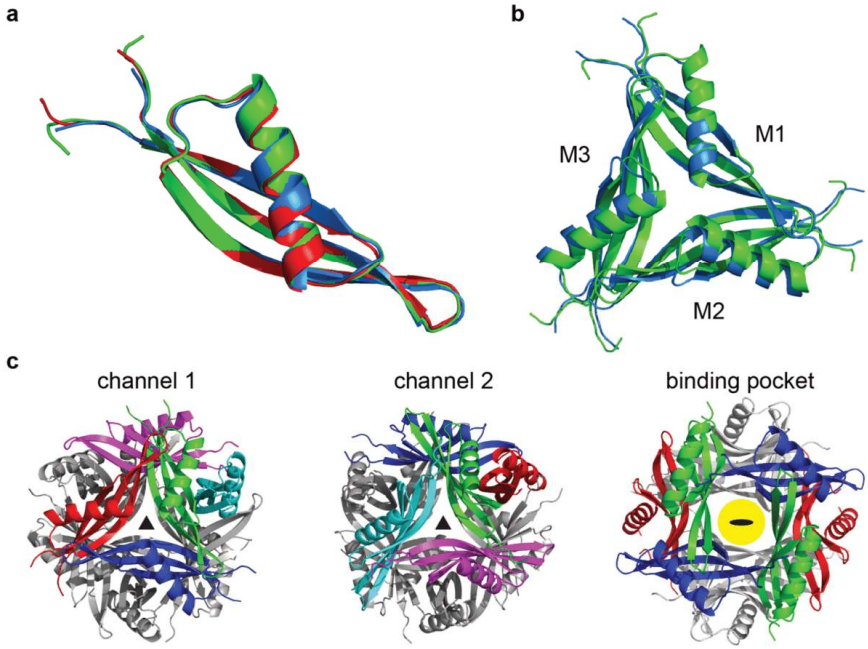


Fig. 3. The X-ray structures of SdDod and ScDod are highly similar. (a) Alignment of SdDod (blue), ScDod (green) and, as a reference, MtDod (red) [11]. The folds of the monomers are highly conserved and show only minor deviations. (b) Superposition of SdDod (blue) and ScDod (green) trimers aligned via monomer 1 (M1). While the monomers themselves align well, the relative positioning of the monomers in the trimer differs. M2 shows a significant deviation between SdDod and ScDod, causing channel 1 of SdDod to be wider than channel 1 of ScDod. In contrast, the trimers of ScDod and MtDod superimpose well (not shown). (c) Core features of the dodecin dodecamer shown for SdDod. The monomers of channel/trimer 1 are coloured in green, blue and red. The monomers in channel/trimer 2 are coloured in magenta, cyan and green. The binding pocket (highlighted in yellow) is located between trimers and embedded between two channel 1 trimers.

dodecin MtDod (K_D=18 nM, pH 5.0) and is one of the lowest found for bacterial dodecins [10, 16].

The three-dimensional structures of SdDod and ScDod are highly similar

As a next step with regard to the characterization of SdDod and ScDod, the three-dimensional structures of SdDod (PDB ID: 6RI3) and ScDod (PDB ID: 6R1E) were determined at resolutions of 2.4 Å and 2.6 Å, respectively and the structures were found to be highly similar. The monomers are composed of three β-sheets that partly enwrap an α-helix ($\beta_1\alpha_1\beta_2\beta_3$ topology) and the alpha-carbon coordinates align well to each other (Fig. 3a), meaning that the root-mean-square deviations (RMSDs) between the compared dodecin monomers are in the range of 1 Å. The quaternary structures of SdDod and ScDod are similar as well, except for the fact that the trimers show differences in the relative positioning of the monomers. Monomers 2 show a significant deviation between SdDod and ScDod, causing channel 1 of SdDod to be wider than channel 1 of ScDod(Fig. 3b). The 12 monomers of SdDod or ScDod assemble to a hollow spherical structure with 6 flavin dimer-binding pockets located at the twofold axes (Fig. 3c). Two types of channel are formed at the threefold axes, termed channel 1 and channel 2, differing with regard to the trimeric interphases and bound ions. In line with the similar binding behaviour of SdDod and ScDod, the nine amino acids relevant for flavin binding (H4, Y6, D38, W39, E41, R46, Q48, Q58 and R66) superimpose well in the structures of both dodecins (not shown). These residues also superimpose well with the corresponding amino acids in MtDod, except for H48 (see Fig. 1b), where SdDod and ScDod contain a glutamine residue (Q48). Owing to the high structural conservation of the flavin-binding pockets, the different binding properties (see the first results section) are best explained by the specific impact of the exchange of H48 in MtDod for Q48 in SdDod and ScDod. Unfortunately, the co-crystallization of dodecins with flavins was not successful and thus we were not able to obtain information on flavin binding at the atomic level.

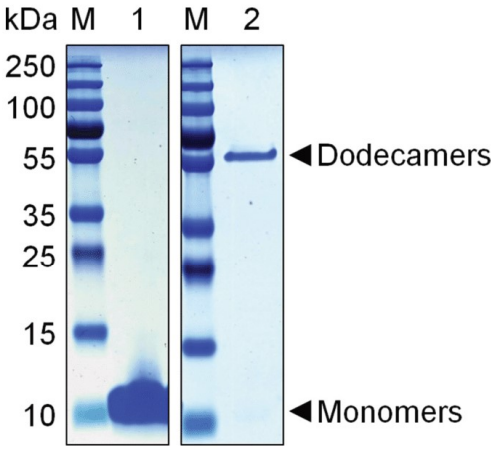


Fig. 4. Denaturing SDS-PAGE analysis of purified His_6-tagged dodecins from *S. davaonensis* (SdDod) (left) and *S. coelicolor* (ScDod) (right) reveals differences with regard to the formation of multimers. Lane M represents the marker. SdDod (lane 1) is present as a monomer under denaturing conditions, whereas ScDod (lane 2) apparently forms a multimer, most likely a dodecamer. The dodecameric form of ScDod appears to be smaller than the expected 144 kDa, which can be explained by its compact spherical structure.

SDS-PAGE analysis of purified ScDod and SdDod reveals differences in the stability of dodecameric multimers

When SdDod and ScDod were analysed by SDS-PAGE under denaturing conditions (2% SDS, 10 mM DTT), SdDod was found to be present only in its monomeric form, whereas ScDod (as expected for a bacterial dodecin) formed a multimer(Fig. 4). ScDod subsequently was challenged with different denaturing agents and conditions (Fig. 5). The monomeric form of ScDod was only obtained when the protein samples were incubated at 95 °C for more than 20 min in the presence of 2% SDS, 10 mM DTT, 2% β-mercaptoethanol, 1 mM EDTA and 8M urea, and we concluded that the ScDod dodecamer was remarkably more stable than ScDod.

The amino acid D10 of SdDod is responsible for the reduced stability of the dodecameric form of SdDod

Structural analysis of the dodecin from *M. tuberculosis* (MtDod) [11] suggested that three salt bridges in particular contribute to stabilization of the dodecameric form of MtDod, namely E10–K62, R29–D69 and E41–R46 (Fig. 1b). Similar salt bridges (E10–K62, R29–E69 and E41–R46) were found to be present in ScDod. In SdDod the corresponding salt bridges were predicted to be D10–K62, R29–E69 and E41–R46, and we hypothesized that the salt bridge D10–K62 in SdDod (replacing E10–K62 in ScDod) was responsible for the strongly reduced stability of the SdDod dodecamer observed in the presence of 2% SDS and 10 mM DTT when compared to ScDod (see preceding paragraph). Site-directed

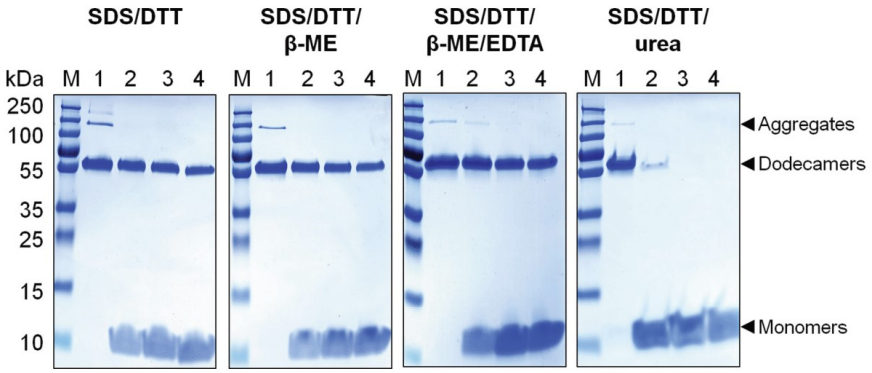


Fig. 5. The dodecameric form of purified His_6-tagged dodecin from *S. coelicolor* (ScDod) is remarkably stable. Pure samples of ScDod were analysed by SDS-PAGE and staining with Coomassie brilliant blue G-250 under different denaturing conditions. Far left panel, 2% SDS (SDS), 10 mM DTT (DTT). Left panel, 2% SDS, 10 mM DTT, 2% *β*-mercaptoethanol (*β*-ME). Right panel, 2% SDS, 10 mM DTT, 2% *β*-mercaptoethanol, 1 mM EDTA (EDTA). Far right panel, 2% SDS, 10 mM DTT, 8M urea (urea). Lane M represents the marker. Lane 1 represents the control samples of ScDod, which were incubated at room temperature. Lane 2 represents the ScDod samples that were incubated at 95 °C for 10 min. Lane 3 represents the ScDod samples that were incubated at 95 °C for 20 min. Lane 4 represents the ScDod samples that were incubated at 95 °C for 30 min. Complete dissociation of the dodecamers was only observed when protein samples were incubated at 95 °C for more than 20 min in the presence of 2% SDS, 10 mM DTT and 8M urea, indicating that ScDod was remarkably stable.

Table 1. Dodecins from various bacterial species and their properties

Organism	Variant	Residue present in[a]	Possible interaction/function	SEC[b]	SDS-PAGE[c]	Thermocyclic assay[d] (°C)
M. tuberculosis	WT	–	–	D*	D*	>95
S. davaonensis	WT	–	–	D	M	87
	D10E	*Hh*, *Mt*, *Tt*, *Sc*	K62	D	D	>95
	E17D	*Mt*, *Sc*	–	D	M	–
	D20E	*Hh*, *Tt*	Ion binding in channel 1	D	M	91
	N28T	*Sc*	–	D	M	–
	H35R	*Hh*, *Mt*, *Tt*, *Sc*	Coenzyme A binding	D	M	85
	N36K	*Sc*	D68	D	M	–
	E44Q	*Sc*	R46	D	M	–
	N50E	*Hh*, *Sc*	–	D	M	–
	T60G	*Sc*	Flavin binding	D	M	75
	M61L	*Sc*	–	D	M	–
	K62A	–	D10	D	M	73
	D68E	*Hh*, *Mt*, *Tt*, *Sc*	N36	D	M	–
	D10E/K62A	*Hh*, *Mt*, *Tt*, *Sc*/–	E10–K62 salt bridge pair	D	M	87
S. coelicolor	WT	–	–	D	D	>95
	D20E	*Hh*, *Tt*	Ion binding channel 1	D	D	>95
	Q44E	*Tt*, *Sd*	R46	D	D	–
	K62A	–	E10	D, more T	M	55
	Q44E/K62A	*Tt*, *Sd*/–	R46/E10	D, mostly T	D, M	–

a, *Hh*, *H. halophila*; *Mt*, *M. tuberculosis*; *Sd*, *S. davaonensis*; *Sc*, *S. coelicolor* and *Tt*, *T. thermophilus*.
b, D, Dodecamer; T, Trimer; M, Monomer.
c, For all SDS-PAGE experiments see Fig. S4.
d, Thermocyclic assay, temperature at which the specific dodecin loses the ability to completely rebind FMN (see also Fig. S6 and Table S3).

mutagenesis experiments were carried out to test this idea. In total, 13 different variant genes of SdDod were generated. The mutated genes and the properties of the corresponding variants of SdDod are listed in Table 1. All dodecin genes were expressed in *E. coli* Rosetta 2(DE3) and the variant proteins were purified to apparent homogeneity by affinity chromatography followed by SEC. While most variants behaved like the wild-type during purification, some ScDod variants showed an additional trimer peak upon SEC, indicating changes in quaternary structure (Fig. S3). Denaturing SDS-PAGE of the SdDod variants in comparison to wild-type SdDod revealed that D10 is indeed the key amino acid with regard to the relative lability of SdDod when compared to ScDod (Fig. S4). The variant SdDod-D10E was the only protein that was able to form multimers (dodecamers) under the applied conditions (Fig. S4, lane D10E). This can be explained by the stronger salt bridge E10–K62 induced by the carboxyl of E10, which is closer to K62 when compared to the carboxyl of the (shorter) amino acid residue D10 of wild-type SdDod. These results were nicely supported upon analysis of the ScDod variant ScDod-K62A (Fig. S4, lane K62A). In this dodecin variant, the formation of a salt bridge between E10 and A62 is not possible and this explains the clearly decreased tendency to form a dodecamer. All other ScDod variants behaved like the wild-type protein. As a next step in our analysis we employed blue native gel electrophoresis (BN-PAGE) and obtained a more differentiated picture with regard to the multimerization of wild-type dodecins and variants thereof (Fig. S5a–c). BN-PAGE was carried out in the presence of Coomassie G-250, which provided the necessary charges for electrophoretic separation of the proteins while being only a very mild denaturant. In contrast to the wild-type (Fig. S5a, lane 1), SdDod-D10E formed a dodecamer (Fig. S5a, lane 2), supporting the idea

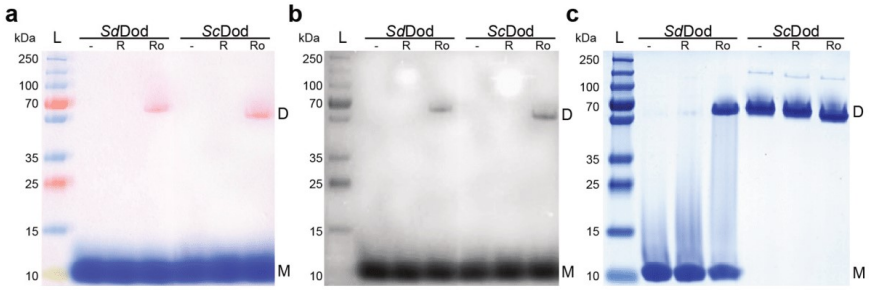


Fig. 6. Riboflavin and roseoflavin bind to ScDod and SdDod, although SdDod changes its quaternary structure in the presence of roseoflavin. The dodecins ScDod and SdDod (50 µM) were incubated in the presence of 250 µM riboflavin (R) or 250 µM roseoflavin (Ro). After incubation in the dark, the samples were mixed with SDS free loading dye and analysed by SDS-PAGE. Samples were loaded to the gel without heat treatment and the running buffer contained SDS to create a mildly denaturing environment. The documentation occurred under (a) visible light, (b) UV light and (c) after staining with Coomassie brilliant blue G-250. Roseoflavin is bound by both dodecins [pink band in (a) between 55–70 kDa and grey band in (b)] and is able to partially stabilize the SdDod dodecamer [monomer and dodecamer band in (c)]. Since ScDod does not denature under these conditions, the oligomer stabilizing effect cannot be tested employing this method. D, dodecamer. M, monomer. Lane L, PageRuler Plus Prestained Protein Ladder.

that D10 is responsible for the relative lability of the *S. davaonensis* dodecin. In addition, upon BN-PAGE analysis of SdDod-E44Q, we found that E44 of SdDod plays a role in multimerization and also that the variant SdDod-E44Q (in contrast to all other variants) formed a dodecamer (Fig. S5a, lane 8). The likely reason for this is that E44 of wild-type SdDod has the potential to weaken the salt bridge E41–R46 that is important for dodecamer formation. When E44 was replaced by Q44, E41–R46 was more stable, resulting in a multimer under the applied conditions. The importance of A62 in salt bridge formation was validated by BN-PAGE analysis of SdDod-K62A and the double mutant SdDod-D10E/K62A (Fig. S5b, lanes 12 and 14). BN-PAGE analysis of the *S. coelicolor* dodecin variants ScDod-Q44E, ScDod-K62A and ScDod-Q44E/K62A gave further insights into the multimerization of dodecins. ScDod-Q44E (Fig. S5c, lane 17) behaved like the wild-type, which can be explained by the fact that although E44 of the mutant (as in SdDod) weakens the salt bridge E41–R46, this does not play a role, as the salt bridges E10–K62 (notably this salt bridge is not present in SdDod) and R29–E69 provide enough stability with regard to the overall complex. When amino acid replacements affect the salt bridge E10–K62, as in the case of ScDod-K62A (Fig. S5c, lane 18), dodecamer formation does not occur, and this is even more pronounced in the double mutant ScDod-Q44E/K62A (Fig. S5c, lane 19). In contrast to the single mutant ScDod-Q44E, E44 in ScDod-Q44E/K62A now markedly interferes with the salt bridge E41–R46, which underscores the importance of E44 of SdDod with regard to stability of the multimer. A selection of mutants was also examined with regard to their thermal stability by a thermocyclic fluorescence assay at different pH values (Fig. S6) [10]. Overall, the data nicely supported the PAGE results and validated that ScDod is more stable than SdDod. Further, the data clearly support the idea that D10 of wild-type SdDod is responsible for this observed difference in stability. Intriguingly, SdDodK62A in this assay was apparently even less stable than the wild-type of SdDod (native D10–K62 interaction), suggesting that D10 in SdDod forms a weak interaction with K62. Introducing the D10E mutation to SdDodK62A, creating the double mutant SdDodD10E/K62A, recovers some stability, presumably through interactions of E10 with other amino acids in the dodecamer. A further interesting feature is that the K62A mutation has a smaller impact on SdDodD10–K62 (native interaction) than on ScDodE10–K62 (native interaction), making SdDodK62A significantly more stable than ScDodK62A. A possible explanation for this is that the weak interaction of D10 and K62 in SdDod needs to be compensated by other interactions and these compensating interactions reduce the impact of the K62A mutations. One of those compensating interactions may be mediated by E44 interacting with R46 (of the same monomer). Indeed, introducing this interaction in ScDod by mutating the low-stability mutant ScDodK62A to ScDodQ44E/K62A again leads to a more stable dodecamer. Notably, this mutation also affects dodecamer assembly, which is reflected by a minor dodecamer fraction in SEC (see Fig. S3). Table 1 summarizes the results of the different experiments.

Roseoflavin affects the dodecameric structure of the dodecin from *S. davaonensis*

The experiments described above showed that ScDod and SdDod have different properties with regard to multimerization and we speculated that the binding of flavins may also be affected by these different properties. Riboflavin is produced by both streptomycetes in similar amounts [30],

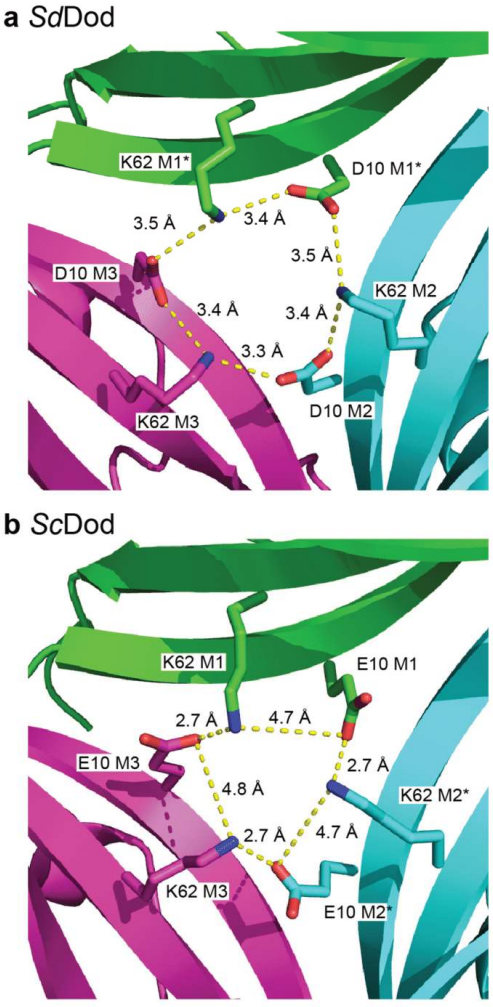


Fig. 7. Interactions in SdDod and ScDod at the atomic level. Channel 2 of SdDod and ScDod is shown and viewed from the inner side of the dodecamer. The monomers SdDod M1* and ScDod M2* were generated by crystallographic symmetry operations. Analysis of SdDod revealed that the inter- and intra-monomer salt bridges are of a similar distance ($D10^{MX}$ to $K62^{MX}$: 3.4±0.0 Å and $D10^{MX}$ to $K62^{MY}$: 3.4±0.1 Å). In ScDod the $E10^{MX}$ to $K62^{MY}$ salt bridge is 2.7±0.0 Å and the distance of the $E10^{MX}$ to $K62^{MX}$ salt bridge is 4.7±0.1 Å.

whereas roseoflavin is an antibiotic that is only produced by *S. davaonensis* [31]. Purified preparations of both dodecins (apo-forms) were incubated in the presence of a fivefold molar excess of riboflavin or roseoflavin in the dark (to avoid flavin degradation) and analysed by SDS-PAGE under denaturing conditions (but without heat treatment)(Fig. 6). The gels were evaluated under visible light (to detect yellow riboflavin and red roseoflavin), under UV light (to detect fluorescent riboflavin) and by staining with Coomassie brilliant blue G-250 (to detect the proteins). It can be clearly seen that formation of dodecamers of SdDod was induced upon the addition of roseoflavin, whereas the structurally very similar riboflavin did not have such an effect. In a subsequent similar experiment different amounts of roseoflavin were added to apo-SdDod and the shift from 9 kDa to 108 kDa was found to occur in the presence of 100 µM roseoflavin (Fig. S7). In case of apo-ScDod no differences could be detected (Fig. S7), since apo-ScDod forms dodecamers even in the absence of flavins.

FMN affects the multimerization of the dodecin from *S. davaonensis*

The following experiments were carried out to evaluate putative differences between the two streptomycete dodecins with regard to the binding of FMN, the preferred ligand of bacterial dodecins. Purified preparations of SdDod and ScDod were incubated in the presence of a fivefold molar excess of FMN and analysed by denaturing SDS-PAGE (Fig. S8). Wild-type dodecin from *S. davaonensis* showed a clear shift towards a higher molecular mass in the presence of FMN, indicating that FMN promoted the formation of multimers (probably trimers). Formation of the dodecamer, however, did not occur (Fig. S8). As expected, *S. davaonensis* SdDod-D10E already showed the dodecameric structure in the absence of FMN (see also Figs S4 and S5) and thus no difference could be observed in the presence of FMN. Some amino acid replacements in SdDod (E17D; D20E; N28T; N36K; N50E) led to an apparent stimulation of multimerization (Fig. S8). SdDod-E44Q apparently formed a dodecamer in the presence of FMN (Fig. S8) underscoring the importance of the salt bridge E41–R46 in multimerization and the role of the competing E44. Wild-type ScDod apparently formed a dodecamer in the absence as well as in the presence of FMN, and the same was true for ScDod-D20E (Fig. S8). The variant ScDod-K62A formed a monomer (and not a dodecamer) in the absence of FMN. In the presence of FMN a trimer (but not a dodecamer) was found. The missing salt bridge E10–K62 apparently affected FMN binding and thus this mutant did not shift to the multimeric structure and rather behaved like the wild-type SdDod, where the salt-bridge D10–K62 is weaker.

DISCUSSION

The streptomycete dodecins SdDod and ScDod show a typical feature of bacterial flavin-binding dodecins as they predominantly bind FMN in contrast to the haloarchaeal dodecins, which prefer riboflavin. Still, SdDod and ScDod are remarkable, because they bind flavins with a very low affinity. FMN and FAD are only bound at low temperatures and/or under acidic conditions, while riboflavin binding is too weak to be observed in our different experimental settings. At present

it is unclear whether these differences, which have been observed *in vitro*, are of physiological relevance. A sequence comparison of SdDod and ScDod to the actinobacterial MtDod (which binds flavins better) immediately suggests that the amino acid residue Q48 in SdDod and ScDod (see Fig. 1b) is the reason for the low flavin affinity of SdDod and ScDod. It is the only exchange in flavin-binding amino acids when compared to MtDod, which harbours a histidine residue at this position (H48). In principle, Q48 of SdDod and ScDod would also be able to able to form a hydrogen bond with the ribityl moiety of flavins. However, we were not able to co-crystallize SdDod and ScDod with flavins and thus could not find evidence for such a hydrogen bond. Notably, the dodecin from *T. thermophilus* carries a threonine at a comparable position (which does not form a hydrogen bond with the ribityl of the flavin) and still binds FMN and riboflavin with similar affinities to MtDod [10, 16]. Another possibility for the low apparent flavin affinity of SdDod and ScDod could be the slightly different arrangement of the dodecin monomers in the dodecamer affecting the geometry of the flavin-binding pocket (Fig. 3b).

An interesting result of our study was the finding that multimerization of SdDod was promoted in the presence of >20 µM roseoflavin. Wild-type *S. davaonesis* produces about 20 µM roseoflavin in a liquid medium under laboratory conditions and it was therefore tempting to speculate that SdDod was involved in resistance to this antibiotic. However, since a dodecin-deficient variant of *S. davaonensis* grows perfectly in the presence of 200 µM roseoflavin [24], a key role for SdDod in resistance to roseoflavin appears to be unlikely. We cannot rule out, of course, the possibility that SdDod plays a role in flavin homeostasis, storing riboflavin and roseoflavin, however, such a role is extremely difficult to show. Alternatively, the dodecins in streptomycetes could play a role in delivering flavins to flavoenzymes and/or to roseoflavin biosynthetic enzymes. FMN is the direct precursor of 8-demethyl-8-aminoriboflavin-5′-phosphate, which is subsequently converted to roseoflavin.

To our surprise, we found strong differences between SdDod and ScDod with regard to dodecamer stability. Thermal melting curves revealed that SdDod is stable up to 88 °C and ScDod does not denature below 95 °C (see Table 1). Our mutagenesis studies revealed that the lower stability of SdDod is caused by the weaker salt bridge between the $D10^{M1}$ and $K62^{M2}$ of two dodecin monomers in channel 2 (the superscript M1 and M2 denote monomers of trimeric units). In ScDod the comparable salt bridge is $E10^{M1}$–$K62^{M2}$. In SdDod neither the intra- nor the inter-monomer salt bridge is dominant(Fig. 7a); both are weaker than the inter-monomer (M1 to M2) salt bridge of ScDod (Fig. 7b). Mutating D10 to E10 in SdDod, creating the mutant SdDodD10E, led to a similar thermal stability to that of ScDod. In line with the important role of this salt bridge, the high thermal stability of ScDod is diminished in the mutant ScDodK62A, where the salt bridge is no longer present. The mutation of E10 to D10 during the course of evolution was tolerated because the protein remained highly stable. The biological reason for the apparent differences in the stability of ScDod and SdDod unfortunately remain unclear.

Funding information
This work was funded by the research training group NANOKAT (FKZ 0316052A) of the German Federal Ministry of Education and Research.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Fischer M, Bacher A. Biosynthesis of flavocoenzymes. *Nat Prod Rep* 2005;22:324–350.
2. Joosten V, van Berkel WJH. Flavoenzymes. *Curr Opin Chem Biol* 2007;11:195–202.
3. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* 2009;5:593–599.
4. García-Angulo VA. Overlapping riboflavin supply pathways in bacteria. *Crit Rev Microbiol* 2017;43:196–209.
5. Jaehme M, Slotboom DJ. Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochim Biophys Acta* 2015;1850:565–576.
6. Serrano A, Ferreira P, Martínez-Júlvez M, Medina M. The prokaryotic FAD synthetase family: a potential drug target. *Curr Pharm Des* 2013;19:2637–2648.
7. Monaco HL. Crystal structure of chicken riboflavin-binding protein. *Embo J* 1997;16:1475–1483.
8. White HB, Merrill AH. Riboflavin-Binding proteins. *Annu Rev Nutr* 1988;8:279–299.
9. Grininger M, Staudt H, Johansson P, Wachtveitl J, Oesterhelt D. Dodecin is the key player in flavin homeostasis of archaea. *J Biol Chem* 2009;284:13068–13076.
10. Bourdeaux F, Hammer CA, Vogt S, Schweighöfer F, Nöll G *et al.* Flavin Storage and Sequestration by *Mycobacterium tuberculosis* Dodecin. *ACS Infect Dis* 2018;4:1082–1092.
11. Liu F, Xiong J, Kumar S, Yang C, Ge S *et al.* Structural and biophysical characterization of *Mycobacterium tuberculosis* dodecin Rv1498A. *J Struct Biol* 2011;175:31–38.
12. Vinzenz X, Grosse W, Linne U, Meissner B, Essen L-O. Chemical engineering of *Mycobacterium tuberculosis* dodecin hybrids. *Chem Commun* 2011;47:11071–11073.
13. Bieger B, Essen LO, Oesterhelt D. Crystal structure of halophilic dodecin: a novel, dodecameric flavin binding protein from *Halobacterium salinarum*. *Structure* 2003;11:375–385.
14. Grininger M, Seiler F, Zeth K, Oesterhelt D. Dodecin sequesters FAD in closed conformation from the aqueous solution. *J Mol Biol* 2006;364:561–566.
15. Staudt H, Oesterhelt D, Grininger M, Wachtveitl J. Ultrafast excited-state deactivation of flavins bound to dodecin. *J Biol Chem* 2012;287:17637–17644.
16. Meissner B, Schleicher E, Weber S, Essen LO. The dodecin from *Thermus thermophilus*, a bifunctional cofactor storage protein. *J Biol Chem* 2007;282:33142–33154.
17. Arockiasamy A, Aggarwal A, Savva CG, Holzenburg A, Sacchettini JC. Crystal structure of calcium dodecin (Rv0379), from *Mycobacterium tuberculosis* with a unique calcium-binding site. *Protein Sci* 2011;20:827–833.
18. Landwehr W, Kämpfer P, Glaeser SP, Rückert C, Kalinowski J *et al.* Taxonomic analyses of members of the *Streptomyces cinnabarinus* cluster, description of *Streptomyces cinnabarigriseus* sp. nov. and *Streptomyces davaonensis* sp. nov. *Int J Syst Evol Microbiol* 2018;68:382-393–.
19. Pedrolli DB, Jankowitsch F, Schwarz J, Langer S, Nakanishi S *et al.* Riboflavin analogs as antiinfectives: occurrence, mode of action, metabolism and resistance. *Curr Pharm Des* 2012.

20. **Langer S**, **Hashimoto M**, **Hobl B**, **Mathes T**, **Mack M**. Flavoproteins are potential targets for the antibiotic roseoflavin in *Escherichia coli*. *J Bacteriol* 2013;195:4037–4045.
21. **Langer S**, **Nakanishi S**, **Mathes T**, **Knaus T**, **Binter A** *et al*. The flavoenzyme azobenzene reductase AzoR from *Escherichia coli* binds roseoflavin mononucleotide (RoFMN) with high affinity and is less active in its RoFMN form. *Biochemistry* 2013;52:4288–4295.
22. **Pedrolli DB**, **Jankowitsch F**, **Schwarz J**, **Langer S**, **Nakanishi S** *et al*. Natural riboflavin analogs. *Methods Mol Biol* 2014;1146:41–63.
23. **Mathes T**, **Vogl C**, **Stolz J**, **Hegemann P**. *In vivo* generation of flavoproteins with modified cofactors. *J Mol Biol* 2009;385:1511–1518.
24. **Ludwig P**, **Sévin DC**, **Busche T**, **Kalinowski J**, **Bourdeaux F** *et al*. Characterization of the small flavin-binding dodecin in the roseoflavin producer *Streptomyces davawensis*. *Microbiology* 2018;164:908–919.
25. **Magill NG**, **Cowan AE**, **Koppel DE**, **Setlow P**. The internal pH of the forespore compartment of *Bacillus megaterium* decreases by about 1 pH unit during sporulation. *J Bacteriol* 1994;176:2252–2258.
26. **Wilks JC**, **Slonczewski JL**. pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. *J Bacteriol* 2007;189:5601–5607.
27. **Quirós L**, **Salas J**. Intracellular water volume and internal pH of *Streptomyces antibioticus* spores. *FEMS Microbiol Lett* 1996;141:245–249.
28. **Chae W-B**, **Kim Y-B**, **Choi S-W**, **Lee H-B**, **Kim E-K**. Enhancing the sporulation of *Streptomyces kasugaensis* by culture optimization. *Korean J Chem Eng* 2009;26:438–443.
29. **Bertani G**. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J Bacteriol* 2004;186:595–600.
30. **Pedrolli DB**, **Matern A**, **Wang J**, **Ester M**, **Siedler K** *et al*. A highly specialized flavin mononucleotide riboswitch responds differently to similar ligands and confers roseoflavin resistance to *Streptomyces davawensis*. *Nucleic Acids Res* 2012;40:8662–8673.
31. **Jankowitsch F**, **Schwarz J**, **Rückert C**, **Gust B**, **Szczepanowski R** *et al*. Genome sequence of the bacterium *Streptomyces davawensis* JCM 4913 and heterologous production of the unique antibiotic roseoflavin. *J Bacteriol* 2012;194:6818–6827.

Edited by: S. V. Gordon and M. Elliot

Comparative biochemical and structural analysis of the flavin-binding dodecins from *Streptomyces davaonensis* and *Streptomyces coelicolor* reveals striking differences with regard to multimerization

Running title: *Streptomyces* dodecins

Florian Bourdeaux[*1], Petra Ludwig[*2], Karthik S. Paithankar[1], Bodo Sander[3], Lars-Oliver Essen[3], Martin Grininger[1] and Matthias Mack[2#]

[*]These authors contributed equally to this work.

[#]Corresponding author.

[1]Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

[2]Institute for Technical Microbiology, Faculty for Biotechnology, Mannheim University of Applied Sciences, 68163 Mannheim, Germany

[3]Unit for Structural Biology, Department of Chemistry and Biology, Philipps University Marburg, 35032 Marburg, Germany

Correspondent footnote:

Matthias Mack

Institute for Technical Microbiology

Mannheim University of Applied Sciences

Paul-Wittsack-Str. 10

68163 Mannheim, Germany

Tel.:+49-621-292-6496

Fax:+49-621-292-6420

E-mail: m.mack@hs-mannheim.de

Supporting Tables and Figures

Table S1: Strains and plasmids used in this report.

Strains	Description	Source or reference
Streptomyces davawensis	wildtype strain (JCM 4913)	Japan Collection of Microorganisms
Streptomyces coelicolor A3(2)	wildtype strain (DSMZ 40783)	German Collection of Microorganisms and Cell Cultures GmbH
Escherichia coli Rosetta 2(DE3)	F^- *ompT* *hsdS*$_B$(r_B^- m_B^-) gal dcm (DE3) pRARE2 (CamR)	Merck
Plasmids	nucleotide sequence (T7 promoter to terminator)	Primary structure
pET24a(+)*Sd*Dod-His$_6$	**taatacgactcactatagg**ggaattgtgagcggataacaat tcccctctagaaataattttgtttaactttaagaaggagatatac atATGTCGAACCACACCTACCGGGTCA CGGACATCGTCGGCACCTCGCCCGAG GGCGTGGACCAGGCGATCCGTAACG GCATCAACCGGGCCTCGCAGACCCTG CACAACCTGGACTGGTTCGAGGTGGT CGAGGTACGCGGCCAGCTCAACGAC GGGCAGATCGCGCACTGGCAGGTGA CCATGAAGGTCGGCTTCCGCCTGGAC GAGACCGGTCTCGAGCACCACCACCA CCACCACTGAgatccgg**ctgctaacaaagcccga aaggaagctgagttggctgctgccaccgctgagcaataa ctagcataacccccttggggcctctaaacgggtcttgaggg gttttttgctgaaaggaggaactatatccggat**	MSNHTYRVTDIVGT SPEGVDQAIRNGINR ASQTLHNLDWFEVV EVRGQLNDGQIAHW QVTMKVGFRLDETG LEHHHHHH*
pET24a(+)*Sc*Dod-His$_6$	**taatacgactcactatagg**ggaattgtgagcggataacaat tcccctctagaaataattttgtttaactttaagaaggagatatac atATGTCGAACCACACCTACCGGGTCA	MSNHTYRVTEVVGT SPDGVDQAVRNAVT RASQTLRKLDWFEV TQVRGQIEDGQVAH

Page 3 Ludwig et al. 6/29/2019 12:42 PM

	CCGAGGTCGTCGGCACCTCGCCCGAC GGAGTCGACCAGGCCGTCCGCAACG CGGTCACGCGCGCCTCGCAGACCCTG CGCAAGCTGGACTGGTTCGAGGTGAC GCAGGTGCGCGGCCAGATCGAGGAC GGGCAGGTCGCGCACTGGCAGGTCG GCCTCAAGCTCGGCTTCCGCCTGGAG GAGTCCGACCTCGAGCACCACCACCA CCACCACTGAgatccgg**ctgctaacaaagcccga** **aaggaagctgagttggctgctgccaccgctgagcaataa** **ctagcataaccccttggggcctctaaacgggtcttgaggg** **gttttttgctgaaaggaggaactatatccggat**	WQVGLKLGFRLEES DLEHHHHHH*
pET28a(+)*Sc*Dod	**taatacgactcactatagg**ggaattgtgagcggataacaat tcccctctagaaataattttgtttaactttaagaaggagatatac cATGTCGAACCACACCTACCGGGTCA CCGAGGTCGTCGGCACCTCGCCCGAC GGAGTCGACCAGGCCGTCCGCAACG CGGTCACGCGCGCCTCGCAGACCCTG CGCAAGCTGGACTGGTTCGAGGTGAC GCAGGTGCGCGGCCAGATCGAGGAC GGGCAGGTCGCGCACTGGCAGGTCG GCCTCAAGCTCGGCTTCCGCCTGGAG GAGTCCGACTGActcgagcaccaccaccaccacc actgagatccgg**ctgctaacaaagcccgaaaggaagctg** **agttggctgctgccaccgctgagcaataactagcataacc** **ccttggggcctctaaacgggtcttgaggggttttttgctga** **aaggaggaactatatccggat**	MSNHTYRVTEVVGT SPDGVDQAVRNAVT RASQTLRKLDWFEV TQVRGQIEDGQVAH WQVGLKLGFRLEES D*

Sequences in bold, T7 promoters and T7 terminators; Capital nucleotides, translated sequence;

Table S2: Primers used for initial plasmid cloning and mutations.

Cloning Primer		
Primer name	Sequence 5'-3'	Used for Plasmid
P1333NdeI-His-FP	ATATATCATATGTCGAACCACACCTACCGGGTCACGGAC	pET24a(+)*Sd*Dod-His_6
P1333XhoI-His-RP	ATATATCTCGAGACCGGTCTCGTCCAGGCGGAAGCC	
PSCO0915NdeI-His-FP	ATATATCATATGTCGAACCACACCTACCGGGTCACCGAGG	pET24a(+)*Sc*Dod-His_6
PSCO0915XhoI-His-RP	ATATATCTCGAGGTCGGACTCCTCCAGGCGGAAGCCG	
PSCO0915PciI-FP*	GGCACATGTCGAACCACACCTACCG	pET28a(+)*Sc*Dod*
PSCO0915XhoI-RP*	CATCTCGAGTCAGTCGGACTCCTCCAG	
Mutation-primer		
Primer name	Sequence 5'-3'	Plasmid
Sd-D10E-FP	GGTGCCGACGATCTCCGTGACCCGG	pET24a(+)*Sd*Dod(D10E)-His_6
Sd-D10E-RP	CCGGGTCACGGAGATCGTCGGCACC	
Sd-E17D-FP	GTCCACGCCGTCGGGCGAGGTGC	pET24a(+)*Sd*Dod(E17D)-His_6
Sd-E17D-RP	GCACCTCGCCCGACGGCGTGGAC	
Sd-D20E-FP	CGGATCGCCTGCTCCACGCCCTC	pET24a(+)*Sd*Dod(D20E)-His_6
Sd-D20E-RP	GAGGGCGTGGAGCAGGCGATCCG	
Sd-N28T-FP	GAGGCCCGGGTGATGCCGTTACGGATC	pET24a(+)*Sd*Dod(N28T)-His_6
Sd-N28T-RP	GATCCGTAACGGCATCACCCGGGCCTC	
Sd-H35R-FP	CCAGTCCAGGTTGCGCAGGGTCTGCGA	pET24a(+)*Sd*Dod(H35R)-His_6
Sd-H35R-RP	TCGCAGACCCTGCGCAACCTGGACTGG	
Sd-N36K-FP	GAACCAGTCCAGCTTGTGCAGGGTCTGCG	pET24a(+)*Sd*Dod(N36K)-His_6
Sd-N36K-RP	CGCAGACCCTGCACAAGCTGGACTGGTTC	
Sd-E44Q-FP	GGCCGCGTACCTGGACCACCTCGAA	pET24a(+)*Sd*Dod(E44Q)-His_6
Sd-E44Q-RP	TTCGAGGTGGTCCAGGTACGCGGCC	
Sd-N50E-FP	CGATCTGCCCGTCCTCGAGCTGGCCGCGT	pET24a(+)*Sd*Dod(N50E)-His_6
Sd-N50E-RP	ACGCGGCCAGCTCGAGGACGGGCAGATCG	

Sd-M61L-FP	GAAGCCGACCTTCAGGGTCACCTGCCAGT	pET24a(+)*Sd*Dod(M6
Sd-M61L-RP	ACTGGCAGGTGACCCTGAAGGTCGGCTTC	1L)-His_6
Sd-K62A-FP	GGCGGAAGCCGACCGCCATGGTCACCTGCC	pET24a(+)*Sd*Dod(K62
Sd-K62A-RP	GGCAGGTGACCATGGCGGTCGGCTTCCGCC	A)-His_6
Sd-D68E-FP	GAGACCGGTCTCCTCCAGGCGGAAG	pET24a(+)*Sd*Dod(D68
Sd-D68E-RP	CTTCCGCCTGGAGGAGACCGGTCTC	E)-His_6
Sc-D20E-FP	CGGACGGCCTGCTCGACTCCGTC	pET24a(+)*Sc*Dod(D20
Sc-D20E-RP	GACGGAGTCGAGCAGGCCGTCCG	E)-His_6
Sc-Q44E-FP	GCCGCGCACCTCCGTCACCTCGA	pET24a(+)*Sc*Dod(Q44
Sc-Q44E-RP	TCGAGGTGACGGAGGTGCGCGGC	E)-His_6
Sc-K62A-FP	GCGGAAGCCGAGCGCGAGGCCGACCTGC	pET24a(+)*Sc*Dod(K62
Sc-K62A-RP	GCAGGTCGGCCTCGCGCTCGGCTTCCGC	A)-His_6

Table S3: Thermocyclic fluorescence data

		denaturation temperature / °C [a]					
Dodecin	heating duration	pH 5.0	pH 5.0 Citrate	pH 6.0	pH 6.5	pH 7.0	Average ± StDev
MtDod	3min	>95	>95	>95	>95	>95	>95
(WT)	6min	>95	>95	>95	>95	>95	>95
SdDod	3min	89	89	91	95	89	91 ± 3
(WT)	6min	87	89	87	91	87	88 ± 2
SdDod	3min	>95	>95	>95	>95	>95	>95
(D10E)	6min	>95	>95	>95	>95	>95	>95
SdDod	3min	89	89	>95	>95	>95	89 to >95
(D20E)	6min	87	87	91	95	91	90 ± 3
SdDod	3min	85	87	91	95	89	89 ± 4
(H35R)	6min	85	85	85	91	87	87 ± 3
SdDod	3min	73	79	83	93	85	83 ± 7
(T60G)	6min	71	75	75	85	85	78 ± 6
D62	3min	75	75	75	79	81	77 ± 3
	6min	73	73	73	75	79	75 ± 3
D1062	3min	87	87	91	91	89	89 ± 2
	6min	87	85	87	87	85	86 ± 1
CWT	3min	>95	>95	>95	>95	>95	>95
	6min	>95	>95	>95	>95	>95	>95
C20	3min	>95	>95	>95	>95	>95	>95
	6min	>95	>95	>95	>95	>95	>95
C62	3min	51	51	57	51	51	52 ± 3

6min	51	50	55	51	50	51 ± 2

[a]At these temperatures a significant increase of fluorescence was detectable upon analysis of FMN:SdDod (wild-type, WT and variants) and FMN:ScDod (wild-type, WT and variants), respectively. Values were determined *via* the first deviation of the fluorescence:temperature curves. First temperature points above a slope of 120 rfu/°C were selected for analysis.

Table S4: Crystallographic data collection and refinement statistics for *Sd*Dod and *Sc*Dod

	*Sd*Dod[a]	*Sc*Dod[a]
X-ray diffraction source	SLS, X06DA PX3	Bruker-Nonius FR591/CuKα
Detector	PILATUS 2M-F	MAR345
Voltage and current	-	50 kV and 80 mA
Wavelength (Å)	1.0718	1.5418
Temperature (K)	100	100
Space group	P43212	H32
a, b, c (Å)	72.29, 72.29, 151.21	93.20, 93.20, 228.29
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution range (Å)	36.1-2.4	12.0-2.6
Total no. of reflections	192150	66956
Unique reflections	16434	11644
Completeness (%)	100 (100)	99 (100)
Redundancy	11 (11)	5.8 (5.8)
<I/σ(I)>	28.7 (2.4)	28.6 (3.7)
R_{meas}	0.18 (1.1)	0.04 (0.4)
Wilson *B* factor (Å^2)	41	23
R_{work}/R_{free}	0.22/0.25	0.20/0.24
Solvent content	0.41	0.59
Monomers per unit cell	6	4
Number of residues	68 (chains A-F)	70 (chains A-D)
r.m.s.d bond lengths (Å), bond angles (°)	0.01, 1.4	0.01, 1.8

[a]Figures in parentheses indicate the values for the outer shells of the data.

Figure legends

Fig. S1 The dodecins from *Streptomyces davaonensis* (SdDod) and *Streptomyces coelicolor* (ScDod) bind FMN at pH 5.0 but not at pH 7.5. a. A constant amount of flavin (1 µM; RbF, riboflavin; FMN, flavin mononucleotide) was titrated with increasing amounts of purified apo-dodecin and fluorescence was measured (n=3). The drop in fluorescence indicated binding of FMN by SdDod and ScDod which occured at pH 5.0 but not at pH 7.5. At pH 7.5 SdDod was neither found to bind FMN nor riboflavin and the same was found for ScDod (data not shown). Since riboflavin (in contrast to FMN) does not change its charge upon protonation, it was not tested with regard to dodecin binding at pH 5.0. b. Normalized tryptophane fluorescence of SdDod titrated with FMN at pH 7.5. Except for the three first data points a linear increase in fluorescence indicated that formation of FMN:dodecin complexes did not occur.

Fig. S2 Titration experiment of SdDod and ScDod with FMN at pH 5.0. A constant amount of flavin (1 µM FMN) was titrated with increasing amounts of purified apo-dodecin (SdDod and ScDod) and fluorescence was measured (n=5). The decrease of flavin fluorescence is due to binding of flavins by dodecins and formation of an FMN:dodecin complex. The titration curve for MtDod at pH 5.0 was published earlier [1] and is shown for comparison. Equilibrium binding data were collected in volumes of 100 µL of mixtures of flavins and apo-dodecins at varying concentrations in 96 well plates. Flavins were used at concentrations of 1 µM. The plates were incubated for 1 h at room temperature in the dark. Flavin binding was recorded by reading out unbound flavin by excitation at 445 ± 9 nm and emission at 520 ± 20 nm. A K_D for FMN binding to ScDod and SdDod was estimated from these data to be between 50 and 150 nM. These apparent binding affinities for FMN are significantly lower when compared to the FMN binding affinity for bacterial dodecin MtDod (K_D = 18 nM, pH 5.0) and the lowest found for bacterial dodecins [1, 2]. Data were fitted with a single step binding model with Origin Pro 8.6. Dissociation constants were determined by non linear curve fitting (Origin Pro 8.6).

Fig. S3 Size exclusion chromatography (SEC) profiles of different wild-type dodecins SdDod (a) and ScDod (b) and variants thereof underscore the importance of the E10-K62 salt bridge. Dodecins were purified from recombinant *Escherichia coli* strains and analyzed by SEC. SdDod proteins eluted as dodecamers under these conditions. In contrast, the dodecamer:trimer ratio of the variants ScDod(K62A) and ScDod(Q44E/K62A) was impaired, underscoring the importance of the E10-K62 salt bridge.

Fig. S4 D10 of SdDod is the key amino acid with regard to multimerization/dodecamer formation under denaturing conditions. SDS-PAGE and Coomassie brilliant blue G-250 staining of purified His_6-tagged dodecin variants from *S. davaonensis* (SdDod; bold) and *S. coelicolor* (ScDod; not bold). Lanes L represent the molecular mass marker. Monomers (M) and putative dodecamers (D) are shown. Lane WT, SdDod wild-type. Lane D10E, SdDod-D10E, replacement of D10 by E allows formation of an additional salt bridge which in turn promotes dodecamer formation. E17D, SdDod-E17D. D20E, SdDod-D20E. N28T, SdDod-N28T. H35R, SdDod-H35R. N36K, SdDod-N36K. E44Q, SdDod-E44Q. N50E, SdDod-N50E. T60G, SdDod-T60G. M61L, SdDod-M61L. K62A, SdDod-K62A. D68E, SdDod-D68E. D10E/K62A, SdDod-D10E/K62A. WT, ScDod wild-type. D20E, ScDod-D20E. Q44E, ScDod-Q44E. K62A, ScDod-K62A. Q44E/K62A, ScDod-Q44E/K62A.

Fig. S5 Blue-native gel electrophoresis (BN-PAGE) reveals more differences with regard to multimerization of the dodecin variants than denaturing SDS-PAGE. Purified His_6-tagged dodecin variants from *S. davaonensis* (SdDod) were analyzed by BN-PAGE and Coomassie brilliant blue G-250 staining. In A. and B. dodecins from *S. davaonensis* were analysed whereas in C. dodecins from *S. coelicolor* were studied. Lanes M represent the marker; Monomers and putative multimers (trimers, dodecamers and even larger protein complexes) are shown. Lane 1, SdDod wild-type. Lane 2, SdDod-

D10E, replacement of D10 by E allows formation of an additional salt bridge which in turn promotes dodecamer formation. Lane 3, SdDod-E17D. Lane 4, SdDod-D20E. Lane 5, SdDod-N28T. Lane 6, SdDod-H35R. Lane 7, SdDod-N36K. Lane 8, SdDod-E44Q, Q44 does not as strongly interfere with salt bridge formation E41-R46 when compared to E44 which in turn promotes dodecamer formation. Lane 9, SdDod-N50E. Lane 10, SdDod-T60G. Lane 11, SdDod-M61L. Lane 12, SdDod-K62A. Lane 13, SdDod-D68E. Lane 14, SdDod-D10E/K62A. Lane 15, ScDod wild-type. Lane 16, ScDod-D20E. Lane 17, ScDod-Q44E. Lane 18, ScDod-K62A. Lane 19, ScDod-Q44E/K62A.

Fig. S6 Analysis of selected dodecin variants by thermocyclic fluorescence assays support the SDS-PAGE data in Fig. S4. Purified His_6-tagged dodecin variants from *S. davaonensis* (SdDod) were analyzed by thermocyclic fluorescence assays. Curves are color coded as indicated in figure insets. As the temperature-dependent dissociation assays (see Fig. 2), thermocyclic fluorescence assays uses the fluorescence of flavins as a read-out. However, cooling steps are integrated in the thermocyclic fluorescence assays allowing flavins to rebind to the dodecin. Thus, each cycle contains a heating and a cooling phase. In each cycle, the temperature increases during the heating step and the flavin fluorescence is measured at the end of the cooling phase. The fluorescence remains quenched until the temperature is reached at which the dodecamer starts to denature and released flavins can be detected in the cooling phase. This assay allows the determination of the thermal stability of the dodecin complexes. Increasing fluorescence indicates loss of the ability to bind FMN and thereby denaturation of the dodecamer. Wild-type SdDod and Wild-type ScDod clearly show differences in stability at all pH values supporting the PAGE results (Fig. S4 and Fig. S5). Analyses of the variant dodecins nicely support the PAGE results as well and show the importance of the E10-K62 salt bridge for the stability of the ScDod dodecamer. The variant ScDodK62A in turn is destabilized and even less stable when compared to wild-type SdDod and SdDod-K62A.

Fig S7 Roseoflavin affects multimerization of SdDod under denaturing conditions. Apo-dodecin (10 µM) from *S. davaonensis* was purified from a recombinant *E. coli* strain and incubated for 15 min in the absence and presence of flavins (RF, riboflavin, yellow; RoF, roseoflavin, red). The protein samples were analyzed by native PAGE with a running buffer containing SDS. The resulting gel was analyzed under visible light (A) and UV-light (B) prior to staining. Subsequently, the gel was stained with Coomassie brilliant blue G-250 and evaluated under visible light (C). Lane M, molecular mass marker. Lane 1, no flavin was added to SdDod. Lane 2, 20 µM RF was added. Lane 3, 100 µM RF. Lane 4, 150 µM RF. Lane 5, 200 µM RF. Lane 6, 20 µM RoF was added. Lane 7, 100 µM RoF. Lane 8, 150 µM RoF. Lane 9, 200 µM RoF. A. The yellow bands indicate that free riboflavin migrates in the gel. Red roseoflavin is present mainly bound to SdDod and thus migrated faster. B. Riboflavin displays a much stronger fluorescence when compared to roseoflavin and only free riboflavin is found [3]. C. The Coomassie brilliant blue G-250 stained gel shows a shift from 9 kDa to 108 kDa in the presence of roseoflavin (>20 µM) indicating formation of a dodecin dodecamer. When the protein was treated with riboflavin such a shift did not occur (lanes 2 to 6). In case of ScDod no differences could be detected (data not shown) since ScDod forms dodecamers in the absence as well as in the presence of flavins (not shown).

Fig. S8 FMN affects multimerization of the dodecin from *S. davaonensis*. Dodecins from *S. davaonensis* and *S. coelicolor* were analyzed by SDS-PAGE and Coomassie brilliant blue G-250 staining in the presence or absence of FMN as indicated (-, no addition of FMN; +, FMN is present). L, molecular mass marker. Lane WT, SdDod wild-type. Lane D10E, SdDod-D10E, replacement of D10 by E allows formation of an additional salt bridge which in turn promotes dodecamer formation. E17D, SdDod-E17D. D20E, SdDod-D20E. N28T, SdDod-N28T. H35R, SdDod-H35R. N36K, SdDod-N36K. E44Q, SdDod-E44Q. N50E, SdDod-N50E. T60G, SdDod-T60G. M61L, SdDod-M61L. K62A, SdDod-K62A. D68E, SdDod-D68E. D10E/K62A, SdDod-D10E/K62A. WT, ScDod wild-type. D20E, ScDod-

D20E. Q44E, ScDod-Q44E. K62A, ScDod-K62A. This variant does not form multimers in the absence of FMN, addition of FMN leads to multimerization as was observed for the dodecins from *S. davaonensis* Q44E/K62A, ScDod-Q44E/K62A.

Fig. S1

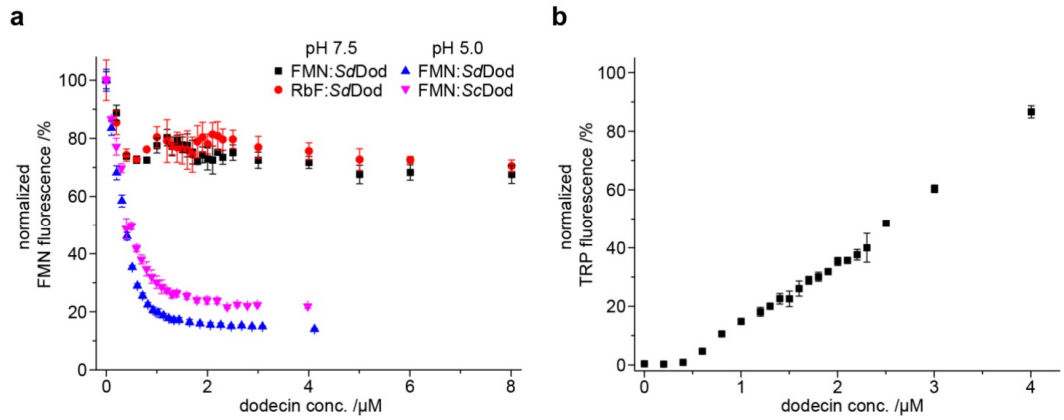

a
pH 7.5
pH 5.0
FMN:SdDod
RbF:SdDod
FMN:SdDod
FMN:ScDod
normalized FMN fluorescence /%
dodecin conc. /µM
b
normalized TRP fluorescence /%
dodecin conc. /µM

Fig. S2

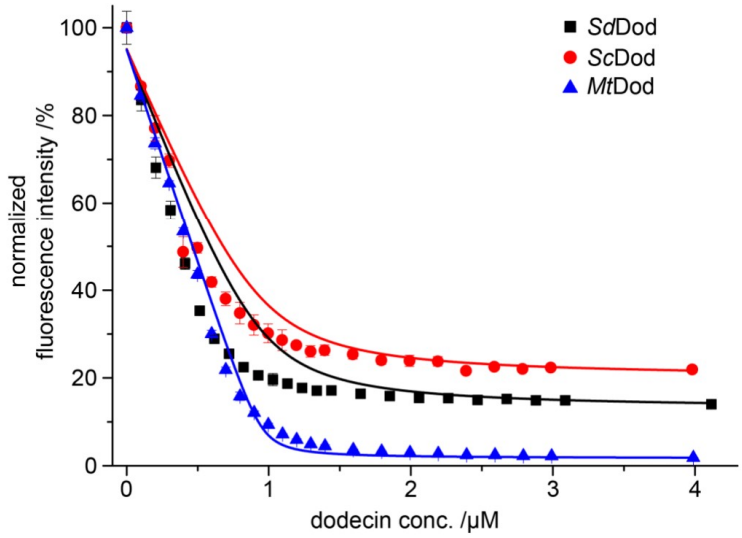

SdDod
ScDod
MtDod
normalized fluorescence intensity /%
dodecin conc. /µM
0
20
40
60
80
100
0
1
2
3
4

Fig. S3

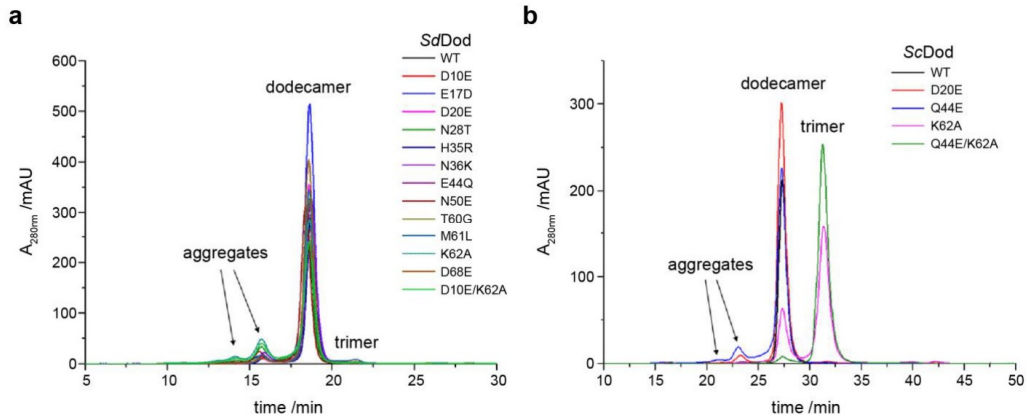

a
SdDod
WT
D10E
E17D
D20E
N28T
H35R
N36K
E44Q
N50E
T60G
M61L
K62A
D68E
D10E/K62A
dodecamer
aggregates
trimer
A280nm /mAU
time /min
b
ScDod
WT
D20E
Q44E
K62A
Q44E/K62A
dodecamer
trimer
aggregates
A280nm /mAU
time /min

Fig. S4

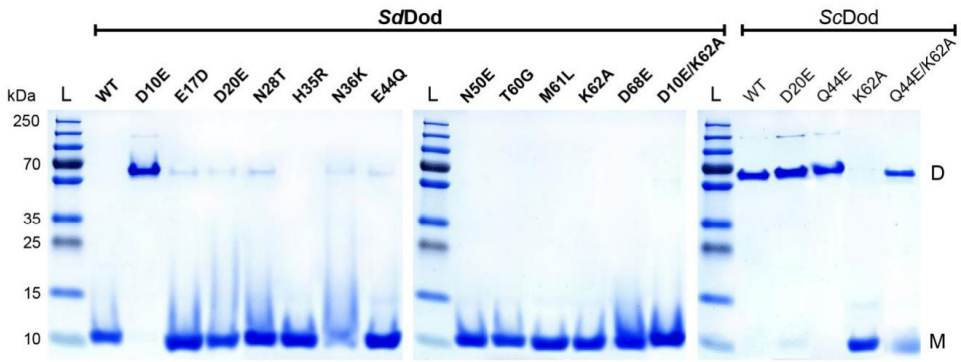

SdDod
ScDod
kDa
L
WT
D10E
E17D
D20E
N28T
H35R
N36K
E44Q
N50E
T60G
M61L
K62A
D68E
D10E/K62A
D20E
Q44E
K62A
Q44E/K62A
250
70
35
25
15
10
D
M

Fig. S5

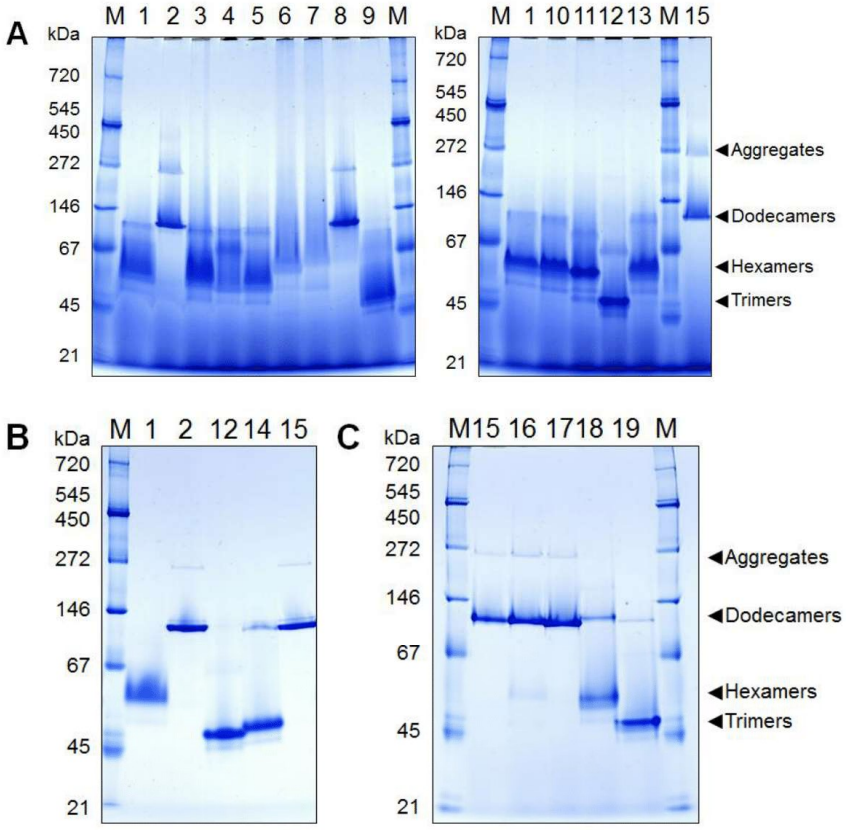

A
kDa
M 1 2 3 4 5 6 7 8 9 M
720
545
450
272
146
67
45
21
M 1 10 11 12 13 M 15
Aggregates
Dodecamers
Hexamers
Trimers
B
M 1 2 12 14 15
C
M 15 16 17 18 19 M

Fig. S6

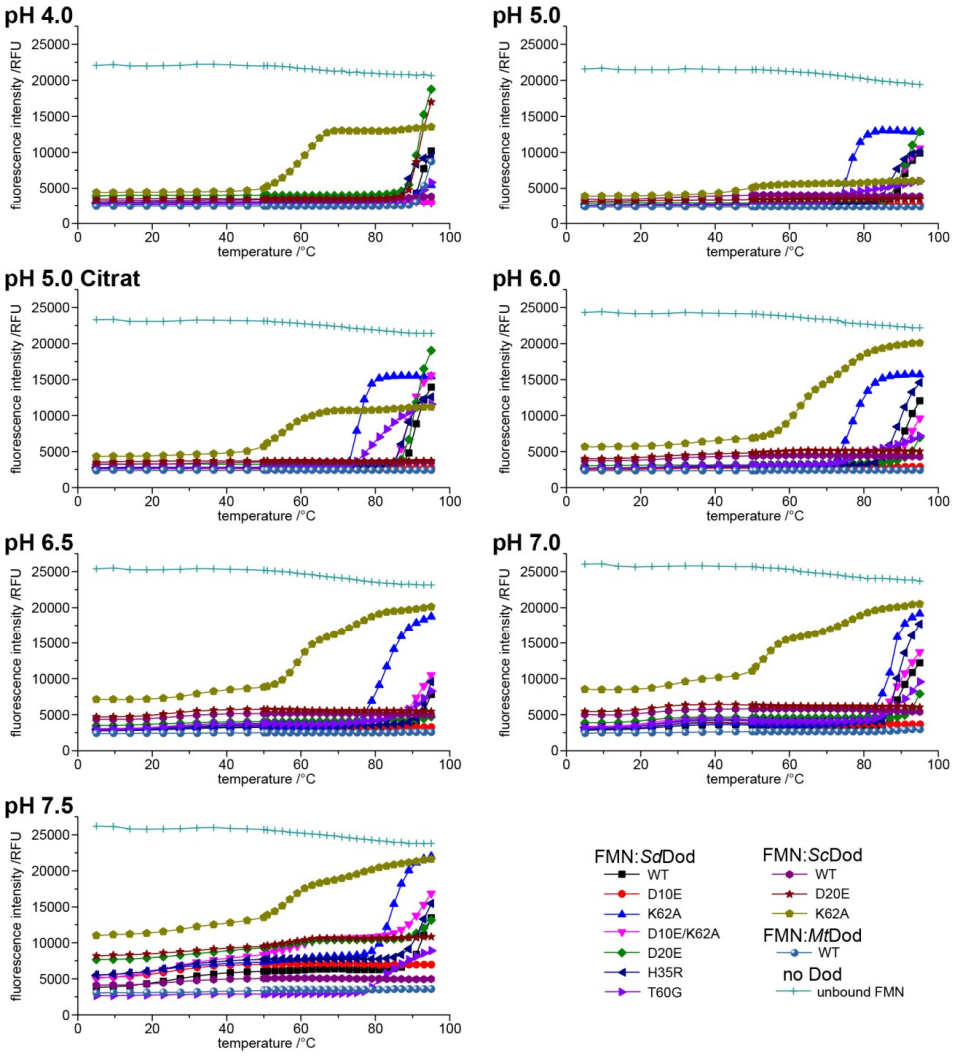

pH 4.0
pH 5.0
pH 5.0 Citrat
pH 6.0
pH 6.5
pH 7.0
pH 7.5
fluorescence intensity /RFU
temperature /°C
FMN:SdDod
WT
D10E
K62A
D10E/K62A
D20E
H35R
T60G
FMN:ScDod
WT
D20E
K62A
FMN:MtDod
WT
no Dod
unbound FMN

Fig. S7

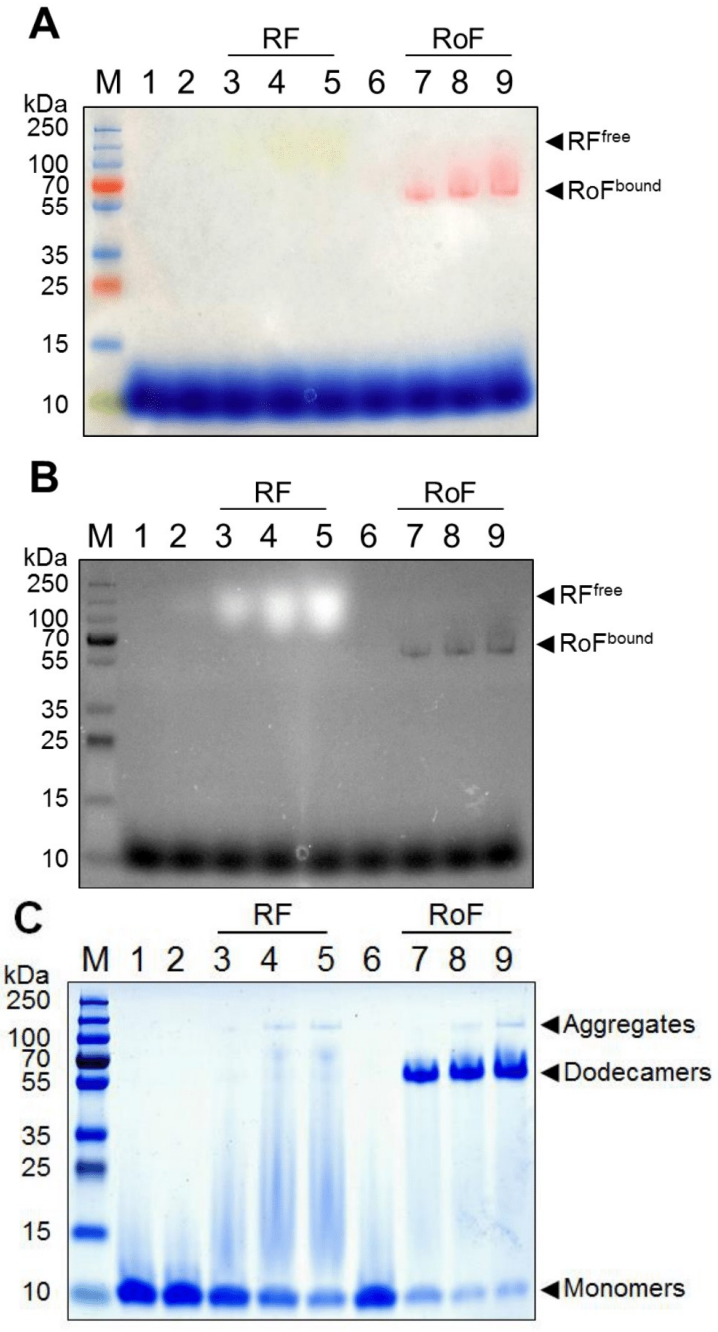
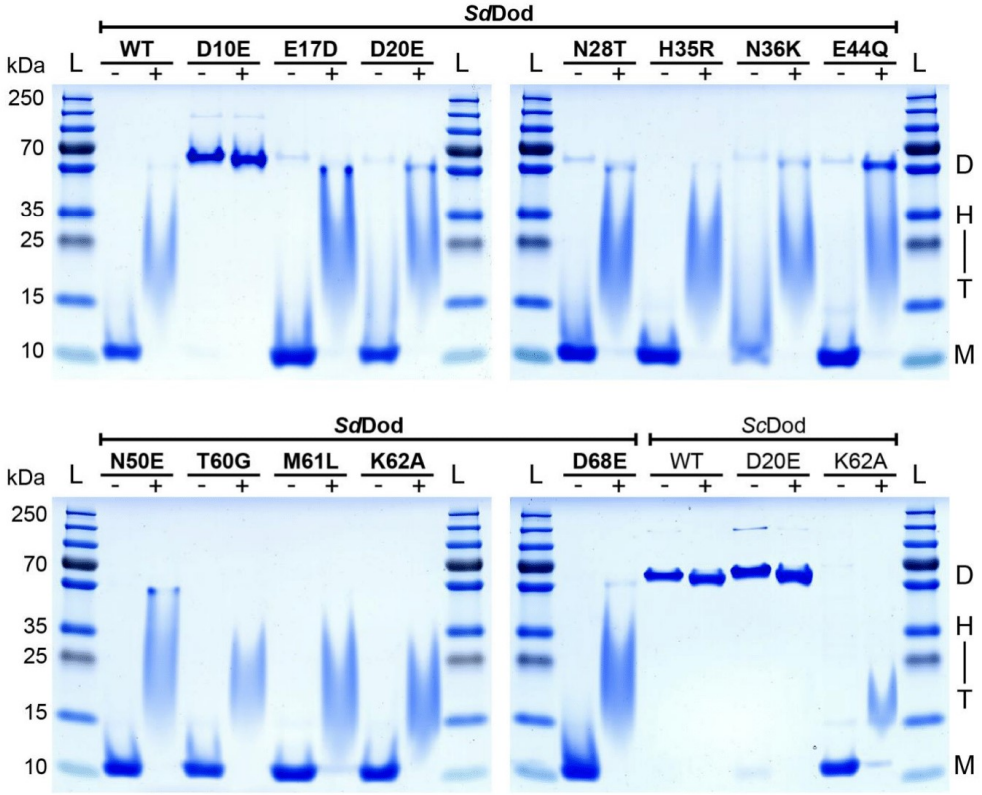

A
RF
RoF
M 1 2 3 4 5 6 7 8 9
kDa
250
100
70
55
35
25
15
10
RF^free
RoF^bound
B
RF
RoF
M 1 2 3 4 5 6 7 8 9
kDa
250
100
70
55
35
25
15
10
RF^free
RoF^bound
C
RF
RoF
M 1 2 3 4 5 6 7 8 9
kDa
250
100
70
55
35
25
15
10
Aggregates
Dodecamers
Monomers

Fig. S8



References of the supplement

1. **Bourdeaux F, Hammer CA, Vogt S, Schweighofer F, Noll G et al.** Flavin Storage and Sequestration by *Mycobacterium tuberculosis* Dodecin. *ACS Infect Dis* 2018;4(7):1082-1092.
2. **Meissner B, Schleicher E, Weber S, Essen LO**. The dodecin from *Thermus thermophilus*, a bifunctional cofactor storage protein. *J Biol Chem* 2007;282(45):33142-33154.
3. **Zirak P, Penzkofer A, Mathes T, Hegemann P**. Absorption and emission spectroscopic characterization of BLUF protein Slr1694 from *Synechocystis* sp. PCC6803 with roseoflavin cofactor. *J Photochem Photobiol B* 2009;97(2):61-70.

ChemPubSoc Europe

DOI: 10.1002/slct.201802364

ChemistrySELECT Full Papers

Materials Science inc. Nanomaterials & Polymers

Diffusion-Ordered NMR Spectroscopy of Guest Molecules in DNA Hydrogels and Related Matrices

Tanja Nöll,[a] Sabine Wenderhold-Reeb,[a] Florian Bourdeaux,[b] Thomas Paululat,*[c] and Gilbert Nöll*[a]

Diffusion-ordered NMR spectroscopy was used to monitor the diffusion of guest molecules in DNA hydrogels and related DNA matrices. As guest molecule the highly symmetric hollow-spherical flavoprotein dodecin was studied. Thermoresponsive hydrogels were formed by self-assembly *via* hybridization of linear double-stranded DNA building blocks of 30 base pairs equipped with sticky ends, i.e. additional overhangs of 15 bases on both ends, which were complementary to each other. This resulted in hydrogels, in which dodecin was freely diffusing. When in contrast self-assembly was performed with rather short building blocks (9 base pairs + sticky ends of 6 bases), the diffusion of the guest molecule was hampered, but as hybridization was reversible within the timescale of the experiment, the resulting DNA matrix did not behave as a true gel. Apparently true DNA hydrogels with small mesh size can be obtained only, when self-assembly of short DNA building blocks *via* hybridization is combined with enzymatic ligation leading to a covalently linked network. In that case, the minimum achievable mesh size should be limited by the diameter of the ligase. While DNA hydrogels are an ideal matrix to host rather large molecules or even living cells, it is a challenge to design pristine DNA hydrogels with mesh sizes sufficiently small to capture guest molecules such as drugs or enzymes in the size of only a few nm.

Introduction

Since the first report on the formation of pristine DNA hydrogels in 2006 by Dan Luo et al.[1] there is a growing interest in this class of materials for biomedical applications such as cell transplant therapy, biomineralization, tissue engineering, gene therapy, drug release, and biosensing.[1–11] As DNA hydrogels are generated by self-assembly of exactly defined building blocks waiving the need for additional chemical reagents, the resulting materials are absolutely biocompatible and biodegradable, they are not immunogenic, and their three-dimensional structures can be precisely controlled. In comparison to conventional hydrogels[12] DNA hydrogels will contain less defects and their material properties can be adjusted over a broad range. The first pristine DNA hydrogels of Luo et al. were formed by self-assembly of flexible branched DNA building blocks equipped with palindromic sticky ends.[1] Self-assembly (*via* hybridization) was followed by enzymatic ligation, which resulted in the formation of covalent bonds between the individual building blocks. When these hydrogels were evaluated for drug capture and release applications, it turned out that not only the enzyme-catalyzed gel formation but also the release of entrapped molecules were rather slow. In 2009 Dongsheng Liu and coworkers published a fast responding pH-triggered DNA hydrogel, which has been developed from Y-shaped DNA building blocks equipped with cytosine-rich interlocking i-motif domains.[13] At low pH a hydrogel was obtained, which was able to trap Au-nanoparticles (Au-NPs). When the pH was changed to pH 8, the gel quickly disassembled, and the Au-NPs were released. Two years later D. Liu and coworkers formed DNA hydrogels by hybridization of Y-scaffolds with linear double-stranded DNA (dsDNA).[14] Hydrogel formation could be reversed by heating above the melting temperature, and the temperature dependent gel-sol transition could be monitored by rheological measurements. While DNA hydrogel formation usually required branched building blocks, we have shown in 2014 that highly viscous hydrogels can be obtained from simple short linear double stranded DNA (dsDNA) building blocks equipped with sticky ends as shown in **Scheme 1**.[15] It was concluded that the high viscosity of this type of DNA hydrogels originates from the flexibility of the linear DNA building blocks, which allows also the formation of cyclic structures including multiple interlocked rings. As the hydrogels were formed by self-assembly of the building blocks *via* hybridization, they were thermoresponsive. To gain more insight in the supramolecular structures present in the DNA hydrogel network, the material was investigated by temperature dependent diffusion ordered NMR spectroscopy (DOSY NMR) measurements. DOSY NMR allows to determine diffusion

[a] *Dr. T. Nöll, S. Wenderhold-Reeb, Dr. G. Nöll*
NRW Nachwuchsforschergruppe für Nanotechnologie, Organische Chemie, Universität Siegen, Fakultät IV, Department für Chemie und Biologie, Adolf-Reichwein-Strasse 2, 57076 Siegen (Germany)
E-mail: noell@chemie.uni-siegen.de

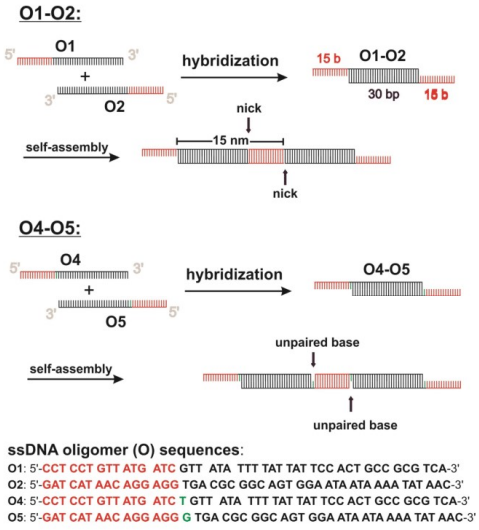
[b] *F. Bourdeaux*
Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438, Frankfurt am Main (Germany)
E-mail: Bourdeaux@chemie.uni-frankfurt.de

[c] *Dr. T. Paululat*
Organische Chemie, Universität Siegen, Fakultät IV, Department für Chemie und Biologie, Adolf-Reichwein-Strasse 2, 57076 Siegen (Germany)
E-mail: thomas.paululat@uni-siegen.de

Supporting information for this article is available on the WWW under https://doi.org/10.1002/slct.201802364

ChemPubSoc Europe

Chemistry SELECT Full Papers



Scheme 1. Strategy for the formation of DNA hydrogels by self-assembly of the linear building blocks **O1-O2** and **O4-O5**. The sequences of the oligomers are given below. For the thermoresponsive hydrogels formed from **O1-O2** and **O4-O5** gel points of 42 °C and 47 °C have been determined, respectively.[15]

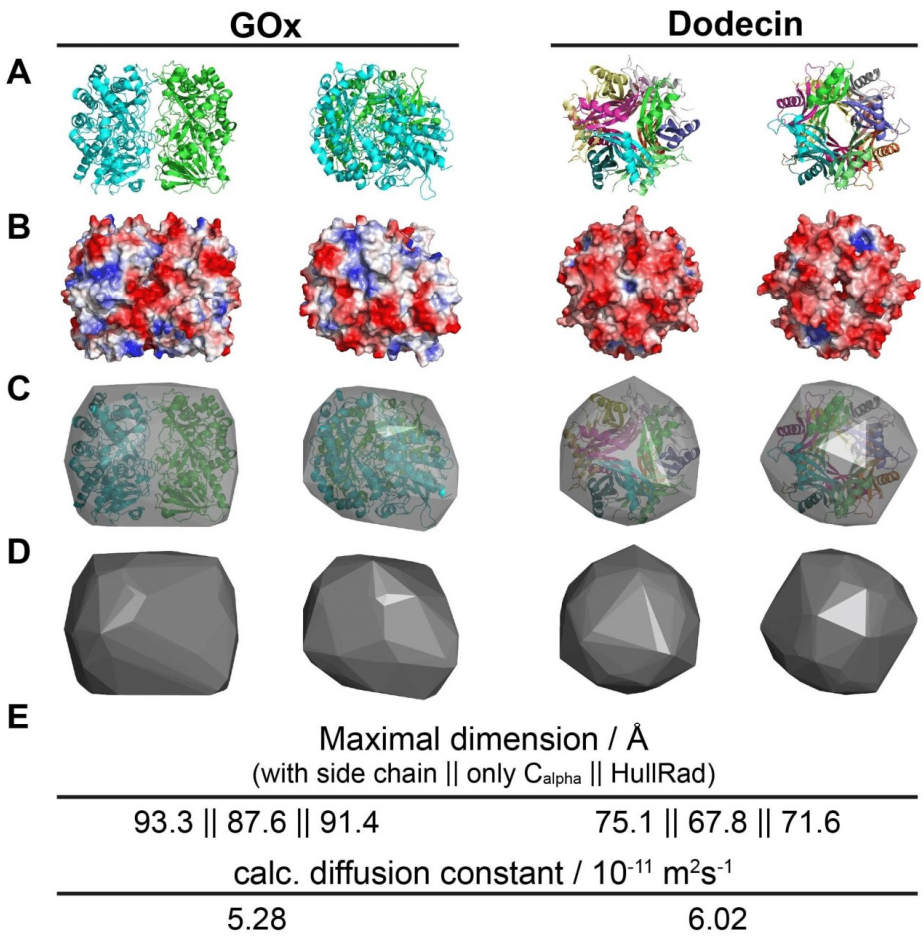
coefficients of different molecules, even in mixtures.[16] According to the DOSY NMR measurements the gel started to melt at elevated temperature leading to supramolecular structures comprising more or less linear dsDNA building blocks (with increasing temperature the average size of the supramolecular structures decreased). At about 70 °C on average monomeric dsDNA building blocks were present, and when the temperature was increased even further, also the monomeric dsDNA building blocks started to dehybridize. By temperature dependent rheological measurements of the hydrogels formed by self-assembly of the linear dsDNA building blocks **O1-O2** and **O4-O5**, gel points were observed at temperatures of 42 °C and 47 °C, respectively. By DOSY NMR measurements at temperatures below the gel point it was not possible to determine a diffusion coefficient for the supramolecular structures present inside the hydrogel network. Apparently either the diffusion of the gel forming substructures is completely supressed (because a permanent network has been formed), or (if from time to time dehybridization between individual fragments may occur) the diffusion of individual substructures is on average too slow to be monitored by DOSY NMR.

Even though DNA hydrogels are highly suited for several biomedical applications, currently the advantages of these materials are not being exploited because these materials are rather expensive and available only in small amounts. This is because DNA hydrogels are usually formed from relatively short synthetically derived building blocks comprising individual oligomeric DNA strands prepared by solid phase synthesis. To overcome this problem we have recently shown that DNA hydrogels can also be formed from much longer linear DNA building blocks derived from plasmid DNA,[17] which can be prepared in large scale by fermentation. By this procedure the costs for hydrogel production can be reduced at least by one or two orders of magnitude. A different strategy to reduce the overall costs for hydrogel production is to replace the polymeric backbone of the hydrogel building blocks by other materials such as polyacrylamide or polypeptide resulting in DNA-hybrid hydrogels.[3,5,8,17–20]

When DNA based hydrogels have been used for drug capture and release applications, the release of a guest or signaling molecule has been monitored most frequently visually by color changes (changes in the UV/Vis absorption spectra), or by fluorescence.[1,3,5,8,13,18] In such studies usually the diffusion of captured guest molecules or of signaling molecules being generated and/or released from the hydrogel matrix (e.g. as a result of an external stimulus) is not monitored in detail. There are also publications in which DNA hydrogels have been used to capture redox-active enzymes (oxidoreductases) such as glucose oxidase (GOx) for analytical or bioelectrocatalytical applications.[17–19,21–23] Sometimes it is assumed that the enzymes can be efficiently captured inside the hydrogel network. However, many oxidoreductases are only a few nanometer in size, and it has been stressed out by Liu and coworkers, that the networks of pristine DNA hydrogels are expected to be quite rigid, as for most pristine DNA hydrogels the mesh size is much shorter than the persistence length of dsDNA.[5] Therefore the smallest meshes are excluded, and the DNA hydrogels are expected to be permeable for small molecules and proteins.[5] Actually this is an advantage for cell culture applications, because commonly used nutrients will be able to reach the cells embedded in the DNA hydrogels, and undesired products will not be accumulated, because they are able to leave the cell culture matrix.[5]

If the mesh size cannot be reduced below a certain level, it seems on the other hand questionable how well DNA hydrogels are suited to capture relatively small guest molecules. Possibly in some cases guest molecules, which are initially located inside the hydrogel, because the hydrogel formation was carried out in the presence of the guests, are going to be released from the network instead of being permanently entrapped inside. Therefore it was our intention to find an experimental approach that allows us to estimate whether a DNA hydrogel will be indeed able to capture a specific type of guest molecule or not. To prove whether the hydrogels formed by self-assembly of the linear dsDNA building blocks **O1-O2** and **O4-O5**, can be used for the binding of guest molecules with a diameter of a few nm we measured the diffusion of the dodecameric flavoprotein dodecin inside the hydrogels by DOSY NMR. Dodecin from *Halobacterium salinarum* is a highly symmetric hollow-spherical riboflavin (RbF) binding protein with a diameter of about 7 nm.[24–26] Even though for dodecin no catalytic function nor any application as a drug is known so far, it is an ideal model system for our studies because of its high symmetry. In **Scheme 2** the structures of dodecin and GOx are shown for comparison.

ChemPubSoc Europe
Chemistry SELECT Full Papers



Scheme 2. Structure of GOx dimer (which has been frequently used as an enzyme embedded in different types of hydrogels) and dodecin dodecamer for comparison. GOx (pdb 3QVP) and dodecin (pdb 2CCB) each depicted in two orientations without any ligands. GOx: left: view along C2 axis, right: side view through dimer. Dodecin: left: view along C3 axis, right: view along C2 axis (binding pocket). (A) Cartoon illustration, colored by chain. (B) Polar surface depiction, red: negative charge, blue: positive charge. (C) Cartoon illustration within convex hull (generated by HullRad[27]). (D) convex hull. (E) Maximal dimension and calculated diffusion constant based on the pdb structures. Maximal dimensions "with side chain" and "only C_{alpha}" determined with PYMOL, maximal dimension "HullRad" calculated by HullRad. HullRad reduces the side chains to spheres of defined diameter, explaining the differences in maximal dimension. Calculated diffusion constant: determined by HullRad for a temperature of 20 °C. The dodecin in this study contains a C-terminal His-Tag (LEHHHHHH), which is presented 12 times at the dodecamer surface and should affect/lower the diffusion constant.

Results and Discussion

In this contribution we use DOSY NMR to determine the diffusion coefficients of guest molecules in DNA hydrogels and related DNA matrices depending on the experimental conditions. For DOSY NMR the hydrogels have to be loaded with the guest molecules at a rather high concentration, but this is not a drawback, as for efficient drug capture and release high loading capacities are desired. As model for a guest molecule we use the DtE (apo)protein variant of dodecin,[34] which is currently the largest spherical guest molecule available to us. Like GOx which catalyzes the oxidation of sugars, and has thus been captured in DNA hydrogels,[17–19,21–23] dodecin is also a flavin-binding protein. The dodecameric dodecin appears to be an ideal model compound for NMR diffusion studies because of its high symmetry giving stronger NMR signals than asymmetric compounds. Furthermore, it is known from previous studies

ChemPubSoc Europe

Chemistry SELECT Full Papers

that dodecin does not show non-specific binding to DNA.[24–25,28–29] In addition dodecin comprises a high stability also at elevated temperatures. Depending on the salt concentration dodecin is stable even at temperatures ≥ 90 °C as shown in the supporting information. To obtain an estimate of the dodecin stability, the ability of dodecin to bind riboflavin (RbF) was used. RbF bound to dodecin shows nearly no fluorescence, allowing one to distinguish between bound and unbound RbF *via* the fluorescence intensity.[26] Assuming only the whole dodecamer can bind RbF efficiently any defolding or disassembly would cause a higher concentration of unbound RbF, which can be measured *via* fluorescence. Dodecin and an equivalent amount of RbF were mixed and incubated at 5 °C to ensure binding. In a cyclic pattern the dodecin:RbF complex solution was heated up and then cooled down to 5 °C, after each heating and cooling cycle the fluorescence was measured. With each cycle the heating step temperature was increased. The increase of fluorescence was plotted against the heating step temperature and showed a correlation between ion strength of the buffer solution and the thermal stability. To determine the longterm stability of dodecin the duration of the heating step was increased to 4 h (less steps). Both measurements show that in the buffer solution used in the current study dodecin is stable up to 50 °C, after 4 h at 50 °C about 8.3% ± 0.2% of the dodecamer denatured and 39% ± 1% after 4 h at 60 °C. We measured the DOSY spectra of dodecin at temperatures between 5 °C and 50 °C. A ^{1}H NMR spectrum of dodecin at a concentration of 1 mM and the corresponding DOSY spectrum measured in a buffered solution of H_2O/D_2O 9/1 at 5 °C is shown in **Figure 1**. The separated signal at about

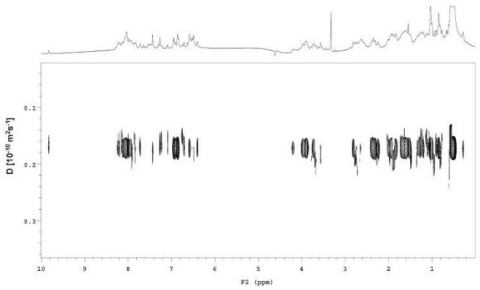


Figure 1. ^{1}H NMR and ^{1}H DOSY NMR spectrum of dodecin at a concentration of 1 mM measured at 5 °C in buffer solution.

10 ppm is probably caused by the H–N of the tryptophans at the dodecin binding sites, which are the only tryptophans in dodecin.[24] Unfortunately this signal broadens at elevated temperature, at temperatures ≥ 40 °C this signal is not visible in the DOSY spectra anymore.

When we compare the measured diffusion coefficient of dodecin at 20 °C ($2.4 \cdot 10^{-11}$ m^2s^{-1}, see **Table 1**) with that calculated with HullRad ($6.0 \cdot 10^{-11}$ m^2s^{-1}), and with that calculated by the Stokes-Einstein Equation ($6.1 \cdot 10^{-11}$ m^2s^{-1}), it becomes obvious that the diffusion coefficient calculated with HullRad is in agreement with that calculated by the Stokes-Einstein Equation, whereas the measured diffusion coefficient of dodecin implies that dodecin is diffusing slower as expected for a molecule that is significantly larger than the actual diameter of dodecin (7 nm). From the diffusion coefficient of dodecin at 20 °C ($2.4 \cdot 10^{-11}$ m^2s^{-1}) we calculated the radius *r* of dodecin in water by the Stokes-Einstein equation (supporting information S4) to be $r \approx 8.9$ nm. When the radius of dodecin was calculated from the diffusion coefficient measured at 5 °C ($1.6 \cdot 10^{-11}$ m^2s^{-1}), a radius of $r \approx 8.0$ nm was calculated. This value is about twice as large as the actual radius of dodecin, which is known from X-ray structural analysis to be $r \approx 3.5$ nm without His-Tag. As the dodecin variant used in the current study contains twelve His-Tags, its radius might be a little larger, i.e. about 4 nm. Also there might be a slight increase because of the hydration layer. Apparently the radius of dodecin determined from the NMR-data by the Stokes-Einstein Equation is about twice as large as the true value. This observation is not surprising since it has been pointed out that size estimates of proteins and nanoparticles obtained from measured diffusivities using the Stokes – Einstein relation are typically twice the true size.[30] As shown below, we measured also the diffusion coefficient of sucrose in a hydrogel formed of **O1-O2** for comparison. From the diffusion coefficient of sucrose $3.5 \cdot 10^{-10}$ m^2s^{-1}, measured at 25 °C we calculated the hydrodynamic radius according to the Stokes-Einstein equation (supporting information, S4). For the hydrodynamic radius a value of $r \approx 0.7$ nm was obtained. Also in this case the value calculated by the Stokes-Einstein equation is larger than the published value of 0.45 nm,[31] but not exactly twice as large.

When the diffusion of two different molecules in mixtures is investigated by DOSY NMR, and the ^{1}H NMR spectra of these molecules are overlapping at certain ppm values, then both molecules will contribute to the overall value for the diffusion coefficient measured in this ppm range. Therefore we could determine in our previous work only combined values for the

Table 1. Temperature dependent diffusion coefficients *D* of dodecin in solution and in the hydrogels made from **O1-O2** and from **O4-O5**.

temperature / °C	*D* / m^2s^{-1} of dodecin in solution[a]	*D* / m^2s^{-1} of dodecin in the hydrogel made of **O1-O2**[a]	*D* / m^2s^{-1} of dodecin in the hydrogel made of **O4-O5**[a]
5	0.16×10^{-10}[c]	0.16×10^{-10}[c]	0.18×10^{-10}[c]
10	0.19×10^{-10}[c]	0.19×10^{-10}[c]	0.21×10^{-10}[c]
15	0.21×10^{-10}[c]	0.22×10^{-10}[c]	0.22×10^{-10}[c]
20	0.24×10^{-10}[c]	0.25×10^{-10}[c]	0.27×10^{-10}[c]
25	0.30×10^{-10}[c]	0.27×10^{-10}[c]	0.30×10^{-10}
30	0.32×10^{-10}[c]	0.33×10^{-10}[c]	0.33×10^{-10}
35	0.35×10^{-10}	0.35×10^{-10}	0.37×10^{-10}
40	0.39×10^{-10}	[b]	0.38×10^{-10}
45	0.43×10^{-10}	[b]	[b]
50	0.47×10^{-10}	[b]	[b]

[a] in H_2O/D_2O 9/1 containing 4 mM Tris buffer, 200 mM NaCl, 1 mM $MgCl_2$, [b] no clear value due to significant peak spreading, [c] value determined for signal from tryptophan-NH at around 10 ppm

ChemPubSoc Europe
Chemistry SELECT Full Papers

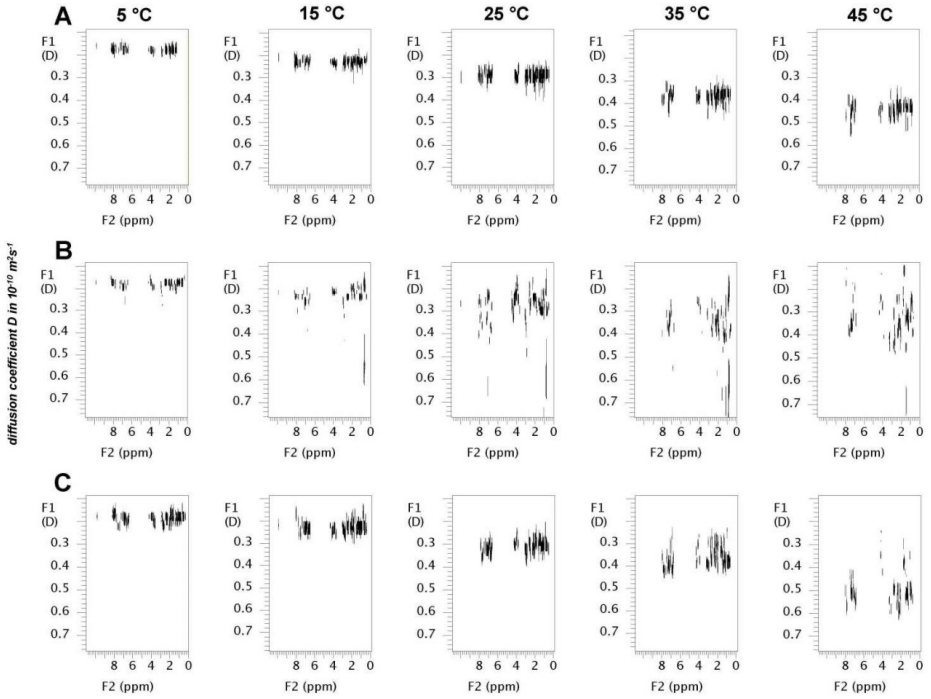


Figure 2. ^{1}H DOSY NMR spectra of dodecin at a concentration of 1 mM in solution (**A**), dodecin in a hydrogel made from **O1-O2** (**B**), and dodecin in a hydrogel made from **O4-O5** (**C**) at temperatures from 5 °C to 45 °C. The concentration of the monomeric building blocks **O1-O2** or **O4-O5**, out of which the hydrogels were formed, was 1 mM.

diffusion coefficients of the supramolecular structures that were present in the hydrogels formed from **O1-O2** at elevated temperature. When for instance at a temperature of 65 °C a diffusion coefficient of $D = 0.45 \cdot 10^{-10}$ m^{2}s^{-1} has been determined, this value was assigned to be caused on average by trimers of **O1-O2**, but it could also be caused from a 1 to 1 mixture of dimers and tetramers.

For this reason, the diffusion coefficient of any guest molecule inside the DNA hydrogel cannot be determined with high accuracy at ppm values, at which the ^{1}H NMR spectra of dodecin (the guest) and the hydrogel overlap. At room temperature the signals in the ^{1}H NMR spectrum of the DNA hydrogel formed from **O1-O2** became rather broad, and as the material forms a gel, it was not possible to determine any value for the diffusion coefficient. Nevertheless, in the ^{1}H NMR spectra of the hydrogel formed from **O1-O2** there were no signals at ppm values close to 10 ppm. Therefore we preferably used the dodecin signal close to 10 ppm (probably the tryptophan H–N signal) to determine the diffusion coefficient of dodecin at temperatures at which a diffusion signal could be detected in this ppm region. The separate ^{1}H NMR spectra of dodecin at a concentration of 1 mM, dodecin in the hydrogel formed from **O1-O2**, (the concentration of the monomeric building block **O1-O2** was 1 mM) and the hydrogel formed from **O1-O2** are shown in the supporting information in **Figures S1A-S1C** for a temperature of 25 °C, respectively.

We started our studies by measuring the diffusion coefficients of dodecin at different temperatures in pure buffer solution and in a buffered hydrogel formed from **O1-O2** and from **O4-O5** shown in **Scheme 1**.[15] In **Table 1** the temperature-dependent diffusion coefficients of dodecin at a concentration of 1 mM freely diffusing in buffer, in a hydrogel made from **O1-O2** and from **O4-O5** in H_2O/D_2O 9/1 (measured at 5, 10, .. 50 °C) are summarized, and the corresponding DOSY spectra measured at 5 °C, 15 °C, 25 °C, 35 °C, and 45 °C are shown in **Figure 2**.

As it can be seen from the spectra of dodecin in a hydrogel made of **O1-O2** at temperatures $\geq$ 15 °C and dodecin in a hydrogel made of **O4-O5** at temperatures $\geq$ 30 °C a signal distribution in the diffusion dimension of the DOSY spectra occurs, and this signal distribution increases at higher temperature, which makes the determination of an exact value for the diffusion coefficient increasingly difficult. Independent from using pulse sequences with or without convection compensa-

ChemPubSoc Europe

Chemistry SELECT Full Papers

tion and different processing protocols the signal distribution remains visible. It is likely that the signal distribution arises from signal overlay of different molecular structures in the mixture (overlay of the intact DNA hydrogel and the dodecin signals). In addition some relatively short oligomeric units of the DNA network could be temporary liberated, possibly due to collision with dodecin molecules (many signals are distributed towards higher diffusion coefficients, which should be caused by the presence of molecules smaller than dodecin). Therefore we used the dodecin NMR signal around 10 ppm (if available) to determine the diffusion coefficient of dodecin. In any case the DOSY spectra show unequivocally that the diffusion of dodecin inside the hydrogel matrices is not hampered. Previously it has been pointed out that for most DNA hydrogels comprising substructures of dsDNA the mesh size is smaller than the persistence length of dsDNA, and therefore the network is expected to be rigid and permeable even for small proteins (or nutrients, if the hydrogels are used as matrix for cell culture).[5] For the hydrogel formed from **O1-O2** previous rheological measurements revealed a mesh size of 15 nm at 20 °C, which is equal to the length of one repeating unit in the DNA network.[15] Since in DNA hydrogels derived from linear building blocks there are no covalent linkages (knots) between individual dsDNA strands it is even more accurate to use the term entanglement length instead of mesh size. The mesh size (or entanglement length) has to be distinguished from the pore size. The latter describes the size of cavities filled with liquid. Many conventional hydrogels comprise pores much larger than their mesh size, because of inhomogeneity and defects in their structure. In contrast for DNA hydrogels, formed by self-assembly *via* reversible hybridization, the number of defects or inhomogeneities is expected to be low. If large pores are present, the maximum size of guest molecules, which are able to diffuse inside a hydrogel, should be limited by the pore size. Thus, the observation that the diffusion of dodecin in the hydrogels made from **O1-O2** and from **O4-O5** is not restrained could be explained either by the fact that dodecin with a diameter of around 7 nm is much smaller than the entanglement length of the hydrogels (as expected due to the entanglement length of 15 nm determined by rheology)[15] or by the presence of defects/pores. Therefore it was our intention to shorten the length of the DNA building blocks leading to a DNA hydrogel with an entanglement length shorter than the diameter of dodecin. For this purpose a short GC-rich building block termed **SO1-SO2**, comprising 9 base pairs (bp) with a length of about 3 nm, as well as a single base (which stays unpaired after hybridization in order to introduce additional flexibility) and a sticky end of six bases at either side was designed. The corresponding sequences and the strategy for self-assembly are shown in **Scheme 3**.

Each of the two ssDNA strands of the **SO1-SO2** building block comprises only 16 bases whereas those of **O1-O2** or **O4-O5** comprised 45 or 46 bases, respectively. To keep the total amount of DNA constant, self-assembly of **SO1-SO2** was carried out at a monomeric building block concentration of 3 mM, whereas for the generation of hydrogels from **O1-O2** and **O4-O5** a concentration of 1 mM had been used. For hydrogel

Scheme 3. Strategy for the formation of DNA hydrogels with short entanglement length by self-assembly of the linear building blocks **SO1-SO2**. The sequences of the corresponding short oligomers are given below.

formation by hybridization (without subsequent enzymatic ligation) it is very important that both oligomers forming the dsDNA building block are mixed at equimolar ratio, because already a small excess of one ssDNA strand may act as end group (chain stopper). In order to prepare equal samples of **SO1-SO2** to be studied with and without dodecin, we have produced a single batch of 400 µL DNA matrix at a concentration of 3 mM in D_2O, which was then divided into two samples of equal volume (200 µL each). By this procedure it is ensured that both samples have the same DNA composition. One sample was lyophilized inside the NMR tube before a solution of dodecin in buffer was added. Heating to 60 °C and subsequent cooling to room temperature resulted in the formation of a dodecin loaded highly viscous matrix. For comparison the second sample was lyophilized and then dissolved and heated in the same type of buffer as the protein before it was analyzed by DOSY NMR. In **Table 2** the temperature-dependent diffusion coefficients (measured at 5, 10, .. 50 °C) for dodecin in buffer solution (as shown before) and in the matrix formed from **SO1-SO2**, as well as those of the supramolecular structures formed from **SO1-SO2** in buffer are

Table 2. Temperature dependent diffusion coefficients *D* of dodecin in buffer solution and in the matrix formed from **SO1-SO2** as well as those of the supramolecular structures present in the matrix formed from **SO1-SO2**.

temp / °C	D / m^2s^{-1} of dodecin[a]	D / m^2s^{-1} of dodecin in the matrix formed from **SO1-SO2**[a]	D / m^2s^{-1} of the supramol. str. of **SO1-SO2**[a]
5	0.16×10^{-10}[b]	0.10×10^{-10}[b]	0.10×10^{-10}
10	0.19×10^{-10}[b]	0.11×10^{-10}[b]	0.13×10^{-10}
15	0.21×10^{-10}[b]	0.12×10^{-10}[b]	0.15×10^{-10}
20	0.24×10^{-10}[b]	0.14×10^{-10}[b]	0.17×10^{-10}
25	0.30×10^{-10}[b]	0.15×10^{-10}	0.20×10^{-10}
30	0.32×10^{-10}[b]	0.15×10^{-10}	0.21×10^{-10}
35	0.35×10^{-10}	0.15×10^{-10}	0.28×10^{-10}
40	0.39×10^{-10}	0.20×10^{-10}	0.35×10^{-10}
45	0.43×10^{-10}	0.25×10^{-10}	0.50×10^{-10}
50	0.48×10^{-10}	0.41×10^{-10}	0.73×10^{-10}

[a] in H_2O/D_2O 9/1 containing 4 mM Tris buffer, 200 mM NaCl, 1 mM $MgCl_2$, [b] value determined for signal from tyrosine-NH at around 10 ppm

ChemPubSoc Europe
Chemistry SELECT Full Papers

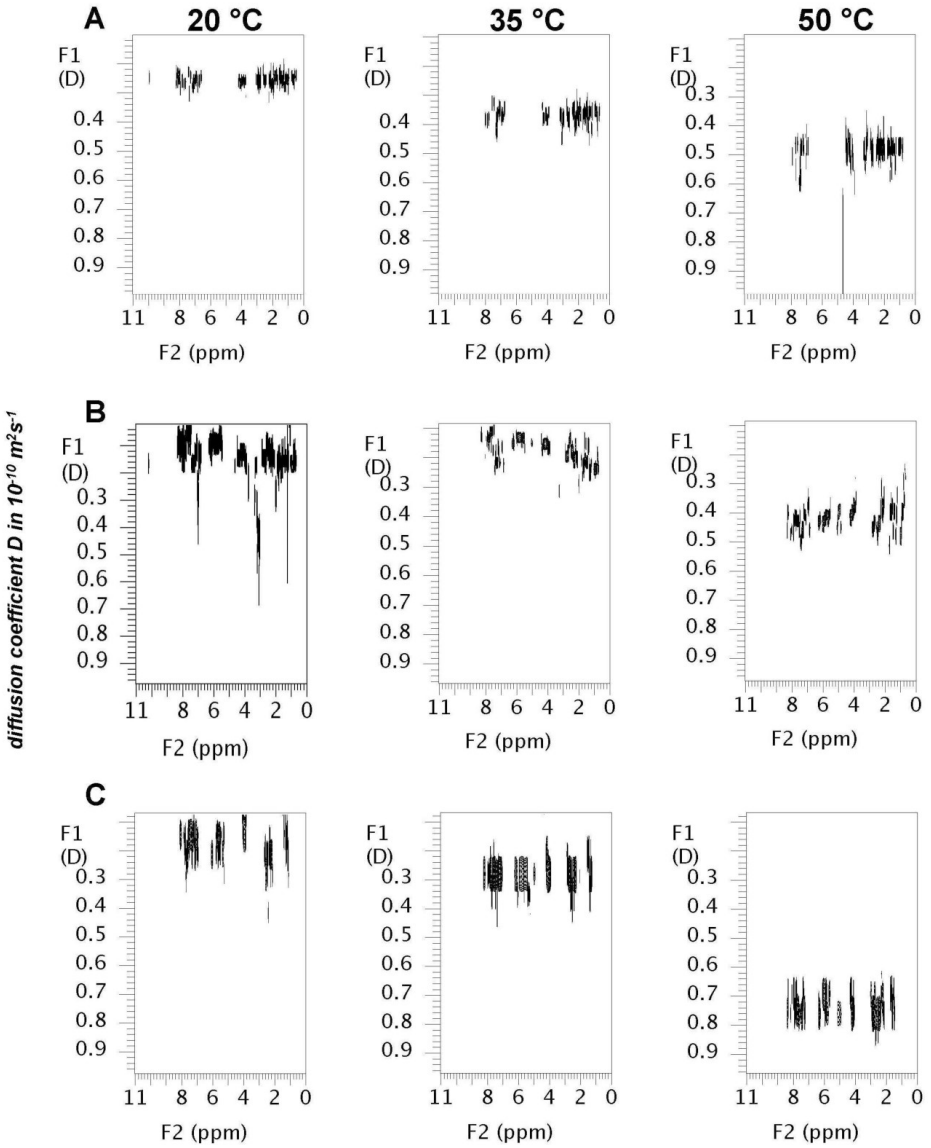


Figure 3. Representative ^{1}H DOSY NMR spectra of dodecin in buffer solution (**A**), dodecin in the matrix formed from **SO1-SO2** (**B**), and the matrix formed from **SO1-SO2** in buffer solution (**C**) at temperatures from 20 °C, 35 °C and 50 °C.

summarized, and representatively the corresponding DOSY spectra measured at 20 °C, 35 °C, and 50 °C are shown in **Figure 3**. For comparison the diffusion coefficients of dodecin in pure buffer solution as well as in the matrices formed from **O1-O2**, from **O4-O5**, and from **SO1-SO2** are plotted *vs.* the temperature in **Figure 4**. As it can be seen from **Figure 4**, there is no significant difference in the temperature-dependent diffusion coefficients of dodecin in buffer solution and the

ChemPubSoc Europe

Chemistry SELECT Full Papers

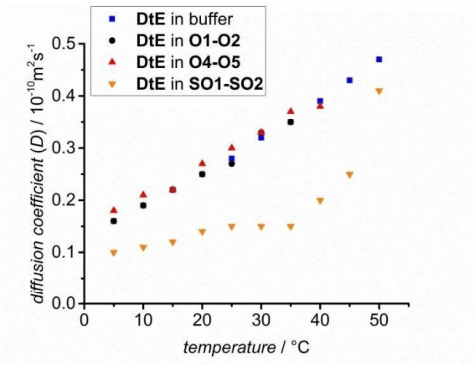


Figure 4. Plot of the diffusion coefficients of apododecin DtE measured in pure buffer solution, as well as in the matrices formed from **O1-O2**, from **O4-O5**, and from **SO1-SO2** *vs.* temperature.

hydrogels made of **O1-O2** and **O4-O5**. Apparently the diffusion of dodecin in these hydrogels is not hampered at all. In contrast inside the matrix formed from **SO1-SO2** the diffusion coefficient of dodecin is significantly decreased in the entire temperature range (5 °C – 50 °C). In the following this observation will be discussed more in detail with respect to the experimental data.

While at temperatures $\leq$ 35 °C the diffusion coefficient of dodecin in the DNA matrix is $\leq 0.15 \times 10^{-10}$ m^2s^{-1}, at 40 °C it is strongly increasing, and at 50 °C the diffusion coefficient of dodecin in the matrix formed from **SO1-SO2** is close to the value of dodecin in buffer solution.

According to the values summarized in **Table 2** and plotted in **Figure 4**, the diffusion of dodecin inside the matrix formed from **SO1-SO2** is significantly decreased at low temperatures and temperatures up to 35 °C (here values of $D \leq 0.15 \cdot 10^{-10}$ m^2s^{-1} were obtained), whereas at higher temperatures the diffusion coefficient of dodecin inside the matrix starts to approach the values for free dodecin in buffer solution. At a temperature of 50 °C there is not much difference between both values any more. Apparently the diffusion of dodecin is indeed restricted in the entire range of temperature under investigation, but it is not completely suppressed. As shown in **Table 2**, again the 10 ppM signal of dodecin could be used to determine the diffusion coefficient of dodecin at low temperatures. As this signal is by about 1.5 ppm separated from the next signal at lower ppm values belonging to **SO1-SO2**, we assume that at least the diffusion coefficients that could be determined for dodecin using the 10 ppm signal are not falsified by overlapping signals from **SO1-SO2**. The separate ^{1}H NMR spectra of dodecin, dodecin in the matrix formed from **SO1-SO2**, and the matrix formed from **SO1-SO2** are shown in the supporting information in **Figures S2A-S2C** for a temperature of 15 °C, respectively. Another important observation is that it was always possible to determine a diffusion coefficient for the supramolecular structures formed from **SO1-SO2**, at room temperature, and even at a temperature of 5 °C. Also the corresponding signals in the ^{1}H NMR spectra are not significantly broadened. Even though the corresponding value of $D = 0.1 \cdot 10^{-10}$ m^2s^{-1} for 5 °C is rather low, the diffusion of the supramolecular structures formed by self-assembly of **SO1-SO2** is not completely suppressed. This implies that the DNA matrix formed by self-assembly of **SO1-SO2** may comprise gel-like properties (depending on the time scale of an experiment), but it is not a true gel. This observation, which could be confirmed by frequency dependent rheological measurements shown in the supporting information, **Figure S4**, is not really surprising, as the overhangs of **SO1-SO2** comprise only six bases, and for dsDNA of six base pairs it is expected that even below room temperature the hybridization is reversible. Hence, different scenarios are possible. Larger supramolecular structures formed by self-assembly of **SO1-SO2** may partially disassemble from time to time due to dehybridization of individual regions of dsDNA with 6 bp formed by hybridization of the overhangs. In principle dehybridization of the centered dsDNA region with 9 bp may occur as well. It thus seems like dodecin can be captured in the DNA matrix for a short period (as long as the matrix around the protein is stable), but always when a dehybridization event occurs in the vicinity of the protein, dodecin is able to move/diffuse to some extent. In addition it might be possible that an oligomeric DNA cocoon is formed around some dodecin molecules, and assemblies of dodecin and oligomeric DNA are diffusing together as larger subunits. The latter interpretation is supported by the observation that in **Figure 3b** (showing dodecin in the matrix formed from **SO1-SO2**) also the diffusion of the supramolecular DNA fragments formed of **SO1-SO2** seems to be retarded due to the presence of dodecin (see **Figure 3c** for comparison).

In our experiments the concentration of dodecin is 1 mM, which is a compromise, i.e. it is rather high for a protein or a large guest molecule in a matrix, but on the other hand, it is rather small for DOSY measurements with decent signal intensity. At this point it makes sense to consider the ratio between dodecin and the dsDNA building blocks with additional overhangs, as well as the volume of both components in the sample. The length of one dsDNA repeating unit (including a single unpaired base) formed by self-assembly of **SO1-SO2** is about 6 nm, and the diameter of dodecin (including His-Tags and hydration layer) is about 8 nm. Let us discuss a scenario in which a three dimensional lattice (with infinite extension in space) is formed of **SO1-SO2** with an edge length of one **SO1-SO2** repeating unit. This grid contains cubic cells or cavities with an edge length of one **SO1-SO2** repeating unit (i.e. 6 nm), and these cells share their edges and planes with neighbored cells (densest packing). Let us now assume that a single cell is just large enough to capture a single dodecin molecule (which is actually not exactly the case, because dodecin with a diameter of 8 nm is a little too large). In this case one dodecin is surrounded by twelve **SO1-SO2** units, and one **SO1-SO2** unit is surrounded by four dodecin molecules. Hence, to form such a lattice a ratio of three to one between the **SO1-SO2** building blocks and dodecin is required, just as in our experiment. The ratio of 3:1 can also be derived by a different approach using

ChemPubSoc Europe

Chemistry SELECT Full Papers

1 m^3 of water surrounded by several other m^3 of water as model for a three-dimensional lattice. It contains 1 million mL with a volume of 1 cm^3, and to form these 1 million cubes, three million edges of 1 cm are necessary (assuming infinite extension in space). Next let us consider the total amount of DNA. As the concentration was 3 mM, and the volume of the sample in the NMR tube was 200 μL, the total number of **SO1-SO2** building blocks is $3 \cdot 1000^{-1} \cdot 5000^{-1} \cdot$ Avogadro's number (N_A), which corresponds to $3.6 \cdot 10^{17}$ molecules. Analogue to the previous model approach considering 1 m^3 of water, a volume of $(600\ nm)^3$ can be formed of 3 million **SO1-SO2** building blocks, and this volume, which is $2.16 \cdot 10^{-19}$ m^3, contains then 1 million cubic cells with an edge length of 6 nm. Accordingly the total number of **SO1-SO2** building blocks is sufficient to form such a cubic grid (edge length = 6 nm) of a volume of ($3.6 \cdot 10^{17}$ molecules / 3 million molecules) times $2.16 \cdot 10^{-19}$ m^3, which is 25 μL. As we are working with a total sample volume of 200 μL, this implies that in future experiments the concentration of dsDNA can be further increased to obtain a denser grid. Of course we have to keep in mind that by our approach an ideal cubic grid will not be formed. Furthermore, one or more guest molecules can also be captured within larger cavities. In addition we have to consider that dsDNA has a diameter of 2 nm, and thus it is not possible to form a grid with edge units of infinite low thickness. Also for these reasons the volume of one cube of $(6\ nm)^3$ would be far too small to capture a single dodecin molecule with a diameter of 8 nm. Let us now calculate the total volume the dsDNA occupies within the 200 μL sample volume, which is $3.6 \cdot 10^{17}$ times 6 nm times $\pi \cdot (1\ nm)^2$, leading to a volume of 6.8 μL (neglecting that in each dsDNA repeating unit formed of **SO1-SO2** one base stays unpaired). Also the total volume of dodecin can be estimated by $1000^{-1} \cdot 5000^{-1} \cdot N_A$ times $4/3 \cdot \pi \cdot (4\ nm)^3$ resulting in a volume of 32.3 μL. These calculations show that in our experiments dodecin molecules occupy more than 16% of the total sample volume, whereas the dsDNA mesh occupies less than 4%. Therefore, by increasing the concentration of dsDNA, it should be possible to further decrease the mesh size (or entanglement length). Of course we have to keep in mind that our calculation was performed based on the assumption that an ideal cubic lattice is formed, which is not the case by the self-assembling strategy examined in the current work.

In any case, in the DNA matrix formed from **SO1-SO2** the diffusion of dodecin in the DNA matrix is hampered, but not completely suppressed. As the frequency of hybridization/dehybridization will increase with temperature, also the diffusion coefficient for dodecin inside the matrix will increase. The observation that the DNA matrix formed from **SO1-SO2** is indeed able to suppress the diffusion of dodecin to some extent leads to the conclusion that by self-assembly of **SO1-SO2** a highly regular network, which is (at least almost) free of pores or larger defects, is formed. Apparently it is possible to tailor the physical properties of DNA hydrogels and related materials by a proper choice of the building blocks. By the current approach it is not possible to generate a DNA matrix that is able to capture small guest molecules such as dodecin permanently, because self-assembly is carried out only *via* hybridization of rather short overhangs. Nevertheless our results show that by a proper choice of the building blocks the physical properties of DNA hydrogels (or more in general of supramolecular DNA matrices) can be tailored. To form a real DNA hydrogel with small mesh size, the building blocks need to be linked covalently by enzymatic ligation. In principle this should be possible, if the building blocks do not contain bases that stay unpaired after self-assembly *via* hybridization, and they need to be equipped with a phosphate at their 5′-ends, as this phosphate group is required for enzymatic ligation. In this case the minimum mesh size that can be obtained is expected to be limited by the size of the ligase, because if the ligase itself will be trapped inside the hydrogel matrix, the ligation reaction cannot be completed. For instance, the T4 ligase, which we have used in our previous work for hydrogel formation from biotechnologically derived linear dsDNA building blocks,[17] shows an elliptic shape. The maximum diameter is about 9 nm, the shorter diameter is about 6.6 nm.

Furthermore it was our intention to estimate how reliable the determination of diffusion coefficients by DOSY NMR can be, especially if dodecin or any other guest molecule is located in a DNA matrix that may contribute to the overall NMR signal. Therefore we carried out further experiments studying the diffusion of sucrose with a ^{13}C enriched fructose subunit in a DNA hydrogel formed from **O1-O2** in pure D_2O. Sucrose is so small that it is not expected to be captured inside a DNA hydrogel, but our intention was to carry out DOSY NMR measurements with ^1H and ^{13}C detection on the same guest molecule to prove whether the diffusion coefficient measured for the sucrose by ^1H DOSY NMR (which may be falsified by interference with the ^1H NMR signals of DNA hydrogel) will differ from that determined by DOSY NMR with carbon detection (which should not be significantly falsified by interference with the ^{13}C NMR signals of DNA hydrogel, as the hydrogel was not ^{13}C enriched). The used DOSY INEPT pulse sequence allows to overcome the sensitivity problem in comparison to ^{13}C DOSY *via* detection of the carbon nucleus while the diffusion is done on proton nucleus giving the information of interest.

In **Figure 5** the ^1H NMR, ^{13}C NMR, the ^1H DOSY NMR, and DOSY INEPT NMR spectra of 10 mM sucrose in a hydrogel formed from **O1-O2** (monomeric building block concentration: 1 mM) measured at a temperature of 25 °C are shown. As it can be seen from **Figure 5**, by both techniques the signals for diffusion coefficients with the core area at $3.5 \cdot 10^{-10}$ m^2s^{-1} were found (the signals are distributed over a range between 3.2 and $3.9 \cdot 10^{-10}$ m^2s^{-1}). Apparently there is not much influence of the DNA hydrogel on the absolute values of the diffusion coefficients of the sucrose guest molecules, even though they were measured at ppm values at which the DNA hydrogel has strong ^1H NMR signals. Therefore we come to the conclusion that also the diffusion coefficients determined for dodecin by ^1H DOSY NMR should not be much influenced by the ^1H NMR signals of the DNA matrices, especially when they are determined from the dodecin signal around 10 ppm.

ChemPubSoc Europe
Chemistry SELECT Full Papers

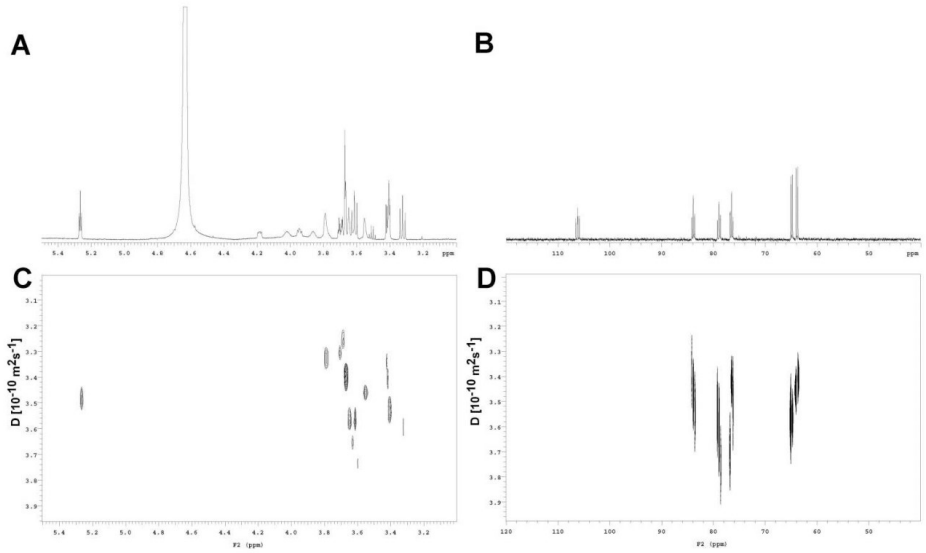


Figure 5. Comparison of different NMR spectra of [$^{13}C_6$]-sucrose (10 mM) in a hydrogel formed from **O1-O2** in D_2O at 25 °C: ^{1}H NMR (**A**), ^{13}C NMR (**B**), ^{1}H DOSY NMR (**C**), DOSY INEPT NMR (**D**).

Conclusions

We have shown by DOSY NMR spectroscopy that it is possible to monitor the diffusion of guest molecules in DNA hydrogels and related materials. When these DNA matrices are formed by self-assembly *via* hybridization of dsDNA building blocks equipped with additional complementary overhangs of ssDNA (i.e. sticky ends), a highly regular DNA network can be obtained. Larger structural inhomogeneities such as pores are either not present or have at least no noticeable impact on the DNA matrix, which is a strong advantage of DNA hydrogels over conventional hydrogels. To obtain a true gel, the sticky ends have to be sufficiently long to minimize the probability for dehybridization at the experimental conditions applied. This was the case for the materials formed from **O1-O2** and **O4-O5**, but not from **SO1-SO2** equipped with sticky ends of only six bases. In the latter case, even at a temperature of 5 °C hybridization was reversible in the timescale of the instrument, and therefore the diffusion of the guest molecule could only be slowed down but not completely suppressed. By a proper choice of the building blocks the physical properties of DNA hydrogels including the mesh size can be adjusted over a broad range. To form a hydrogel with small mesh size small building blocks with short sticky ends have to be employed. In that case hybridization has to be combined with subsequent enzymatic ligation to obtain a permanent DNA network behaving as a true gel. As for complete enzymatic ligation the ligase has to be able to diffuse within the DNA network, we anticipate that the minimum mesh size that can be obtained for this type of hydrogels is limited by the diameter of the ligase, which is $>$ 6 nm. Hence, while DNA hydrogels are an ideal matrix to host larger molecules such as living cells, it is a challenge to design pristine DNA hydrogels with mesh sizes sufficiently small to capture guest molecules such as drugs or small enzymes.

Besides by self-assembly of proper building blocks DNA hydrogels have been formed also by rolling circle amplification (RCA), but this approach results in rather soft gels with values for the storage modulus $G' \approx 10$ Pa in the frequency range from 0.1 Hz - 100 Hz.[32–33] Furthermore, it has been reported that these gels comprise many differently shaped pores with sizes from 42 nm to 161 nm.[33] Hence these gels generated by RCA form less regular structures, and are not suited to capture small guest molecules in the size of only a few nm (as typical for many enzymes).

When ^{1}H DOSY NMR spectroscopy is applied to monitor the diffusion of guest molecules in hydrogels, and there is some overlap between the NMR signals of the hydrogel and those of the guest molecule, the values for the diffusion coefficients may be falsified. As we have shown, this effect can be minimized when ^{13}C enriched guest molecules are studied by DOSY INEPT NMR. An alternative strategy might be the performance of ^{2}H or ^{19}F DOSY measurements using deuterated or fluorinated guest molecules.

ChemPubSoc Europe

Chemistry SELECT Full Papers

Supporting Information Summary

In the supporting information file the ^{1}HNMR spectra of dodecin, dodecin in **O1-O2**, and the **O1-O2** hydrogel at 25 °C, as well as the ^{1}HNMR spectra of dodecin, dodecin in **SO1-SO2**, and the **SO1-SO2** matrix at 15 °C, and the signal decays (DOSY spectra) of the aromatic region from dodecin, dodecin in the **O1-O2** sample, and dodecin in the **O4-O5** sample at 25 °C are shown. In addition the calculation of the radius of dodecin and sucrose from the measured diffusion coefficients, frequency dependent rheological measurements of the dodecin containing matrix formed from **SO1-SO2** at 20 °C, stability measurements of dodecin at elevated temperatures, and the Experimental Section are presented.

Acknowledgements

This work has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement n° 240544, from the state North Rhine-Westphalia, and the University of Siegen. G. N. thanks Dr. Martin Grininger for continuous support.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: diffusion-ordered NMR (DOSY NMR) · DNA hydrogel · guest-host systems · nanostructures · rheology

[1] S. H. Um, J. B. Lee, N. Park, S. Y. Kwon, C. C. Umbach, D. Luo, *Nat. Mater.* **2006**, *5*, 797–801.
[2] N. Park, S. H. Um, H. Funabashi, J. Xu, D. Luo, *Nat. Mater.* **2009**, *8*, 432–437.
[3] D. Wang, Y. Hu, P. Liu, D. Luo, *Acc. Chem. Res.* **2017**, *50*, 733–739.
[4] Y. Huang, W. Xu, G. Liu, L. Tian, *Chem. Commun.* **2017**, *53*, 3038–3041.
[5] Y. Shao, H. Jia, T. Cao, D. Liu, *Acc. Chem. Res.* **2017**, *50*, 659–668.
[6] A. J. Simon, L. T. Walls-Smith, M. J. Freddi, F. Y. Fong, V. Gubala, K. W. Plaxco, *ACS Nano* **2017**, *11*, 461–468.
[7] J. Wang, J. Chao, H. Liu, S. Su, L. Wang, W. Huang, I. Willner, C. Fan, *Angew. Chem., Int. Ed.* **2017**, *56*, 2171–2175; *Angew. Chem.* **2017**, *129*, 2203–2207.
[8] J. Li, L. Mo, C.-H. Lu, T. Fu, H.-H. Yang, W. Tan, *Chem. Soc. Rev.* **2016**, *45*, 1410–1431.
[9] Y. Dong, Z. Yang, D. Liu, *Acc. Chem. Res.* **2014**, *47*, 1853–1860.
[10] D. Yang, M. R. Hartman, T. L. Derrien, S. Hamada, D. An, K. G. Yancey, R. Cheng, M. Ma, D. Luo, *Acc. Chem. Res.* **2014**, *47*, 1902–1911.
[11] X. Du, J. Zhou, B. Xu, *Chem. - Asian J.* **2014**, *9*, 1446–1472.
[12] J. Kopecek, *J. Polym. Sci., Part A Polym. Chem.* **2009**, *47*, 5929–5946.
[13] E. Cheng, Y. Xing, P. Chen, Y. Yang, Y. Sun, D. Zhou, L. Xu, Q. Fan, D. Liu, *Angew. Chem., Int. Ed.* **2009**, *48*, 7660–7663; *Angew. Chem.* **2009**, *121*, 7796–7799.
[14] Y. Xing, E. Cheng, Y. Yang, P. Chen, T. Zhang, Y. Sun, Z. Yang, D. Liu, *Adv. Mater.* **2011**, *23*, 1117–1121.
[15] T. Nöll, H. Schönherr, D. Wesner, M. Schopferer, T. Paululat, G. Nöll, *Angew. Chem., Int. Ed.* **2014**, *53*, 8328–8332; *Angew. Chem.* **2014**, *126*, 8468–8472.
[16] H. Barjat, G. A. Morris, S. Smart, A. G. Swanson, S. C. R. Williams, *J. Magn. Reson., Ser. B* **1995**, *108*, 170–172.
[17] T. Nöll, S. Wenderhold-Reeb, H. Schönherr, G. Nöll, *Angew. Chem., Int. Ed.* **2017**, *56*, 12004–12008; *Angew. Chem.* **2017**, *129*, 12167–12171.
[18] J. S. Kahn, Y. Hu, I. Willner, *Acc. Chem. Res.* **2017**, *50*, 680–690.
[19] S. Lilienthal, Z. Shpilt, F. Wang, R. Orbach, I. Willner, *ACS Appl. Mater. Interfaces* **2015**, *7*, 8923–8931.
[20] C. Li, A. Faulkner-Jones, A. R. Dun, J. Jin, P. Chen, Y. Xing, Z. Yang, Z. Li, W. Shu, D. Liu, R. R. Duncan, *Angew. Chem., Int. Ed.* **2015**, *54*, 3957–3961; *Angew. Chem.* **2015**, *127*, 4029–4033.
[21] B. Xiang, K. He, R. Zhu, Z. Liu, S. Zeng, Y. Huang, Z. Nie, S. Yao, *ACS Appl. Mater. Interfaces* **2016**, *8*, 22801–22807.
[22] K. Van Nguyen, S. D. Minteer, *Chem. Commun.* **2015**, *51*, 13071–13073.
[23] K. V. Nguyen, Y. Holade, S. D. Minteer, *ACS Catal.* **2016**, *6*, 2603–2607.
[24] Y. Yu, B. Heidel, L. Parapugna Tamara, S. Wenderhold-Reeb, B. Song, H. Schönherr, M. Grininger, G. Nöll, *Angew. Chem., Int. Ed.* **2013**, *52*, 4950–4953; *Angew. Chem.* **2013**, *125*, 5040–5054.
[25] C. Gutierrez Sanchez, Q. Su, H. Schönherr, G. Nöll, M. Grininger, *ACS Nano* **2015**, *9*, 3491–3500.
[26] M. Grininger, K. Zeth, D. Oesterhelt, *J. Mol. Biol.* **2006**, *357*, 842–857.
[27] P. J. Fleming, K. G. Fleming, *Biophys. J.* **2018**, *114*, 856–869.
[28] S. Vogt, S. Wenderhold-Reeb, G. Nöll, *Electrochim. Acta* **2017**, *232*, 1–6.
[29] C. G. Sanchez, Q. Su, S. Wenderhold-Reeb, G. Nöll, *RSC Adv.* **2016**, *6*, 56467–56474.
[30] X. Zhang, S. Tran, A. Gray-Weale, *J. Phys. Chem. C* **2016**, *120*, 21888–21896.
[31] Y. Chenyakin, D. A. Ullmann, E. Evoy, L. Renbaum-Wolff, S. Kamal, A. K. Bertram, *Atmos. Chem. Phys.* **2017**, *17*, 2423–2435.
[32] J. B. Lee, S. Peng, D. Yang, Y. H. Roh, H. Funabashi, N. Park, E. J. Rice, L. Chen, R. Long, M. Wu, D. Luo, *Nat. Nanotechnol.* **2012**, *7*, 816–820.
[33] H. Stoll, H. Steinle, K. Stang, S. Kunnakattu, L. Scheideler, B. Neumann, J. Kurz, I. Degenkolbe, N. Perle, C. Schlensak, H. P. Wendel, M. Avci-Adali, *Macromol. Biosci.* **2017**, *17*, n/a.
[34] Electrochemical Switching of the Flavoprotein Dodecin at Gold Surfaces Modified by Flavin-DNA Hybrid Linkers, M. Grininger and G. Nöll (contributed equally), S. Trawöger, E.-K. Sinner, and D. Oesterhelt*, *Biointerphases* **2008**, *3*, 51–58.

Submitted: July 30, 2018
Accepted: September 17, 2018

Supporting Information

Diffusion-Ordered NMR Spectroscopy of Guest Molecules in DNA Hydrogels and Related Matrices

Tanja Nöll, Sabine Wenderhold-Reeb, Florian Bourdeaux, Thomas Paululat,* and Gilbert Nöll*

Table of Contents

Figure S1A: ^{1}H NMR (600 MHz, buffer solution with 10% D_2O, T = 25 °C) of Dodecin at a concentration of 1 mM

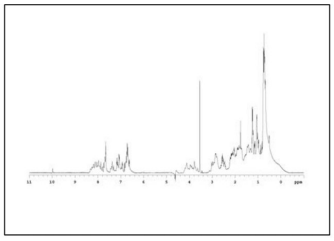


Figure S1C: ^{1}H NMR (600 MHz, D_2O, T = 25 °C) of the hydrogel formed from 1 mM **O1-O2**.

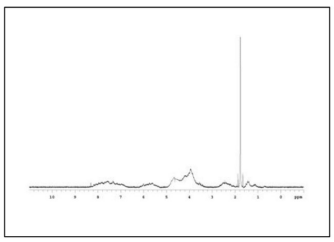


Figure S1B: ^{1}H NMR (600 MHz, buffer solution with 10% D_2O, T = 25 °C) of Dodecin (1mM) in the hydrogel formed from 1 mM **O1-O2**.

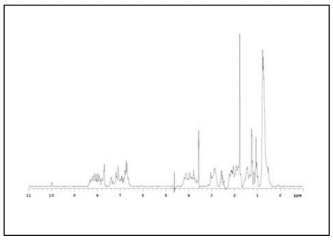


Figure S2A: ^{1}H NMR (600 MHz, buffer solution with 10% D_2O, T = 15 °C) of Dodecin at a concentration of 1 mM

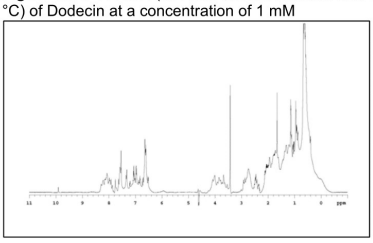


Figure S2B: ^{1}H NMR (600 MHz, buffer solution with 10% D_2O, T = 15 °C) of Dodecin (1mM) in the matrix formed from 3 mM **SO1-SO2**

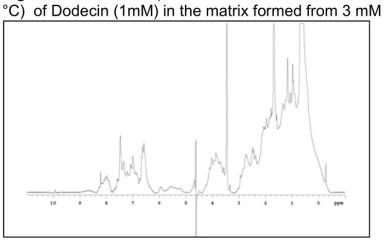
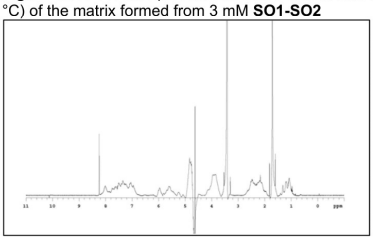


Figure S2C: ^{1}H NMR (600 MHz, buffer solution with 10% D_2O, T= 15 °C) of the matrix formed from 3 mM **SO1-SO2**



3

Figure S3A: Example of a signal decay (DOSY spect.) of aromatic region from dodecin sample (600 MHz, buffer solution with 10% D_2O, T = 25 °C)

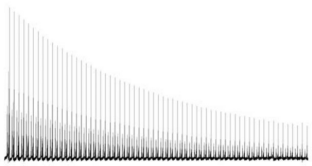


Figure S3B: Example of a signal decay (DOSY spect.) of aromatic region from dodecin in **O1-O2 sample** (600 MHz, buffer solution with 10% D_2O, T = 25 °C)

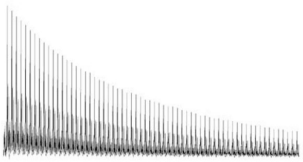


Figure S3C: Example of a signal decay (DOSY spect.) of aromatic region from **O4-O5** sample (600 MHz, buffer solution with 10% D_2O, T = 25 °C)

4

Calculation S4: Calculation of the radius of dodecin and sucrose from the measured diffusion coefficients

Stokes-Einstein equation: $r = (k\ T)/(6\ \pi\ \eta\ D)$

η = 1.52 x 10^{-3} [N s m^{-2}] = 1.52 x 10^{-3} [kg m^{-1} s^{-1}] for water at 5 °C[1],

η = 1.00 x 10^{-3} [N s m^{-2}] = 1.00 x 10^{-3} [kg m^{-1} s^{-1}] for water at 20 °C[1],

η = 8.9 x 10^{-4} [N s m^{-2}] = 8.9 x 10^{-4} [kg m^{-1} s^{-1}] for water at 25 °C[1],

$D_{\text{dodecin at 5 °C}}$ = 0.16 x 10^{-10} m^2 s^{-1}

$D_{\text{dodecin at 20 °C}}$ = 0.24 x 10^{-10} m^2 s^{-1}

→ $r_{\text{calculated}}$ **≈ 8.9 nm** calculated from the diffusion coefficient measured at 20 °C; and $r_{\text{calculated}}$ **≈ 8.0 nm** calculated from the diffusion coefficient measured at 5 °C; this value is a bit larger than twice the actual radius (3.5 nm) of dodecin.[2]

The calculated value for *r* may become a little smaller if we take into account that the DOSY NMR measurements were performed in H_2O/D_2O 9/1, containing 4 mM Tris buffer, 200 mM NaCl, 1 mM $MgCl_2$, because the viscosity of D_2O (1.25 x 10^{-3} [kg m^{-1} s^{-1}])[3] at 20 °C is little larger than that of H_2O, and also the NaCl content may slightly increase the viscosity.[4] Nevertheless the calculated radius is nearly twice as large as the actual radius.

For sucrose a value of *r* ≈ 0.7 nm was calculated from

$D_{\text{sucrose at 25 °C}}$ = 0.24 x 10^{-10} m^2 s^{-1}.

[1] value taken from:
J.C. Crittenden, R. R. Trussell, D. W. Hand, K. J. Howe, G. Tchobanoglous: "MWH's Water Treatment: Principles and Design", 3rd ed., "Appendix C: Physical Properties of water"
Wiley, 2012
Print ISBN: 9780470405390
Online ISBN: 9781118131473
DOI: 10.1002/9781118131473

[2] C. Gutierrez Sanchez, Q. Su, H. Schönherr, G. Nöll, M. Grininger, *ACS Nano* **2015**, *9*, 3491-3500.

[3] R. C. Hardy and R. L. Cottington, Viscosity of Deuterium Oxide and Water, *Journal of Research of the National Bureau of Standards* **1949**, *42*, 573.

[4] H. Ozbek, I. A. Fair, S. L. Phillips, Viscosity of aqueous sodium chloride solutions from 0 - 150 °C, Energy and Environment Division, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720, **1977**.

Frequency dependent rheological measurements of the dodecin containing matrix formed from SO1-SO2

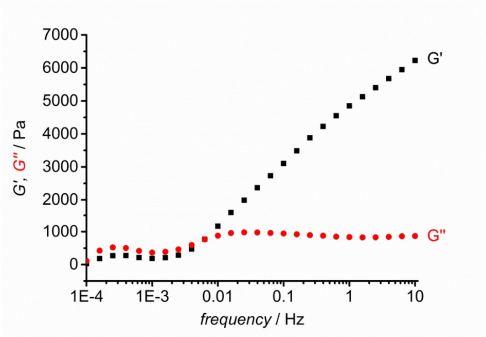


Figure S4. Frequency dependent rheological measurements of the dodecin (1 mM) containing matrix formed from **SO1-SO2** (3 mM) at 20 °C.

To transfer the highly viscous dodecin containing matrix formed from **SO1-SO2** from the NMR tube into a rheometer, the NMR-tube first had to be heated to reduce the viscosity of the DNA matrix. The rheometer was assembled with the matrix at elevated temperature und then cooled to 20 °C before the frequency dependent rheological measurement could be started going from high to low frequencies. As during the transfer of the matrix some water might have been evaporated, the absolute values for *G'* and *G''* have to be taken with care. In any case the rheological measurements show that at frequencies ≥ 0,01 Hz the matrix comprises gel-like properties (*G'* > *G''*), whereas at lower frequencies the matrix behaves as a liquid.
For the rheological measurements a Malvern Kinexus Ultra Rheometer equipped with a 20 mm, 1° cone geometry and a lower, Peltier tempered plate was employed. To prevent drying and further loss of solvent a solvent trap filled with oil was used. The matrix was studied at an amplitude of 1%.

Stability measurements of dodecin at elevated temperatures

The thermal stability of dodecin was determined by the ability of dodecin to rebind riboflavin (RbF) after heat induced dissociation of the dodecin:RbF complex. After each cycle of a heating step (5-95 °C, 6 min) and a cooling step (5 °C, 30 min) the heating step temperature was increased (5-50 °C 4.5 °C steps and 51-95 °C 2 °C steps). After each cooling step the fluorescence intensity is measured (excitation/emission filter bandwidth: 450-490/560-580 nm). Buffer solutions vary in NaCl concentration (see **Figure S5**), Tris (4 mM, pH 7.5) and $MgCl_2$ (1 mM) concentration are fixed for all solutions with 2 exceptions: "200 mM NaCl pH ~6.0" buffer solution contains maleic acid (3.5 mM), MES (3.5 mM) and Tris (4 mM), final pH ~6.0. "90% standard buffer" contains 907 mM NaCl, 4.5 mM $MgCl_2$ and 18.1 mM Tris (pH 7.5).

RbF bound to dodecin shows only marginal fluorescence intensities, allowing one to distinguish between bound and unbound RbF *via* fluorescence, any increase in fluorescence thereby is caused by not rebound RbF. With the assumption that only the dodecamer can efficiently bind RbF, the increase of fluorescence corresponds to the disassembly/denaturing of the dodecamer. The fluorescence intensity plotted against the heating step temperature shows that with higher ion strength the higher temperatures are needed to reach similar fluorescence intensities (curves shifted to the right). Under the condition used in this study (200 mM NaCl) dodecin shows up to 50 °C only slightly increased fluorescence intensities, suggesting the dodecamer is stable under these conditions.

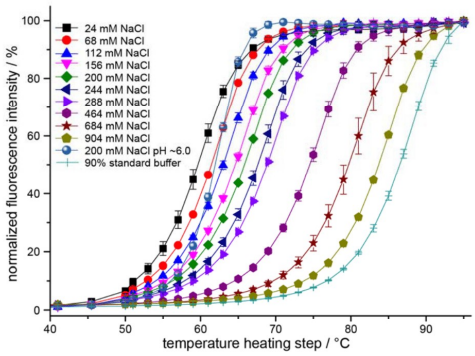


Figure S5. Stability measurement. Normalized RbF fluorescence intensity plotted against heating step temperature. Fluorescence intensity of each well was corrected by the corresponding fluorescence intensity minimum and maximum, and then averaged (3 replicates).

To determine the longtime thermal stability of dodecin the measurements were repeated with a 4 h heating step (only 5 steps, see Table S1). The longtime measurement is in agreement with the first measurement and suggests that dodecin is under the here used condition stable up to 50 °C.

Table S1. Longtime thermal stability of dodecin in different NaCl concentrations. Fluorescence was corrected by the initial fluorescence and normalized to the maximal value of each measurement/well, for most measurements this was obtained after the 80 °C heating step. If the fluorescence intensity was below the corresponding control (only RbF no dodecin added) the value of the control was used for normalization, for more stable conditions this method of normalization was needed. With the assumption that only the dodecamer can bind efficiently RbF, the normalized fluorescence can be seen as the amount of disassembled/denatured dodecamer.

	Normalized fluorescence / %				
Temperature /°C	**30**	**40**	**50**	**60**	**80**
NaCl concentration					
24 mM NaCl	0.7 ± 0.2	3.4 ± 0.1	20.2 ± 0.3	78.9 ± 1.2	98.8 ± 1.8
68 mM NaCl	0.65 ± 0.03	3.06 ± 0.03	15.7 ± 0.2	65.1 ± 0.3	100 ± 0
112 mM NaCl	0.4 ± 0.1	2.5 ± 0.1	12.8 ± 0.3	55.3 ± 1.3	100 ± 0
156 mM NaCl	0.5 ± 0.1	2.2 ± 0.1	10.3 ± 0.3	46.6 ± 0.5	100 ± 0
200 mM NaCl	0.4 ± 0.1	1.74 ± 0.04	8.3 ± 0.2	38.8 ± 1.0	100 ± 0
244 mM NaCl	0.6 ± 0.5	1.5 ± 0.5	6.7 ± 0.4	31.5 ± 0.5	100 ± 0
288 mM NaCl	0.7 ± 0.2	1.5 ± 0.2	6.0 ± 0.1	26.5 ± 0.7	100 ± 0
464 mM NaCl	1.4 ± 0.2	1.8 ± 0.3	4.1 ± 0.5	14.4 ± 1.5	92.7 ± 6.6
684 mM NaCl	1.2 ± 0.1	1.5 ± 0.2	2.5 ± 0.2	7.2 ± 0.6	55.7 ± 1.4
904 mM NaCl	0.9 ± 0.2	1.3 ± 0.2	3.1 ± 1.3	5.7 ± 0.5	26.6 ± 2.5
200 mM NaCl pH ~6.0	1.7 ± 0.2	2.0 ± 0.2	7.5 ± 0.2	73.7 ± 0.7	100 ± 0
90% standard buffer	1.3 ± 0.2	1.9 ± 0.5	2.8 ± 0.7	5.5 ± 1.1	16.4 ± 1.0

Experimental Section

NMR. Spectra were recorded using a Varian VNMR-S 600 MHz spectrometer equipped with a 3 mm triple resonance inverse probe. All NMR measurements were carried out in buffer solution. As buffer we used 4mM tris-buffer at pH = 7.5 containing 200 mM NaCl and 1 mM $MgCl_2$ prepared with a mixture of H_2O/D_2O at a ratio of 9:1.

DOSY measurements were carried out with temperature regulation using the 'Dbppste' or 'Dbppste_cc' pulse sequence from Varian Pulse sequence library. Water suppression was done using WET. The gradient ramp was built in either 32 or 64 steps with gradient strength from 0.2 G · cm^{-1} to 41 G · cm^{-1} and 2 milliseconds gradient length. The diffusion delay was set to 250 ms (only **SO1-SO2**) or 800 ms. Per step 32 to 128 transients are recorded using 2 s relaxation delay, 2.7-3.1 s acquisition time, line broadening (lb = 1 or 3), solvent subtraction and baseline correction were used.

For DOSY measurements of **O1-O2** with 10mM ^{13}C labelled sucrose in D_2O a dual broadband 3 mm probe has been used. The disaccharide was purchased from Omicron Biochemicals Inc., South Bend, Indiana, USA with [$^{13}C_6$] fructose. The 'Dbppste' pulse sequence has been used for ^{1}H detection with a gradient ramp of 32 steps and 32 transient per step. 'Dbppsteinept' has been used for ^{13}C detection using 16 steps and 512 transient per step. The diffusion delay was set to 100 ms. 2 s relaxation delay, line broadening (lb = 3, lb = 10 for DOSYInept) and baseline correction has been used.

The DOSY data were processed with VNMRJ4.2A according to discrete multi-exponential fitting (SPLMOD). Data were corrected for non-uniform gradients. Gradient calibration was done using the "doped D_2O sample" delivered with the spectrometer.

DNA samples. Highly purified (ion exchange HPLC, IEX-HPLC followed by demineralization including a quality check by capillary electrophoresis CGE) oligonucleotides were purchased from Eurogentec, Belgium.

Dodecin. Dodecin was prepared and purified as described elsewhere.[5] Dodecin stability measurement were carried out similar as described in ref.[6], instead of flavin mononucleotide (FMN) in the current study RbF was used and 3 µL of a 40 µM dodecin.RbF complex solution instead of 2 µL of a 50 µM solution were used (buffer solution volume adjusted to 22 µL). For the longtime stability measurement the heating step duration was increased from 6 min to 4 h (Heating step temperatures see supporting information Table S1) See supporting information for buffer solutions.

[5] C. Gutierrez Sanchez, Q. Su, H. Schönherr, G. Nöll, M. Grininger, *ACS Nano* **2015**, *9*, 3491-3500.

[6] F. Bourdeaux, C. A. Hammer, S. Vogt, F. Schweighoefer, G. Nöll, J. Wachtveitl, M. Grininger, *ACS Infectious Diseases* **2018**, *4*, 1082-1092.

www.nature.com/scientificreports

SCIENTIFIC REPORTS
nature research

OPEN

Dodecin as carrier protein for immunizations and bioengineering applications

Florian Bourdeaux[1], Yannick Kopp[2], Julia Lautenschläger[1], Ines Größner[1], Hüseyin Besir[3,5], R. Martin Vabulas[4] & Martin Grininger[1]✉

In bioengineering, scaffold proteins have been increasingly used to recruit molecules to parts of a cell, or to enhance the efficacy of biosynthetic or signalling pathways. For example, scaffolds can be used to make weak or non-immunogenic small molecules immunogenic by attaching them to the scaffold, in this role called carrier. Here, we present the dodecin from *Mycobacterium tuberculosis* (*mt*Dod) as a new scaffold protein. *Mt*Dod is a homododecameric complex of spherical shape, high stability and robust assembly, which allows the attachment of cargo at its surface. We show that *mt*Dod, either directly loaded with cargo or equipped with domains for non-covalent and covalent loading of cargo, can be produced recombinantly in high quantity and quality in *Escherichia coli*. Fusions of *mt*Dod with proteins of up to four times the size of *mt*Dod, e.g. with monomeric superfolder green fluorescent protein creating a 437 kDa large dodecamer, were successfully purified, showing *mt*Dod's ability to function as recruitment hub. Further, *mt*Dod equipped with SYNZIP and SpyCatcher domains for post-translational recruitment of cargo was prepared of which the *mt*Dod/SpyCatcher system proved to be particularly useful. In a case study, we finally show that *mt*Dod-peptide fusions allow producing antibodies against human heat shock proteins and the C-terminus of heat shock cognate 70 interacting protein (CHIP).

Abbreviations

ACP	Acyl carrier protein
AB	Antibody
BSA	Bovine serum albumin
Catcher	small protein fold (SpyCatcher or SnoopCatcher) that binds and reacts with Tag
CellSig.	Cell Signaling Technology
CHIP	C-terminus of heat shock cognate 70 interacting protein
EU	Endotoxin units
FMN	Riboflavin-5′-phosphate
GFP	Green fluorescent protein
GSG, PAS, PAS2 GPAS, GPAS2, PASG, PAS2G, linker systems	See Table 1
HB-EGF	Proheparin-binding EGF-like growth factor
HSP	Heat shock protein
IPTG	Isopropyl-β-D-thiogalactopyranoside
KLH	Keyhole limpet hemocyanin
L	Ladder (only used in figures)

[1]Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Cluster of Excellence for Macromolecular Complexes, Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt am Main, Germany. [2]Institute of Biophysical Chemistry, Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, Max-von-Lauer Str. 15, 60438 Frankfurt am Main, Germany. [3]European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [4]Institute of Biochemistry, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. [5]Present address: PROGEN Biotechnik GmbH, 69123 Heidelberg, Germany. ✉email: grininger@chemie.uni-frankfurt.de

LAL	*Limulus* amebocyte lysate
Lys	Lysate (only used in figures)
MAP	Multiple antigen peptides
*mm*ACP	*Mus musculus* acyl carrier protein
msfGFP	Monomeric superfolder green fluorescent protein
*mt*Dod	*Mycobacterium tuberculosis* dodecin
*mt*Dod(WT)	*Mycobacterium tuberculosis* dodecin wild type
OD600	Optical density at 600 nm
OE	Over expressing cells
RSA	Rabbit serum albumin
SCBT	Santa Cruz Biotechnology
*se*ACP	*Saccharopolyspora erythraea* acyl carrier protein
SEC	Size exclusion chromatography
Sfp	4′-Phosphopantetheine transferase from *Bacillus subtilis*
Sigma	Sigma-Aldrich
SnpC	SnoopCatcher
SnpT	SnoopTag
SpyC	SpyCatcher
SpyT	SpyTag
SZ	SYNZIP domain
Tag	Small peptide sequence that interacts with Catcher's (SpyTag or SnoopTag)
TB	Terrific broth
TBS	Tris-HCl buffered saline
TBST	Tris-HCl buffered saline with Tween-20
TT	Tetanus toxoid
VLP	Virus-like particle

For being suited as scaffolds, proteins need to meet an array of requirements. Depending on the actual use, multiple features of the protein can be important; e.g., particle size, achievable purity, expression level, robustness of fold/assembly, general stability and immunogenicity (if used for immunizations). Two key features are obligatory, in addition. Scaffolds need to form a stable and water-soluble structure that is best insensitive to the attached cargo, and they should further allow the dense packing of the cargo in homovalent and ideally also in heterovalent fashion[1–5].

One application of scaffold proteins is their conjugation with peptides for the generation of antibodies (AB), utilizing the increased immunogenicity of the protein-peptide conjugate (in this role often called carrier proteins)[6]. Such ABs can be used to identify proteins, which contain the peptides used for AB generation, in complex samples, and allow the specific labelling of proteins of interest in their spatiotemporal distribution, e.g. by immunofluorescence imaging or western blotting. For the reactivity of the ABs, the selection of the peptide is critical, since the ABs derived from the conjugate can only recognize the peptide as presented (or similar) on the carrier[7]. For the recognition of the protein in its native form, the correct sequence, but also the structure and surface exposure of the selected peptide need to be considered[7]. For B-cell activation, the conjugated peptide needs to be exposed on the surface of the carrier, and it is thought that a dense packing of the conjugated peptide is advantageous for this, because highly repetitive epitopes on the particle/carrier surface facilitate B-cell receptor oligomerization[1,2].

Usually peptide-carrier conjugates for AB production are formed by coupling an about 20 amino acid-long peptide to residues at the surface of a carrier protein via a chemical reaction[8–10]. Commonly used carrier proteins are keyhole limpet hemocyanin (KLH), bovine serum albumin (BSA) and rabbit serum albumin (RSA), but also other proteins, e.g. tetanus toxoid (TT), and artificial carrier-systems, e.g. multiple antigen peptides (MAP) or virus-like particles (VLP, not limited to chemical conjugations), are used[11,12]. While BSA bears typical carrier properties (likely also other albumins), and exposes the peptides at the surface at a potentially high density[13,14], KLH is often preferred as a carrier-protein due to its high immunogenicity[15,16]. Notably, the immune system reacts to the entire conjugate, and, therefore, ABs are not just raised against the peptide of interest, but also against the carrier protein and the linker (peptide or remnant of the coupling agent/crosslinker). To avoid cross-reactivity by anti-carrier or anti-linker ABs, it is beneficial to use carrier-linker systems for immunization that have no or only very low similarity with the inventory of cells and tissues that are supposed to be analysed[16].

Although the method of chemically coupling peptides of interest to carrier proteins is commonly used, it is not without weaknesses. Besides limitations arising from solid support synthesis of the peptide themselves, e.g. limited solubility of hydrophobic sequences or amyloid forming sequences, the spectrum of peptides that can be coupled to the above listed carrier proteins is constrained by its compatible with the coupling agent. For example, internal cysteine residues are avoided, because they are commonly terminally introduced for the coupling to the carrier protein[8–10]. Further, in some cases, the stability of the conjugation product or intermediate (activated carrier) can be problematic[10,17]. An alternative method to the coupling approach is the direct expression of self-assembling peptide-carrier conjugates, provided as encoding DNA sequence to the recombinant expression host. This approach allows more flexibility in the design of epitopes and linkers, since the limitations of coupling reactions or peptide synthesis need not be taken into account. Self-assembling carrier proteins can also be produced with tags or proteins that allow post-translational covalent linking of cargo, thereby not relying

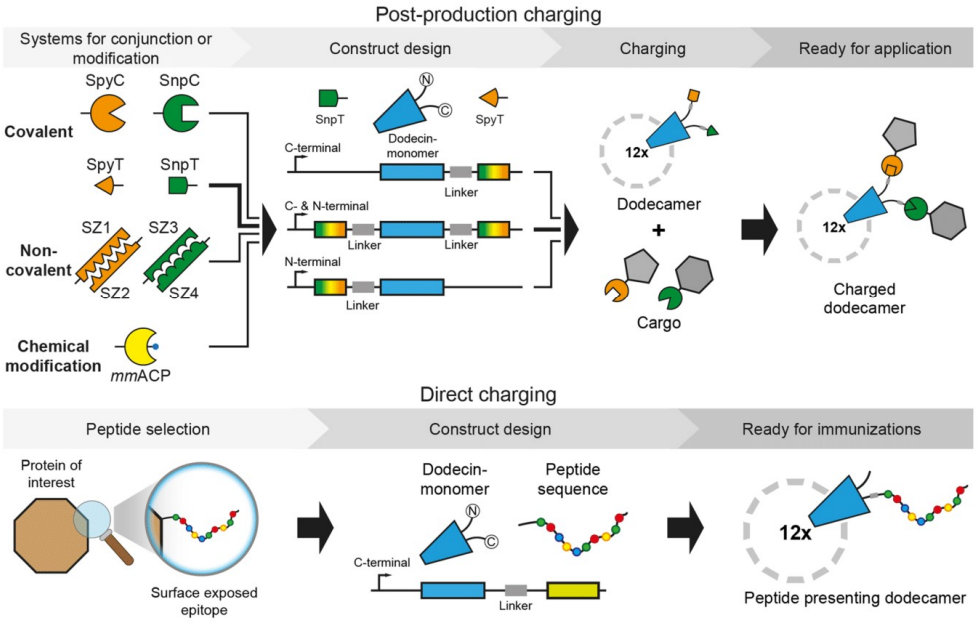


Figure 1. Schematic depiction of *mt*Dod constructs and workflow presented in this study. In this study, termini were modified directly at gene level with peptides, domains or proteins for subsequent charging with cargo (top) or direct coupling of peptides for immunizations (bottom). The dodecamers expose their termini at the outer surface. (Top) Selected example is highlighted by arrows in bold. SpyT and SpyC: SypTag and SpyCatcher[25,26]. SnpT and SnpC: SnoopTag and SnoopCatcher[27]. SZ1-SZ4: helical domains that bind to their specific counterpart, called SYNZIP[28]. *Mm*ACP: *Mus musculus* acyl carrier protein (ACP), gene *Fasn*.

on chemical coupling[18,19]. While this carrier method has high potential, it is reliant on the availability of stable and robust self-assembling protein- and peptide-scaffolds.

The dodecin protein family was recently discovered as a flavin storage and buffering system that occurs in bacteria and archaea, but not in eukaryotes[20–23]. Dodecins are 8 kDa small proteins of βαββ-topology. Although forming a small antiparallel β-sheet that partly enwraps the helix, the dodecin fold is unique. Dodecins largely meet the requirements of protein scaffolds. In the native dodecameric state, dodecins are of spherical shape with 23-cubic symmetry, and the N- and C-termini are exposed at the protein surface. Dodecins show pronounced thermostability (> 95 °C)[22–24], which likely originates from an extensive antiparallel β-sheet that is built upon protomer assembly.

Here, we present dodecin from *Mycobacterium tuberculosis* (*mt*Dod) as a new carrier protein for peptides and scaffold for bioengineering applications. To evaluate *mt*Dod's suitability as a carrier/scaffold protein, we expressed and purified diverse *mt*Dod fusion constructs, and analysed the obtained dodecamers. The robustness of the dodecamer assembly was probed by the ability to express the diverse constructs as dodecamers in *Escherichia coli*. Further, we analysed the stability of the obtained dodecamers, and found that it is largely unaffected by the attached tag, linker and/or cargo. Finally, we demonstrate that the use of *mt*Dod as a carrier benefits from its accessibility in high amounts via by a simple heat denaturation protocol. *Mt*Dod conjugates with heat sensitive cargo were purified via conventional affinity chromatography.

The exposed termini of *mt*Dod can be harnessed for charging/functionalization in diverse ways of which we used two in this study: First, a cargo was directly fused by attaching the peptide/protein-encoding sequence at the gene level. Second, *mt*Dod was terminally modified with conjugation sites that allow post-translational covalent and non-covalent fusions of the peptide/protein as well as other chemical entities to the intact dodecin carrier (Fig. 1). In a case study, we directly tested the suitability of *mt*Dod as carrier for producing anti-peptide ABs for laboratory use. ABs were raised in rabbits against *mt*Dod-peptide fusions, and showed comparable labelling capability as commercially available ABs judged by western blotting.

Results and discussion

Dodecin can be recombinantly produced in high yields. To evaluate the suitability of *mt*Dod as a carrier protein, several *mt*Dod constructs were designed and purified. All constructs were expressed in *E. coli* BL21 (DE3). Cells were grown in terrific broth (TB) medium to an optical density at 600 nm (OD_{600}) of about

Construct name	Linker system	Molar mass/Da	Expression state[#]
mtDod-peptides			
*mt*Dod(WT)	–	7,497	Soluble
*mt*Dod-GSG-Lys	GSG	8,411	Soluble
*mt*Dod-PAS-Met	PAS	8,876	Soluble
*mt*Dod-SpyT	PASG	10,458	Soluble
SpyT-*mt*Dod	GPAS	10,215	Soluble
*mt*Dod-PAS2-SpyT	PAS2G	11,447	Soluble
SpyT-PAS2-*mt*Dod	GPAS2	11,205	Soluble
SpyT-*mt*Dod-SnpT	GPAS / PASG	13,142	Soluble
mtDod-proteins			
*mt*Dod-*mm*ACP	PAS	17,994	Soluble
*mt*Dod-*mm*ACP-H8	PAS	19,334	Soluble
*mt*Dod-msfGFP-H8	PAS	36,388	Soluble
*mt*Dod-SpyC-H8*	PAS	22,413	Inclusion body
H8-SpyC-*mt*Dod	PAS	22,072	Inclusion body
*mt*Dod-SZ1**	PAS	14,232	Inclusion body
*mt*Dod-SZ3**	PAS	13,396	Inclusion body
*mt*Dod-*se*ACP***	PAS	19,966	Inclusion body

Linker details			
GSG:	GGGGSGGGG	PAS:	SPAAPAPASPAS
PASG:	SPAAPAPASPASGGSG	GPAS:	GGSGSPAAPAPASPAS
PAS2G:	SPAAPAPASPASPAPSAPAASPAAGGSG	GPAS2:	GGSGSPAAPAPASPASPAPSAPAAS-PAA

Table 1. Selection of *mt*Dod constructs used for expression studies are divided into two groups: *mt*Dod-peptides (constructs with only short peptides fused to *mt*Dod) and *mt*Dod-proteins (constructs with domains or entire proteins fused to *mt*Dod). *mt*Dod(WT): wild type *mt*Dod. *se*ACP: *Saccharopolyspora erythraea* ACP, gene *chlB2*. msfGFP: monomeric superfolder green fluorescent protein[29]. For a full description of constructs, see Supplementary Table S1. The constructs *mt*Dod-GSG-Lys and *mt*Dod-PAS-Met have been used as control; *Mt*Dod-GSG-Lys was used for its flexible linker, and *mt*Dod-PAS-Met for its rigid linker. The PAS linker were based on sequences (slight alterations) shown in ref[30]. [#]Describes whether the major fraction of the expressed construct is soluble or forms inclusion bodies. *Mt*Dod constructs with low solubility form often non-classical inclusion bodies (correctly folded protein)[31], leading to their yellowish colouring (flavin binding). Since flavin binding only requires intact *mt*Dod, it is possible that inclusion bodies are yellow although the protein cargo is misfolded. **Mt*Dod-SpyC-H8 seems to be soluble in cellular environment but forms yellow aggregates after cell lysis. **SZ1-*mt*Dod and SZ3-*mt*Dod also formed inclusion bodies and behaved similarly as the C-terminal constructs (data not shown). ***Mt*Dod-*se*ACP could not be obtained in soluble form; under all applied refolding conditions yellow aggregate was formed.

0.6–0.8 at 37 °C before induction with isopropyl-β-D-thiogalactopyranoside (IPTG; 0.5 mM final concentration), and expression was performed over night at 20 °C. Since *mt*Dod is a flavin binding protein (preferred flavin-ligand is riboflavin-5′-phosphate (FMN))[23,24], its overexpression causes increased amounts of cellular flavin, leading to a yellowish colouring of the cells. Cells were lysed by French press, and the cell debris was removed by centrifugation. Depending on the *mt*Dod construct different purification strategies were applied. In the following, construct names are underlying the nomenclature: peptides or proteins fused to the N-terminus of a protein are written in front of the protein (e.g. peptide-*mt*Dod) and C-terminal fusions are written after the protein (e.g. *mt*Dod-peptide). Most *mt*Dod constructs were produced as soluble proteins, but some proteins, such as *mt*Dod-SZ1, *mt*Dod-SZ3 (SYNZIP constructs)[28], H8-SpyC-*mt*Dod and *mt*Dod-SpyC-H8 (SpyC constructs)[25,26] accumulated as inclusion bodies (Table 1).

For soluble *mt*Dod constructs, most cytosolic *E. coli* proteins were removed by heat denaturation at about 75 °C. *Mt*Dod itself is stable to temperatures above 95 °C under standard conditions (pH ~ 7.5 and ionic strength > 100 mM, e.g., in PBS), and the thermal stability can be further increased by adding the native FMN ligand in excess[23,24]. Depending on the stability of the fused cargo, lower temperatures during the heat denaturation may be necessary, or different purification approaches need to be applied (affinity chromatography). For example, *mt*Dod-*mm*ACP started to precipitate at about 55–60 °C in spite of *mt*Dod staying intact, as indicated by maintained FMN binding and preserved dodecameric stability (Supplementary Fig. S1). In this case, heat denaturation was conducted at about 55 °C. Lower temperatures during the heat denaturation step can affect the purity of preparations, because some *E. coli* proteins remain soluble. Following heat treatment, *mt*Dod constructs were generally further purified by two cycles of DMSO-induced precipitation (50% final DMSO concentration). Finally, size-exclusion chromatography (SEC) was performed to select for dodecameric fractions, which can be

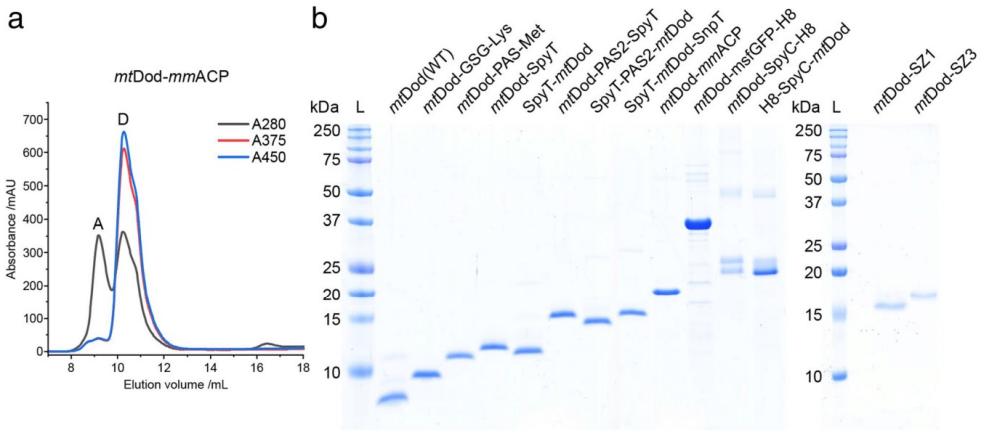


Figure 2. SEC profile of *mt*Dod-*mm*ACP and SDS-PAGE gel of various *mt*Dod constructs. (**a**) SEC chromatogram of *mt*Dod-*mm*ACP after heat denaturation and DMSO precipitation. Used column: Superdex 200 increase 10/300 column (GE Healthcare). A: peak representing aggregates. D: peak representing the dodecamer. The oligomerization status was assigned based on the absorption at 375 nm (A375) and 450 nm (A450) (indicating bound flavin) and their elution volume (species at lowest molecular weight eluting with bound flavin). Although *mt*Dod-*mm*ACP in general formed yellow coloured aggregates, the here observed aggregate does not show flavin absorption bands, indicating that this aggregate does not contain dodecameric species. (**b**) SDS-PAGE gel of purified *mt*Dod constructs. L: Ladder. For full dissociation of the dodecamers in SDS-PAGE, an acidic loading buffer containing 3.3% SDS was used during the heat treatment (5 min 60 °C). After the heat treatment, the pH was increased to about 6.8 using a glycerol- and Tris–HCl-containing buffer, followed by a second heat treatment (5 min 95 °C). *Mt*Dod-SZ1 and *mt*Dod-SZ2 were denatured with loading buffer containing ~ 7 M urea and 2.5% SDS (prolonged heat treatment: 15 min 95 °C). Denaturation with acidic loading buffer was more reliable for most constructs and easier in handling compared to urea-based protocols (in some cases even 7–8 M urea failed to dissociate the protein completely). Of note, when treated with acidic loading buffer, some constructs showed additional bands (mainly SpyC constructs, Supplementary Fig. S5). The origin of this behaviour was not further investigated.

easily identified by the absorption bands of bound flavin (375 nm and 450 nm). For *mt*Dod-peptide fusions, the dodecamer turned out to be the main oligomeric species. While only minor peaks representing lower oligomeric states were detected, aggregation peaks were observed at high concentrations (Supplementary Fig. S2). For larger *mt*Dod constructs with fused proteins, like *mt*Dod-*mm*ACP (Fig. 2a), significant aggregation was observed in the SEC profiles (see Supplementary Fig. S2).

For the construct *mt*Dod-msfGFP-H8, purification by heat denaturation (70 °C, above that aggregation was observed) and purification by affinity chromatography were compared. GFP is a suitable cargo for this test, because GFP is highly thermostable[32]. The dodecameric structure of dodecin causes a high density of surface exposed affinity tags, allowing vigorous washing without severe protein loss during Ni-chelating affinity chromatography. Accordingly, *mt*Dod-msfGFP-H8 was washed with two column volumes of a 200 mM imidazole-containing wash buffer, and elution was performed at 400 mM imidazole. While with both purification strategies *mt*Dod-msfGFP-H8 dodecamer was obtained, the sample purified by heat denaturation showed severe aggregation in SEC (Supplementary Fig. S3).

*Mt*Dod constructs that aggregate in inclusion bodies can be refolded by dialysis, as previously described[23], under conditions optimized for the respective fused cargo. All inclusion bodies were first washed and then dissolved by denaturation using 6 M guanidinium chloride. *Mt*Dod was refolded without further purification at different conditions ranging from pH 5.0[23] to pH 8.5. Refolding was possible for all constructs obtained as inclusion bodies in this study, although the resolubilized proteins remained aggregation-prone, particularly during protein concentration and filtration. For a screen of buffer conditions for refolding constructs *mt*Dod-SpyC-H8 and H8-SpyC-*mt*Dod, see Supplementary Fig. S4. Notably, for both constructs, a glycerol-containing buffer was found to be best suited for refolding.

Overall, all constructs presented in Table 1, except *mt*Dod-*se*ACP, were obtained in high purity (see Fig. 2b).

We thought that the insolubility and aggregation problems observed for some constructs may be solved by the formation of *mt*Dod-heterododecamers, because then the density of entities on the surface could be reduced. To probe heterododecamer formation with *mt*Dod in vitro and in vivo, we worked with the two species *mt*Dod-PAS-Strep and *mt*Dod(WT). We note that *mt*Dod-PAS-Strep was used for its availability in the lab and is not compromised in solubility. We assume that other *mt*Dod-PAS-peptide constructs than *mt*Dod-PAS-Strep could have been used, too. For in vitro heterododecamer formation, *mt*Dod(WT) and *mt*Dod-PAS-Strep were jointly refolded in different relative concentrations, while for the formation of heterododecamers in vivo,

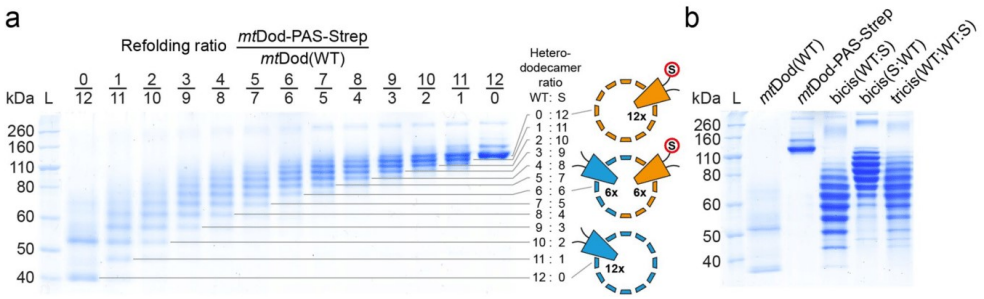


Figure 3. SDS-PAGE gel of purified heterododecamers of *mt*Dod(WT) (WT, light blue in cartoon) and *mt*Dod-PAS-Strep (S, orange with red encircled "S"). L: Ladder. For analysis of heterododecamer composition, the Tris–glycine (Lämmli) SDS-PAGE system was used, and samples were prepared without heat treatment and SDS (for full length gel images see Supplementary Fig. S6). For *mt*Dod(WT) two bands at about 40 kDa and slightly above 50 kDa are observable (both below the weight of the dodecamer, about 90 kDa). The band at about 40 kDa represents likely the dodecamer and the band above seems to be caused by the mild denaturation conditions during sample preparation. The appearance of additional bands at higher molecular weight is also observable in other lanes. We want to note that these bands don't depict the hexamer and the dodecamer, as all protein–protein interactions, which would stabilize the hexamer, are present in higher numbers in the dodecamer with also other additional stabilizing interactions. The origin of the *mt*Dod dodecamer migration behaviour and the molecular mechanism behind the bands at higher molecular weight is not clear. (**a**) SDS-PAGE gel of heterododecamers obtained by refolding *mt*Dod(WT) and *mt*Dod-PAS-Strep at different ratios. Each band represents a heterododecamer with a defined composition of *mt*Dod-PAS-Strep and *mt*Dod(WT) (13 bands for 13 possible compositions, higher molecular weight bands above about 160 kDa excluded). (**b**) SDS-PAGE gel of purified heterododecamers formed during polycistronic expression of *mt*Dod(WT) and *mt*Dod-PAS-Strep. Bicis(WT:S): Bicistronic expression vector design: *mt*Dod(WT) encoding gene first and *mt*Dod-PAS-Strep encoding gene second. Bicis(S:WT): Bicistronic expression vector design: *mt*Dod-PAS-Strep encoding gene first and *mt*Dod(WT) encoding gene second. Tricis(WT:WT:S): Tricistronic expression vector design: *mt*Dod(WT) encoding gene first and second and *mt*Dod-PAS-Strep encoding gene third.

three combinations of the *mt*Dod constructs were expressed polycistronically. The analysis of heterododecamer compositions was possible by the high stability of the dodecamers (see Supplementary Fig. S5 for *mt*Dod(WT) dodecamer observable in SDS-PAGE) and the different migration behaviour of constructs/species in SDS-PAGE (Fig. 3). The slower migration of *mt*Dod-PAS-Strep is likely caused by the limited binding of SDS to the exposed PAS sequence (see Fig. 2b *mt*Dod-GSG-Lys compared to *mt*Dod-PAS-Met)[30]. Data indicates that the composition of heterododecamers is controlled by the relative concentration of species in the refolding solution,i.e., the higher the concentration of a construct, the more abundant it is in the refolded dodecamer. Of note, assuming that the heterododecamer formation is just controlled by the concentration of each construct (see Fig. 3a), the band patterns for heterododecamers assembled in vivo can be used to estimate gene order related expression strength (see Fig. 3b), as described in the literature for other methods, e.g. FRET[33]. The estimated relative expression strength of each gene is for the bicistronic vector: first > second, and for the tricistronic vector first > third > second.

Dodecin is highly stable. We have recently established the cyclic thermal shift assay, termed thermocyclic fluorescence assay, to determine the stability of dodecins[23]. This assay is based on the fluorescence quenching that is observed when flavins bind to dodecin. In each binding pocket of the dodecamer, the two isoalloxazine ring systems of two bound flavins are embedded between symmetry-related tryptophans.[22–34] Since dodecins can only bind flavins in the dodecameric state, the fluorescence intensity of flavins can be used to estimate the amount of dodecameric *mt*Dod in solution. In contrast to standard melting analysis, in which the temperature is continuously increased, the thermocyclic fluorescence assay runs cyclic temperature profiles that contain a heating phase (temperature increased per cycle) and a cooling phase (for all cycles cooled to 5 °C). At the heating phase, FMN is released from the binding pocket and the fluorescence intensity increases. During cooling, FMN can rebind to the dodecamer (cooling phase) restoring initial low fluorescence values. As soon as the dodecamer denatures irreversibly, the fluorescence intensity remains at elevated levels. By plotting the fluorescence intensity of the cooling phase against the heating phase temperature, the thermal stability of the dodecamer of the *mt*Dod constructs can be determined.

Since all constructs, except *mt*Dod-SZ1 and *mt*Dod-SpyC-H8, proved to be stable in PBS buffer throughout the entire temperature range, we identified the slightly destabilizing conditions of pH 4.2 as suited to sense the impact of the cargo on the integrity of the *mt*Dod dodecameric scaffold (see Fig. 4). Under this condition, the thermally stable constructs *mt*Dod(WT) and *mt*Dod-peptides started to denature at 75–80 °C. Of note, we considered a protein to denature when the fluorescence curve reached its knee right before going into a steep increase in fluorescence reaching values of above 30%. It is further important to note that the thermocyclic fluorescence assay does only monitor the dodecameric stability, which may be influenced by the attached cargo. For

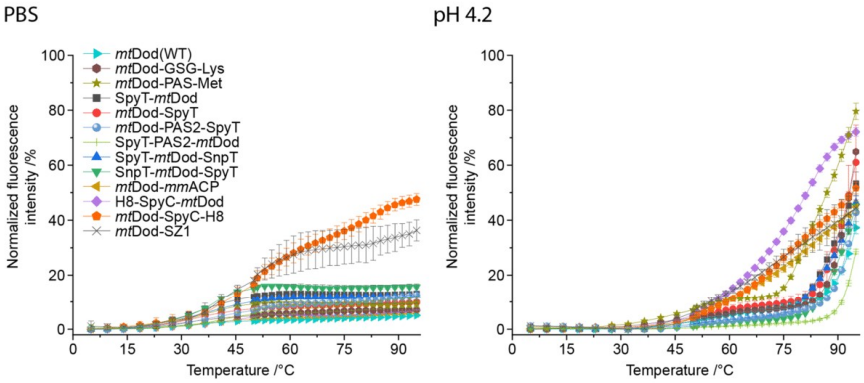


Figure 4. Thermal stablity of *mt*Dod constructs. The FMN fluorescence at the rebinding/cooling phase is plotted against the heating phase temperature. The increase of FMN fluorescence indicates disassembly of the dodecamer at the heating phase, as the flavin cannot rebind in the cooling phase and its fluorescence is not quenched. In PBS, only a negligible increase of fluorescene is observed within the entire temperature range, except for *mt*Dod-SpyC and *mt*Dod-SZ1, indicating that the dodecameric *mt*Dod core structures do not disassemble. The minor increase of fluorescence of up to 20% around 45–50 °C might be caused by hindered rebinding of FMN and not by disassembling of the dodecamer. At pH 4.2, for all constructs, a steep (compared to PBS data) increase of fluorescence is observable indicating the dodecamer disassambly. Most constructs behave like *mt*Dod(WT), and are stable to about 80 °C, except *mt*Dod-PAS-Met, *mt*Dod-*mm*ACP, *mt*Dod-SpyC-H8 and H8-SpyC-*mt*Dod of which the latter three start denaturing already at 50 °C. *mt*Dod-PAS-Met is only slightly less stable, and starts to denature around 75 °C. Note that *mt*Dod-SpyC-H8 and *mt*Dod-SZ1 suffer from strongly impaired FMN binding leading to non-saturated binding sites, and data needs to be treated with care. The FMN binding may also be altered by denaturing/aggregation of the fused fold, causing different curve profiles, e.g. *mt*Dod-*mm*ACP. For all measurements, fluorescence was normalized to the maximum values recorded in the heating phase, corrected by the temperature-induced fluorescence decline of FMN. Curves connect the averages of three technical replicates. Standard deviations are shown as error bars.

example, in screening temperatures for the heat denaturation of *mt*Dod-*mm*ACP, we observed the formation of yellowish agglomerates above 55–60 °C, indicating that the construct is intact in the *mt*Dod scaffold, as capable of FMN binding, but precipitated by the thermally unfolded *mm*ACP (band representing the intact dodecamer observable in SDS PAGE, see Supplementary Fig. S7).

The high dodecameric stability of *mt*Dod is also observed in SDS-PAGE using the standard loading buffer (2.5% SDS, pH 6.8) (see Supplementary Fig. S5). Under these conditions, further depending on the heat treatment for sample preparation, a dodecameric fraction remains intact, as indicated by the high molecular weight band representing the dodecamer. In accordance to the lower stability at pH 4.2, observed in the thermocyclic fluorescence assay, a two-component acidic loading buffer (3.3% SDS and pH < 4.2 during heat treatment, afterwards 2.5% SDS and pH 6.8) was applied to fully denature the dodecamer (Fig. 2).

While we did not study the effects of freezing and thawing explicitly, we would like to note that we did not observe noticeable aggregation for most *mt*Dod constructs after freezing and thawing (all constructs presented here were frozen and thawed at least once). However, constructs that are prone to aggregation might be problematic during freezing and thawing. Accordingly, we noticed aggregation for *mt*Dod-msfGFP-H8, indicated by green fluorescent aggregates after thawing, and *mt*Dod SYNZIP constructs, forming yellowish precipitate. For SpyC *mt*Dod constructs, glycerol containing buffer could prevent noticeable aggregation after freezing and thawing.

Peptides/proteins fused to *mt*Dod remain functional. The accessibility and functionality of folds and peptides fused to *mt*Dod were tested by the reactivity of the SpyT/-C and SnpT/-C pairs[25–27]. These systems allow the covalent conjugation between two entities of which one is equipped with a peptide tag (Tag) and the other with a small protein fold (Catcher)[26]. Applications range from attaching proteins from pathogens to scaffolds, like VLPs and IMX313 (heptamer forming coiled coils), for immunizations[18,19], to recruiting enzymes to a scaffold hub for creating assemblies with elevated substrate turnover[35]. In this study, *se*ACP-SpyC and mClover3-SnpC were prepared as cargo for performing SpyT/-C and SnpT/-C reactions with the respective Tag-labelled *mt*Dod constructs. For the inverse reaction, *mt*Dod SpyC constructs and SpyT-*se*ACP were used. For all reactions, the scaffold was saturated with two molar equivalents of cargo. The reactions were incubated for 20 h at 22 °C, and analysed by SDS-PAGE (Fig. 5).

For all combinations of *mt*Dod scaffold and cargo, the expected product band(s) of *mt*Dod and the specific cargo(s) were observed in SDS-PAGE. While for *mt*Dod SpyT/SnpT constructs no unreacted scaffold proteins was observed, for the inverse setting, with *mt*Dod SpyC constructs, bands of unreacted scaffold monomer were visible (possibly caused by aggregation problems of the *mt*Dod SpyC constructs). We note that *mt*Dod SpyT/

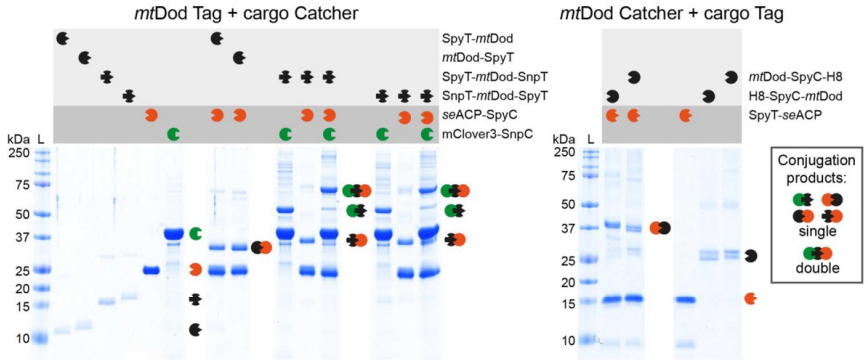


Figure 5. SDS-PAGE of the SpyT/-C and SnpT/-C reactions. Left: Reactions of *mt*Dod Spy-/SnpT constructs with *se*ACP-SpyC and/or mClover3-SnpC. Right: Inverse reactions of *mt*Dod SpyC constructs with SpyT-*se*ACP. In all reactions, bands of higher mass representing the conjugation products are observed. As mentioned above, the acidic loading buffer causes the appearance of double bands (*mt*Dod SpyC constructs) and smearing bands (*se*ACP-SpyC-H8) for some constructs.

SnpT constructs are lower in molecular mass than the *mt*Dod SpyC constructs, and traces of unreacted scaffold protein may be less visible on SDS-PAGE gels. This data shows that a high degree of saturation was achieved, indicating that SpyT/-C and SnpT/-C are well accessible at the *mt*Dod dodecamer scaffold. Double-tagged constructs SpyT-*mt*Dod-SnpT or SnpT-*mt*Dod-SpyT, heterovalently loaded with *se*ACP-SpyC and mClover3-SnpC, revealed bands of single-charged *mt*Dod monomers in SDS-PAGE. We explain this observation by an increased density at the surface of *mt*Dod that sterically constrains the conjugation with both cargos. Similar as the SpyT/-C and SnpT/-C constructs, also the SYNZIP constructs can be used for recruiting proteins to the *mt*Dod scaffold (although non-covalently). Due to the limited solubility and high aggregation tendencies of SYNZIP constructs, we only tested if *mt*Dod SYNZIP constructs are able to interact with the respective SYNZIP counterpart (e.g., *mt*Dod-SZ1 with SZ2-mClover3). For both *mt*Dod SYNZIP constructs, we observed the formation of *mt*Dod cargo adducts, indicated by higher apparent molecular mass peaks in SEC (Supplementary Fig. S8). This shows that also SYNZIP domains fused to *mt*Dod are functional and accessible. However, we deemed the SpyT/-C and SnpT/-C systems more suitable for *mt*Dod constructs, and did not further investigated the SYNZIP system.

In order to probe the accessibility and functionality of linked folds further, we tested the labelling of *mm*ACP linked to *mt*Dod with a 4′-phosphopantetheine CoA fluorophore mediated by the 4′-phosphopantetheine transferase from *Bacillus subtilis* (Sfp). The Sfp-mediated modification of ACP with CoA-modified fluorophores (CoA-488; ATTO-TEC dye ATTO 488) has been frequently used for the labelling of cellular compounds[36]. All reactions were conducted at 25 °C for 1 h in triplicates, and stopped by the addition of acidic loading buffer and analysed by SDS-PAGE. To determine the relative accessibility of *mm*ACP linked to *mt*Dod, fluorescence intensities of *mt*Dod-*mm*ACP and *mt*Dod-*mm*ACP-H8 were compared to free *mm*ACP after labelling (Fig. 6).

By comparing the fluorescence intensities of CoA-488-labeled *mt*Dod-*mm*ACP and *mt*Dod-*mm*ACP-H8 with CoA-488-labeled free *mm*ACP, the relative degree of labelling was determined to about 31% ± 8% and 36% ± 8% respectively. After an additional hour of labelling, in-gel fluorescence of *mt*Dod-*mm*ACP and of *mt*Dod-*mm*ACP-H8 further increased by 14% ± 8% and 24% ± 12%, respectively. The overall low relative degree of labelling and the increase after an additional hour of reaction time indicates a reduced accessibility of *mm*ACP fused to *mt*Dod. It cannot be ruled out that the *mm*ACP fold fused to *mt*Dod is instable or partly unfolded. Note that in SDS-PAGE, *mt*Dod-*mm*ACP runs at just two different apparent molecular weights corresponding to labelled and non-labelled protein (see Fig. 6). It seems that SDS-PAGE is limited in its efficiency of separating mixtures of unlabelled and labelled monomers.

_Mt_Dod-PAS-pep constructs for AB production. Protein carriers are generally used for the production of ABs against peptides or proteins[9]. In the standard approach, the peptide or the protein of interest is linked to the carrier, usually BSA or KLH, by chemical ligation[8–10]. While the method is well-established and broadly used for AB production, problems can arise during conjugating the peptide/hapten to the carrier, e.g., owing to the low stability or solubility of the conjugate (or even for the peptide alone) or altered antigenic properties of the peptide[17]. The dodecameric structure with the exposed termini allows *mt*Dod to be charged with 12 or 24 peptides/proteins on its surface by simply fusing the peptide/protein encoding sequence to the *mt*Dod gene. In order to evaluate the suitability of *mt*Dod for AB production, 11 fusion constructs were produced in *E. coli* of which each is comprised of *mt*Dod, a PAS linker and a peptide of interest, termed *mt*Dod-PAS-Pep (Table 2). Peptide sequences originated from human heat shock proteins (HSP), proheparin-binding EGF-like growth factor (HB-EGF) and C-terminus of the heat shock cognate protein 70 interacting protein (CHIP) (for detailed peptide origin see Supplementary Table S2). Peptides/epitopes were selected, because of a specific scientific interest in

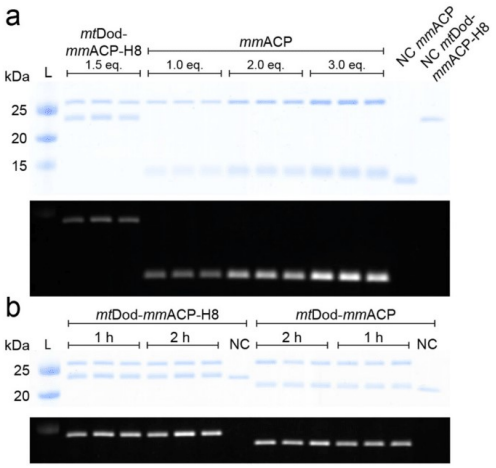


Figure 6. Modification of *mt*Dod-*mm*ACP, *mt*Dod-*mm*ACP-H8 and *mm*ACP by Sfp with fluorescent CoA. L: ladder. NC: negative control reaction without Sfp. Eq.: molar equivalents of *mm*ACP loaded onto the SDS-PAGE gel. (**a**) Top: Coomassie stained SDS-PAGE gel of the reaction solution and negative controls. Sfp, runs at an apparent molecular weight of slightly above 25 kDa, *mt*Dod-*mm*ACP-H8 slightly below 25 kDa and *mm*ACP slightly below 15 kDa. Unmodified *mm*ACP and *mt*Dod-*mm*ACP-H8 (negative controls) show lower apparent molecular weights indicating a different running behaviour between CoA-488 modified and unmodified proteins. Bottom: In-gel fluorescence taken before Coomassie staining. Only proteins modified with the CoA-488 are visible. Higher bands represent labelled *mt*Dod-*mm*ACP-H8 and lower bands labelled *mm*ACP. (**b**) Top: Coomassie stained SDS-PAGE gel of the reaction solutions of *mt*Dod-*mm*ACP-H8 and *mt*Dod-*mm*ACP after 1 h and 2 h. Bottom: in-gel fluorescence taken before Coomassie staining. For both constructs, fluorescence intensity increases at longer reaction times. For uncropped images and *mt*Dod-*mm*ACP blots see Supplementary Fig. S9.

*mt*Dod constructs	Peptide sequence	Calculated mass without start-Met/Da	Measured mass by MS (+1 H^+)/Da
*mt*Dod-PAS-Pep1	PKGGSGSGPTIEEVD	10,155	10,156.7
*mt*Dod-PAS-Pep2	PLEGDDDTSRMEEVD	10,434	10,435.0
*mt*Dod-PAS-Pep3	ECYPNEKNSVNMDLD	10,497	10,803.4*
*mt*Dod-PAS-Pep4	VPSDSDKKLPEMDID	10,415	10,416.1
*mt*Dod-PAS-Pep5	DSSQHTKSSGEMEVD	10,363	10,363.9
*mt*Dod-PAS-Pep6	EQSTGQKRPLKNDEL	10,469	10,470.2
*mt*Dod-PAS-Pep7**	ALMVYRCAPPRSSQF	10,453	–
*mt*Dod-PAS-Pep8	LVTGESLEQLRRGLA	10,368	10,369.1
*mt*Dod-PAS-Pep9	MKGKEEKEGGARLGA	10,287	10,288.0
*mt*Dod-PAS-Pep10	EERRIHQESE	10,038	10,039.6
*mt*Dod-PAS-Pep11	NHEGDEDDSH	9,880	9,881.3
*mt*Dod-PAS-H7	HHHHHHH	9,704	9,705.2

Table 2. *Mt*Dod constructs for AB production. After purification by the heat treatment protocol, the correct size of the proteins was verified by ESI–MS. *Difference of mass is about 305 Da and could be caused by S-glutathionylation[37,38], No mass for the unmodified *mt*Dod-PAS-Pep3 was observed. ***Mt*Dod-PAS-Pep7 formed inclusion bodies and was not purified.

the proteins carrying the peptides, and not for their sequence composition or the thermal stability properties of the source proteins (e.g. thermal stability of HSP). In that light, the case study presented here is also a "real case" for the applicability of the dodecin matrix.

Pep-encoding sequences, provided on oligonucleotide primers, were introduced in single-step by ligation-free cloning. Recombinant expressions and purifications followed the established protocols described above. All constructs were received as soluble proteins, except *mt*Dod-PAS-Pep7 that formed inclusion bodies (see Table 2).

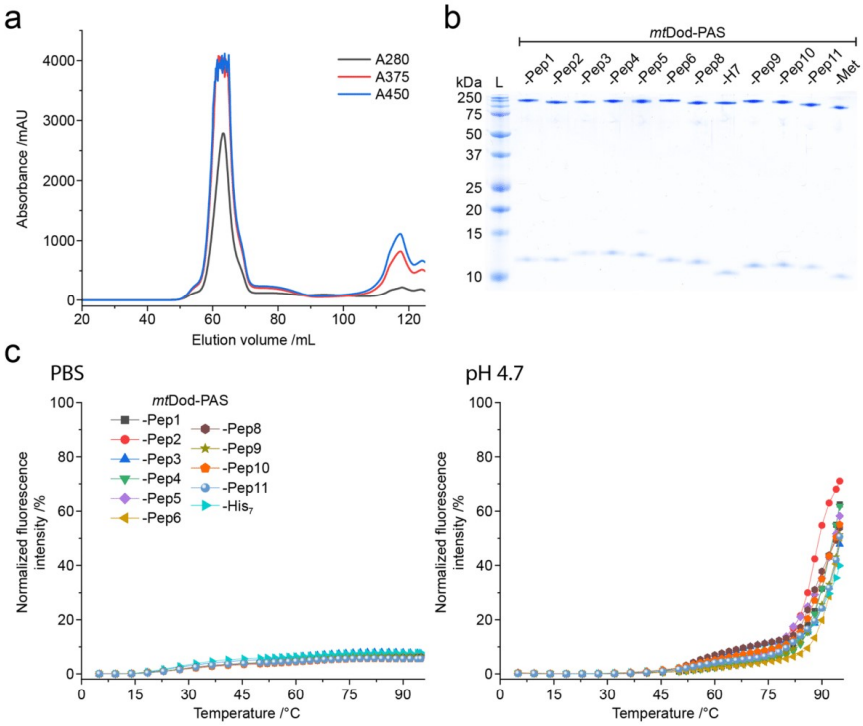


Figure 7. *Mt*Dod-PAS-Pep purity and stability. (**a**) SEC profile of FMN:*mt*Dod-PAS-Pep1. The dodecameric species can be identified by the FMN absorption at 375 nm and 450 nm. In addition, unbound FMN is visible at ~ 117 mL. All constructs, except *mt*Dod-PAS-Pep3, showed comparable chromatographic profiles. (**b**) SDS-PAGE gel of all purified *mt*Dod-PAS-Pep constructs. Use of standard loading buffer (pH 6.8, 2.5% SDS) at prolonged heat incubation (30 min, 95 °C) allows the observation of monomer and dodecamer. (**c**) Thermocyclic fluorescence assay of *mt*Dod-PAS-Pep constructs. Measurements were performed in PBS and at pH 4.7 as described above (see Fig. 4).

The yellow colour of the inclusion bodies indicated assembled dodecamer, and we assume that aggregation of *mt*Dod-PAS-Pep7 was induced by the cysteine in Pep7, forming disulfide-bridges between the dodecamers. All constructs, except *mt*Dod-PAS-Pep7, were further purified by two cycles of DMSO-induced precipitations. FMN was added before constructs were eventually forwarded to SEC to remove unbound FMN and remaining DMSO as well as to select for dodecameric species (Fig. 7a). We assumed that *mt*Dod remains saturated with FMN due to its high affinity and cooperative binding mode.[23] FMN-saturated *mt*Dod constructs (FMN:*mt*Dod constructs) can be determined in concentration by absorbance at 450 nm, and are amenable to stability measurements by the thermocyclic fluorescence assay. All constructs were received as dodecamers, as indicated by SEC (Supplementary Fig. S10). *Mt*Dod-PAS-Pep3 shows in addition to the dodecamer species higher oligomeric states in SEC, which we assume to result from disulfide-bridges formed by the cysteine in Pep3. The dodecamer containing fractions were pooled, and the purity was controlled by SDS-PAGE (see Fig. 7b). The thermocyclic fluorescence assay revealed the high thermal stability of all *mt*Dod-PAS-Pep constructs, similar as the wild type protein (Fig. 7c)[23]. Molecular masses of the constructs were measured with ESI–MS and confirmed full-length protein (see Table 2).

The amount of purified protein/dodecamer of *mt*Dod-PAS-Pep constructs were between 20–50 mg, purified from about 100 mL of the 500 mL expression culture (calculated yields of 200–500 mg per litre). The construct *mt*Dod-PAS-Pep3 was received at the lowest amount (about 20 mg; 200 mg/litre), which may be explained by agglomeration of dodecamers due to disulfide bridges. Constructs *mt*Dod-PAS-Pep3 and *mt*Dod-PAS-Pep7 indicate that cysteine containing peptides can cause problems when processed via the described purification strategy. A changed protocol, in which the oxidative conditions imposed by the high concentrations of FMN are avoided, could lead to improved results.

Endotoxin concentrations, measured in endotoxin units (EU) via a *Limulus* amebocyte lysate (LAL) test, were determined to avoid an endotoxin shock in immunizations. *Mt*Dod-PAS-Pep3 and Dod-PAS-Pep6 contained

ABs derived from *mt*Dod construct	Binding region of the ABs	AB class
*mt*Dod-PAS-Pep1	HSP-70 C-terminus	Class 1
*mt*Dod-PAS-Pep2	HSP-90 C-terminus	Class 1
*mt*Dod-PAS-Pep3	HSP-110 C-terminus	Class 1
*mt*Dod-PAS-Pep4	HSP-A4 C-terminus	Class 1
*mt*Dod-PAS-Pep5	H3 C-terminus	Class 2*
*mt*Dod-PAS-Pep6	H4 C-terminus	Class 3
*mt*Dod-PAS-Pep8	HB-EGF** Pos. 20–34	Class 3
*mt*Dod-PAS-Pep9	CHIP N-terminus	Class 1
*mt*Dod-PAS-Pep10	CHIP Broken helix Pos. 151–161	Class 2***
*mt*Dod-PAS-Pep11	CHIP Tip of helix Pos. 183–192	Class 1

Table 3. Classification of *mt*Dod-PAS-Pep derived ABs. ABs were classified as followed: "Class 1" ABs recognize the proteins of interest provided as recombinantly purified protein (produced in *E. coli*). "Class 2" ABs do not recognize recombinant protein of interest, but protein of expected apparent molecular weight in HEK293T human cell lysates. "Class 3" ABs only recognize *mt*Dod-PAS constructs (like the respective *mt*Dod-PAS-Pep or *mt*Dod-PAS-Met). Western blots of ABs rated as "class 1" or "class 2" are shown in Fig. 8. The AB derived from *mt*Dod-PAS-Pep6 and *mt*Dod-PAS-Pep8 only recognized *mt*Dod-PAS-Pep constructs and were considered "class 3" (Supplementary Fig. S11). **mt*Dod-PAS-Pep5 derived ABs showed only a very weak signal for 1 µg and 500 ng of recombinant protein with no intensity difference (see Supplementary Fig. S11). Thus, the AB preparation was regarded as "class 2". **Proheparin-binding EGF-like growth factor of *Chlorocebus aethiops* (green monkey). ****mt*Dod-PAS-Pep10 derived ABs didn't recognize purified CHIP but seem to recognize a protein in CHIP-overexpressing cells; no detection range was determined.

the highest amount of endotoxin with 73 EU/mg and 55 EU/mg, respectively; all other samples showed values less than 30 EU/mg (average of all constructs 30 ± 23 EU/mg). Since about 100 µg protein was used per injection, none of the samples were critical in endotoxin levels (above 5–10 EU/kg of rabbit per injection)[39,40].

Purified *mt*Dod constructs were eventually submitted to an AB production company for immunization in rabbits and AB purification (Davids Biotechnologie GmbH, Germany). AB productions were induced in one rabbit for each construct by 5 injections (about 100 µg each, *mt*Dod-PAS-Pep solution concentration: 2.2–7.5 mg/mL (average 4.5 ± 1.3 mg)) over 63 days using the adjuvant MF59/AddaVax. The ABs were purified from the collected serum by affinity chromatography with the respective *mt*Dod-PAS-Pep construct immobilized on the column matrix. For all the 10 *mt*Dod-PAS-Pep constructs, which were submitted to immunizations, purified ABs were obtained and their binding behaviour analysed by western blotting (summarized in Table 3).

The produced ABs did not create any reoccurring background signals indicating that the *mt*Dod-PAS carrier matrix does not cause the generation of ABs that recognize proteins present in lysate of HEK293T cells (see Fig. 8a, Supplementary Fig. S11 and Supplementary Fig. S12). This agrees with the low sequence identity of *mt*Dod-PAS with human proteins (protein–protein BLAST with default settings finds no human protein with significant similarity).

The six ABs rated as "class 1" were tested on different concentrations of target protein (1 µg to 60 ng) to estimate their general labelling capability in western blots (see Fig. 8b). Further, the same dilutions of target protein were used to give a rough comparison for *mt*Dod-PAS-Pep derived ABs with commercially available ABs (exception *mt*Dod-PAS-Pep2, for which we had no commercial AB to hand). In general, the *mt*Dod-PAS-Pep derived ABs were comparable with commercial ABs (see Fig. 8b), and are well suited for specific labelling of target protein in lysate samples.

The *mt*Dod-PAS-Pep-derived ABs show that *mt*Dod-PAS is a well-suited carrier system for the production of peptide-specific ABs. *Mt*Dod-PAS-based fusions benefit from the easy cloning, uncomplicated production/purification and the high protein yields. In our case study, problems in the purification of cysteine containing constructs emerged from the oxidative conditions induced by high FMN concentrations, which may, however, be overcome when working under reducing conditions. Alternatively, cysteine could be replaced by serine residues, as recommended by AB producing companies for epitopes containing internal cysteines, although this might alter the epitopes and affect the specificity of the produced ABs.

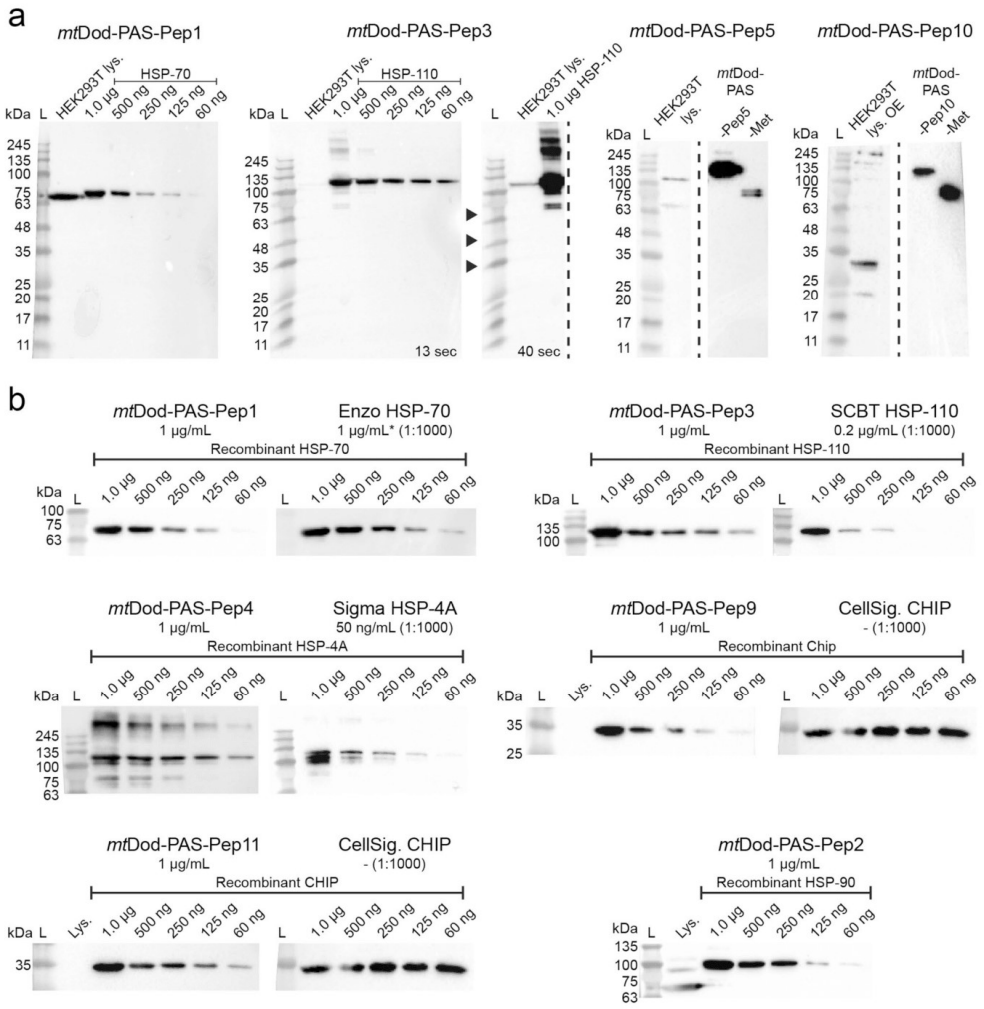


Figure 8. Western blots with selected *mt*Dod-PAS-Pep construct ABs. L: Ladder. Lys.: Lysate. OE: protein overexpressing cells. (**a**) Detection of the target protein in purified form and in the lysate. ABs derived from *mt*Dod-PAS-Pep1 and *mt*Dod-PAS-Pep3 detect target proteins in the lysate. For *mt*Dod-PAS-Pep3, the same blot is shown with two exposure times (13 s and 40 s) to make the potential HSP-110 representing band in the lysate observable. ABs derived from *mt*Dod-PAS-Pep5 and *mt*Dod-PAS-Pep10 did not recognize purified target protein, but seem to recognize a protein in overexpressing cells (uncropped western blots see Supplementary Fig. S11) (**b**) Comparison with commercially available ABs. Different amounts (1 µg to 60 ng) of purified target protein were loaded and analysed by *mt*Dod-PAS-Pep derived ABs (1 µg/mL) and commercially available ABs (recommended dilutions were used; concentration given when available). All used commercial ABs should recognize different epitopes than the *mt*Dod-PAS-Pep derived ABs, since the given epitope amino acid regions are not overlapping with the selected peptides. Enzo HSP-70: Monoclonal anti HSP70 AB (epitope region amino acids 436–503) from Enzo Life Science, GmbH. *: Concentration assumed based on older aliquots, since information is not given for the newer ones. SCBT HSP-110: Monoclonal anti HSP-105/HSP-110 AB (epitope region amino acids 187–512) from Santa Cruz Biotechnology, Inc. Sigma HSP-4A: Polyclonal anti HSP-4A AB (epitope region amino acids 639–748) from Sigma-Aldrich, Merck KGaA. CellSig. CHIP: Monoclonal anti CHIP AB (epitope region L36 surrounding amino acids (synthetic peptide, length not given)) from Cell Signaling Technology, Inc. For uncropped western blots, see Supplementary Fig. S12. ABs derived from *mt*Dod-PAS-Pep2 also recognize HSP-70. For a comparison of *mt*Dod-PAS-Pep1-3 derived ABs, see Supplementary Fig. S13.

Conclusion

During the last years, dodecins have been characterized as flavin binding proteins involved in flavin homeostasis[21,23]. In addition, the unique protein fold and particularly the exceptional flavin binding mode were harnessed in technological applications, although exclusively on the archaeal protein from *Halobacterium salinarum*[41,42]. In this study, we present *mt*Dod as a versatile scaffold protein to attach peptides and small proteins.

The *mt*Dod dodecamer tolerates high temperatures and various chemical conditions, which allows protein purification by quick heat-induced denaturation and protein precipitation with solvents, and holds out the prospect that *mt*Dod broadly accept conditions for chemical ligation reactions. In addition to its high stability, *mt*Dod can be produced in high amounts in *E. coli*. For soluble constructs, yields of up to several hundred milligrams of *mt*Dod-peptide fusions per litre of bacterial culture can be expected. Proteins fused to *mt*Dod are presented at the *mt*Dod outer surface, and have been shown to remain accessible and functional. Both the SpyT/-C and the ACP/Sfp system allowed attaching the cargo at the *mt*Dod surface. In this respect, *mt*Dod is comparable to the recently presented IMX313 scaffold, suggested for use in vaccine development[19]. When evaluating *mt*Dod as a scaffold, we observed that constructs can suffer from low solubility in response to the properties of the attached cargo. While agglomeration by disulfide formation, observed for *mt*Dod-PAS-Pep3 and *mt*Dod-PAS-Pep7, could simply be avoided by reducing conditions during protein preparation or replacing of cysteine with serine residues, solubility problems induced by hydrophobic and structurally unstable fold may be solved by heterododecamer formation to dilute the aggregation-inducing species on the *mt*Dod surface. For the construct *mt*Dod-PAS-Strep, we demonstrated that heterododecamer formation with the wild type protein is readily possible in vitro and in vivo by simply providing both proteins during refolding or recombinant protein production, respectively (see Fig. 3).

As a pilot run for evaluating the suitability of *mt*Dod as a carrier matrix for AB production, we chose 11 peptides originating from different human proteins like CHIP or HSP-70, and fused them to *mt*Dod. One of the 11 peptide constructs formed inclusion bodies, while all other constructs were purified by the standard heat-denaturation purification protocol without any need for individual optimization. From all immunizations performed in this study, ABs were received that at least recognized the *mt*Dod-PAS scaffold in western blotting. Overall, 8 of the 10 ABs recognized proteins in HEK293T human cell lysate at expected molecular weight. For 6 of them, correct target recognition could be confirmed with the recombinantly purified protein as reference. No AB preparation showed any unspecific reactivity in HEK293T cell lysate, demonstrating that *mt*Dod is a suitable matrix for the production of ABs that specifically label the proteins of interest (in the HEK293T lysate) without the need to remove anti-carrier ABs. The low sequence identity of *mt*Dod to human proteins and eukaryotic proteins in general suggests that ABs derived from *mt*Dod will also be specific to proteins of interest in other human samples/cells[20,24].

The here presented AB production strategy with *mt*Dod may be attractive for labs that are experienced in protein expression, and want to produce ABs targeting peptides without relying on peptide synthesis and chemical crosslinking. We expect that the exposed termini are also suited for chemical ligation of haptens or antigens, following standard immunization protocols, or for Click chemical modification[43]. However, in proofing the concept of dodecin for peptide immunizations, we did not elaborate on this further. Finally, we note that the availability of dodecins with similar features (e.g. *Streptomyces coelicolor*, *Streptomyces davaonensis* and *Thermus thermophilus* dodecins)[22,44] is advantageous when aiming for heterologous prime/boost protocols by using two dodecin scaffolds with low sequence identity fused with the same antigen[45].

While the *mt*Dod has been mainly tested as carrier matrix for AB production in this study, the properties of *mt*Dod call for its application as a scaffold in a broad range of biotechnological and bioengineering applications. We encourage to explore the *mt*Dod as a scaffold when defined particles with specific surface properties are required. Such constructs can be valuable in e.g. diffusion measurements,[46] for formation of biomaterials[47–50] and in creating enzyme scaffolds[35,51,52]. *Mt*Dod heterododecamers may be applied for pull down assays when combining a *mt*Dod construct bearing a protein recruiting peptide and a *mt*Dod construct with a purification tag.

Material and methods

Cloning. Expression constructs were cloned using standard PCR methods and In-Fusion HD Cloning (TaKaRa Bio Europe). Primers were ordered from Sigma-Aldrich. Inserts were verified by Sanger sequencing (by Microsynth Seqlab, Göttingen, Germany). For polycistronic constructs spacer DNA sequences (between genes) were designed with EGNAS (version 1,158)[53]. For a list of all constructs see Supplementary Table S1.

Expression and cell lysis. Plasmids were transformed into BL21 (DE3) Gold cells and cells were plated onto LB-agar plates containing 100 ng/µL ampicillin and 1 g/mL glucose. 10 mL LB medium with 100 ng/µL ampicillin and 1 g/mL glucose were inoculated with a single colony and incubated at 37 °C and 180 rpm overnight. 1 L TB medium with 100 ng/µL ampicillin was inoculated with 10 mL overnight LB culture, and incubated at 37 °C and 180 rpm until the OD_{600} reached about 0.8. The cultures were cooled to 20–30 °C, and the expression was induced with 1 mL 1 M IPTG solution. The cultures were incubated overnight at 20 °C 160 rpm for protein production. *Mt*Dod-PAS-Pep constructs were expressed in 500 mL TB medium induced with 500 µL 1 M IPTG. Cells were harvested at 4,000 rcf and frozen in liquid nitrogen or directly processed. For purification by heat denaturation (*mt*Dod constructs), cell pellets were resuspended in 30 mL standard dodecin buffer: 300 mM NaCl, 5 mM $MgCl_2$ and 20 mM Tris–HCl (pH 7.4, adjusted with HCl). For purification by His-tag affinity chromatography, cell pellets were resuspended in 30 mL Ni–NTA wash buffer I: 200 mM NaCl, 35 mM K_2HPO_4 and 15 mM KH_2PO_4 (pH 7.4, adjusted with NaOH or HCl) and 40 mM imidazole. To the resuspended cells, PMSF and DNase I were added, and cells were disrupted by French press. Cell debris was removed by centrifugation (50,000 rcf, 20 min). All steps after cell harvest were conducted at 4 °C or on ice.

Purification by heat denaturation. The cell debris free lysates (about 30 mL) were divided in about 10 mL aliquots and incubated at 75 °C for 15 min. Lysate containing *mt*Dod-*mm*ACP or *mt*Dod-msfGFP-H8 was incubated at 55 °C or 70 °C, respectively. Heat-denatured proteins were removed by centrifugation (15,000 rcf, 10 min). Proteins in the supernatant (combined from 3 aliquots 20–25 mL) were precipitated with 50% (v/v) DMSO (final concentration) and pelleted by centrifugation (15,000 rcf, 10 min). Depending on the construct, higher concentrations or other organic solvents (MeOH and acetone) might be needed (precipitation with 75% (v/v) acetone (final concentration) turned out to be fastest and most reliable but might cause aggregation). The obtained protein pellets were dissolved in about 20 mL standard dodecin buffer at RT and afterwards cooled on ice. Precipitation was repeated once, and the pellets were dissolved in about 5 mL. Insoluble precipitate was removed by centrifugation (15,000 rcf, 10 min). To the protein solutions, FMN was added in excess (above its solubility limit) (F6750, Sigma-Aldrich: 70% pure, free RbF ≤ 6%) and samples were incubated on ice for at least 1 h. About 250–500 µL of protein solution were further purified by SEC. For loading the column, the concentration of the protein solution was judged by the yellow tone of the solution (depending on the dodecin concentration FMN solution turn from yellow to orange-brown).

Purification by His-tag affinity chromatography. The cleared lysate was poured on 5 mL packed Ni–NTA agarose (His60 Superflow, TaKaRa Bio Europe) gravity flow columns, pre-equilibrated with Ni–NTA wash buffer. The loaded resin was washed with 15 mL Ni–NTA wash buffer I and with 15 mL Ni–NTA wash buffer II (Ni–NTA wash buffer I with 80 mM imidazole). *Mt*Dod-msfGFP-H8 was additionally washed with 10 mL Ni–NTA wash buffer III (Ni–NTA wash buffer I with 200 mM imidazole). Proteins were eluted in 15 mL elution buffer (200 mM NaCl, 35 mM K_2HPO_4 and 15 mM KH_2PO_4 (pH 7.4, adjusted with NaOH or HCl) and 500 mM imidazole). Eluted proteins were concentrated with ultra centrifugal filters (Amicon, Merck) with the appropriated mass cut-off and further purified by SEC.

Refolding of *mt*Dod-*se*ACP, *mt*Dod SpyC and *mt*Dod SYNZIP constructs. Yellowish inclusion bodies obtained after cell disruption and centrifugation (50,000 rcf, 30 min) were manually separated from other solid cell debris and then washed three times (resuspended and centrifuged) with inclusion body wash buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 0.5 mM KH_2PO_4 (PBS, pH 7.4 not adjusted), 5 mM EDTA and 2% (v/v) Triton X-100). The washed inclusion bodies were then dissolved in 10 mL GdmCL buffer (6 M guanidinium hydrochloride, 20 mM Tris–HCl (pH 8.0, adjusted with HCl)). For refolding, 0.5 mL protein solution were diluted with 4.5 mL GdmCL buffer containing L-Arginine (1 M final concentration). Refolding was performed by dialyzing twice against the 100-fold volume of the respective buffer containing 1 mM FMN (omitted for *mt*Dod SYNZIP constructs). For SpyC constructs, a phosphate borate buffer (100 mM NaCl 25 mM Na_2HPO_4, 25 mM H_3BO_3 (pH 8.5 adjusted with NaOH) and 20% glycerol) was used, and for all other constructs the standard dodecin buffer was used. Dialysis was conducted at 4 °C and without stirring for the first ~ 12 h (overnight). Aggregated protein was removed from the solution by centrifugation (3,000 rcf, 10 min). Refolded proteins were concentrated with ultra centrifugal filters (Amicon, Merck) with the appropriate mass cut-off and further purified by SEC.

SEC. Prior to injection, all samples were filtered with 0.22 µm membrane filters (Durapore, Merck). All proteins in this study, except *mt*Dod-PAS-Pep constructs, were purified by using an equilibrated Superdex 200 increase 10/300 column (GE Healthcare) on an ÄKTA Explorer or ÄKTA Basic device. For *mt*Dod-PAS-Pep constructs, a HiLoad Superdex 200 16/600 pg was used. Running buffer for all constructs, except *mt*Dod SpyC constructs and *mt*Dod-msfGFP-H8, was the standard dodecin buffer. The flow rate was 0.5 mL/min for the Superdex 200 increase 10/300 column or 1.0 mL/min for the HiLoad Superdex 200 16/600 pg column with fraction resolution of 0.3 mL or 2.0 mL, respectively. The running buffer for the *mt*Dod SpyC constructs and *mt*Dod-msfGFP-H8 was phosphate borate buffer (as used for refolding), and the flow rate was reduced to 0.45 mL. All runs were conducted at 4 °C. Fractions were pooled and analysed by SDS-PAGE with Coomassie staining. Pooled fractions were aliquoted, frozen in liquid nitrogen and stored at − 80 °C.

Protein concentrations. The concentration of FMN saturated *mt*Dod constructs was determined by the FMN absorption and the corresponding extinction coefficients (375 nm, 450 nm, 473 nm respective extinction coefficients 10,000 $M^{-1} \times cm^{-1}$, 12,000 $M^{-1} \times cm^{-1}$ and 9,200 $M^{-1} \times cm^{-1}$). The concentration of fluorescent proteins were determined by the chromophore absorption and the corresponding extinction coefficient (mClover3 constructs: 506 nm, 109,000 $M^{-1} \times cm^{-1}$ (ref[54]), mRuby3 constructs: 558 nm, 128,000 $M^{-1} \times cm^{-1}$ (ref[54]), msfGFP constructs: 485 nm 82,400 $M^{-1} \times cm^{-1}$ (ref[29])). For other proteins, the concentrations were determined by the absorption at 280 nm and applying the calculated extinction coefficient. For proteins with low amounts of bound flavin, the absorption at 280 nm was corrected using the absorptions at 450 nm and the 280/450 nm ratio of pure FMN.

SDS-PAGE. SDS-PAGE was performed at initial 70 V (15 min) followed by 200 V (about 70 min) on Tris-tricine gels (self-casted 10% gels, as described in ref[55] using an Mini-PROTEAN Tetra Cell system (Bio-Rad Laboratories, Inc.). For heterododecamer samples Tris-tricine (Laemmli) gels (self-casted 12% gels) and SDS free loading buffer (4×: 50% glycerol, 5 mM FMN) were used. Laemmli SDS-PAGE was performed at initial 70 V (15 min) followed by 200 V (about 150 min). For non-*mt*Dod construct samples or if full dodecamer denaturation was not necessary/wanted, standard SDS-PAGE loading buffer (4×: 50% (v/v) glycerol, 375 mM Tris–HCl (pH 6.8, adjusted with HCl), 10% (w/v) SDS, 10% (v/v) 2-mercaptoethanol, 50 mM EDTA, bromophenol

blue) was used. For full denaturation of the *mt*Dod dodecamer, a 2-component acidic SDS-PAGE loading buffer (4 × acidic part 1: 10% (w/v) SDS, 300 mM acetic acid; 4 × acidic part 2: 50% (v/v) glycerol, 300 mM Tris (unbuffered), 200 mM Tris–HCl (pH 6.8, adjusted with HCl), 10% (v/v) 2-mercaptoethanol, 50 mM EDTA, bromophenol blue) was used. Amount of acetic acid and unadjusted Tris can be varied together if needed. Samples were mixed with the standard SDS-PAGE loading buffer or acidic SDS-PAGE loading buffer component 1 and heat treated at 95 °C (10 min). After heat treatment, acidic SDS-PAGE loading buffer component 2 was added to the acidic samples and a second heat treatment was applied. Gels were stained over night with InstantBlue Coomassie stain (Expedeon) and imaged using a scanner (Epson Expression 1,680 Pro, Seiko Epson Corporation).

Thermocyclic fluorescence assay. For stability measurements, a clear 96 well PCR plate (MLL9601; Bio-Rad Laboratories, Inc.) was prefilled with 23 µL of respective buffer per well; i.e., PBS or acetate buffer (150 mM NaCl, 100 mM acetic acid (pH 4.2, adjusted with NaOH)). Plates were placed on ice and 2 µL of the corresponding 50 mM *mt*Dod construct solution were added to the wells. Plates were then sealed with optical tape (iCycler iQ; Bio-Rad Laboratories, Inc.), centrifuged (3,000 rcf, 2 min) and placed into a precooled (5 °C) real-time PCR instrument (C1000 Thermal Cycler and CFX96 Real-Time System; Bio-Rad Laboratories, Inc.). For the fluorescence detection, excitation/emission filter bandwidth of 450–490/560–580 nm was used. After 1 h incubation at 5 °C, the heating and cooling cycles were started, with each cycle containing a heating phase for 6 min and a cooling (5 °C) phase for 30 min. The heating phase temperature was raised stepwise from 5 °C to 95 °C. Until 50 °C, the step size was 4.5 °C while at higher temperatures the step size was reduced to 2.0 °C. Data points were taken after each phase. The complete temperature protocol was applied to every sample. For the stability measurements of *mt*Dod-PAS-Pep constructs, the acetate buffer was replaced with a MES buffer (150 mM NaCl, 100 mM MES (pH 4.7, adjusted with HCl) and the heating phase duration was prolonged to 10 min.

Modification of *mm*ACP, *mt*Dod-*mm*ACP and *mt*Dod-*mm*ACP-H8 with Sfp. Sfp used for phosphopantetheinylation was produced in *E. coli* (Strain M15; helper plasmid pREP4 and pQE-60 encoding Sfp) and purified by His-tag affinity chromatography, dialysis overnight against imidazole free buffer (250 mM NaCl, 2 mM $MgCl_2$, 10% glycerol, 50 mM HEPES (pH 8.0 adjusted with HCl)) and SEC (HiLoad Superdex 200 16/60 pg)[56]. Holo *mm*ACP was produced in *E. coli* and purified by His-tag affinity chromatography (basic buffer: 200 mM NaCl, 35 mM K_2HPO_4 and 15 mM KH_2PO_4 (pH 7.4, adjusted with NaOH or HCl), 10% glycerol; washing buffer contained 20 mM imidazole and elution buffer contained 300 mM imidazole) and SEC (HiLoad Superdex 75 16/60 pg), similar as described for ACP-GFP by Heil and Rittner et al.[57] Pooled SEC fractions of Sfp and *mm*ACP were aliquoted, frozen in liquid nitrogen and stored at − 80 °C. The modification of *mm*ACP and *mt*Dod-*mm*ACP (both variants with H8-Tag and without) with CoA 488 (CoA modified with ATTO-TEC dye ATTO 488, NEB #S9348) by Sfp was conducted in phosphate borate buffer (pH 7.4, adjusted with HCl) at 25 °C for 1 h at following conditions: 10 µM *mt*Dod-*mm*ACP or *mm*ACP, 4 µM Sfp, 1 mM DTT, 10 mM $MgCl_2$ and 10 µM CoA 488. The final reaction volume was 50 µL, and reactions were carried out in triplicates. For calibration of the fluorescent signals, 3 different amounts of each *mm*ACP-H8 reaction solution (1 µL, 2 µL, 3 µL) were analysed.1.5 µL of the corresponding *mt*Dod-*mm*ACP construct reaction solution were used for the determination of the degree of labelling. The reaction mixtures were by analysed by SDS-PAGE, in-gel fluorescence, using the Fusion SL2 Xpress imaging system (Vilber Lourmat GmbH) with the emission filter F-595 Y3 (520–680 nm), and Coomassie staining. For quantification of fluorescence signals, the Fusion-Capt Advance SL2 Xpress 16.08a software (Vilber Lourmat GmbH) was used. The degree of modification of *mt*Dod-*mm*ACP was determined by linear regression of the normalized (by molar amount) fluorescence signals.

SpyC and SnpC reactions. The reactions were carried out in phosphate borate buffer (pH 8.5) at 25 °C for 20 h. Concentration of the respective carrier constructs were 10 µM (1 eq.) and 20 µM of the respective cargo constructs (2 eq.), reaction volume was 50 µL. Each protein also was separately prepared in the same concentration as used in the reaction and incubated under the same conditions. 5 µL of each reaction and control were analysed by SDS-PAGE and Coomassie staining, for all samples the acidic loading buffer was used.

LC–MS. For the LC–MS analysis of the *mt*Dod-PAS-Pep constructs, 500 µg protein were precipitated with 75% (v/v) acetone (− 20 °C, final concentration), pelleted by centrifugation (20,000 rcf, 5 min) and dissolved in 100 µL water. After removal of undissolved aggregates by centrifugation (20,000 rcf, 5 min), the solution was diluted with 5% (v/v) acetonitrile:water to a final concentration of 0.1 mg/mL. The injection/sample size was 2.5 µL (250 ng). Samples were analysed by using a Dionex UltiMate 3,000 RSLC (Thermo Fischer Scientific) coupled to a micrOTOF-Q II (Bruker Daltonik GmbH) equipped with an electrospray ionization source. Chromatographic separation (further desalting) was performed on a Discovery BIO Wide Pore C5 column (100 × 2.1 mm, particle size 3 µm, Sigma-Aldrich) at 55 °C with a mobile-phase system consisting of water and acetonitrile (each containing 0.1% formic acid). A linear gradient ranging from 5 to 95% acetonitrile over 14 min at a flow rate of 0.4 mL min^{-1} was used. MS data was acquired in positive mode in the range from 200–2,500 m/z and later analysed using Compass DataAnalysis 4.0 software (Bruker Daltonik GmbH).

Western blots. Samples separated by SDS-PAGE were transferred to 0.45 µm nitrocellulose membranes at 100 V for 35–40 min in a cooled Criterion Blotter filled with transfer buffer at 4 °C (transfer buffer: 25 mM Tris, 192 mM glycine, 20% (v/v) methanol). Membranes were blocked in 5% (w/v) milk powder in TBS (50 mM Tris–HCl pH 7.6, 150 mM NaCl) with 0.1% Tween-20 added (TBST) for 1 h at RT, before addition of primary AB in indicated dilutions and incubation over night at 4 °C. Used primary ABs: Enzo HSP-70: #ADI-SPA-810 Enzo Life Science GmbH; SCBT HSP-110: #sc-74550 Santa Cruz Biotechnology, Inc.; Sigma HSP-4A: #HPA010023

Sigma-Aldrich, Merck KGaA; CellSig. CHIP: #2080 Cell Signaling Technology, Inc. Membranes were washed twice with TBST, followed by addition of secondary AB (#A4416 or #A9169 Sigma-Aldrich, Merck KGaA) in a 1:3,000 dilution in blocking buffer for 1 h at RT. Finally, membranes were washed twice with TBST. Chemiluminescence was developed with SuperSignal West Pico PLUS, and images were acquired with a ChemiDoc MP imaging system and quantified using Image Lab 5.0 software (Bio-Rad Laboratories, Inc.). Recombinant proteins used for AB comparison were produced in *E. coli* and purified by affinity chromatography and SEC. Proteins HSP-70, HSP-90, HSP-110 and HSP-A4 were expressed based on pPROEX vector system with a N-terminal His-tag that was removed by a TEV protease after affinity chromatography. CHIP was expressed based on a pGEX-6P1 vector system with a N-terminal GST-tag that was removed by the PreScission protease after affinity chromatography. All ABs were stored at − 20 °C prior use.

Received: 19 March 2020; Accepted: 13 July 2020
Published online: 06 August 2020

References

1. Tolar, P., Hanna, J., Krueger, P. D. & Pierce, S. K. The constant region of the membrane immunoglobulin mediates B cell-receptor clustering and signaling in response to membrane antigens. *Immunity* **30**, 44–55 (2009).
2. Bachmann, M. F. & Zinkernagel, R. M. Neutralizing antiviral B cell responses. *Annu. Rev. Immunol.* **15**, 235–270 (1997).
3. Chauhan, V., Rungta, T., Goyal, K. & Singh, M. P. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Sci. Rep.* **9** (2019).
4. Schubert, B. & Kohlbacher, O. Designing string-of-beads vaccines with optimal spacers. *Genome Med.* **8** (2016).
5. Bennett, N. R., Zwick, D. B., Courtney, A. H. & Kiessling, L. L. Multivalent antigens for promoting B and T cell activation. *ACS Chem. Biol.* **10**, 1817–1824 (2015).
6. Mariani, M. *et al.* Immunogenicity of a free synthetic peptide: Carrier-conjugation enhances antibody affinity for the native protein. *Mol. Immunol.* **24**, 297–303 (1987).
7. Sanchez-Trincado, J. L., Gomez-Perosanz, M. & Reche, P. A. Fundamentals and methods for T- and B-cell epitope prediction. *J. Immunol. Res.* **2017**, 2680160 (2017).
8. Grant, G. A. Synthetic peptides for production of antibodies that recognize intact proteins. *Curr. Protoc. Mol. Biol.* **59**, 11.16.1-11.16.19 (2002).
9. Trier, N. H., Hansen, P. R. & Houen, G. Production and characterization of peptide antibodies. *Methods* **56**, 136–144 (2012).
10. Peeters, J. M., Hazendonk, T. G., Beuvery, E. C. & Tesser, G. I. Comparison of four bifunctional reagents for coupling peptides to proteins and the effect of the three moieties on the immunogenicity of the conjugates. *J. Immunol. Methods* **120**, 133–143 (1989).
11. Sakarellos-Daitsiotis, M., Krikorian, D., Panou-Pomonis, E. & Sakarellos, C. Artificial carriers: a strategy for constructing antigenic/immunogenic conjugates. *Curr. Top. Med. Chem.* **6**, 1715–1735 (2006).
12. Hume, H. K. C. *et al.* Synthetic biology for bioengineering virus-like particle vaccines. *Biotechnol. Bioeng.* **116**, 919–935 (2019).
13. Singh, K. V., Kaur, J., Varshney, G. C., Raje, M. & Suri, C. R. Synthesis and characterization of hapten–protein conjugates for antibody production against small molecules. *Bioconjug. Chem.* **15**, 168–173 (2004).
14. Adamczyk, M. *et al.* Characterization of protein-hapten conjugates. 1. matrix-assisted laser desorption ionization mass spectrometry of immuno BSA-hapten conjugates and comparison with other characterization methods. *Bioconjug. Chem.* **5**, 631–635 (1994).
15. Harris, J. R. & Markl, J. Keyhole limpet hemocyanin (KLH): a biomedical review. *Micron* **30**, 597–623 (1999).
16. Swaminathan, A., Lucas, R. M., Dear, K. & McMichael, A. J. Keyhole limpet haemocyanin—a model antigen for human immunotoxicological studies. *Br. J. Clin. Pharmacol.* **78**, 1135–1142 (2014).
17. Briand, J. P., Muller, S. & Van Regenmortel, M. H. V. Synthetic peptides as antigens: Pitfalls of conjugation methods. *J. Immunol. Methods* **78**, 59–69 (1985).
18. Brune, K. D. *et al.* Plug-and-display: decoration of virus-like particles via isopeptide bonds for modular immunization. *Sci. Rep.* **6**, 19234 (2016).
19. Brune, K. D. *et al.* Dual plug-and-display synthetic assembly using orthogonal reactive proteins for twin antigen immunization. *Bioconjug. Chem.* **28**, 1544–1551 (2017).
20. Bieger, B., Essen, L.-O. & Oesterhelt, D. Crystal structure of halophilic Dodecin: a novel, dodecameric flavin binding protein from *Halobacterium salinarum*. *Structure* **11**, 375–385 (2003).
21. Grininger, M., Staudt, H., Johansson, P., Wachtveitl, J. & Oesterhelt, D. Dodecin is the key player in flavin homeostasis of Archaea. *J. Biol. Chem.* **284**, 13068–13076 (2009).
22. Meissner, B., Schleicher, E., Weber, S. & Essen, L. O. The dodecin from *Thermus thermophilus*, a bifunctional cofactor storage protein. *J. Biol. Chem.* **282**, 33142–33154 (2007).
23. Bourdeaux, F. *et al.* Flavin storage and sequestration by *Mycobacterium tuberculosis* dodecin. *ACS Infect. Dis.* **4**, 1082–1092 (2018).
24. Liu, F. *et al.* Structural and biophysical characterization of *Mycobacterium tuberculosis* Dodecin Rv1498A. *J. Struct. Biol.* **175**, 31–38 (2011).
25. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci.* **109**, E690–E697 (2012).
26. Li, L., Fierer, J. O., Rapoport, T. A. & Howarth, M. Structural analysis and optimization of the covalent association between spycatcher and a peptide tag. *J. Mol. Biol.* **426**, 309–317 (2014).
27. Veggiani, G. *et al.* Programmable polyproteams built using twin peptide superglues. *Proc. Natl. Acad. Sci.* **113**, 1202–1207 (2016).
28. Thompson, K. E., Bashor, C. J., Lim, W. A. & Keating, A. E. SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth Biol* **1**, 118–129 (2012).
29. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79 (2006).
30. Schlapschy, M. *et al.* PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins. *Protein Eng. Des. Sel.* **26**, 489–501 (2013).
31. Jevševar, S. *et al.* Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. *Biotechnol. Prog.* **21**, 632–639 (2005).
32. Ishii, M., Kunimura, J. S., Jeng, H. T., Penna, T. C. V. & Cholewa, O. Evaluation of the pH-and thermal stability of the recombinant green fluorescent protein (GFP) in the presence of sodium chloride. In *Applied Biochemistry and Biotecnology*, 555–571 (Springer, Berlin, 2007).
33. Shieh, Y.-W. *et al.* Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science* **350**, 678–680 (2015).
34. Grininger, M., Zeth, K. & Oesterhelt, D. Dodecins: a family of lumichrome binding proteins. *J. Mol. Biol.* **357**, 842–857 (2006).

35. Lili, J., Kosuke, M., Hirofumi, I., Kouhei, T. & Noriho, K. Polymeric SpyCatcher scaffold enables bioconjugation in a ratio-controllable manner. *Biotechnol. J.* **12**, 1700195 (2017).
36. Yin, J., Lin, A. J., Golan, D. E. & Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **1**, 280–285 (2006).
37. Hill, B. G., Ramana, K. V., Cai, J., Bhatnagar, A. & Srivastava, S. K. Measurement and identification of s-glutathiolated proteins. *Methods Enzymol.* **473**, 179–197 (2010).
38. Zhang, H. *et al.* Glutathionylation of the bacterial Hsp70 chaperone dnak provides a link between oxidative stress and the heat shock response. *J. Biol. Chem.* **291**, 6967–6981 (2016).
39. Wachtel, R. E. & Tsuji, K. Comparison of limulus amebocyte lysates and correlation with the United States pharmacopeial pyrogen test. *Appl. Environ. Microbiol.* **33**, 1265–1269 (1977).
40. Pearson, F. C. *et al.* Comparison of several control standard endotoxins to the National Reference Standard Endotoxin–an HIMA collaborative study. *Appl. Environ. Microbiol.* **50**, 91–93 (1985).
41. Nöll, G., Trawöger, S., von Sanden-Flohe, M., Dick, B. & Grininger, M. Blue-light-triggered photorelease of active chemicals captured by the flavoprotein Dodecin. *ChemBioChem* **10**, 834–837 (2009).
42. Gutiérrez Sánchez, C., Su, Q., Schönherr, H., Grininger, M. & Nöll, G. Multi-ligand-binding flavoprotein Dodecin as a key element for reversible surface modification in nano-biotechnology. *ACS Nano* **9**, 3491–3500 (2015).
43. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* **40**, 2004–2021 (2001).
44. Bourdeaux, F. *et al.* Comparative biochemical and structural analysis of the flavin-binding Dodecins from *Streptomyces davaonensis* and Streptomyces coelicolor reveals striking differences with regard to multimerization. *Microbiology* https://doi.org/10.1099/mic.0.000835 (2019).
45. Lu, S. Heterologous prime-boost vaccination. *Curr. Opin. Immunol.* **21**, 346–351 (2009).
46. Gutiérrez Sánchez, C., Su, Q., Wenderhold-Reeb, S. & Nöll, G. Nanomechanical properties of protein–DNA layers with different oligonucleotide tethers. *RSC Adv.* **6**, 56467–56474 (2016).
47. Chen, A. Y. *et al.* Synthesis and patterning of tunable multiscale materials with engineered cells. In *Nature materials* (2014). https://doi.org/10.1038/nmat3912.
48. Botyanszki, Z., Tay, P. K. R., Nguyen, P. Q., Nussbaumer, M. G. & Joshi, N. S. Engineered catalytic biofilms: site-specific enzyme immobilization onto *E. coli* curli nanofibers. *Biotechnol. Bioeng.* **112**, 2016–2024 (2015).
49. Sun, F., Zhang, W.-B., Mahdavi, A., Arnold, F. H. & Tirrell, D. A. Synthesis of bioactive protein hydrogels by genetically encoded SpyTag-SpyCatcher chemistry. *Proc. Natl. Acad. Sci.* **111**, 11269–11274 (2014).
50. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).
51. Giessen, T. W. & Silver, P. A. A catalytic nanoreactor based on in vivo encapsulation of multiple enzymes in an engineered protein nanocompartment. *ChemBioChem* **17**, 1931–1935 (2016).
52. Siu, K.-H. *et al.* Synthetic scaffolds for pathway enhancement. *Curr. Opin. Biotechnol.* **36**, 98–106 (2015).
53. Kick, A., Bönsch, M. & Mertig, M. EGNAS: an exhaustive DNA sequence design algorithm. *BMC Bioinform.* **13**, 138 (2012).
54. Bajar, B. T. *et al.* Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci. Rep.* **6**, 20889 (2016).
55. Schägger, H. Tricine–SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006).
56. Mofid, M. R., Marahiel, M. A., Ficner, R. & Reuter, K. Crystallization and preliminary crystallographic studies of Sfp: a phosphopantetheinyl transferase of modular peptide synthetases. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1098–1100 (1999).
57. Heil, C. S., Rittner, A., Goebel, B., Beyer, D. & Grininger, M. Site-specific labelling of multidomain proteins by Amber Codon suppression. *Sci. Rep.* **8**, B08206 (2018).

Acknowledgements
We thank Kim Remans for helpful discussions. We are grateful to Ilka Siebels for providing Sfp and Alexander Rittner for providing holo *mm*ACP-H8. We thank Emily Hensch and Dominik Scheliu, who assisted in this project during her bachelor thesis and his master thesis, respectively. This work was supported by intramural funds of the Goethe University Frankfurt to M.G., and was associated to the LOEWE program (Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz) of the state of Hesse conducted within the framework of the MegaSyn Research Cluster. Open access funding provided by Projekt DEAL.

Author contributions
F.B. cloned several protein expression constructs and prepared several proteins. F.B. further established protocols, designed experiments, analysed data, prepared all figures and tables, and designed research. Y.K. cloned protein expression constructs and prepared proteins used for western blotting (except *mt*Dod constructs). Y.K. also conducted the western blots and analysed data. J.L. (under supervision of F.B.) cloned several protein expression constructs, purified *mt*Dod-PAS-Pep constructs, prepared several proteins, and analysed data. I.G. cloned *mt*Dod-PAS-Pep constructs and expressed them in *E. coli*. H.B. designed research. R.M.V. selected peptide sequences used for AB production. M.G. analysed data and designed research. F.B. and M.G. wrote the paper. Y.K., R.M.V., and H.B. edited the paper. All authors reviewed the paper.

Competing interests
F.B., H.B., I.G., and M.G. are inventors of EP patent application "Carrier Matrix Comprising Dodecin Protein", EP19220117.6 (filed on December 30th, 2019) for the use of dodecins for antibody production. The remaining co-authors have no conflicts of interest with the content of this article.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-69990-0.

Correspondence and requests for materials should be addressed to M.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supporting Information: Dodecin as carrier protein for immunizations and bioengineering applications

*Authors: Florian Bourdeaux [a], Yannick Kopp [b], Julia Lautenschläger [a], Ines Gößner [a], Hüseyin Besir[c], R. Martin Vabulas [d], Martin Grininger [a,]**

Author affiliations:
(a) Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Cluster of Excellence for Macromolecular Complexes, Goethe University Frankfurt, Max-von-Laue-Str. 15, D-60438 Frankfurt am Main, Germany
(b) Institute of Biophysical Chemistry, Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, Max-von-Lauer Str. 15, D-60438 Frankfurt am Main, Germany
(c) European Molecular Biology Laboratory, 69117 Heidelberg, Germany; currently PROGEN Biotechnik GmbH, 69123 Heidelberg, Germany
(d) Charité - Universitätsmedizin Berlin, Institute of Biochemistry, Charitéplatz 1, D-10117 Berlin

*Corresponding author: grininger@chemie.uni-frankfurt.de

Contents:

Abbreviations:
ACP, acyl carrier protein; AB, antibody; BSA, bovine serum albumin; Catcher, small protein fold (SpyCatcher or SnoopCatcher) that binds and reacts with Tag; CDS, coding sequence; CellSig., Cell Signaling Technology; CHIP, C-terminus of heat shock cognate 70 interacting protein; EU, endotoxin units; FMN, riboflavin-5'-phosphate; GFP, green fluorescent protein; GSG, PAS, PAS2 GPAS, GPAS2, PASG, PAS2G, linker systems, see Table 1; HB-EGF, proheparin-binding EGF-like growth factor; HSP, heat shock protein; IPTG, isopropyl-β-D-thiogalactopyranoside; KLH, keyhole limpet hemocyanin; L, Ladder (only used in figures); LAL, *Limulus* amebocyte lysate; Lys., lysate (only used in figures); MAP, multiple antigen peptides; mCHIP, middle fragment of C-terminus of heat shock cognate 70 interacting protein; MG, proteasome inhibitor MG-132; *mm*ACP, *Mus musculus* acyl carrier protein; msfGFP, monomeric superfolder green fluorescent protein; *mt*Dod, *Mycobacterium tuberculosis* dodecin; *mt*Dod(WT), *Mycobacterium tuberculosis* dodecin wild type; OD600, optical density at 600 nm; OE, over expressing cells; RSA, rabbit serum albumin; SCBT, Santa Cruz Biotechnology; *se*ACP, *Saccharopolyspora erythraea* acyl carrier protein; SEC, size exclusion chromatography; Sfp, 4'-phosphopantetheine transferase from *Bacillus subtilis*; Sigma, Sigma-Aldrich; SnpC, SnoopCatcher; SnpT, SnoopTag; SpyC, SpyCatcher; SpyT, SpyTag; SZ, SYNZIP domain; Tag, small peptide sequence that interacts with Catcher's (SpyTag or SnoopTag); TB, terrific broth; TBS, Tris-HCl buffered saline; TBST, Tris-HCl buffered saline with Tween-20; TT, tetanus toxoid; VLP, virus-like particle.

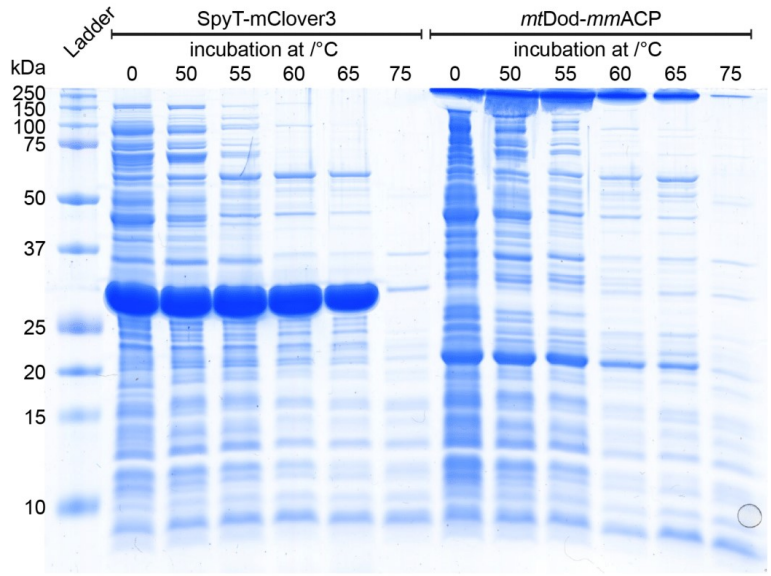


Fig. S1: SDS-PAGE gel of supernatant after heat denaturation of lysate at different temperatures. Standard loading buffer (pH 6.8, 2.5% SDS) was used for sample preparation with 15 min heat treatment (95 °C). Lysate samples were incubated at depicted temperatures for 25 min and aggregated protein was removed by centrifugation (15.000 rcf, 10 min). After centrifugation, for all samples, the same amount of supernatant was mixed with loading buffer, incubated as stated and loaded onto the gel. Left: Bands show heat denaturation of SpyT-mClover3 at different temperatures. At 60 °C, most *E. coli* proteins are aggregated and removed by centrifugation, while SpyT-mClover (thick band above 25 kDa) is still present in high concentration. At 75 °C, also SpyT-mClover is denatured and aggregated, as indicated by the intensity decrease of the respective band. Right: Supernatant samples of heat denaturation of *mt*Dod-*mm*ACP. Since standard loading buffer is not able to denature the dodecamer (only a small fraction of the dodecamer denatures under these conditions), the dodecamer hardly migrated in the gel, as indicated by bands at very top edge of the gel. The band representing the *mt*Dod-*mm*ACP monomer is visible slightly above 20 kDa. At 60 °C, the intensity of the *mt*Dod-*mm*ACP band decreases (best observable for the monomer band slightly above 20 kDa), indicating that *mt*Dod-*mm*ACP starts to precipitate during the heat denaturing step. At 75 °C, only weakly stained bands *mt*Dod-*mm*ACP are visible, showing that most *mt*Dod-*mm*ACP precipitated at this temperature and was removed by centrifugation (see Fig. S7 for SDS-PAGE of the pellet).

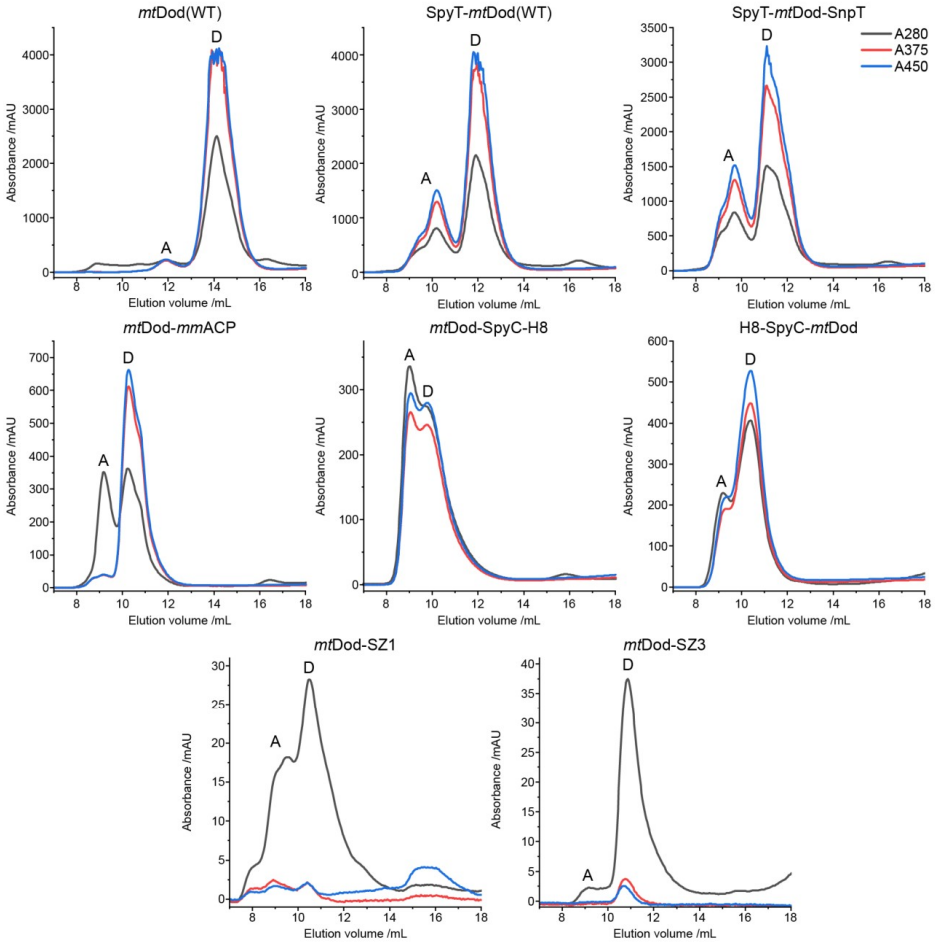


Fig. S2: SEC chromatograms of various *mt*Dod constructs. Column: Superdex 200 increase 10/300. A: Peak/peaks representing aggregates. D: Peak representing the dodecamer. Top row: Example of *mt*Dod constructs purified by the heat denaturation protocol. The second DMSO precipitation step was used to concentrate samples as high as possible (buffer added in small portions, until the pellet was nearly fully dissolved but not completely). Under these high concentrations, some *mt*Dod constructs seem to form aggregates that still bind FMN. Chromatogram of *mt*Dod-*mm*ACP purified by the heat denaturation strategy shows a prominent aggregation peak at about 9 mL. Chromatograms of refolded *mt*Dod SpyC constructs form FMN binding dodecamers, but tend to aggregate (peak at about 9 mL). H8-SpyC-*mt*Dod seems to have lower aggregation tendencies than *mt*Dod-SpyC-H8 (the separation of dodecamer and aggregates was not possible). *Mt*Dod SYNZIP constructs were refolded and purified without additional FMN. The high aggregation tendencies of SYNZIP constructs (especially *mt*Dod-SZ1) made the purification of higher amounts challenging, as samples tended to suddenly precipitate during concentration and filtration at higher concentrations.

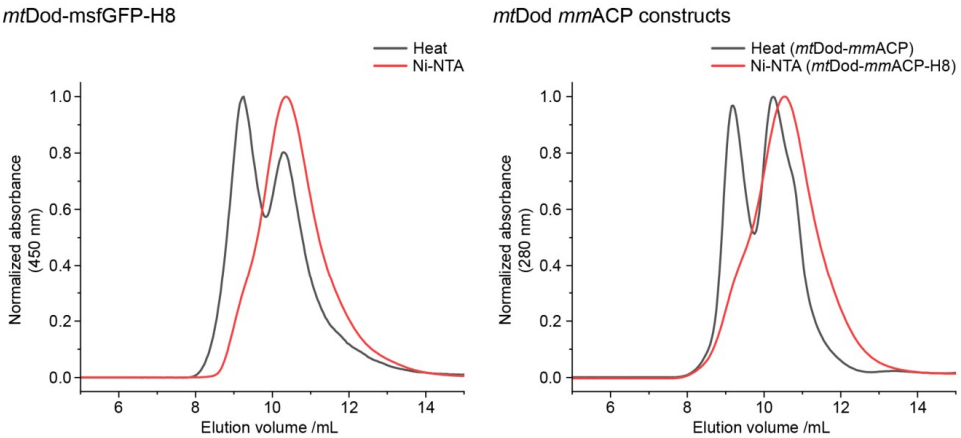


Fig. S3: Comparison of SEC chromatograms of *mt*Dod-msfGFP-H8 and *mt*Dod *mm*ACP constructs purified by Ni-NTA affinity chromatography and/or by heat denaturation. In both cases, the Ni-NTA affinity chromatography purification caused less aggregation. However, dodecameric fractions of *mt*Dod-msfGFP-H8 and *mt*Dod-*mm*ACP could also be received from the heat denaturation protocol.

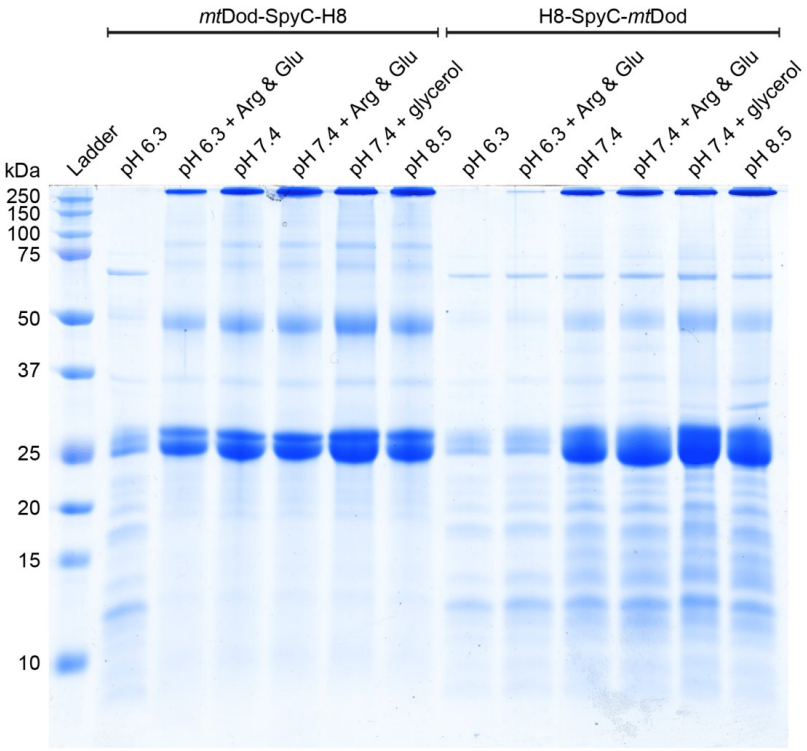


Fig. S4: SDS-PAGE gel of *mt*Dod SpyC constructs refolded under different conditions via dialysis. Acidic loading buffer used for sample preparation contained 50 mM acetic acid, which did not allow a full denaturation of the dodecamer. General buffer solution for refolding (final concentrations): 100 mM NaCl, 25 mM Na_2HPO_4 and 25 mM boric acid. The pH was adjusted to the shown values with HCl or NaOH. Refolding additives were 50 mM arginine and 50 mM glutamic acid (Arg & Glu)[1] and 20% glycerol (glycerol). Except under slightly acidic conditions, refolding of *mt*Dod SpyC constructs is possible. Arginine and glutamic acid seem to be beneficial for refolding (clearly seen for *mt*Dod-SpyC-H8 at pH 6.3). Overall, the best tested condition seems to be pH 7.4 with 20% glycerol (highest band intensity and lowest amount of precipitate was observed).

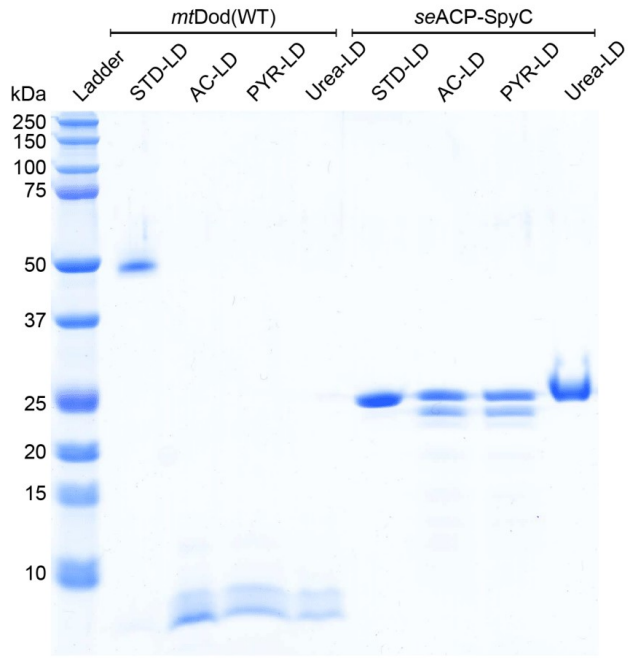


Fig. S5: SDS-PAGE gel for comparison of loading buffers. *Mt*Dod(WT) and *se*ACP-SpyC were used. STD-LD: standard loading buffer (denaturing step conditions: pH 6.8, 2.5% SDS). AC-LD: Acetic acid two-component loading buffer (denaturing step conditions: pH < 5.0, 3.3% SDS). PYR-LD: Pyridine two-component loading buffer (denaturing step conditions: pH < 5.0, 3.3% SDS). Urea-LD: standard loading buffer with 8 M urea (denaturing step conditions: pH 6.8, 2.5% SDS, 8 M Urea). All samples were denatured for 5 min at 95 °C. After denaturation, the second component buffer was added to the two-component loading buffer samples (final pH, SDS and glycerol content as in standard loading buffer). All loading buffers, except the standard loading buffer, are able to denature the *mt*Dod(WT) dodecamer, indicated by the monomer band (below 10 kDa). The smearing double bands seem to be an artefact of this specific SDS-PAGE gel, as also the lower molecular weight bands of the ladder show this phenomenon. The gel shows that the *mt*Dod dodecamer tolerates SDS, as long the conditions do not become too acidic. For seACP-SpyC, the acidic denaturing conditions cause the appearance of a band at lower molecular weight. The formation of this band is also temperature dependent. Heat treatment at 60°C with acidic loading buffer did not cause the formation of double bands (see Fig 5).

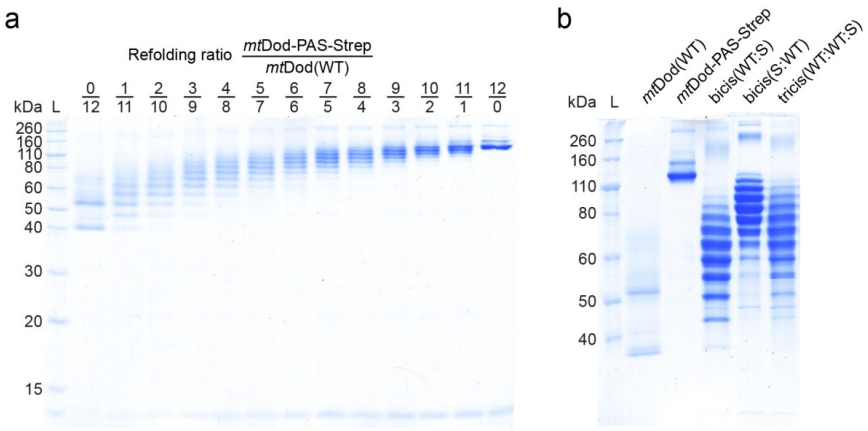


Fig. S6: Uncropped SDS-PAGE gel of Fig. 3. L: Ladder. S: *mt*Dod-PAS-Strep. WT: *mt*Dod(WT). a) Uncropped SDS-PAGE gel of Fig. 3 a. Heterododecamers obtained by refolding *mt*Dod(WT) and mtDod-PAS-Strep at different ratios. Bands below 15 kDa represent the mtDod-PAS-Strep monomer (9916 Da), while *mt*Dod(WT) monomer (7497 Da) migrated through the full gel so that no bands are visible. b) Full length gel image of Fig. 3 b. Heterododecamers formed during polycistronic expression of *mt*Dod(WT) and *mt*Dod-PAS-Strep. Gene order of polycistronic expression vectors given in brackets (bicis: bicistronic(first gene : second gene); tricis: tricistronic(first gene : second gene : third gene). For improved separation of higher molecular weight bands running time was prolonged to about 4 h. The monomers of *Mt*Dod constructs and lower weight proteins migrated through the entire gel so that protein and ladder bands (below 40 kDa) are not visible. The gel is cropped at the right to remove not relevant samples.

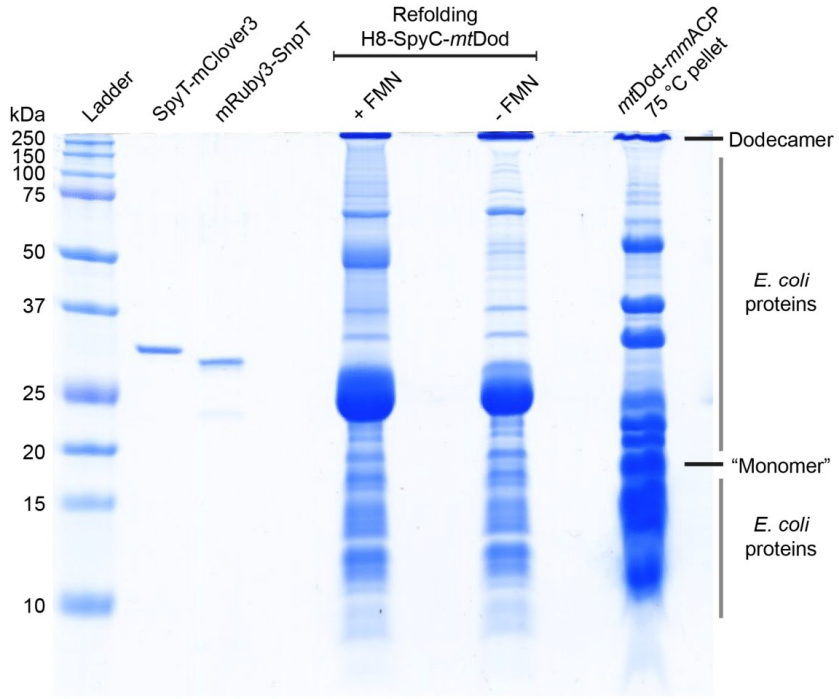


Fig. S7: SDS-PAGE gel of *mt*Dod-*mm*ACP precipitating during heat denaturation and other samples. Standard loading buffer was used to preserve dodecameric species. Right sample: Cytosol containing *mt*Dod-*mm*ACP was incubated at 75 °C for 25 min and aggregated proteins were pelleted by centrifugation (15.000 rcf, 10 min). A small amount of the pellet was dissolved in standard SDS loading buffer at 95 °C (about 15 min) and loaded to the gel. In the stained gel, a strong band at high molecular weight (top of the gel, labelled dodecamer) is visible indicating intact dodecamer of *mt*Dod-*mm*ACP. The potential band representing the monomer of *mt*Dod-*mm*ACP (about 18 kDa) cannot be selected with certainty (most fitting band labelled with "Monomer"). Other bands are the during the heat incubation denatured *E. coli* proteins resolubilized with SDS (grey bar). Other samples: SpyT-Clover-H8 and mRuby3-SnpT: pooled SEC fractions (purified by a heat denaturation based protocol). Refolding H8-SpyC-*mt*Dod: first refolding attempt of H8-SpyC-*mt*Dod in our standard buffer, aggregated protein was removed by centrifugation (15,000 rcf, 10 min).

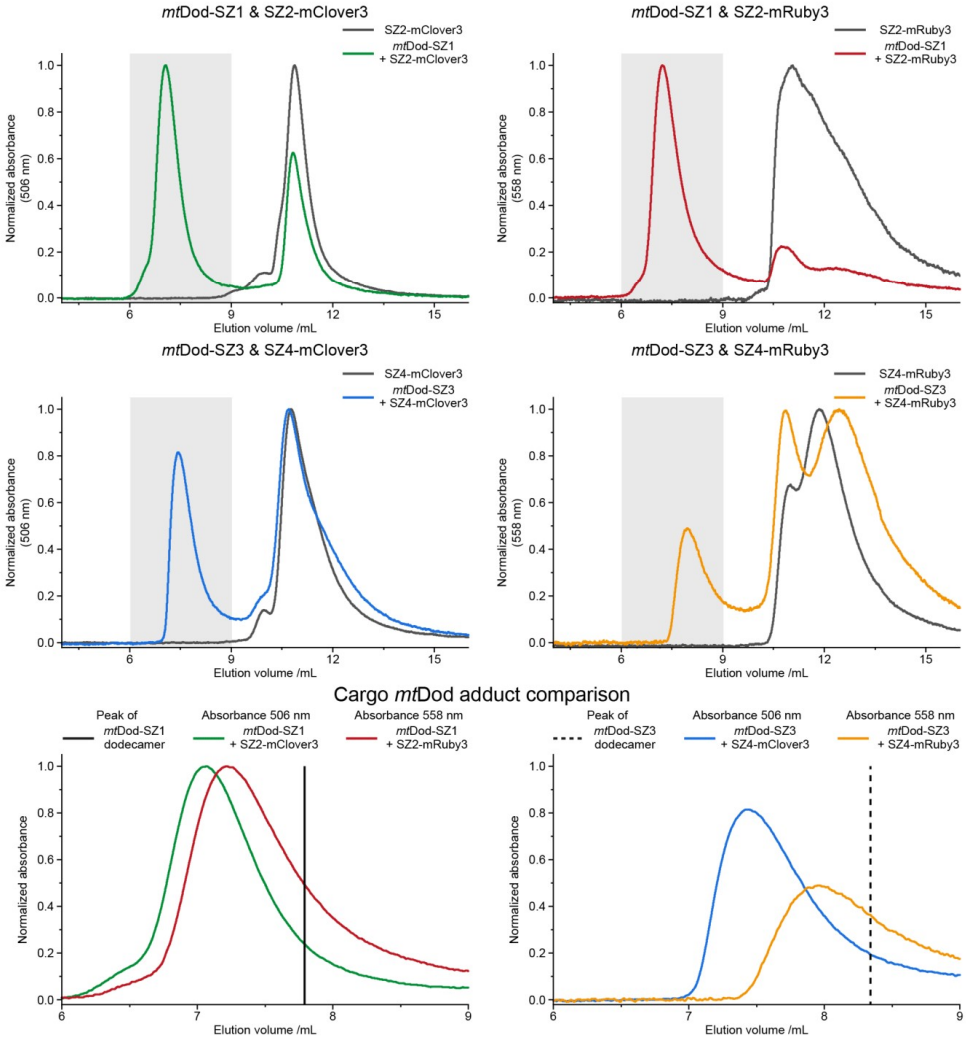


Fig. S8: HPLC-SEC chromatograms of *mt*Dod SYNZIP constructs and SYNZIP fluorescence protein constructs. Chromophore absorption of mRuby3 or mClover3 constructs was measured to observe the formation of high molecular weight species. Peaks of the formed adducts (grey highlighted areas) are overlaid in the bottom row. For comparison, elution volumes of the SYNZIP *mt*Dod constructs are shown as a straight (*mt*Dod-SZ1) or dashed (*mt*Dod-SZ3) line (based on absorbance at 280 nm). While for all combinations adduct species were observed, there is a clear difference in dodecamer to cargo composition. The SZ1-SZ2 pair overall performed better than the SZ3-SZ4 pair. For adduct formation, equimolar amounts of carrier and cargo (67 µM) were used and first incubated at 37 °C for 1 h and then on ice for 1 h. Single protein samples were treated the same way. As reaction buffer and running buffer, the standard dodecin buffer (300 mM NaCl, 5 mM $MgCl_2$ and 20 mM Tris-HCl (pH 7.4) was used. Samples were filtered (0.22 µm) and 8 µL were loaded. Runs were performed at 0.3 mL/min and 22 °C. Used column: bioZen 1.8 µm SEC-3 (300 × 4.6 mm, Phenomenex). Absorbance at 280 nm, 375 nm, 506 nm and 558 nm was measured.

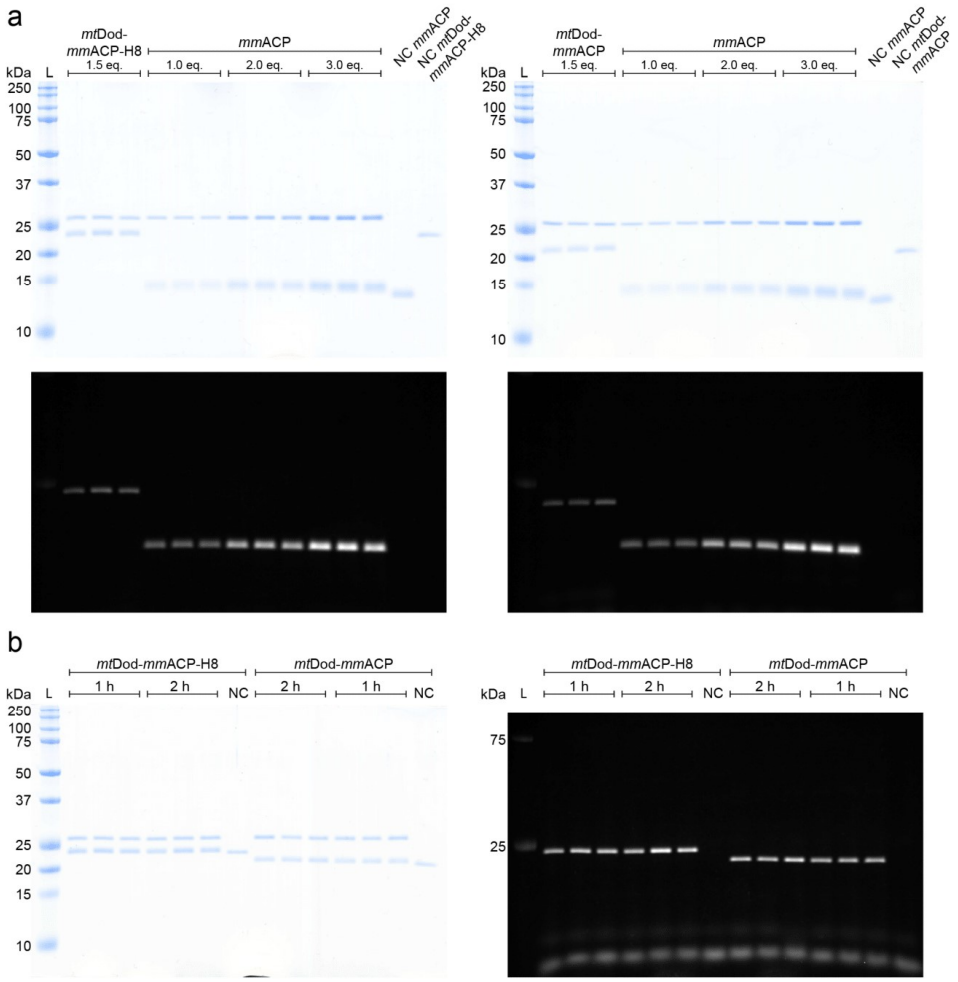


Fig. S9: Uncropped SDS-PAGE gels of the modification of *mt*Dod-*mm*ACP and *mt*Dod-*mm*ACP-H8 by Sfp. a) Coomassie stained gels at top, fluorescence images at the bottom. b) (left panel) Coomassie stained gels, (right) fluorescence images. At the bottom of the gel, free fluorophore is observable. In a) no free fluorophore is visible, because of the shorter exposure time used for the images

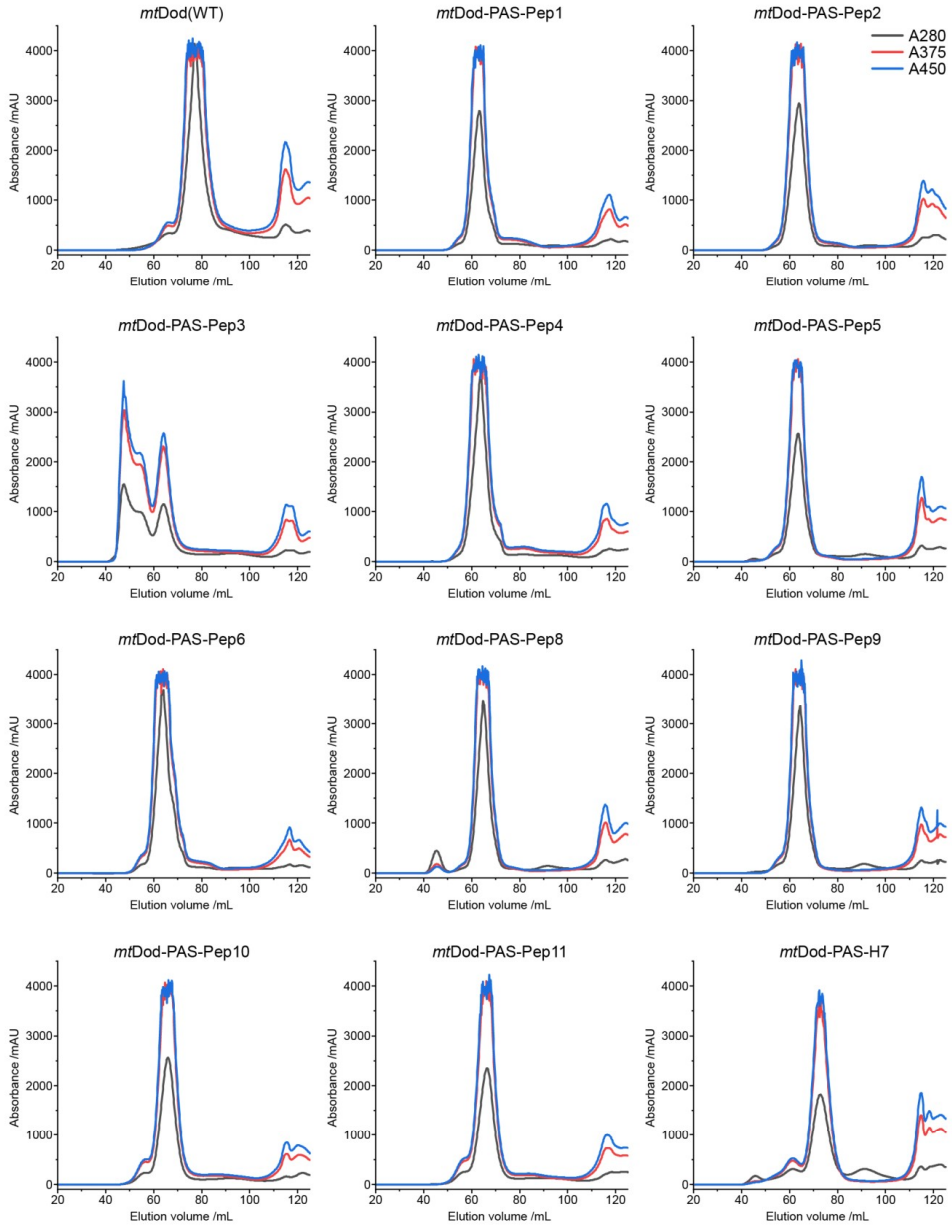


Fig. S10: SEC chromatograms of *mt*Dod-PAS-Pep constructs, *mt*Dod-PAS-H7 and *mt*Dod(WT). Used column: Superdex 200 16/60 pg column. Except for *mt*Dod-PAS-Pep3, only minor amounts of aggregates are visible. At about 117 mL, unbound FMN is eluting.

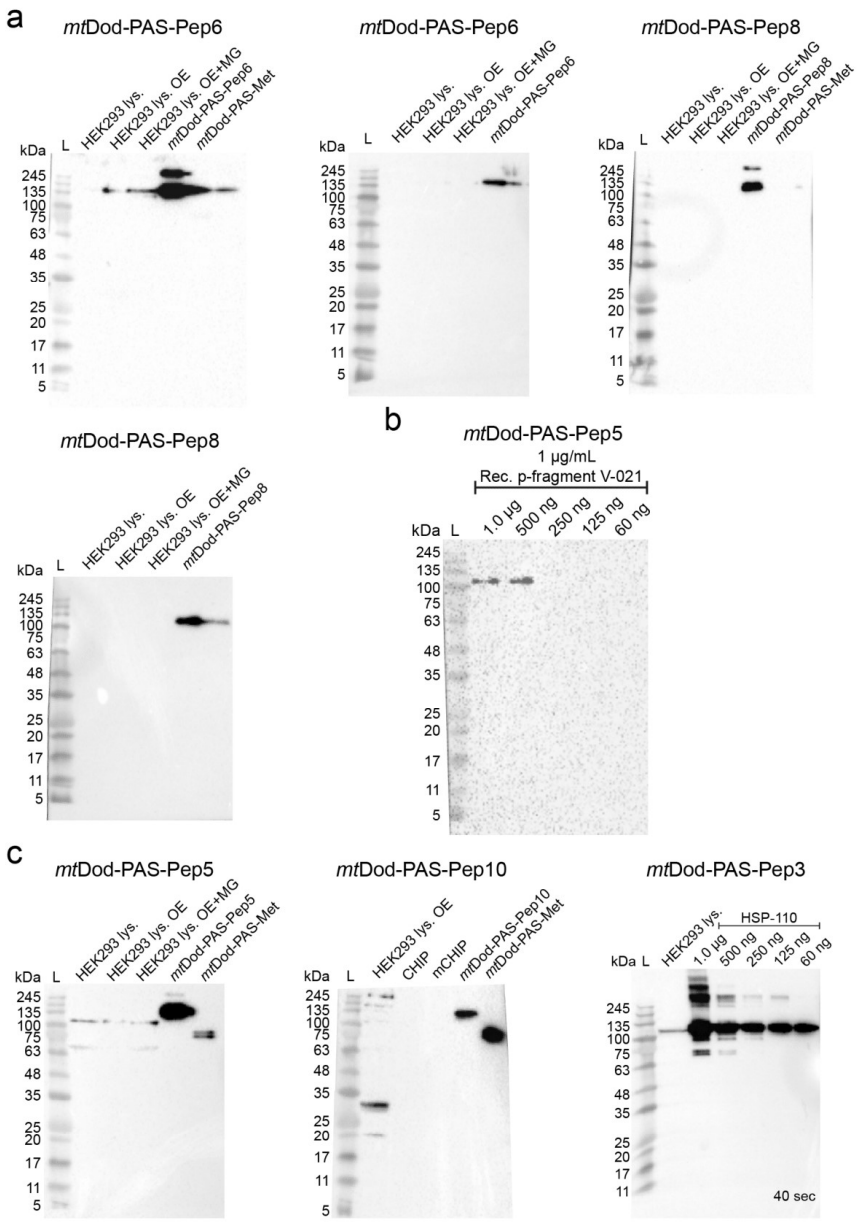


Fig. S11: Additional western blots for class assignment. L: Ladder. Lys.: Lysate. OE: protein overexpressing cells. MG: proteasome inhibitor MG-132 added. mCHIP: fragment of CHIP. a) Western blots of "class 3"ABs. Both ABs seem not to recognize *mt*Dod-PAS-Met (band expected at about 75 kDa) b) Western blot with ABs derived from *mt*Dod-PAS-Pep5. Bands were observed after an exposure time of about 300 sec. c) Uncropped western blots of Fig. 8 a.

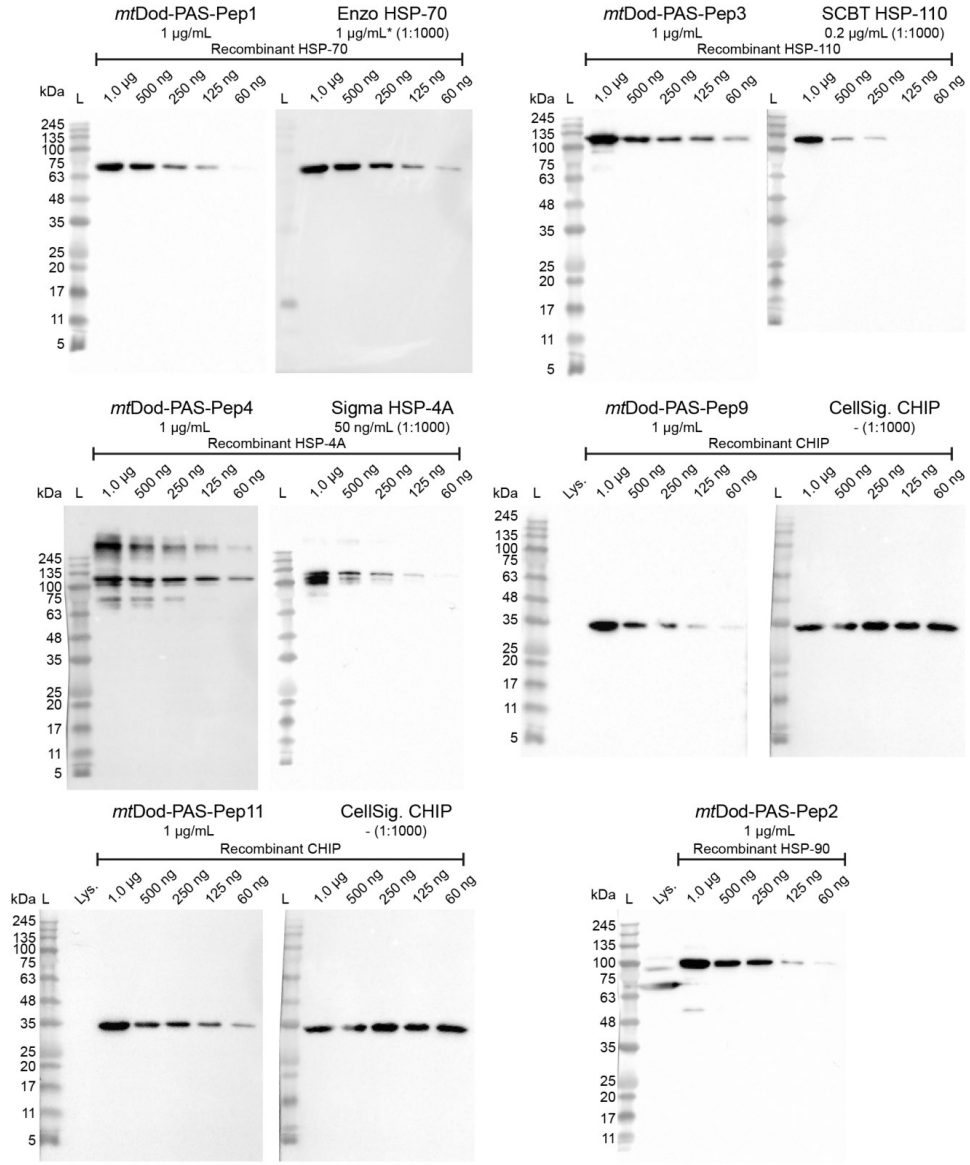


Fig. S12: Uncropped western blots of Fig 8 b. L: Ladder. Lys.: Lysate.

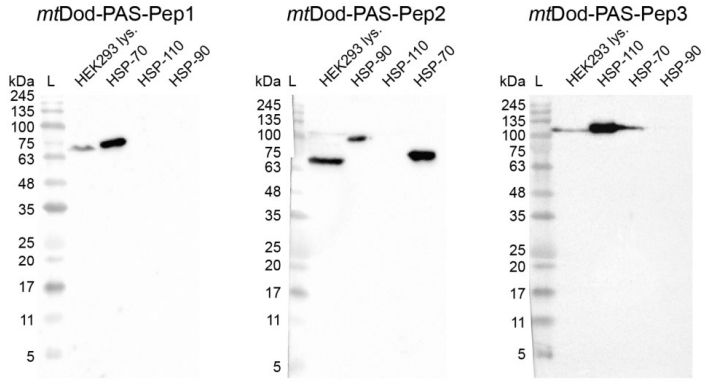


Fig. S13: Western blots of HSP recognizing ABs in comparison. L: Ladder. ABs derived from *mt*Dod-PAS-Pep2 (designed for HSP-90) recognizes also HSP-70. AB derived from *mt*Dod-PAS-Pep1 and *mt*Dod-PAS-Pep3 recognize only the protein of interest (HSP-90 and HSP-110, respectively).

Table S1: Table of encoding sequences for all constructs used in this study, except recombinant proteins used in western blotting. All plasmids used for expression were based on a pET22b vector backbone (*lacI* coding sequence, ampicillin resistance, pBR322 origin and f1 origin). Sequences encoding mClover3, mRuby3, SpyC and SnpC were cloned from vectors obtained from Addgene (Plasmid: #74252, pKanCMV-mClover3-mRuby3[2]; Plasmid: #72324, pET28a SpyCatcher-SnoopCatcher[3]). For the polycistronic constructs, spacer DNA sequences between stop codon of the previous gene and the +42 upstream bases of the next gene (based on the pET22b vector and an added restriction side) were designed with EGNAS (version 1158, to minimize secondary structures).[4] These spacer regions were used for cloning (annealing area for In-Fusion HD Cloning (TaKaRa Bio Europe)). Amino acid sequences of linkers and restriction sites are highlighted in yellow.

Construct ***mt*Dod-peptides**	DNA sequence T7 promotor to T7 terminator (CDS uppercase)	Amino acid sequence (linker)
*mt*Dod(WT)	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCTGAccgagcacca ccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataacta gcataaccctggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDS*
*mt*Dod-GSG-Lys	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGGGGG GTGGCGGCAGTGGTGGCGGCGGTAAATGAGGTGACTCTCTGTCTTGGCTGCTGCGTC TGCTGAACTgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagc ataaccctggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLEGGGGSGGGG K*
*mt*Dod-PAS-Met	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCATGTGAggatcccaccaccaccaccaccactgag atccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcc tctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SM*
*mt*Dod-SpyT	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGGTGGCAGCGGTGCACATATCGTCAT GGTTGATGCGTACAAACCGACCAAATGAggatcccaccaccaccaccaccaccactgagatccggctgctaac aaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtctt gaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SGGSGAHIVMVDAYKPTK*
SpyT-*mt*Dod	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGGCACATATCGTCATGGTTGATGCGTACAAACCGACCAAAGGTGGCAGC GGTTCTCCAGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCAGCAATCACACCTACCGA GTGATCGAGATCGTCGGGACCTCGCCCGACGGCGTCGACGCGGCAATCCAGGGCGG TCTGGCCCGAGCTGCGCAGACCATGCGCGCGCTGGACTGGTTCGAAGTACAGTCAAT TCGAGGCCACCTGGTCGACGGAGCGGTCGCGCACTTCCAGGTGACTATGAAAGTCGG CTTCCGCCTGGAGGATTCCTGActgagcaccaccaccaccaccaccactgagatccggctgctaacaaagcccg aaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttt tg	MAHIVMVDAYKPTKGGSGSPAAPAPAS PASSNHTYRVIEIVGTSPDGVDAAIQGGL ARAAQTMRALDWFEVQSIRGHLVDGAV AHFQVTMKVGFRLEDS*
*mt*Dod-PAS2-SpyT	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGTCTCCGGCACCGTCTGCGCCAGCTGCATC TCCAGCAGCGGGTGGCAGCGGTGCACATATCGTCATGGTTGATGCGTACAAACCGAC CAAATGAggatcccaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgct gccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SPAPSAPAASPAAGGSGAHIVMVDAYKP TK*
SpyT-PAS2-*mt*Dod	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGGCACATATCGTCATGGTTGATGCGTACAAACCGACCAAAGGTGGCAGC GGTTCTCCAGCTGCGCCTGCTCCGGCAAGCCCTGCGTCTCCGGCACCGTCTGCGCCA GCTGCATCTCCAGCAGCGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCT CGCCCGACGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACC ATGCGCGCGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGA GCGGTCGCGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCTGAc tcgagcaccaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctga gcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MAHIVMVDAYKPTKGGSGSPAAPAPAS PASPAPSAPAASPAASNHTYRVIEIVGTS PDGVDAAIQGGLARAAQTMRALDWFEV QSIRGHLVDGAVAHFQVTMKVGFRLED S*

SpyT-*mt*Dod-SnpT	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGGCACATATCGTCATGGTTGATGCGTACAAACCGACCAAAGGTGGCAGC GGTTCTCCAGGTGCGCCTGCTCCGGCAAGCCCTGCGAGCAGCAATCACACCTACCGA GTGATCGAGATCGTCGGGACCTCGCCCGACGGCGTCGACGCGGCAATCCAGGGCGG TCTGGCCCGAGCTGCGCAGACCATGCGCGCGCTGGACTGGTTCGAAGTACAGTCAAT TCGAGGCCACCTGGTCGACGGAGCGGTCGCGCACTTCCAGGTGACTATGAAAGTCGG CTTCCGCCTGGAGGATTCCTCTCCAGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGG TGGCAGCGGTGGCAAACTGGGCGATATTGAATTTATTAAAGTGAACAAAGGCTATTGAg gatcccaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctga gcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MAHIVMVDAYKPTKGGSGSPAAPAPAS PASSNHTYRVIEIVGTSPDGVDAAIQGGL ARAAQTMRALDWFEVQSIRGHLVDGAV AHFQVTMKVGFRLEDSSPAAPAPASPA SGGSGGKLGDIEFIKVNKGY*
SnpT-*mt*Dod-SpyT	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGGGCAAACTGGGCGATATTGAATTTATTAAAGTGAACAAAGGCTATGGT GGCAGCGGTTCTCCAGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCAGCAATCACACCC TACCGAGTGATCGAGATCGTCGGGACCTCGCCCGACGGCGTCGACGCGGCAATCCAG GGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCGCGCTGGACTGGTTCGAAGTACAG TCAATTCGAGGCCACCTGGTCGACGGAGCGGTCGCGCACTTCCAGGTGACTATGAAA GTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCCAGCTGCGCCTGCTCCGGCAAGC CCTGCGAGCGGTGGCAGCGGTGCACATATCGTCATGGTTGATGCGTACAAACCGACC AAATGAggatcccaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctg ccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MGKLGDIEFIKVNKGYGGSGSPAAPAPA SPASSNHTYRVIEIVGTSPDGVDAAIQGG LARAAQTMRALDWFEVQSIRGHLVDGA VAHFQVTMKVGFRLEDSLESPAAPAPAS PASGGSGAHIVMVDAYKPTK*
*mt*Dod-PAS-StrepII	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCCACCCGCAGTTCGAAAAATG Agatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggg gcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SWSHPQFEK*
*mt*Dod-PAS-H7	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCCACCACCACCACCACCACCACTGAgat ccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctc taaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SHHHHHHH*
*mt*Dod-PAS-Pep1	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCCCGAAAGGTGGCAGCGGTTCTGGCC CGACCATTGAAGAAGTGGATTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccac cgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SPKGGSGSGPTIEEVD*
*mt*Dod-PAS-Pep2	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCCCGCTGGAAGGCGATGACGATACCAG CCGCATGGAAGAAGTGGATTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccacc gctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SPLEGDDDTSRMEEVD*
*mt*Dod-PAS-Pep3	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGAATGCTATCCGAACGAGAAGAACAG CGTGAACATGGATCTGGATTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccacc gctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SECYPNEKNSVNMDLD*
*mt*Dod-PAS-Pep4	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGTGCCGAGCGACAGCGATAAGAAACT GCCGGAAATGGATATTGATTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccacc gctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SVPSDSDKKLPEMDID*
*mt*Dod-PAS-Pep5	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGATAGCAGCCAGCATACCAAGAGCTC TGGCGAAATGGAAGTGGATTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccacc gctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SDSSQHTKSSGEMEVD*
*mt*Dod-PAS-Pep6	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGAACAGAGCACCGGCCAGAAACGCC CGCTGAAGAACGATGAACTGTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccac cgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SEQSTGQKRPLKNDEL*
*mt*Dod-PAS-Pep7	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGCGCTGATGGTGTATCGCTGCGCGCC GCCGCGCAGCAGCCAGTTTTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccacc gctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SALMVYRCAPPRSSQF*

*mt*Dod-PAS-Pep8	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCCTGGTGACCGGCGAAAGCCTGGAACA GCTGCGCCGCGGCCTGGCGTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccac cgctgagcaataactagcataacccctttggggcctctaaacgggtcttgagggtttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SLVTGESLEQLRRGLA*
*mt*Dod-PAS-Pep9	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCATGAAAGGCAAAGAAGAGAAAGAAGG CGGCGCGCGCCTGGGCGCGTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgcca ccgctgagcaataactagcataacccctttggggcctctaaacgggtcttgagggtttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SMKGKEEKEGGARLGA*
*mt*Dod-PAS-Pep10	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGAAGAACGCCGCATTCATCAGGAAAG CGAATGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataa cccctttggggcctctaaacgggtcttgagggtttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SEERRIHQESE*
*mt*Dod-PAS-Pep11	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCAACCATGAAGGCGATGAAGATGATAG CCATTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataac ccctttggggcctctaaacgggtcttgagggtttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SNHEGDEDDSH*
poly-cistronic constructs	DNA sequence T7 promotor to T7 terminator (CDS uppercase, spacer sequence)	Amino acid sequence (linker)
bici (*mt*Dod(WT); *mt*Dod-PAS-StrepII)	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCTGAcggtttgaaacc ttccgaggagaattcaataatttttgtttaacttaagaaggagatatacatATGAGCAATCACACCTACCGAGTGA TCGAGATCGTCGGGACCTCGCCCGACGGCGTCGACGCGGCAATCCAGGGCGGTCTG GCCCGAGCTGCGCAGACCATGCGCGCGCTGGACTGGTTCGAAGTACAGTCAATTCGA GGCCACCTGGTCGACGGAGCGGTCGCGCACTTCCAGGTGACTATGAAAGTCGGCTTC CGCCTGGAGGATTCCCTCGAGTCTCCAGCTGCGCCTGCTCCGGCAAGCCCTGCGAGC TGGAGCCACCCGCAGTTCGAAAAATGAgaatcgcgattcaccatggtgggatccgatccggctgctaacaaa gcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccctttggggcctctaaacgggtcttgag gggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDS* & MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SWSHPQFEK*
bici (*mt*Dod-PAS-StrepII; *mt*Dod(WT))	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCTGGAGCCACCCGCAGTTCGAAAAATG AcggtttgaaacctccgaggagaattcaataatttttgtttaacttaagaaggagatatacatATGAGCAATCACACCT ACCGAGTGATCGAGATCGTCGGGACCTCGCCCGACGGCGTCGACGCGGCAATCCAGG GCGGTCTGGCCCGAGCTGCGCAGACCATGCGCGCGCTGGACTGGTTCGAAGTACAGT CAATTCGAGGCCACCTGGTCGACGGAGCGGTCGCGCACTTCCAGGTGACTATGAAAG TCGGCTTCCGCCTGGAGGATTCCTGAgaatcgcgattcaccatggtgggatccgatccggctgctaacaaag cccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccctttggggcctctaaacgggtcttgagg ggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SWSHPQFEK* & MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDS*
tricis (*mt*Dod(WT); *mt*Dod(WT); *mt*Dod-PAS-StrepII)	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCTGAcggtttgaaacc ttccgaggagaattcaatattttgtttaacttaagaaggagatatacatATGAGCAATCACACCTACCGAGTGA TCGAGATCGTCGGGACCTCGCCCGACGGCGTCGACGCGGCAATCCAGGGCGGTCTG GCCCGAGCTGCGCAGACCATGCGCGCGCTGGACTGGTTCGAAGTACAGTCAATTCGA GGCCACCTGGTCGACGGAGCGGTCGCGCACTTCCAGGTGACTATGAAAGTCGGCTTC CGCCTGGAGGATTCCTGAacctgtatcgatgcagctaggaagcttaataatttttgtttaacttaagaaggagatatac atATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGACGGCGT CGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCGCGCTGG ACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCGCGCACTT CCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCCAGCTGCG CCTGCTCCGGCAAGCCCTGCGAGCTGGAGCCACCCGCAGTTCGAAAAATGAgaatcgcga ttcaccatggtgggatccgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataact agcataacccctttggggcctctaaacgggtcttgagggtttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDS* & MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDS* & MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SWSHPQFEK
***mt*Dod-proteins**	DNA sequence T7 promotor to T7 terminator (CDS uppercase)	Amino acid sequence (linker)

*mt*Dod-*mm*ACP	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGACGGGGACACCCAGAGGGATCTGG TGAAAGCTGTAGCACACATCCTAGGCATCCGAGACCTCGCAGGTATTAACCTGGACAG CACGCTGGCAGACCTCGGCCTGGACTCGCTCATGGGTGTGGAAGTTCGTCAGATCCT GGAACGAGAACACGATCTGGTGCTGCCCATGCGTGAGGTGCGGCAGCTCACGCTGCG GAAACTTCAGGAAATGTCCTCCAAGACTGACTCGGCTACTGACACGACAGCCCCCCTG GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGC CTCTAAACGGGTCTTGAGGGGTTTTTTG	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SDGDTQRDLVKAVAHILGIRDLAGINLDS TLADLGLDSLMGVEVRQILEREHDLVLP MREVRQLTLRKLQEMSSKTDSATDTTAP *
*mt*Dod-*mm*ACP-H8	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGACGGGGACACCCAGAGGGATCTGG TGAAAGCTGTAGCACACATCCTAGGCATCCGAGACCTCGCAGGTATTAACCTGGACAG CACGCTGGCAGACCTCGGCCTGGACTCGCTCATGGGTGTGGAAGTTCGTCAGATCCT GGAACGAGAACACGATCTGGTGCTGCCCATGCGTGAGGTGCGGCAGCTCACGCTGCG GAAACTTCAGGAAATGTCCTCCAAGACTGACTCGGCTACTGACACGACAGCCCCCCTC GAGCATCATCACCACCACCACCACCACTGAgatccggctgctaacaaagcccgaaaggaagctgagttg gctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgagggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SDGDTQRDLVKAVAHILGIRDLAGINLDS TLADLGLDSLMGVEVRQILEREHDLVLP MREVRQLTLRKLQEMSSKTDSATDTTAP LEHHHHHHHH*
*mt*Dod-msfGFP-H8	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGTGAGCAAGGGCGAGGAGCTGTTCAC CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG CGTGCGCGGCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCAT CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTA CGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGC ACCTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG TACAACTTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCA AGGCGAACTTCAAGATCCGCCACAACGTCGAGGACGGCAGCGTGCAGCTCGCCGACC ACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT ACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTCGAGC ATCATCACCACCACCACCACCACTGAtgagatccggctgctaacaaagcccgaaaggaagctgagttggctg ctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgagggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SVSKGEELFTGVVPILVELDGDVNGHKF SVRGEGEGDATNGKLTLKFICTTGKLPV PWPTLVTTLTYGVQCFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTYKTRA EVKFEGDTLVNRIELKGIDFKEDGNILGH KLEYNFNSHNVYITADKQKNGIKANFKIR HNVEDGSVQLADHYQQNTPIGDGPVLL PDNHYLSTQSKLSKDPNEKRDHMVLLEF VTAAGITLGMDELEHHHHHHHH*
*mt*Dod-SpyC-H8*	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCGCTATGGTGGACACCCTGTCGGGCCT CTCTAGTGAACAGGGGCAAAGCGGCGATATGACTATCGAAGAAGATAGTGCTACCCAT ATTAAATTCTCAAAACGTGATGAGGACGGCAAAGAGTTAGCTGGTGCAACTATGGAGTT GCGTGATTCATCTGGTAAAACTATTAGTACATGGATTTCAGATGGACAAGTGAAAGATTT CTACCTGTATCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGTTATGAGG TAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAA GCAACTAAAGGTGACGCTCATATTCTCGAGCATCATCACCACCACCACCACCACTGAgat ccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctc taaacgggtcttgagggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SAMVDTLSGLSSEQGQSGDMTIEEDSA THIKFSKRDEDGKELAGATMELRDSSGK TISTWISDGQVKDFYLYPGKYTFVETAAP DGYEVATAITFTVNEQGQVTVNGKATKG DAHILEHHHHHHHH*
H8-SpyC-*mt*Dod	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaag gagatatacatATGCATCATCACCACCACCACCATCACGGTTCTGCTATGGTGGACACCCTG TCGGGCCTCTCTAGTGAACAGGGGCAAAGCGGCGATATGACTATCGAAGAAGATAGTG CTACCCATATTAAATTCTCAAAACGTGATGAGGACGGCAAAGAGTTAGCTGGTGCAACT ATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATTTCAGATGGACAAGT GAAAGATTTCTACCTGTATCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACG GTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTA AATGGCAAAGCAACTAAAGGTGACGCTCATATTTCTCCAGCTGCGCCTGCTCCGGCAA GCCCTGCGAGCAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCG ACGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGC GCGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTC GCGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCTGActcgagcacc accaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataact agcataaccccttggggcctctaaacgggtcttgagggttttttg	MHHHHHHHHGSAMVDTLSGLSSEQGQ SGDMTIEEDSATHIKFSKRDEDGKELAG ATMELRDSSGKTISTWISDGQVKDFYLY PGKYTFVETAAPDGYEVATAITFTVNEQ GQVTVNGKATKGDAHISPAAPAPASPAS SNHTYRVIEIVGTSPDGVDAAIQGGLARA AQTMRALDWFEVQSIRGHLVDGAVAHF QVTMKVGFRLEDS*
*mt*Dod-SZ1	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCAATCTAGTCGCTCAGCTAGAGAACGA AGTAGCATCATTAGAGAATGAAAACGAAACCTTGAAAAAGAAGAATCTACACAAAAAGG ATCTTATAGCCTACCTAGAAAAGGAAATTGCTAACTTAAGGAAAAAGATTGAGGAATGAg gatcccaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctga gcaataactagcataaccccttggggcctctaaacgggtcttgagggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SNLVAQLENEVASLENENETLKKKNLHK KDLIAYLEKEIANLRKKIEE*
*mt*Dod-SZ3	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTCGAGTCTCC AGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCAACGAAGTTACAACACTTGAGAATGAC GCTGCCTTTATCGAAAATGAAAATGCTTATCTAGAAAAAGAGATAGCACGTTTGAGAAA GGAGAAAGCAGCATTGAGAAATAGACTGGCACACAAAAAGTGAggatcccaccaccaccaccac cactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccct tggggcctctaaacgggtcttgagggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SNEVTTLENDAAFIENENAYLEKEIARLR KEKAALRNRLAHKK*

SZ1-*mt*Dod	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAATCTAGTCGCTCAGCTAGAGAACGAAGTAGCATCATTAGAGAATGAA AACGAAACCTTGAAAAAGAAGAATCTACACAAAAAGGATCTTATAGCCTACCTAGAAAA GGAAATTGCTAACTTAAGGAAAAAGATTGAGGAATCTCCAGCTGCGCCTGCTCCGGCA AGCCCTGCGAGCAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCC GACGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCG CGCGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGT CGCGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCTGAggatccca ccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaata actagcataacccttggggcctctaaacgggtcttgaggggttttttg	MNLVAQLENEVASLENENETLKKKNLHK KDLIAYLEKEIANLRKKIEESPAAPAPASP ASSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDS*
SZ3-*mt*Dod	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAACGAAGTTACAACACTTGAGAATGACGCTGCCTTTATCGAAAATGAAA ATGCTTATCTAGAAAAAGAGATAGCACGTTTGAGAAAGGAGAAAGCAGCATTGAGAAAT AGACTGGCACACAAAAAGTCTCCAGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCAGC AATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGACGGCGTCGACGCG GCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCGCGCTGGACTGGTTC GAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCGCGCACTTCCAGGTG ACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCTGAggatcccaccaccaccaccaccactgagatc cggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccctggggcctct aaacgggtcttgaggggttttttg	MNEVTTLENDAAFIENENAYLEKEIARLR KEKAALRNRLAHKKSPAAPAPASPASSN HTYRVIEIVGTSPDGVDAAIQGGLARAAQ TMRALDWFEVQSIRGHLVDGAVAHFQV TMKVGFRLEDS*
*mt*Dod-*se*ACP**	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGAGCAATCACACCTACCGAGTGATCGAGATCGTCGGGACCTCGCCCGA CGGCGTCGACGCGGCAATCCAGGGCGGTCTGGCCCGAGCTGCGCAGACCATGCGCG CGCTGGACTGGTTCGAAGTACAGTCAATTCGAGGCCACCTGGTCGACGGAGCGGTCG CGCACTTCCAGGTGACTATGAAAGTCGGCTTCCGCCTGGAGGATTCCCTcgagTCTCCA GCTGCGCCTGCTCCGGCAAGCCCTGCGAGCCGCGTAGACGAGATCGAGGACAAGTTG GGAAACTATATCCGCAGGCACCTGCTGACTGAGGACCCTCCAGAGGAATTCACTTACT CCACCGCCCTCTTCGGCGATGGGGTGCTGGATTCGCTCCGGCTGGCGATGCTGATCA ACTTCATCCGCAACGAGCTGGCCGTGGAGATCCCGTACGAGCACGTGAACCGGGACG ACTTCCACGATGTGCACACTATCGCCAAGATGGTGGTCGGCCTGTCGAGCGAAGCGAA ACTCGAGCATCATCACCACCACCACCACCACTGAgatccggctgctaacaaagcccgaaaggaagct gagttggctgctgccaccgctgagcaataactagcataacccttggggcctctaaacgggtcttgaggggttttttg	MSNHTYRVIEIVGTSPDGVDAAIQGGLA RAAQTMRALDWFEVQSIRGHLVDGAVA HFQVTMKVGFRLEDSLESPAAPAPASPA SRVDEIEDKLGNYIRRHLLTEDPPEEFTY STALFGDGVLDSLRLAMLINFIRNELAVEI PYEHVNRDDFHDVHTIAKMVVGLSSEAK LEHHHHHHHH*
non-dodecin *constructs*	DNA sequence T7 promotor to T7 terminator (CDS uppercase)	Amino acid sequence (linker)
SpyT-*se*ACP	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGGTCGTCATGGTTGATGCGTACAAACCGACCAAAGGTGGCAGC GGTTCTCCAGCTGCGCCTGCTCCGGCAAGCCCTGCGAGCCGCGTAGACGAGATCGAG GACAAGTTGGGAAACTATATCCGCAGGCACCTGCTGACTGAGGACCCTCCAGAGGAAT TCACTTACTCCACCGCCCTCTTCGGCGATGGGGTGCTGGATTCGCTCCGGCTGGCGAT GCTGATCAACTTCATCCGCAACGAGCTGGCCGTGGAGATCCCGTACGAGCACGTGAAC CGGGACGACTTCCACGATGTGCACACTATCGCCAAGATGGTGGTCGGCCTGTCGAGC GAAGCGAAACTCGAGCATCATCACCACCACCACCACCACTGAgatccggctgctaacaaagcccg aaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccttggggcctctaaacgggtcttgaggggttttt tg	MAHIVMVDAYKPTKGGSGSPAAPAPAS PASRVDEIEDKLGNYIRRHLLTEDPPEEF TYSTALFGDGVLDSLRLAMLINFIRNELA VEIPYEHVNRDDFHDVHTIAKMVVGLSS EAKLEHHHHHHHH*
*se*ACP-SpyC	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGCGCGTAGACGAGATCGAGGACAAGTTGGGAAACTATATCCGCAGGCA CCTGCTGACTGAGGACCCTCCAGAGGAATTCACTTACTCCACCGCCCTCTTCGGCGAT GGGGTGCTGGATTCGCTCCGGCTGGCGATGCTGATCAACTTCATCCGCAACGAGCTG GCCGTGGAGATCCCGTACGAGCACGTGAACCGGGACGACTTCCACGATGTGCACACT ATCGCCAAGATGGTGGTCGGCCTGTCGAGCGAAGCGAAAGGTGGCAGCGGTGCTATG GTGGACACCCTGTCGGGCCTCTCTAGTGAACAGGGGCAAAGCGGCGATATGACTATC GAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGACGGCAAAGAGTT AGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATTT CAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAATATACATTTGTCGAAACC GCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGG TCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCTCATATTCTCGAGCATCATC ACCACCACCACCACCACTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgct gagcaataactagcataacccttggggcctctaaacgggtcttgaggggttttttgctgaaag	MRVDEIEDKLGNYIRRHLLTEDPPEEFTY STALFGDGVLDSLRLAMLINFIRNELAVEI PYEHVNRDDFHDVHTIAKMVVGLSSEAK GGSGAMVDTLSGLSSEQGQSGDMTIEE DSATHIKFSKRDEDGKELAGATMELRDS SGKTISTWISDGQVKDFYLYPGKYTFVE TAAPDGYEVATAITFTVNEQGQVTVNGK ATKGDAHILEHHHHHHHH*
mClover3-SnpC	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG TCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTCCGCGGCGAGGGCGAG GGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG CCCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCGTGGCCTGCTTCAGC CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCT ACGTCCAGGAGCGCACCATCTCTTTCAAGGACGACGGTACCTACAAGACCCGCGCCG AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACT TCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACTA CGTCTATATCACGGCCGACAAGCAGAAGAACTGCATCAAGGCTAACTTCAAGATCCGC CACAACGTTGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCCATCAGTCCAAG CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACC GCCGCCGGCATTACCCATGGCATGGATGAACTGTATAAAGGTGGCAGCGGTAGCGGT AGCGGCAAGCCGCTGCGTGGTGCCGTGTTTAGCCTGCAGAAACAGCATCCCGACTAT CCCGATATCTATGGCGCGATTGATCAGAATGGGACCTATCAAAATGTGCGTACCGGCG AAGATGGTAAACTGACCTTTAAGAATCTGAGCGATGGCAAATATCGCCTGTTTGAAAAT AGCGAACCCGCTGGCTATAAACCGGTGCAGAATAAGCCGATTGTGGCGTTTCAGATTG TGAATGGCGAAGTGCGTGATGTGACCAGCATTGTGCCGCAGGATATTCCGGCTACATA TGAATTTACCAACGGTAAACATTATATCACCAATGAACCGATACCGCCGAAACTCGAGC ATCATCACCACCACCACCACCACTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctg ccaccgctgagcaataactagcataacccttggggcctctaaacgggtcttgaggggttttttgctgaaag	MVSKGEELFTGVVPILVELDGDVNGHKF SVRGEGEGDATNGKLTLKFICTTGKLPV PWPTLVTTFGYGVACFSRYPDHMKQHD FFKSAMPEGYVQERTISFKDDGTYKTRA EVKFEGDTLVNRIELKGIDFKEDGNILGH KLEYNFNSHYVYITADKQKNCIKANFKIR HNVEDGSVQLADHYQQNTPIGDGPVLL PDNHYLSHQSKLSKDPNEKRDHMVLLE FVTAAGITHGMDELYKGGSGSGSGKPL RGAVFSLQKQHPDYPDIYGAIDQNGTYQ NVRTGEDGKLTFKNLSDGKYRLFENSEP AGYKPVQNKPIVAFQIVNGEVRDVTSIVP QDIPATYEFTNGKHYITNEPIPPKLEHHH HHHHH*

SpyT-mClover3	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGGCACATATCGTCATGGTTGATGCGTACAAACCGACCAAAGGTGGCAGC GGTTCTGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAG CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTCCGCGGCGAGGGCGAGGGCGA TGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGT GCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCGTGGCCTGCTTCAGCCGCTA CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC CAGGAGCGCACCATCTCTTTCAAGGACGACGGTACCTACAAGACCCGCGCCGAGGTG AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG GAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACTACGTCT ATATCACGGCCGACAAGCAGAAGAACTGCATCAAGGCTAACTTCAAGATCCGCCACAA CGTTGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGG CGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCCATCAGTCCAAGCTGAG CAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC CGGCATTACCCATGGCATGGATGAACTGTATAAAcTCGAGCATCATCACCACCACCACC ACCACTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcata accccttggggcctctaaacgggtcttgaggggtttttg	MAHIVMVDAYKPTKGGSGSVSKGEELFT GVVPILVELDGDVNGHKFSVRGEGEGD ATNGKLTLKFICTTGKLPVPWPTLVTTFG YGVACFSRYPDHMKQHDFFKSAMPEGY VQERTISFKDDGTYKTRAEVKFEGDTLV NRIELKGIDFKEDGNILGHKLEYNFNSHY VYITADKQKNCIKANFKIRHNVEDGSVQL ADHYQQNTPIGDGPVLLPDNHYLSHQSK LSKDPNEKRDHMVLLEFVTAAGITHGMD ELYKLEHHHHHHHH*
SZ2-mRuby3	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGGCCAGAAATGCATACTTAAGGAAAAAGATTGCTAGATTGAAAAAGGAC AACTTACAATTAGAAAGAGATGAGCAAAATCTTGAAAAGATCATTGCCAATTTGAGAGAT GAAATCGCCAGACTTGAAAATGAGGTGGCCTCTCATGAACAAGGTGGCAGCGGTGTGT CTAAGGGCGAAGAGCTGATCAAGGAAAATATGCGTATGAAGGTGGTCATGGAAGGTTC GGTCAACGGCCACCAATTCAAATGCACAGGTGAAGGAGAAGGCAGACCGTACGAGGG AGTGCAAACCATGAGGATCAAAGTCATCGAGGGAGGACCCCTGCCATTTGCCTTTGAC ATTCTTGCCACGTCGTTCATGTATGGCAGCCGTACCTTTATCAAGTACCCGGCCGACAT CCCTGATTTCTTTAAACAGTCCTTTCCTGAGGGTTTTACTTGGGAAAGAGTTACGAGATA CGAAGATGGTGGAGTCGTCACCGTCACGCAGGACACCAGCCTTGAGGATGGCGAGCT CGTCTACAACGTCAAGGTCAGAGGGGTAAACTTTCCCTCCAATGGTCCCGTGATGCAG AAGAAGACCAAGGGTTGGGAGCCTAATACAGAGATGATGTATCCAGCAGATGGTGGTC TGAGAGGATACACTGACATCGCACTGAAAGTTGATGGTGGTGGCCATCTGCACTGCAA CTTCGTGACAACTTACAGGTCAAAAAAGACCGTCGGGAACATCAAGATGCCCGGTGTC CATGCCGTTGATCACCGCCTGGAAAGGATCGAGGAGAGTGACAATGAAACCTACGTAG TGCAAAGAGAAGTGGCAGTTGCCAAATACAGCAACCTTGGTGGTGGCATGGACGAGCT GTACAAGCTCGAGCATCATCACCACCACCACCACCACTGAgatccggctgctaacaaagcccgaa aggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggtttttg	MARNAYLRKKIARLKKDNLQLERDEQNL EKIIANLRDEIARLENEVASHEQGGSGVS KGEELIKENMRMKVVMEGSVNGHQFKC TGEGEGRPYEGVQTMRIKVIEGGPLPFA FDILATSFMYGSRTFIKYPADIPDFFKQSF PEGFTWERVTRYEDGGVVTVTQDTSLE DGELVYNVKVRGVNFPSNGPVMQKKTK GWEPNTEMMYPADGGLRGYTDIALKVD GGGHLHCNFVTTYRSKKTVGNIKMPGV HAVDHRLERIEESDNETYVVQREVAVAK YSNLGGGMDELYKLEHHHHHHHH*
SZ2-mClover3	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGGCCAGAAATGCATACTTAAGGAAAAAGATTGCTAGATTGAAAAAGGAC AACTTACAATTAGAAAGAGATGAGCAAAATCTTGAAAAGATCATTGCCAATTTGAGAGAT GAAATCGCCAGACTTGAAAATGAGGTGGCCTCTCATGAACAAGGTGGCAGCGGTGTGA GCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC GACGTAAACGGCCACAAGTTCAGCGTCCGCGGCGAGGGCGAGGGCGATGCCACCAA CGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCC CACCCTCGTGACCACCTTCGGCTACGGCGTGGCCTGCTTCAGCCGCTACCCCGACCA CATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCG CACCATCTCTTTCAAGGACGACGGTACCTACAAGACCCGCGCCGAGGTGAAGTTCGAG GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC AACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACTACGTCTATATCACGG CCGACAAGCAGAAGAACTGCATCAAGGCTAACTTCAAGATCCGCCACAACGTTGAGGA CGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCC CGTGCTGCTGCCCGACAACCACTACCTGAGCCATCAGTCCAAGCTGAGCAAAGACCCC AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGCATTACC CATGGCATGGATGAACTGTATAAACTCGAGCATCATCACCACCACCACCACCACTGAgat ccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctc taaacgggtcttgaggggtttttg	MARNAYLRKKIARLKKDNLQLERDEQNL EKIIANLRDEIARLENEVASHEQGGSGVS KGEELFTGVVPILVELDGDVNGHKFSVR GEGEGDATNGKLTLKFICTTGKLPVPWP TLVTTFGYGVACFSRYPDHMKQHDFFK SAMPEGYVQERTISFKDDGTYKTRAEVK FEGDTLVNRIELKGIDFKEDGNILGHKLE YNFNSHYVYITADKQKNCIKANFKIRHNV EDGSVQLADHYQQNTPIGDGPVLLPDN HYLSHQSKLSKDPNEKRDHMVLLEFVTA AGITHGMDELYKLEHHHHHHHH*
SZ4-mRuby3	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGCAGAAAGTGGCTGAATTGAAAAACAGAGTTGCTGTTAAACTTAACAGAA ATGAACAATTGAAAAACAAGGTAGAAGAGTTGAAAAACCGTAATGCTTACCTGAAAAAC GAACTGGCTACATTAGAAAATGAAGTCGCCAGATTGGAGAACGATGTTGCTGAAGGTG GCAGCGGTGTGTCTAAGGGCGAAGAGCTGATCAAGGAAAATATGCGTATGAAGGTGGT CATGGAAGGTTCGGTCAACGGCCACCAATTCAAATGCACAGGTGAAGGAGAAGGCAGA CCGTACGAGGGAGTGCAAACCATGAGGATCAAAGTCATCGAGGGAGGACCCCTGCCA TTTGCCTTTGACATTCTTGCCACGTCGTTCATGTATGGCAGCCGTACCTTTATCAAGTAC CCGGCCGACATCCCTGATTTCTTTAAACAGTCCTTTCCTGAGGGTTTTACTTGGGAAAG AGTTACGAGATACGAAGATGGTGGAGTCGTCACCGTCACGCAGGACACCAGCCTTGAG GATGGCGAGCTCGTCTACAACGTCAAGGTCAGAGGGGTAAACTTTCCCTCCAATGGTC CCGTGATGCAGAAGAAGACCAAGGGTTGGGAGCCTAATACAGAGATGATGTATCCAGC AGATGGTGGTCTGAGAGGATACACTGACATCGCACTGAAAGTTGATGGTGGTGGCCAT CTGCACTGCAACTTCGTGACAACTTACAGGTCAAAAAAGACCGTCGGGAACATCAAGAT GCCCGGTGTCCATGCCGTTGATCACCGCCTGGAAAGGATCGAGGAGAGTGACAATGA AACCTACGTAGTGCAAAGAGAAGTGGCAGTTGCCAAATACAGCAACCTTGGTGGTGGC ATGGACGAGCTGTACAAGCTCGAGCATCATCACCACCACCACCACCACTGAgatccggctg ctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgg gtcttgaggggtttttg	MQKVAELKNRVAVKLNRNEQLKNKVEE LKNRNAYLKNELATLENEVARLENDVAE GGSGVSKGEELIKENMRMKVVMEGSVN GHQFKCTGEGEGRPYEGVQTMRIKVIE GGPLPFAFDILATSFMYGSRTFIKYPADI PDFFKQSFPEGFTWERVTRYEDGGVVT VTQDTSLEDGELVYNVKVRGVNFPSNG PVMQKKTKGWEPNTEMMYPADGGLRG YTDIALKVDGGGHLHCNFVTTYRSKKTV GNIKMPGVHAVDHRLERIEESDNETYVV QREVAVAKYSNLGGGMDELYKLEHHHH HHHH*
SZ4-mClover3	cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaag gagatatacatATGCAGAAAGTGGCTGAATTGAAAAACAGAGTTGCTGTTAAACTTAACAGAA ATGAACAATTGAAAAACAAGGTAGAAGAGTTGAAAAACCGTAATGCTTACCTGAAAAAC GAACTGGCTACATTAGAAAATGAAGTCGCCAGATTGGAGAACGATGTTGCTGAAGGTG GCAGCGGTGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTCCGCGGCGAGGGCGAGGG CGATGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCC CGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCGTGGCCTGCTTCAGCCG CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC GTCCAGGAGCGCACCATCTCTTTCAAGGACGACGGTACCTACAAGACCCGCGCCGAG GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTC AAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACTAC GTCTATATCACGGCCGACAAGCAGAAGAACTGCATCAAGGCTAACTTCAAGATCCGCC ACAACGTTGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA TCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCCATCAGTCCAAGC TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCG CCGCCGGCATTACCCATGGCATGGATGAACTGTATAAACTCGAGCATCATCACCACCA CCACCACCACTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataac tagcataaccccttggggcctctaaacgggtcttgaggggtttttg	MQKVAELKNRVAVKLNRNEQLKNKVEE LKNRNAYLKNELATLENEVARLENDVAE GGSGVSKGEELFTGVVPILVELDGDVNG HKFSVRGEGEGDATNGKLTLKFICTTGK LPVPWPTLVTTFGYGVACFSRYPDHMK QHDFFKSAMPEGYVQERTISFKDDGTYK TRAEVKFEGDTLVNRIELKGIDFKEDGNI LGHKLEYNFNSHYVYITADKQKNCIKANF KIRHNVEDGSVQLADHYQQNTPIGDGPV LLPDNHYLSHQSKLSKDPNEKRDHMVLL EFVTAAGITHGMDELYKLEHHHHHHHH*

*mm*ACP	cccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaag gagatatacatATGAGCGCTTGGAGCCACCCGCAGTTCGAAAAAGGCGCCGGGGACGGGG ACACCCAGAGGGATCTGGTGAAAGCTGTAGCACATCCTAGGCATCCGAGACCTCGC AGGTATTAACCTGGACAGCACGCTGGCAGACCTCGGCCTGGACTCGCTCATGGGTGT GGAAGTTCGTCAGATCCTGGAACGAGAACACGATCTGGTGCTGCCCATGCGTGAGGT GCGGCAGCTCACGCTGCGGAAACTTCAGGAAATGTCCTCCAAGACTGACTCGGCTACT GACACGACAGCCCCCCTCGAGCATCATCACCACCACCACCACTGAgatccggctgctaa caaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccctggggcctctaaacgggtctt gagggggttttttg	MSAWSHPQFEKGAGDGDTQRDLVKAV AHILGIRDLAGINLDSTLADLGLDSLMGV EVRQILEREHDLVLPMREVRQLTLRKLQ EMSSKTDSATDTTAPLEHHHHHHHH*
Sfp	tcataaaaaatttatttgctttgtgagcggataacaattataatagattcaattgtgagcggataacaatttcacacagaattctgcag acggaggatctagaATGAAGATTTACGGAATTTATATGGACCGCCCGCTTTCACAGGAAGAAA ATGAACGGTTCATGACTTTCATATCACCTGAAAAACGGGAGAAATGCCGGAGATTTTAT CATAAAGAAGATGCTCACCGCACCCTGCTGGGAGATGTGCTCGTTCGCTCAGTCATAA GCAGGCAGTATCAGTTGGACAAATCCGATATCCGCTTTAGCACGCAGGAATACGGGAA GCCGTGCATCCCTGATCTTCCCGACGCTCATTTCAACATTTCTCACTCCGGCCGCTGG GTCATTGGTGCGTTTGATTCACAGCCGATCGGCATAGATATCGAAAAAACGAAACCGAT CAGCCTTGAGATCGCCAAGCGCTTCTTTTCAAAAACAGAGTACAGCGACCTTTTAGCAA AAGACAAGGACGAGCAGACAGACTATTTTTATCATCTATGGTCAATGAAAGAAAGCTTT ATCAAACAGGAAGGCAAAGGCTTATCGCTCCCGCTTGATTCCTTTTCAGTGCGCCTGCA TCAGGACGGACAAGTATCCATTGAGCTTCCGGACAGCCATTCCCCATGCTATATCAAAA CGTATGAGGTCGATCCCGGCTACAAAATGGCTGTATGCGCCGCACACCCTGATTTCCC CGAGGATATCACAATGGTCTCGTACGAAGAGCTTTTAAGATCTCATCACCATCACCATC ACTAAgcttaattagctgagcttggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgctc	MKIYGIYMDRPLSQEENERFMTFISPEKR EKCRRFYHKEDAHRTLLGDVLVRSVISR QYQLDKSDIRFSTQEYGKPCIPDLPDAH FNISHSGRWVIGAFDSQPIGIDIEKTKPIS LEIAKRFFSKTEYSDLLAKDKDEQTDYFY HLWSMKESFIKQEGKGLSLPLDSFSVRL HQDGQVSIELPDSHSPCYIKTYEVDPGY KMAVCAAHPDFPEDITMVSYEELLRSHH HHHH*

Table S2: Amino acid sequence of proteins from which the peptide sequences were selected for the *mt*Dod-PAS-Pep constructs. Selected peptide sequences highlighted in red.

Construct & peptide	Identifier & name	Amino acid sequence (peptide)
*mt*Dod-PAS-Pep1 PKGGSGSGPTIEEVD	AAD21816.1 HSP70-1 [Homo sapiens]	MAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVALNPQNTVFDAKRLIGRKF GDPVVQSDMKHWPFQVINDGDKPKVQVSYKGDTKAFYPEEISSMVLTKMKEIAEAYLGYPVTNAVITVPAYFNDSQ RQATKDAGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDLGGGTFDVSILTIDDGIFEVKATAGDTHLGGEDF DNRLVNHFVEEFKRKHKKDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEELCSDLFR STLEPVEKALRDAKLDKAQIHDLVLVGGSTRIPKVQKLLQDFFNGRDLNKSINPDEAVAYGAAVQAAILMGDKSENV QDLLLLDVAPLSLGLETAGGVMTALIKRNSTIPTKQTQIFTTYSDNQPGVLIQVYEGERAMTKDNNLLGRFELSGIPP APRGVPQIEVTFDIDANGILNVTATDKSTGKANKITITNDKGRLSKEEIERMVQEAEKYKAEDEVQRERVSAKNALES YAFNMKSAVEDEGLKGKISEADKKKVLDKCQEVISWLDANTLAEKDEFEHKRKELEQVCNPIISGLYQGAGGPGPG GFGAQGPKGGSGSGPTIEEVD
*mt*Dod-PAS-Pep2 PLEGDDDTSRMEEVD	NP_001017963.2 heat shock protein HSP90-alpha isoform 1 [Homo sapiens]	MPPCSGGDGSTPPGPSLRDRDCPAQSAEYPRDRLDPRPGSPSEASSPPFLRSRAPVNWYQEKAQVFLWHLMVS GSTTLLCLWKQPFHVSAFPVTASLAFRQSQGAGQHLYKDLQPFILLRLLMPEETQTQDQPMEEEEVETFAFQAEIA QLMSLIINTFYSNKEIFLRELISNSSDALDKIRYESLTDPSKLDSGKELHINLIPNKQDRTLTIVDTGIGMTKADLINNLGT IAKSGTKAFMEALQAGADISMIGQFGVGFYSAYLVAEKVTVITKHNDDEQYAWESSAGGSFTVRTDTGEPMGRGTK VILHLKEDQTEYLEERRIKEIVKKHSQFIGYPITLFVEKERDKEVSDDEAEEKEDKEEEKEKEEKESEDKPEIEDVGSD EEEEKKDGDKKKKKKIKEKYIDQEELNKTKPIWTRNPDDITNEEYGEFYKSLTNDWEDHLAVKHFSVEGQLEFRALL FVPRRAPFDLFENRKKKNNIKLYVRRVFIMDNCEELIPEYLNFIRGVVDSEDLPLNISREMLQQSKILKVIRKNLVKKC LELFTELAEDKENYKKFYEQFSKNIKLGIHEDSQNRKKLSELLRYYTSASGDEMVSLKDYCTRMKENQKHIYYITGET KDQVANSAFVERLRKHGLEVIYMIEPIDEYCVQQLKEFEGKTLVSVTKEGLELPEDEEEKKKQEEKKTKFENLCKIMK DILEKKVEKVVVSNRLVTSPCCIVTSTYGWTANMERIMKAQALRDNSTMGYMAAKKHLEINPDHSIIETLRQKAEAD KNDKSVKDLVILLYETALLSSGFSLEDPQTHANRIYRMIKLGLGIDEDDPTADDTSAAVTEEMPPLEGDDDTSRMEE VD
*mt*Dod-PAS-Pep3 ECYPNEKNSVNMDLD	NP_006635.2 heat shock protein 105 kDa isoform 1 [Homo sapiens]	MSVVGLDVGSQSCYIAVARAGGIETIANEFSDRCTPSVISFGSKNRTIGVAAKNQQITHANNTVSNFKRFHGRAFND PFIQKEKENLSYDLVPLKNGGVGIKVMYMGEEHLFSVEQITAMLLTKLKETAENSLKKPVTDCVISVPSFFTDAERRS VLDAAQIVGLNCLRLMNDMTAVALNYGIYKQDLPSLDEKPRIVVFVDMGHSAFQVSACAFNKGKLKVLGTAFDPTLG GKNFDEKLVEHFCAEFKTKYKLDAKSKIRALLRLYQECEKLKKLMSSNSTDLPLNIECFMNDKDVSGKMNRSQFEEL CAELLQKIEVPLYSLLEQTHLKVEDVSAVEIVGGATRIPAVKERIAKFFGKDISTTLNADEAVARGCALQCAILSPAFKV REFSVTDAVPFPISLIWNHDSEDTEGVHEVFSRNHAAPFSKVLTFLRRGPFELEAFYSDPQGVPYPEAKIGRFVVQN VSAQKDGEKSRVKVKVRVNTHGIFTISTASMVEKVPTEENEMSSEADMECLNQRPPENPDTDKNVQQDNSEAGTQ PQVQTDAQQTSQSPPSPELTSEENKIPDADKANEKKVDQPPEAKKPKIKVVNVELPIEANLVWQLGKDLLNMYIETE GKMIMQDKLEKERNDAKNAVEEYVYEFRDKLCGPYEKFICEQDHQNFLRLLTETEDWLYEEGEDQAKQAYVDKLE ELMKIGTPVKVRFQEAEERPKMFEELGQRLQHYAKIAADFRNKDEKYNHIDESEMKKVEKSVNEVMEWMNNVMNA QAKKSLDQDPVVRAQEIKTKIKELNNTCEPVVTQPKPKIESPKLERTPNGPNIDKKEEDLEDKNNFGAEPPHQNGEC YPNEKNSVNMDLD
*mt*Dod-PAS-Pep4 VPSDSDKKLPEMDID	EAW62295.1 heat shock 70kDa protein 4, isoform CRA_b [Homo sapiens]	MSVVGIDLGFQSCYVAVARAGGIETIANEYSDRCTPACISFGPKNRSIGAAAKSQVISNAKNTVQGFKRFHGRAFSD PFVEAEKSNLAYDIVQLPTGLTGIKVTYMEEERNFTTEQVTAMLLSKLKETAESVLKKPVVDCVVSVPCFYTDAERR SVMDATQIAGLNCLRLMNETTAVALAYGIYKQDLPALEEKPRNVVFVDMGHSAYQVSVCAFNRGKLKVLATAFDTT LGGRKFDEVLVNHFCEEFGKKYKLDIKSKIRALLRLSQECEKLKKLMSANASDLPLSIECFMNDVDVSGTMNRGKFL EMCNDLLARVEPPLRSVLEQTKLKKEDIYAVEIVGGATRIPAVKEKISKFFGKELSTTLNADEAVTRGCALQCAILSPA FKVREFSITDVVPYPISLRWNSPAEEGSSDCEVFSKNHAAPFSKVLTFYRKEPFTLEAYYSSPQDLPYPDPAIEKMQ VDQEEPHVEEQQQQTPAENKAESEEMETSQAGSKDKKMDQPPQAKKAKVKTSTVDLPIENQLLWQIDREMLNLYI ENEGKMIMQDKLEKERNDAKNAVEEYVYEMRDKLSGEYEKFVSEDDRNSFTLKLEDTENWLYEDGEDQPKQVYV DKLAELKNLGQPIKIRFQESEERPKLFEELGKQIQQYMKIISSFKNKEDQYDHLDAADMTKVEKSTNEAMEWMNNKL NLQNKQSLTMDPVVKSKEIEAKIKELTSTCSPIISKPKPKVEPPKEEQKNAEQNGPVDGQGDNPGPQAAEQGTDTA VPSDSDKKLPEMDID
*mt*Dod-PAS-Pep5 DSSQHTKSSGEMEVD	NP_055093.2 heat shock 70 kDa protein 4L isoform 1 [Homo sapiens]	MSVVGIDLGFLNCYIAVARSGGIETIANEYSDRCTPACISLGSRTRAIGNAAKSQIVTNVRNTIHGFKKLHGRSFDDPI VQTERIRLPYELQKMPNGSAGVKVRYLEEERPFAIEQVTGMLLAKLKETSENALKKPVADCVISIPSFFTDAERRSV MAAAQVAGLNCLRLMNETTAVALAYGIYKQDLPPLDEKPRNVVFIDMGHSAYQVLVCAFNKGKLKVLATTFDPYLG GRNFDEALVDYFCDEFKTKYKINVKENSRALLRLYQECEKLKKLMSANASDLPLNIECFMNDLDVSSKMNRAQFEQ LCASLLARVEPPLKAVMEQANLQREDISSIEIVGGATRIPAVKEQITKFFLKDISTTLNADEAVARGCALQCAILSPAFK VREFSITDLVPYSITLRWKTSFEDGSGECEVFCKNHPAPFSKVITFHKKEPFELEAFYTNLHEVPYPDARIGSFTIQNV FPQSDGDSSKVKVKVRVNIHGIFSVASASVIEKQNLEGDHSDAPMETETSFKNENKDNMDKMQVDQEEGHQKCHA EHTPEEEIDHTGAKTKSAVSDKQDRLNQTLKKGKVKSIDLPIQSSLCRQLGQDLLNSYIENEGKMIMQDKLEKERND AKNAVEEYVYDFRDRLGTVYEKFITPEDLSKLSAVLEDTENWLYEDGEDQPKQVYVDKLQELKKYGQPIQMKYMEH EERPKALNDLGKKIQLVMKVIEAYRNKDERYDHLDPTEMEKVEKCISDAMSWLNSKMNAQNKLSLTQDPVVKVSEI VAKSKELDNFCNPIIYKPKPKAEVPEDKPKANSEHNGPMDGQSGTETKSDSTKDSSQHTKSSGEMEVD
*mt*Dod-PAS-Pep6 EQSTGQKRPLKNDEL	XP_005271449.1 hypoxia up-regulated protein 1 isoform X1 [Homo sapiens]	MADKVRRQRPRRRVCWALVAVLLADLLALSDTLAVMSVDLGSESMKVAIVKPGVPMEIVLNKESRRKTPVIVTLKE NERFFGDSAASMAIKNPKATLRYFQHLLGKQADNPHVALYQARFPEHELTFDPQRQTVHFQISSQLQFSPEEVLGM VLNYSRSLAEDFAEQPIKDAVITVPVFFNQAERRAVLQAARMAGLKVLQLINDNTATALSYGVFRRKDINTTAQNIMF YDMGSGSTVCTIVTYQMVKTKEAGMQPQLQIRGVGFDRTLGGLEMELRLRERLAGLFNEQRKGQRAKDVRENPR AMAKLLREANRLKTVLSANADHMAQIEGLMDDVDFKAKVTRVEFEELCADLFERVPGPVQQALQSAEMSLDEIEQV ILVGGATRVPRVQEVLLKAVGKEELGKNINADEAAAMGAVYQAAALSKAFKVKPFVVRDAVVYPILVEFTREVEEEP GIHSLKHNKRVLFSRMGPYPQRKVITFNRYSHDFNFHINYGDLGFLGPEDLRVFGSQNLTTVKLKGVGDSFKKYPD YESKGIKAHFNLDESGVLSLDRVESVFETLVEDSAEEESTLTKLGNTISSLFGGGTTPDAKENGTDTVQEEEESPAE GSKDEPGEQVELKEEAEAPVEDGSQPPPPEPKGDATPEGEKATEKENGDKSEAQKPSEKAEAGPEGVAPAPEGE KKQKPARKRRMVEEIGVELVVLDLPDLPEDKLAQSVQKLQDLTLRDLEKQEREKAANSLEAFIFETQDKLYQPEYQE VSTEEQREEISGKLSAASTWLEDEGVGATTVMLKEKLAELRKLCQGLFFRVEERKKWPERLSALDNLLNHSSMFLK GARLIPEMDQIFTEVEMTTLEKVINETWAWKNATLAEQAKLPATEKPVLLSKDIEAKMMALDREVQYLLNKAKFTKP RPRPKDKNGTRAEPPLNASASDQGEKVIPPAGQTEDAEPISEPEKVETAGSEPGDTEPLELGGPGAEPEQKEQST GQKRPLKNDEL
*mt*Dod-PAS-Pep8 LVTGESLEQLRRGLA	sp|Q09118.1|HBEGF_CHLAE	MKLLPSVVLKLLLAAVLSALVTGESLEQLRRGLAAGTSNPDPSTGSTDQLLRLGGGRDRKVRDLQEADLDLLRVTL SSKPQALATPSKEEHGKRKKKGKGLGKKRDPCLRKYKDFCIHGECKYVKELRAPSCICHPGYHGERCHGLSLPVE NRLYTYDHTTILAVVAVVLSSVCLLVIVGLLMFRYHRRGGYDVENEEKVKLGMTNSH
*mt*Dod-PAS-Pep9 MKGKEEKEGGARLGA	NP_005852.2 E3 ubiquitin-protein ligase CHIP isoform a [Homo sapiens]	MKGKEEKEGGARLGAGGGSPEKSPSAQELKEQGNRLFVGRKYPEAAACYGRAITRNPLVAVYYTNRALCYLKMQ QHEQALADCRRALELDGQSVKAHFFLGQCQLEMESYDEAIANLQRAYSLAKEQRLNFGDDIPSALRIAKKKRWNSI EERRIHQESELHSYLSRLIAAERERELEECQRNHEGDEDDSHVRAQQACIEAKHDKYMADMDELFSQVDEKRKKR DIPDYLCGKISFELMREPCITPSGITYDRKDIEEHLQRVGHFDPVTRSPLTQEQLIPNLAMKEVIDAFISENGWVEDY
*mt*Dod-PAS-Pep10 EERRIHQESE	NP_005852.2 E3 ubiquitin-protein ligase CHIP isoform a [Homo sapiens] Pos. 151-160	MKGKEEKEGGARLGAGGGSPEKSPSAQELKEQGNRLFVGRKYPEAAACYGRAITRNPLVAVYYTNRALCYLKMQ QHEQALADCRRALELDGQSVKAHFFLGQCQLEMESYDEAIANLQRAYSLAKEQRLNFGDDIPSALRIAKKKRWNSI EERRIHQESELHSYLSRLIAAERERELEECQRNHEGDEDDSHVRAQQACIEAKHDKYMADMDELFSQVDEKRKKR DIPDYLCGKISFELMREPCITPSGITYDRKDIEEHLQRVGHFDPVTRSPLTQEQLIPNLAMKEVIDAFISENGWVEDY
*mt*Dod-PAS-Pep11 NHEGDEDDSH	NP_005852.2 E3 ubiquitin-protein ligase CHIP isoform a [Homo sapiens] Pos. 183-192	MKGKEEKEGGARLGAGGGSPEKSPSAQELKEQGNRLFVGRKYPEAAACYGRAITRNPLVAVYYTNRALCYLKMQ QHEQALADCRRALELDGQSVKAHFFLGQCQLEMESYDEAIANLQRAYSLAKEQRLNFGDDIPSALRIAKKKRWNSI EERRIHQESELHSYLSRLIAAERERELEECQRNHEGDEDDSHVRAQQACIEAKHDKYMADMDELFSQVDEKRKKR DIPDYLCGKISFELMREPCITPSGITYDRKDIEEHLQRVGHFDPVTRSPLTQEQLIPNLAMKEVIDAFISENGWVEDY

Literature:

1. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L.-Y. A Simple Method for Improving Protein Solubility and Long-Term Stability. *J. Am. Chem. Soc.* **126**, 8933–8939 (2004).
2. Bajar, B. T. *et al.* Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci. Rep.* **6**, (2016).
3. Veggiani, G. *et al.* Programmable polyproteams built using twin peptide superglues. *Proc. Natl. Acad. Sci.* **113**, 1202–1207 (2016).
4. Kick, A., Bönsch, M. & Mertig, M. EGNAS: an exhaustive DNA sequence design algorithm. *BMC Bioinformatics* **13**, 138 (2012).

5 Discussion and Conclusion

This chapter is divided into two core topics to avoid redundancy and to give a clearer picture of the aspects of the presented papers (the in Chapter 4 listed papers are cited with first author name and publication year in addition to the reference numbering).

The first theme is: the discussion of the biological function of dodecins based on the studied flavin binding process of *Mt*Dod (Bourdeaux *et al.* 2018[55]) and to a lesser degree of *Sd*Dod and *Sc*Dod (Bourdeaux *et al.* 2019[53]). The biochemical analysis of these dodecins is further supported by a study on the effect of a dodecin gene deletion in *S. davaonensis* showing a specific phenotype (Ludwig *et al.* 2018[235]).

The second theme is: the use of dodecins as scaffold proteins or nanoparticles. Here, based on the example of *Mt*Dod as a scaffold protein (Bourdeaux *et al.* 2020[236]) and *Hs*Dod as a diffusion probe (Nöll *et al.* 2018[237]), the potential of bacterial dodecins as highly modifiable nanoparticles will be discussed. By comparing *Mt*Dod with other scaffold proteins/peptides the strengths and weaknesses will be highlighted.

5.1 Flavin Binding Model and Biological Function

The main question of this chapter can be summarized into the following statement:

“We know that dodecins bind flavins, but we don’t know why.”

While there is no definitive answer to this question so far and there also might be distinct biological roles for dodecins dependent on the individual species, the establishment of the two-step binding model allows to gain a deeper mechanistic understanding of the flavin binding, which can be used to decipher the potential impact of dodecins on the cell.

Beforehand, as stated in Chapter 2.2.1, dodecins are not considered to be enzymes, since no catalytic function was reported so far and the design of the binding pocket prevents that the reactive site (N5-position)[43] of the bound flavins can be accessed. Further, *Mt*Dod can only bind oxidised flavins, (Bourdeaux *et al.* 2018[55]), which would require a new flavin molecule to be bound after every catalysed reaction, which is rather untypical for flavoenzymes.[49]

Therefore, the here proposed and discussed biological function of bacterial dodecins is:

Bacterial dodecins protect cells from potential oxidative damage caused by high cytosolic concentrations of free oxidised flavins. This is accomplished by acting as an interim storage device or buffer for oxidised flavins, which captures excessive amounts of unbound flavins, while keeping the amount of free flavins on a level that allows the flavoproteome to function. In addition, under certain conditions, bacterial dodecins shift to a tighter binding mode capturing most available flavins to keep them bound until the conditions change again (storage). (Bourdeaux *et al.* 2018[55])

For the sake of the argument, the functions as a buffer or as a storage will be divided more strictly. Here, the "buffer function" means to keep flavin levels constant during fluctuating flavin demand, while the "storage function" is only utilized when more drastic changes appear, like for example switching from dormancy to an active state. This said, a buffer is always able to also function as a storage, although limited, but a strict storage cannot function as a buffer. In general, a buffer has moderate affinity for its ligand, while a storage binds its ligand tightly. Since both systems will bind flavins, the cells are likely to produce more flavin to compensate for bound flavin, here referred to as flavin sequestering, since the cellular flavin content increases.

Before the likelihood of this role will be discussed, it will be explained how the bacterial dodecins fulfil this buffering function by focussing on the flavin binding mechanism of dodecins.

5.1.1 The Flavin Binding Mechanism of Dodecins

The two-step binding model discussed in the publication of Bourdeaux *et al.* 2018[55] will be compared to the single-step model, which is typically presented and used so far in literature.[5,13,14] The focus here is to highlight how the binding mechanism itself can affect the function of a protein.

The dodecin dodecamer contains six identical flavin binding pockets with two flavin binding sites each, allowing a single dodecamer to bind twelve flavins or simplified one flavin per monomer. The simplification "one flavin per monomer" instead of the actual "twelve flavins per dodecamer" allows the binding mechanism to be described in a single binding step (**Figure 18** a). While this simplification is practical, it would require that both bound flavins

are not interacting with each other, which can directly be declared as false, based on the formed aromatic tetrad in the binding pocket.[1,5,13] It would also be very unlikely that the half-filled binding pocket is sterically not more restricted than the empty binding pocket. That the single-step binding model clearly has a problem describing the flavin binding, can be seen in **Table 1**, since depending on the used method, vastly different dissociation constants were obtained.

Table 1: Dissociation constants of diverse dodecins reported in literature based on the single-step model. Because of the vast differences between methods, errors are omitted, and actual values depicted here should only be seen as an estimate of affinity. Methods: Fluor. Titration: Constant amount of flavin with dodecin titrated. Kinetic analysis: time-dependent fluorescence decrease of different dodecin:flavin ratios measured. ITC: Isothermal titration calorimetry; constant amount of dodecin with flavin titrated. N.b.: no binding observed.

Dodecin (method and reference)	pH	K_D (single-step) /nM			
		LmF	RbF	FMN	FAD
*Mt*Dod (fluor. titration) Bourdeaux *et al.* 2018[55]	7.5	-	393	118	n.b.
	5.0	-	598	18	157
*Mt*Dod (kinetic analysis) Bourdeaux *et al.* 2018[55]	7.5	-	900-1530	263-1473	-
	5.0	-	860-1059	35-149	-
*Mt*Dod (ITC) ref. [14]	8.0	-	-	920	n.b.
ScDod/SdDod (fluor. titration) Bourdeaux *et al.* 2019[53]	7.5	-	n.b.	n.b.	n.b.
	5.0	-	n.b.	50-150*	-**
*Tt*Dod (fluor. titration) ref. [13]	8.0	141	233	311	589
*Tt*Dod (ITC) ref. [13]	8.0	-	-	920	-
*Hh*Dod (fluor. titration) ref. [5]	7.5	-	20×10^3	6×10^3	24×10^3
*Hs*Dod (fluor. titration) ref. [5]	7.5	18	36	14×10^3	439
*Hs*Dod (kinetic analysis) Bourdeaux *et al.* 2018[55]	7.5	-	76	-	-

*: single-step model based fit was not able to describe the measurements, only rough estimate was made.
**: under acidic pH FAD binds to *Sc*Dod and *Sd*Dod but no flour. titration was conducted.

A binding model that accounts for the structure of the binding pocket is the two-step binding model (**Figure 18**).

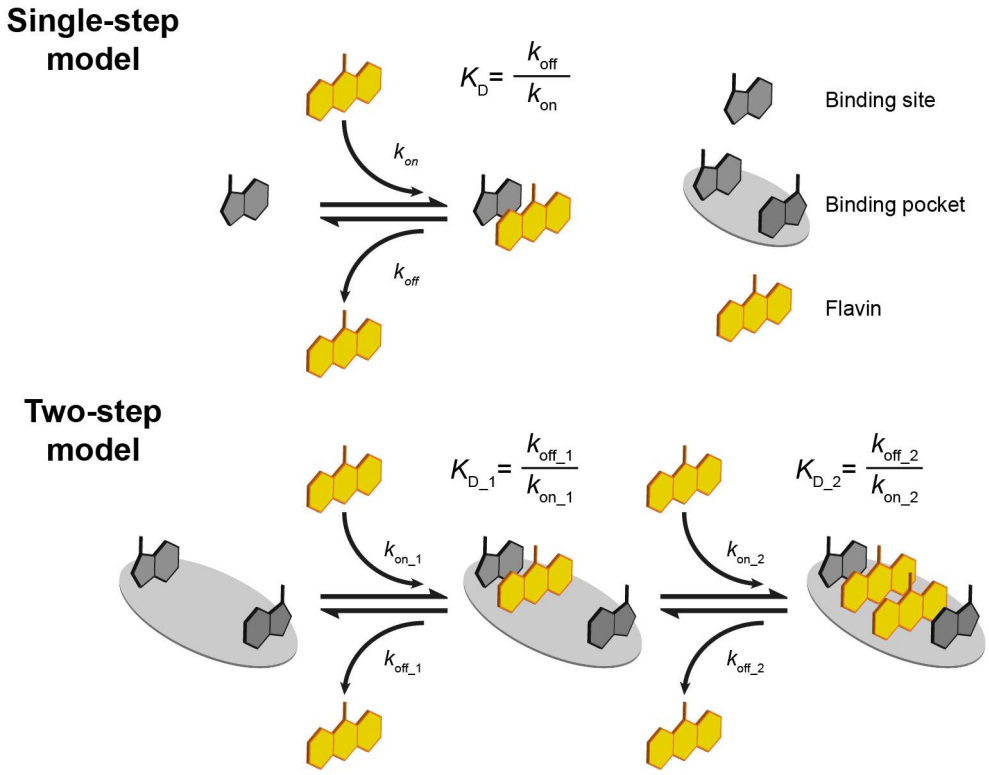


Figure 18: Single-step and two-step binding model.

While the two-step model describes a single binding pocket much better than the single-step model, it still separates the dodecamer into six independent binding pockets. Before the actual mathematical model will be presented, a short explanation is given, why the two-step model and not a more realistic model was chosen.

Dodecamer *versus* six independent binding pockets

The mentioned independence of the binding pockets is only problematic, if the 13 levels of flavin saturation of the dodecamer, ranging from empty to fully saturated, affect the rate constants of the discrete binding events. Effects that could cause such a behaviour, could be for example the overall change in charge of the dodecamer by binding FMN (dependent on pH, the charge changed by −1 or −2 for each bound FMN molecule) or changes in the structure of the dodecamer complex by the binding of flavins, which could alter the arrangement of other binding pockets. While these effects likely exist, the strength of their impact is hard to analyse. Assuming their impact is neglectable, the shown two-step binding model can well describe the

dodecamer, as the six binding pockets are independent and only spatially connected within the dodecamer complex. In contrast, if the saturation level needs to be included, the 13 flavin saturation levels must be separated into their 28 species, which still ignores the relative arrangement of the binding pockets. By binding an additional flavin molecule or by releasing an already bound flavin molecule, each species can transform into one or multiple specific species (**Figure 19**), which the mathematical model would need to account for.

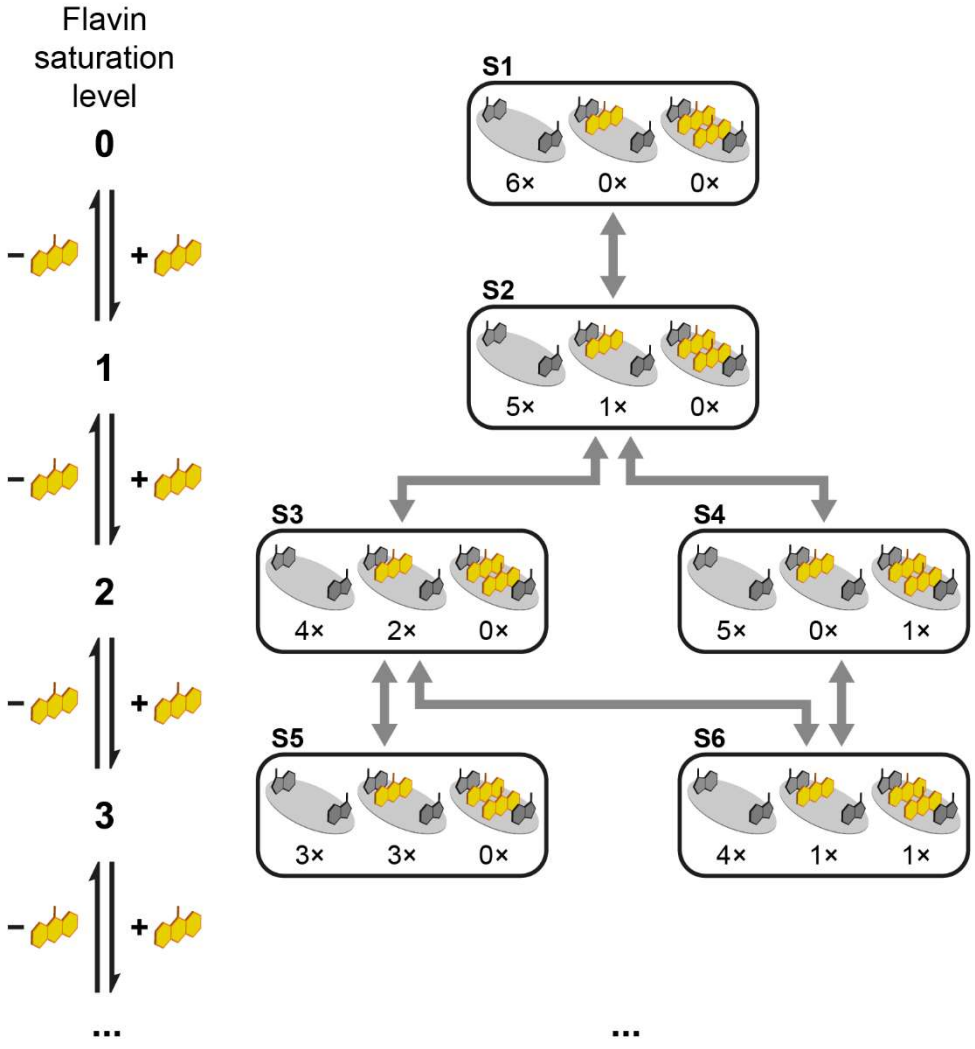


Figure 19: Number of dodecamer species dependent on the flavin saturation level. Species shown for zero to three flavins bound to the dodecamer (four first levels of the overall 13 levels).

While it is possible to design the network of species and the related functions, there is no practical way to obtain enough or even any data about the single species to fully verify this complex model. Therefore, using more complex and also more realistic models above the two-step model is only possible, if the obtained data contains enough information to fully satisfy this complexity, or, the problem or rather the data set will simply be overparameterized.

Two-step binding model requires a numerical solution

Focusing on the mathematical aspects of the model, the biggest advantage of the single-step model is that it can be solved analytically, meaning, it has a distinct solution, while the two-step model requires a numerical approach, as the direct analytical solution is not possible anymore. For the numerical solution, the flavin binding is calculated by utilizing rate constants instead of equilibrium constants. This allows to simulate how the complexes would form over the selected timeframe based on the chosen rate constants. By selecting a long enough timeframe, it can be assumed that the actual simulated state approached close enough to the equilibrium state and thereby allowing a method to describe the equilibrium without the need of the equilibrium constants. To put it simple, instead of directly calculating the equilibrium the system of interest is stepwise brought to the equilibrium. For these types of simulations, the timeframe and the step size are critical factors, since large step sizes can overdrive the simulation and lower the accuracy, causing high concentration fluctuations or even negative concentrations. On the other hand, short timeframes will not allow the system to approach the equilibrium. In general, the smaller the step size and the longer the timeframe, the better is the simulation. A big problem of small step sizes and long timeframes is the increasing amount of calculations needed for the simulation. While the increasing amount of calculations is unproblematic for a single simulation of this complexity, it is laborious to determine the rate constants from actual measurements, as simulations have to be repeated with differently set rate constants for each (simulation) cycle until the simulation fits the measurement. To put it simple, the numerical approach here means that different sets of rate constants are guessed and used to simulate binding curves over and over again until the deviation between the simulation and the actual measurement reached a minimum. The methods how these rate constants are "guessed" will not be discussed, but it needs to be noted that a problem of the here described numerical solving process is that the received rate constants might not be the actual constants and just reflect the best solution of this round/attempt. Therefore, multiple solving rounds need to be conducted to obtain reliable constants, which increases the amount of required calculations even further. The simple description of how the numerical solving process works should highlight, why the single-step model was used in general, since it is more practical. The problem here is that the single-step and the two-step model have distinct differences and are only comparable on a superficial level.

Comparison of the single-step and two-step model

Because it was shown in the publication of Bourdeaux *et al.* 2018[55] that the two step model is far better in describing the observed binding processes, the underlying reason for this will be discussed in more detail here. To show how both models differ, the determined rate constants or dissociation constants and functions in the publication of Bourdeaux *et al.* 2018[55] are used to simulate endpoints (equilibrium states) of different *Mt*Dod:FMN ratios. Values used for the simulation/calculations are chosen from the data set of the kinetic analysis (**Table 2**) and it needs to be mentioned that only the dissociation constants are shown, since the rate constants themselves (as long as the dissociation constants are not changed) don't affect the endpoint/equilibrium states. The concentration of dodecin or flavin are 1 µM, if not stated differently.

Table 2: Dissociation constants and simulation parameters for the show simulations. The dissociation constants were chosen from the kinetic measurements published by Bourdeaux *et al.* 2018[55] The selected single-step constants are the average of both shown biological replicates and have the lowest difference between single-step calculation and two-step simulation under the conditions of a constant *Mt*Dod concentration and increasing FMN concentrations (**Figure 20**). The dissociation constant of the single-step calculation at pH 5.0 and pH 7.5 is based on the 1.5:1 *Mt*Dod:FMN ratio measurement and 2.0:1 *Mt*Dod:FMN ratio measurement respectively. For the two-step model simulations, the timeframe was prolonged to 450 s to allow the simulations to approach the equilibrium close enough (prolonging the timeframe did not further improve the simulations noticeable), while keeping the step size at the published value (Bourdeaux *et al.* 2018[55]).

Model	pH 5.0	pH 7.5	Simulation parameters for the two-step model
single-step	K_D = 74 nM	K_D = 1085 nM	
two-step	K_{D_1} = 332 nM K_{D_2} = 22 nM	K_{D_1} = 7884 nM K_{D_2} = 69 nM	Timeframe = 450 s Step size; Δt = 0.03 s (15,000 steps)

For comparing the single-step and the two-step model, the equilibrium states of different FMN to *Mt*Dod ratios are used. The first question is: can the two-step model actually simulate similar binding curves as depicted in the literature?[5,13] In those experiments the flavin concentration was kept constant and for each measurement more dodecin was added, resulting in decreasing fluorescence signals, since the fluorescence of bound flavin is quenched.[5,13] Instead of the fluorescence signal the concentration of unbound flavin is simulated with the single-step and the two-step model (see **Figure 20**).

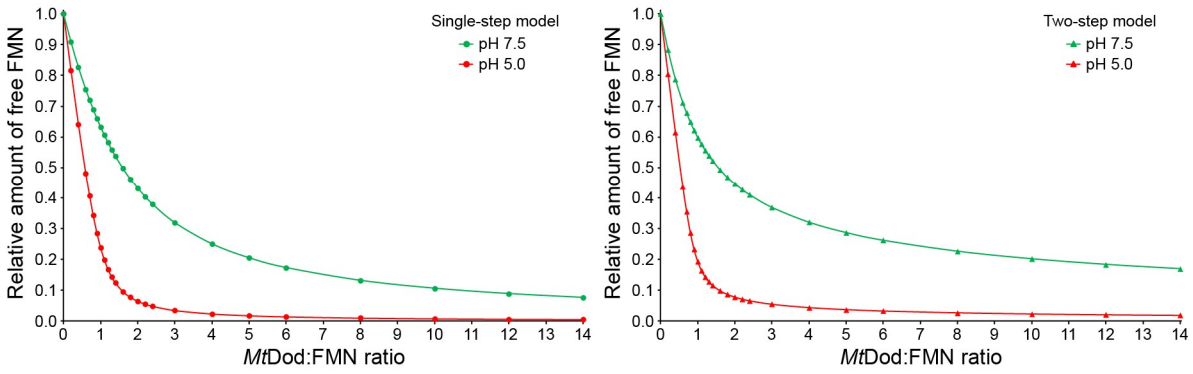


Figure 20: Concentration of unbound FMN against the absolute *Mt*Dod concentration based on one-step and two-step model simulations with constant absolute FMN (1 µM) concentrations and increasing absolute *Mt*Dod concentrations.

While the amount of unbound flavin is not identical, which can easily be seen by comparing the two curves at pH 7.5 at higher dodecin concentrations, it is obvious that both models can produce in general similar shapes of the curves, explaining how the single-step model was deemed appropriate to describe the binding event. Comparing both curves simulated with the lower dissociation constants at pH 5.0, the difference is nearly not noticeable, showing that at high affinities the impact of the different models is less dominant. Since the single-step model was first tested with *Hs*Dod and RbF, for which the dissociation constant is even lower (K_D(*Hs*Dod:RbF) = 36 ± 4 nM) than the one at pH 5.0 here,[5] it is understandable that the model was chosen. For ligands with a relative low affinity the discrepancy between the model and the measurement was masked by allowing the fluorescence coefficient of the bound flavin to adapt freely without restraints (Equation (1); $F_{meas.}$: measured fluorescence. f_{free}: fluorescence coefficient of free flavin. f_{bound}: fluorescence coefficient of bound flavin. $[L]$: concentration of the free flavin. $[D{:}L]$: concentration of the flavin dodecin complex or bound flavin).[5,13,55]

$$F_{meas.} = f_{free} \times [L] - f_{bound} \times [D{:}L] \quad (1)$$

This basically pushes the endpoint of the curve upwards without changing the maximum and compresses the curve until the single-step model seems to fit. Such a correction can also be done by using a correction factor (here $b_{corr.}$) that artificially increases the amount of free flavin dependent of the concentration of bound flavin (Equation (2)).

$$[L_{corr.}] = [L] + [D{:}L] \times b_{corr.} \quad (2)$$

By simply defining $b_{corr.}$ as the difference between the concentration of free flavin of both simulations at pH 7.5 with the highest amount of dodecin added ($[D]_0$: here 14 times the amount

of flavin; basically last point of the curve) (Equation (3); Endpoint is defined by 14-times more dodecin than flavin and 450 s calculation time) the corrected single-step model curve looks much more comparable to the two-step model curve (**Figure 21**).

$$b_{corr.} = \left([L_{2-\mathrm{Step}}](endpoint) - [L_{1-\mathrm{Step}}](endpoint) \right) / [L]_0$$
$$b_{corr.} = (0.17\ \mu\mathrm{M} - 0.08\ \mu\mathrm{M})/1\ \mu\mathrm{M} \qquad (3)$$
$$b_{corr.} = 0.09$$

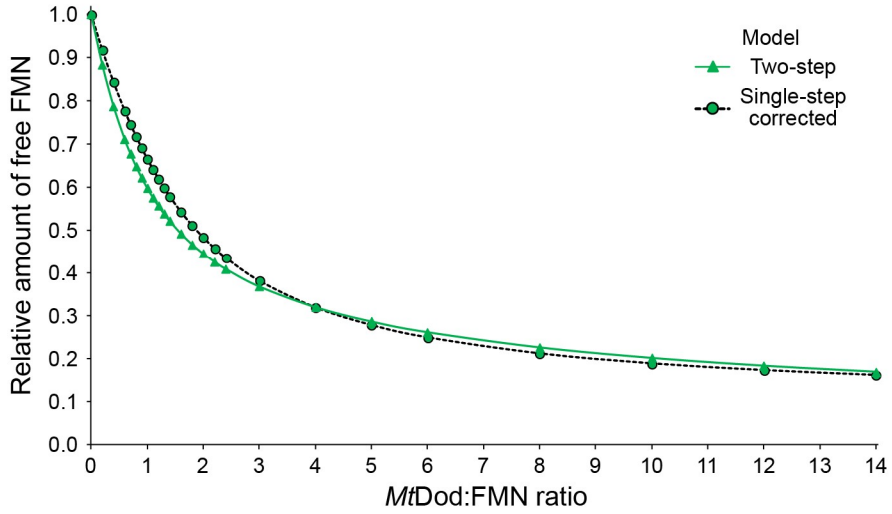


Figure 21: Effect of a correction on the single-step model calculation at pH 7.5 shown in **Figure 20**.

Even without adjusting the dissociation constant of the corrected curve, the differences between the single-step and the two-step simulation can easily be overlooked and would likely be deemed to measurement errors in an actual experiment. While in a simulation such a correction is clearly artificial, in fluorescence measurements different quenching mechanisms/efficiencies might be a possible explanation. In the publication by Bourdeaux *et al.* 2018[55], the quenching process of bound FMN at pH 5.0 and pH 7.5 was analysed by steady-state spectroscopy and showed that under both conditions the quenching processes are nearly identical and would not support the above describe correction. In spite of this, at pH 7.5 the measured binding data required fluorescence correction to be described by the single-step model (Bourdeaux *et al.* 2018[55]). In agreement with this finding is the fact that fluorescence intensities of FAD bound to *Hs*Dod ((K_D(*Hs*Dod:FAD) = 439 ± 48 nM), which binds as a monomer to *Hs*Dod and therefore the single-step model is likely appropriate, are comparable to fluorescence intensities of bound RbF (both approach values close to zero) in spite of the different binding modus.[5] While it is still possible that in some dodecins bound flavins are quenched with different efficiencies, so far, the two-step model seems more plausible and the one-step model should

not be used anymore to describe the binding of flavins to dodecins as long as they don't bind as monomers like FAD to *Hs*Dod and *Hh*Dod.[5]

Effect of the two-step model

With the two-step binding model established, the question remains, if the two-step model actual supports the potential role of dodecins as a flavin buffer. The first indicator for this is that the binding efficiency of dodecin at concentrations that vastly exceed the flavin concentration is lower in the two-step binding model simulations than in simulations with the single-step binding model (higher amounts of flavin remained unbound; see **Figure 20**). It might be argued that this is only caused by an overall lower affinity used in the two-step simulations, but since the curve shapes are so similar (see **Figure 21**) this is not a plausible explanation. This means that the flavin binding efficiency of the two-step model compared to the one-step model depends on the ratio of dodecin to flavin (here at excessive dodecin concentrations). To highlight the effect of this behaviour a series of simulations with a constant dodecin concentration and increasing flavin concentrations was conducted with the same constants as shown in **Table 2** (**Figure 22**).

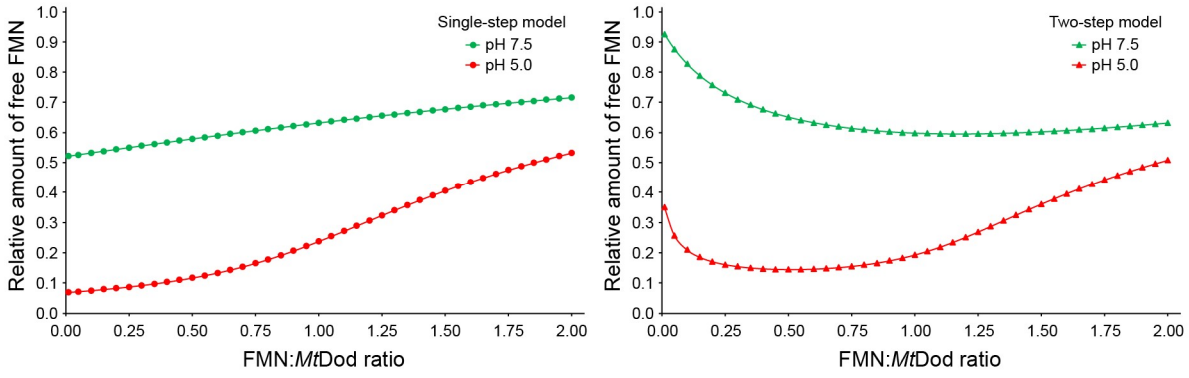


Figure 22: Normalized concentration of unbound FMN against the absolute FMN concentration based on single-step and two-step model simulations with a constant absolute *Mt*Dod concentration (1 µM) and increasing absolute FMN concentrations.

The relative amounts of available flavin, depicted in **Figure 22**, show clearly that the two-step model and the single-step model behave differently in situations where the dodecin concentration is higher than the flavin concentration. While in the one-step model the relative amount of available flavin is lowest at the smallest dodecin:flavin ratio (minimum of the curve), in the two-step model the lowest relative amount of available flavin is above the smallest dodecin:flavin ratio.

The potential effect of this will be explained with FMN, although it affects all binding flavins. To keep it simple, the RbF synthesis is ignored here and FMN bound to flavoenzymes are treated as strictly non-available anymore. The higher relative amount of available FMN at low absolute FMN concentrations means that, if the FMN demand is rising, e.g. by expression of flavoenzymes, and therefore the FMN levels are decreasing, dodecins become less efficient in binding and allow the cell to utilize close to all of the present FMN without strong competition of the dodecins. If the one-step model would be correct, with lower FMN levels dodecins would compete more and more with the other processes requiring FMN.

Concluding the flavin binding model

Although these examples here are extreme simplifications of the flavin homeostasis they still show that a two-step flavin binding model supports the role of bacterial dodecins as a flavin (likely FMN) buffering agents as proposed by Bourdeaux *et al.* 2018[55] for *Mt*Dod. This shows that the binding mechanism itself can be a critical factor for the function of a protein and that simplifications of such mechanisms need to be treated carefully. Additionally, because of the improvement of personal computers and thereby increased availability of computational resources, numerical simulations should be started to be preferred other simple models.

5.1.2 Indications that Dodecins are Actually Part of the Flavin Homeostasis

While the chapter above discussed what mechanisms potentially allow dodecins to function as a flavin buffer systems, such mechanisms do not necessarily mean that dodecins is actually utilized as such systems. In the publication by Ludwig *et al.* 2018[235] the effects of a removal of the dodecin gene in *S. davaonensis* on the metabolome were analysed. The *Sd*Dod knockout strain showed increased concentrations of potential stress protectants, like polysaccharide and carotenoids, and was resistant to plumbagin.(Ludwig *et al.* 2018[235]) So far, also other dodecin genes (*Mt*Dod: Gene-ID: *Rv1498A* and *Tt*Dod: Gene-ID: *TTHA1431*) were found to be non-essential in their respective organism,[238,106] which might be expected since even directly affecting the flavin homeostasis by influencing the regulation or the activity of the bifunctional flavokinase/FAD-synthetase is non-lethal (both cause RbF overproduction).[41,239] While the effect of high cellular riboflavin levels seems to be unproblematic or at least is not discussed

in literature, insufficient amounts of riboflavin slow down the growth showing the essential need for sufficient riboflavin and thereby other flavins.[240,241]

Indicators for flavin sequestering by dodecins

The species *Amphibacillus xylanus*, which lacks a respiratory chain, increases its cellular flavin content upon exposure to higher oxygen concentrations (flavin content more than doubled from 0% to 21% oxygen concentration) and also utilizes high amounts of unbound flavin in the cytosol.[29] In contrast, there are only 4% of the cellular flavins free in *E. coli* and excessive flavins are secreted.[25,242] The reason for *E. coli* to secret flavins might be to avoid over-accumulation of flavins, but so far this is just a hypothesises.[242] Other bacteria secrete flavins for specific purposes, e.g. the flavin release of *Sinorhizobium meliloti* is related to the symbiosis with plants (main secreted flavin is FMN and has also high intracellular FMN levels; carries a dodecin gene (Gene-ID: *SM_b20068*)).[27,243] Other examples are some *Shewanella* species like *Shewanella oneidensis* (strain MR-1; type strain), which use secreted FMN and RbF for the reduction of poorly soluble Fe^{3+} to Fe^{2+} (some/few *Shewanella* species carry a dodecin gene like *Shewanella putrefaciens* (Gene-ID: *NCTC10738_02649*)).[244,245] These examples show how diverse the flavin homeostasis in bacteria is and that without precise data about the specific organism assumption are quite problematic. One common feature of most bacteria (based on the examples here) is that FAD and FMN are the important intracellular flavins and RbF is normally only present in low concentrations.[25,27–29,242,244,246] Even in *H. salinarum* (archaea), which has a dodecin that prefers RbF compared to FMN and FAD,[5] accumulates RbF only at the end of the early stationary phase (about 15 days, *Hs*Dod expression notable after 2-3 days).[7] This late accumulation of RbF seems to be connected to *Hs*Dod, since the *Hs*Dod knockout strains did not accumulate RbF.[7] A similar flavin accumulation, here FMN, was observed by Ludwig *et al.* 2018[235] when comparing the *Sd*Dod knockout *S. davaonensis* strain to the wild type – after 7 days the wild type showed higher intercellular FMN content than the knockout strain. Based on the sequence similarity of the dodecin of *S. meliloti* with other bacterial dodecins (**Figure 23**) known to bind FMN with higher affinity than RbF and FAD (*Mt*Dod[55], *Sc*Dod[53] and *Sd*Dod[53]; *Tt*Dod is suggested to bind FMN as native ligand[13]), also the dodecin of *S. meliloti* likely prefers FMN and this could be related to the reported high amounts of cellular FMN in *S. meliloti*.[27]

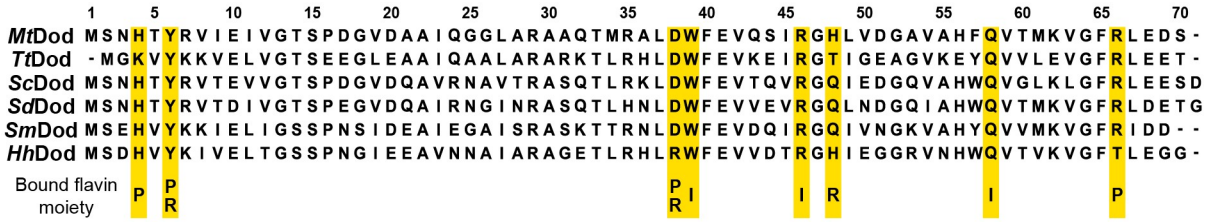


Figure 23: Alignment of studied bacterial dodecins and the dodecin of *S. meliloti*. Amino acid residues related to the flavin binding are highlighted yellow (Bourdeaux *et al.* 2018[55]). Interactions of the amino acid residues with FMN are labelled with: I: Isoalloxazine ring system; R: ribityl chain; P: phosphate group.

All these findings suggest that dodecins allow cells to sequester flavins and thereby might cause an increased cellular flavin content, as reported for *Hs*Dod in *H. salinarum*.[7] For most bacterial dodecins the sequestered flavin is likely FMN (based on the alignment published by Bourdeaux *et al.* 2018[55]), but *Hh*Dod and similar dodecins might sequester other flavins.

Flavin sequestering in *M. tuberculosis*

For *M. tuberculosis* data of cellular flavin levels/concentrations could not be found, but it was suggested that FAD sequestering proteins protect *M. tuberculosis* against hypoxic and oxidative stress (article focusses on Fsq of *M. smegmatis*; *M. tuberculosis* homologue Gene-ID: *Rv3129*),[247] further indicating the importance of flavins for *M. tuberculosis*.[105] Another protein related to flavin sequestering in *M. tuberculosis* is Acg (Gene-ID: *Rv2032*), which evolved from a nitroreductase, but seems to have lost its enzymatic function/role and now might sequester FMN.[248] It was shown that Acg is important for the virulence and survival of *M. tuberculosis* in macrophages, but its deletion did not affect viability when facing nitrosative and oxidative stress *in vitro* and *in vivo*.[249] Acg (Gene-ID: *Rv2032*) and Fsq (Gene-ID: *Rv3129*) are found to be non-essential under *in vitro* conditions like *Mt*Dod (Gene-ID: *Rv1498A*).[106,107] In contrast to Acg (Gene-ID: *Rv2032;* expression low under normal conditions; part of DosR regulon)[250] and Fsq (Gene-ID: *Rv3129*) the gene disruption of *Mt*Dod (Gene-ID: *Rv1498A*) was even reported to be beneficial for growth under *in vitro* conditions.[107] This could be explained by the fact that *Mt*Dod is not part of the DosR regulon and is constantly expressed (cellular concentration about 5 µM; estimated using the cellular copy number).[250] Ignoring the observed growth advantage of the dodecin gene disruption, it seems that *M tuberculosis* has a need for sequestering flavins, which would fit to the findings in other species and further suggests that bacterial dodecins at least function as storage/sequestering device for flavins.

Before discussing possible biological reasons for flavin sequestering by dodecins, the question is raised, if there are indicators for flavin buffering by dodecin, as suggested by the two-step mechanism.

Dodecin knockouts and dodecin related stress

The only study that focussed on the effects of a dodecin deletion was done by Ludwig *et al.* 2018[235] and showed that the knockout strain had an increased resistance to plumbagin, which causes oxidative stress.[251] A comparison of metabolites between the wild type and the dodecin knockout strain revealed that the knockout had higher concentrations of carotenoids (e.g. astaxanthin), diverse glutathione derivatives (e.g. 2-(S-glutathionyl)acetyl glutathione; glutathione itself is lower concentrated, likely because it reacted, higher amounts of glutathione disulfide were found), a large number of polysaccharides (e.g. trehalose) and an osmoprotectant (hydroxyectoine).[235] Glutathione is interesting here as it is directly related to maintaining the reducing environment of the cell and detoxification.[252,253] To elucidate the actual role of all these diverse compounds is well beyond the scope of this thesis and therefore, they are simply seen as part of the anti-stress mechanism of *S davaonensis* here (the upregulation might explain the increased plumbagin resistance of the deletion strain).[235] While the actual role of *Sd*Dod or more precisely how the FMN binding of *Sd*Dod is involved in these processes is not clear, but the study by Ludwig *et al.* 2018[235] shows that *Sd*Dod participates in maintaining the cellular balance (likely redox homeostasis). Considering these findings, it seems contradictory that the *Mt*Dod gene (Gene-ID: *Rv1498A*) disruption in *M. tuberculosis* is beneficial for growth under *in vitro* conditions, but since *M. tuberculosis* seems to naturally proliferate only in humans, the conditions of the *in vitro* experiment might be too different from the natural environment.[65,77] Other data that indicates dodecin's role as more than a simple flavin sequestering device was collected in a transcriptome analysis of a *T. thermophilus Tt*Dod gene (Gene-ID: *TTHA1431*) disruption strain.[238] By comparing this strain with the wild type (data was analysed with NCBI GEO[254], used datasets (ref. [238]): WT: GSM532185-GSM532214; Gene-ID: *TTHA1431* deficient strain: GSM536231-GSM536233), an expression profile of *Tt*Dod (Gene-ID: *TTHA1431*) (**Figure 24**) and other upregulated genes in the knockout strain could be identified.

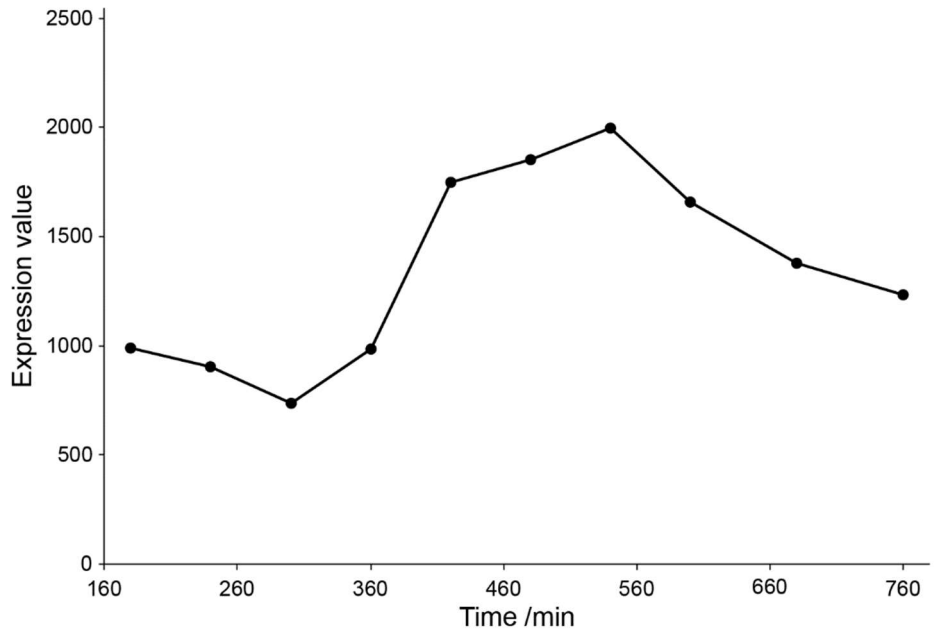


Figure 24: Expression profile of *Tt*Dod (Gene-ID: *TTHA1431*) over 180 min to 760 min. After 680 min the stationary phase was reached. Graph based on data of ref. [238] submitted to NCBI GEO[254].

*Tt*Dod (Gene-ID: *TTHA1431*) is expressed during the whole experiment with a peak in expression around 420-540 min, which starts to decline entering the stationary phase. This shows that *Tt*Dod is constantly present in the cell with some concentration fluctuations. Comparing the knockout with the wild type strain, the three top (sorted by significance) upregulated genes (only hits labelled with a Gene-ID of *T. thermophilus* are considered) are Gene-ID: *TTHA1500* (phosphoenolpyruvate synthase), Gene-ID: *TTHA1836* (isocitrate lyase) and Gene-ID: *TTHA1499* (MoxR-related protein). Without an in-depth analysis of all alterations in gene regulation caused by the knockout and deep insides into the cellular processes of *T. thermophilus*, the actual effects are hard to estimate, which limits the analysis here. The phosphoenolpyruvate synthase (Gene-ID: *TTHA1500*) turns pyruvate with ATP into phosphoenolpyruvate (often called PEP) and might be related to glycogenesis and therefore the production of carbohydrates, but PEP is also involved in other pathways.[255–257] The isocitrate lyase (Gene-ID: *TTHA1836*) is one of the two key enzymes of the glyoxylate shunt, an alteration of the tricarboxylic acid cycle (also called glyoxylate cycle), which bypasses relevant steps for the release of CO_2.[258] By preventing the release of CO_2, the glyoxylate shunt conserves carbon atoms for gluconeogenesis and is related to the oxidative stress response and survival of pathogens (e.g. *M. tuberculosis*) in macrophages.[258,259] The MoxR-related protein (Gene-ID: *TTHA1499*) is likely a MoxR-like ATPase and might have a chaperone-like function by supporting insertion of cofactors into proteins (proposed function of MoxR ATPases).[260] MoxR proteins were found to be important factors of diverse stress responses in several

organisms.[260] Interestingly the expression values of the riboflavin synthase (α subunit: Gene-ID: *TTHA1063*; β subunit: Gene-ID: *TTHA0336*) and the bifunctional RbF kinase/FMN adenylyltransferase (Gene-ID: *TTHA0527*) are not significantly different between the wild type and the *Tt*Dod gene (Gene-ID: *TTHA1431*) disruption strain.

The upregulation of genes related to the gluconeogenesis, as observed in *T. thermophilus* for the *Tt*Dod gene disruption, might be the reason, why the *Sd*Dod deficient *S. davaonensis* strain shows increased polysaccharide levels as reported by Ludwig *et al.* 2018[235]. In general, their proposed conclusion that the *Sd*Dod deficient *S. davaonensis* strain has an increased plumbagin resistance because of an upregulation of the stress response system fits to the findings in *T. thermophilus* (datasets from ref. [238]). But the growth advantage reported for the gene disruption of *Mt*Dod (Gene-ID: *Rv1498A*) under *in vitro* conditions does not fit into this idea of upregulated stress response systems.[107]

Unclear points and unanswered questions

While this data clearly shows that dodecins are involved in keeping the cellular balance and the flavin buffering is likely the reason for this, there are still some uncertainties. For example, it is not clear how much flavin can actually bind to *Tt*Dod at the culturing temperature of 70 °C.[238] Temperature dissociation measurements of the dodecin:flavin complex of *Mt*Dod, *Sc*Dod and *Sd*Dod by Bourdeaux *et al.* 2018[55] and Bourdeaux *et al.* 2019[53] showed that at such temperatures the amount of bound flavin is quite low. While no actual data for *Tt*Dod is available the similar structure and comparable flavin binding residues (see **Figure 23** and Bourdeaux *et al.* 2018[55]) would suggest a similar temperature behaviour of the dodecin:flavin complex. An indicator that *Tt*Dod might behave differently under elevated temperatures, is the fact that the flavin binding process is reported to be faster under such temperatures.[13] But without actual data of the flavin binding efficiency under those conditions, it is not possible to make any estimations about the affinity at higher temperatures. An alternative explanation of how *Tt*Dod still might affect the flavin levels, even if the high temperatures also lower its affinity to flavins, could be that the cytosolic conditions in *T. thermophilus* strongly facilitate the flavin binding of *Tt*Dod and therefore compensate the temperature effect. So far these are just hypothesis, as no experimental data is available to support them.

Another point that seems questionable is the observed growth benefit of *M. tuberculosis in vitro* when the dodecin gene is disrupted.[107] While the *in vitro* culturing conditions might be

the reason for this, temperature can be ruled out, as the experiment was conducted at 37 °C.[107] Too low culturing temperature was the first assumption, since low temperatures basically increase the affinity and would cause *Mt*Dod to bind more flavin, which could make dodecins a disadvantage at growth below the temperature optimum. The last point is the ability of some bacterial dodecins to bind CoA, reported for *Mt*Dod and *Tt*Dod,[13,51] since the observed effects might also be partly related to the CoA binding. But so far, the CoA binding is not at all studied and therefore here not further considered.

Concluding the buffering role of bacterial dodecins

While there is clear evidence that bacterial dodecins play a role in the cellular well-being (here described/listed effects of dodecin gene knockouts) and are able to buffer flavin concentrations (two-step model), it is still not clear how bacterial dodecins are actually involved in the cellular homeostasis. This shows that even such a simple appearing protein like dodecin can still be a mystery in the dynamic context of a living cell.

Dodecins as flavin storage

Another function of dodecins could be the tight binding of flavins under growth limiting conditions to preserve flavins and protect the cells during this time from free flavins. It is assumed, that the stored flavins could give a growth benefit, when the conditions become more favourable again. The above mentioned flavin sequestering could suggest such a function, although a buffer could also increase the overall cellular flavin content, which can be seen as sequestering. An example for an dodecin that might function as a storage is *Hs*Dod, which sequesters RbF in the late stationary growth phase.[7]

In the natural habitat of *H. salinarum* (salt lakes) unbeneficial growth conditions, under which *Hs*Dod sequesters RbF, could be low salinity or nutrition scarcity.[8,9] The low salinity is often followed by an algae bloom that causes a phase of high nutrition when the algae died through increasing salt concentrations, which can cause a *H. salinarum* bloom.[8] In such a scenario dodecin as storage system clearly could help to ensure rapid growth after the change back from unfavourable to favourable conditions, by ensuring enough RbF for the FMN and FAD synthesis.

A comparable scenario is the switching between dormancy and active growth of *M. tuberculosis* in its complex lifecycle, which is related to the survival in macrophages and the tuberculosis typical granuloma (see Introduction Chapter 2.3.3).[82] Here *Mt*Dod tightly binds FMN in the dormant phase to release it when *M. tuberculosis* enters the active state again. Considering that *M. tuberculosis* is not simply immune to the host defence systems and the infection is more of a question, which side can gain and maintain the upper hand,[85,101,102] even small growth advantages like an FMN storage to kick-start the metabolism might be an important factor.

Flavin storage of *Sc*Dod and *Sd*Dod could be beneficial for sporulation, as it would give the spores an enhanced flavin pool for germination and further growth.

Affinity problems: You can't store, what you barley can bind

A problem for the theory that bacterial dodecins generally act as flavin storages is their overall lower affinity compared to *Hs*Dod (see **Table 1**), which makes tight flavin binding questionable. Without tight binding a rather large portion of flavins would be freely available during the storage period, which would either cause constant loss of flavins (e.g. by secretion or degradation) or cellular damage by the free flavins during low activity. If both cases are unproblematic for the cell, flavin storage in general would be pointless and therefore not a biological function of dodecins. This means that if storage is important, there must be conditions under which bacterial dodecins have increased affinities towards flavins to fulfil the tight binding requirement.

Surprisingly under acidic conditions *Mt*Dod, *Sc*Dod and *Sd*Dod showed strongly increased affinity to FMN and FAD, which would make a storage function at low pH possible (Bourdeaux *et al.* 2018[55] and Bourdeaux *et al.* 2019[53]). Under acidic conditions the affinity of *Mt*Dod to FMN even surpasses the affinity of *Hs*Dod to RbF, while *Sc*Dod and *Sd*Dod turn into mediocre to strong FMN binders compared to no measurable binding at pH 7.5 and room temperature (see **Table 1**). For *Tt*Dod, *Hh*Dod and *Hs*Dod no data for the effect of lower pH is available, but because of the high structural similarity of the bacterial dodecin binding pocket, it seems plausible that FMN and FAD binding of *Tt*Dod is also affected by a lower pH. It was shown that RbF binding of *Mt*Dod is not substantiality affected by lower pH, indicating that the increased affinity is likely related to the phosphate group of FMN or the diphosphate group

of FAD, but the molecular principle behind this phenomena was not found.(Bourdeaux *et al.* 2018[55])

The biological relevance of the pH depending affinity of *Mt*Dod seems on the first view plausible, since *M. tuberculosis* encounters acidic conditions in the phagolysosome and the granuloma,[91,94,261,262] but it was shown that *M. tuberculosis* can maintain its cellular pH even when facing external pH-values of 4.5 or inside activated macrophages.[263,264] Still, under certain conditions, dormant *M. tuberculosis* can lose its acid-fast character,[85,94–96,265] which indicates a compromised protection against acidity, possibly caused by an altered/weakened cell wall. In such a case, *Mt*Dod would be a stronger FMN and FAD binder and the storage function would be more dominant, but there is no data that shows a lower cytosolic pH. The relation of *Mt*Dod's function and acidity is also indicated by the slight upregulation of *Mt*Dod under acidic conditions (higher *Mt*Dod counts compared to standard conditions: 1.85-fold at pH 6.5 and 1.83-fold at pH 5.5).[266]

The increased affinity of *Sc*Dod and *Sd*Dod to FMN and FAD under acidic conditions could be related to sporulation, since it was shown for some *Streptomyces* species and *Bacillus* species that their spores (here comparable to dormant cells) are slightly acidic.[267–270] This would make *Sc*Dod and *Sd*Dod FMN and FAD storage devices for spores and could be beneficial for germination (reactivation from a dormant/low activity state).

Since *H. halophila* lives like *H. salinarum* in saline to hypersaline lakes, although more alkaline ones (soda lakes), and season depending blooms were reported,[271] a similar storage function as *Hs*Dod would be plausible for *Hh*Dod, but its low affinities make flavin storage questionable and so far no conditions are known which enhance its affinity.

While for *Mt*Dod, *Sc*Dod and *Sd*Dod acidic dependent affinities could be correlated to specific stages in their life cycles, this is likely not the case for all dodecins and their respective species. A more general factor, which increases the binding affinity is low temperature and therefore some dodecins could be flavin storages for low temperature phases. For organisms that have to endure/tolerate temperature fluctuations, like soil bacteria, such a function might be plausible. The general temperature dependency of ligand binding can be described with the van 't Hoff equation (Equation (4), adapted for K_D; T: absolute temperature. $\Delta H^{\ominus}$: standard enthalpy change. R: gas constant), dodecins will bind flavins with higher affinities at lower temperatures (see Chapter 5.1.3).

$$\frac{d}{dT}\ln\frac{1}{K_D} = \frac{\Delta H^{\ominus}}{R \times T^2} \quad (4)$$

A simple example where such a storage system could be beneficial, would be the day and night cycle. During the day increased temperature cause higher cellular activity and no notable flavin binding by dodecins, while at the night, decreased temperatures lower cellular activity and cause dodecins to bind flavins. While this assumption seems plausible, since it fits in the same area as storage for other low activity phases, there is no data to backup this hypothesis.

Concluding the role of bacterial dodecins as flavin storages

While the storage or rather tight binding of flavins seems not to be the dominant role of bacterial dodecins under normal physiological conditions (pH 7.4), since the affinities here are more or less mediocre, it seems that under certain conditions tight binding of flavins is facilitated and therefore a storage function might be possible. For *Mt*Dod, *Sc*Dod and *Sd*Dod this condition seems to be acidity, which correlates with events in their life cycle. *Mt*Dod encounters acidic environments and goes into a dormant state, in which a flavin storage would be beneficial, as it allows to utilize those flavins when it recovers from its dormancy. For *Sc*Dod and *Sd*Dod the lower pH found in their spores, would allow their dodecins to function as flavin storages for spores, which could be beneficial for germination and growth. While in these cases acidity dependent binding seems beneficial, it is not likely that this is true for all organisms with a dodecin encoding gene and flavin storage might therefore be more of an extra feature of some dodecins.

5.1.3 How to Characterize Dodecins without the Single-step Model Dissociation Constants

While the newly established two-step binding model describes the flavin binding in detail and therefore allows to gain insights into the biological role of dodecins, it is not suited to make easy comparisons between different dodecins. In the characterization with a single-step model, it is directly clear which ligand has the highest affinity simply by looking for the lowest dissociation constant. In contrast, with a two-step binding event both dissociation constants and the potential dodecin to flavin ratio needs to be considered to find out which ligand is preferred at the condition of interest. Of course there are also clear cases, if both dissociation

constants are lower than the affinity is higher, but for cases where only one dissociation constant is lower but the other is higher an easy answer is not possible (Dod:L_1 K_{D_1} < Dod:L_2 K_{D_1} and Dod:L_1 K_{D_2} > Dod:L_2 K_{D_2}). Further the correct characterization of a two-step binding event requires higher experimental effort and computational resources than of a single-step binding event, which also makes it less practical.

To avoid these problems, a method was designed that does not require the inclusion of a binding model. Instead of determining any dissociation constants, the inflection point of the temperature dependent dissociation of the dodecin:flavin complex is used to compare affinities by Bourdeaux *et al.* 2018[55]. This method can be compared to the thermal shift assay (also called "Thermoflour"), where the inflection point (melting point) of protein melting curves at different conditions is compared.[272,273] The "ThermoFAD" method, which uses the release of bound FAD of flavoenzymes upon melting/denaturing to measure their stability,[274] is similar, except that in the dodecins release flavins without melting/denaturing. Similar to how melting points are employed for stability comparisons (the higher the melting point temperature, the more stable is the protein), the inflections points of the dodecin:flavin dissociation curve can be used to compare affinities. The higher the inflection point temperature, the higher is the affinity, or more precisely, the more stable is the dodecin:flavin complex (see van 't Hoff equation (Equation (4)) for K_D temperature relation). The inflection point temperatures can be determined by using the first derivation (maximal slope) or by fitting the curves with a Boltzmann sigmoid equation (Equation (5); $F(T)$: Fluorescence intensity at temperature T. F_{lowest}: lowest fluorescence intensity (normally background). $F_{highest}$: Highest fluorescence intensity (normally maximum of the curve). $T_{Inf.}$: Temperature of the inflection point).

$$F(T) = F_{lowest} + \frac{(F_{highest} - F_{lowest})}{1 + e^{\frac{(T_{Inf} - T)}{slope}}} \quad (5)$$

This method can be used to measure many affinities in parallel at different conditions or for the parallel screening of multiple mutants (method was used to determine if in *Mt*Dod the histidine at position 4 (H4) is the origin for the acidic binding behaviour (Bourdeaux *et al.* 2018[55])) without a lot of effort and allows a quantitative comparison of these obtained affinities (**Table 3**).

Table 3: Inflection point temperatures of *Mt*Dod, ScDod and SdDod at different pH-values. Above certain pH-values was F_{lowest} for *Sc*Dod and *Sd*Dod too high to allow comparison of the data with lower pH-value. Since the RbF binding to dodecin is not affected by lower pH-values or at least not noticeable, no binding to *Sc*Dod or *Sd*Dod observed. In all measurements, the concentration of the used dodecin:flavin complex was about 4 µM.

Dodecin (reference)	Inflection point temperature /°C								
	RbF			FMN			FAD		
	pH 5.0	pH 6.0	pH 7.0	pH 5.0	pH 6.0	pH 7.0	pH 5.0	pH 6.0	pH 7.0
*Mt*Dod Bourdeaux *et al.* 2018[55]	36.2	36.5	34.0	74.3	69.0	55.7	52.6	45.4	31.2
ScDod/ Bourdeaux *et al.* 2019[53]	n.b.	n.b.	n.b.	55.6	53.8	39.3	38.6	26.0	-
SdDod Bourdeaux *et al.* 2019[53]	n.b.	n.b.	n.b.	53.0	54.7	39.1	35.6	22.9	-

Binding "affinities" measured with the temperature dependent dissociation method are comparable to the dissociation constants shown in **Table 1** (the same trends are observed). Acid dependent binding was confirmed, and the same preferred ligands were determined (*Mt*Dod: FMN, *Sc*Dod: FMN *Sd*Dod: FMN), showing that the method has the same potential to analyse the affinities of dodecins. It needs to be noted that the obtained inflection point temperature is not a constant like the dissociation constant and depends also on the used concentration during the measurement. So, for comparison of different data sets the method requires the same dodecin:flavin complex concentrations to be used in all experiments.

5.2 Applications of dodecin

In this chapter the benefits of bacterial dodecins as carriers/scaffolds or as un-/functionalized nanoparticles will be discussed. The dodecin dodecamer is referred to as a carrier or scaffold when it is utilized as a core for assemblies, but without any modifications, it is called a nanoparticle (ignoring its biological function). In general, all dodecins can be used as nanoparticles since they share the dodecin fold and thereby have the ability to form the dodecin typical dodecamer, but not all dodecins are suited for all applications.

In the first subchapter the suitability of dodecins as nanoparticles will be discussed based on the example of *Hs*Dod, which was used by Nöll *et al.* 2018[237] to measure via diffusion-ordered NMR the ability of DNA hydrogels to trap guest particles/molecules. *Hs*Dod's only role in this setup was to function as a diffusion probe, making it an ideal system to discuss the main features of a protein nanoparticle: stability and size conformity. With the stability measurement

methods created to study *Mt*Dod, *Sc*Dod and *Sd*Dod (Bourdeaux *et al.* 2018[55] and Bourdeaux *et al.* 2019[53]) it could be shown that *Hs*Dod still was suitable for the diffusion measurements and sample preparation, although its high salt dependency showed that bacterial dodecins might be an better option.

In the second subchapter it will be shown how *Mt*Dod can be used as a carrier protein with a focus on antibody (AB) production against selected epitopes of target proteins (Bourdeaux *et al.* 2020[236]) and its properties will be compared to other carrier proteins. That *Mt*Dod fulfils the basic needs of a carrier or scaffold protein was shown by producing diverse *Mt*Dod fusions constructs, highlighting that the dodecamer can tolerate a vast array of attachments. Since for biotechnological applications also production costs are an important factor, a simple purification protocol was established that allowed to easily purify high amounts of dodecamer.

5.2.1 The Dodecin Nanoparticle a.k.a. Dodecamer

A benefit of using proteins as nanoparticles or cores for scaffold-/carrier-designs is that their structure is defined and can be obtained by crystallography. So, each protein-nanoparticle has the same structure and thereby size and shape. While this seems trivial, size distribution and shape are important factors of nanoparticle production, since they affect how nanoparticles behave, e.g. the toxicity and distribution of PEG-coated gold nanoparticles in the body depends on its size.[275,276] The importance of size and shape distribution is even greater, when the particles are used as probes to study other objects, like in the here described case of measuring the ability of DNA hydrogels to capture particles (Nöll *et al.* 2018[237]).

Linear building block DNA hydrogels

While the publication of Nöll *et al.* 2018[237] mainly focuses on DNA hydrogels and the measurement method, here, the *Hs*Dod as a diffusion probe will be the centre of the discussion. Nevertheless, to understand the role of *Hs*Dod as a diffusion probe a short description of DNA hydrogels is given. The used DNA hydrogels were build up from linear DNA double strands with 5'-3' complementary overlaps (sticky ends) on both sides.[277] These linear building blocks can form long linear bending structures that likely result in circular structures when both ends meet (nicks in the backbone increases the flexibility). The final DNA hydrogel is a 3-dimensional network of interlocked and entangled DNA rings, of which the diameter depends

on the length of the used building blocks (or more precisely, the repeating units after hybridization).[277] With a repeating unit length of 45 bases (approx. 15 nm) the average diameter of the formed rings was about 15.7 ± 4.8 nm and the calculated mesh size of the hydrogel was about 15 nm.[277] Since mesh size is not the same as pore size in hydrogels, because large pores (compared to the pores formed by the regular gel matrix) can form through irregularities in the hydrogel, also methods need to be considered to judge the quality of the generated hydrogel, which was done by the diffusion of *Hs*Dod entrapped in the hydrogel in this case.[237]

In a hydrogel with a mesh size below the size of the entrapped particles and no larger pores, the diffusion of the entrapped particle is greatly reduced, while with large pores or larger mesh sizes the particle can diffuse rather freely. In Nöll *et al.* 2018[237], three different types of DNA hydrogels were studied this way, two hydrogels based on approx. 15 nm repeating units (45 bases and 46 bases, the latter contains a single unpaired base (higher flexibility)) and one hydrogel based on approx. 5 nm repeating units (16 bases, containing single unpaired base). Assuming that the mesh size depends only on the repeating unit length, the mesh size for the hydrogel with the 5 nm repeating unit should be in the range of 5-6 nm. Therefore, the three tested hydrogels should have twice about 15 nm (15 nm repeating units) and once about 6 nm meshes (5 nm repeating units).

The particle size of dodecins is defined by its structure

Based on the crystal structure of *Hs*Dod (PDB ID: 2CCB), the maximal diameter of the dodecamer is about 7 nm, but the diameter slightly varies depending on how the distance is measured: Maximal Cα-Cα distance is 6.8 nm, while the maximal distance including side chains is 7.5 nm. Because of the high symmetry *Hs*Dod can be seen as a sphere-like structure with the radius of about 3.5 nm or programs like HullRad can be used to create a polyhedron based on the structure (which basically treats dodecins as a sphere; diameter 7.1 nm).[278]

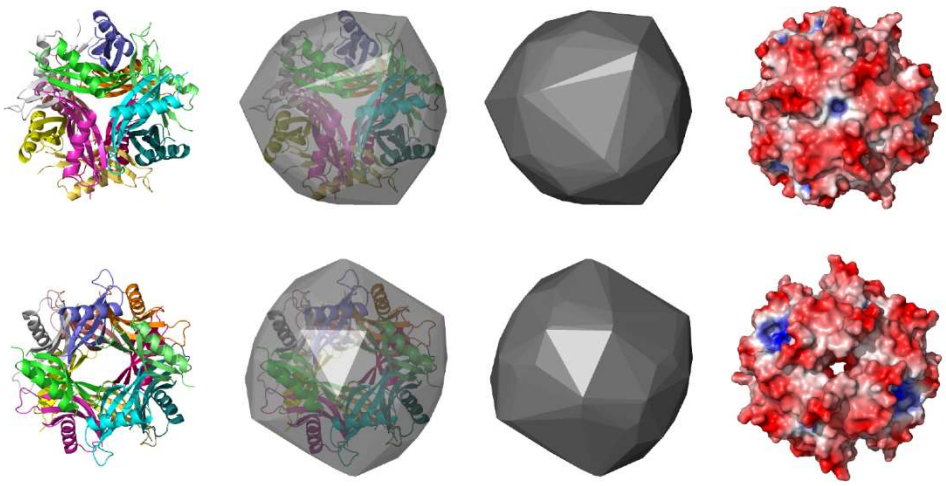


Figure 25: *Hs*Dod dodecamer in different presentation styles. Left to right: cartoon, cartoon and polyhedron overlay, polyhedron and charged surface (red negative, blue positive). Polyhedron created by using HullRad based Pymol scripts.[278]

This sphere-like structure of the dodecamer allows to ignore different orientations of the particle, and makes diffusion through a mesh only size dependent, although in this measurement the accuracy to observe such effects might not be sufficient.

A problem here is that the crystal structure does not contain any structural data of the used His-Tag (LEH_6), which are present 12-times on the surface of the dodecamer. Since the His-Tag is presented on the surface, it is very likely that the tag influences diffusion. Due to the flexibility of the His-Tag, its contribution to diffusion it is not easy to calculate. As another factor, the calculations are based on the Stokes-Einstein equation (radius to diffusion coefficient relation), which was shown to produce too large radii or too fast diffusion coefficient, since it does not include the hydrodynamic drag (diffusion particles can increase the local viscosity and thereby diffuse slower).[279] Accordingly, the diffusion coefficients measured by Nöll *et al.*[237] via diffusion ordered spectroscopy (DOSY) NMR suggested a double particle size (about 16 nm diameter) for free *Hs*Dod than measurements based on the crystal structure would have suggested. Therefore, the actual measured diffusion constants are of limited value, but the relative change of diffusion is still a valid tool to analyse the entrapment in the DNA hydrogel. This already shows the value of dodecins as a particle since their core size is defined by the structure and due to their structure similarity all dodecins can be treated as about 7-8 nm particles.[1,13,14,52,53,7] *Hs*Dod was picked in this study since it does not bind to DNA (non-specific binding),[280,281] likely because of the high density of negatively charged amino acid residues on the surface and the low pI (3.8 calculated with ExPASy-ProtParam[282]).[1]

High particle stability allows freedom for experimental design

To measure the actual quality of the DNA hydrogel, it is important that the diffusion probe is present during the formation of the hydrogel, since the particle cannot enter the hydrogel afterwards if the mesh size is to narrow.[283] Because DNA hydrogels require hybridization of DNA strands for the formation, the solution containing the building blocks and *Hs*Dod needed to be heated up to 60 °C and then cooled down to form the *Hs*Dod containing DNA hydrogel (Nöll *et al.* 2018[237]). In the cases of the two larger building blocks, the diffusion of *Hs*Dod in the respective DNA hydrogels was unaltered, indicating that these gels do not contain any significant amount of small meshes blocking *Hs*Dod diffusion. It needs to be noted, that in theory also large pores, originating from gel irregularities, could cause this behaviour. For the DNA hydrogel formed with shorter building blocks, the diffusion coefficient of *Hs*Dod was reduced substantially, but not completely. Up to 35 °C the diffusion coefficient of trapped *Hs*Dod stayed relatively constant (about 0.15×10^{-10} m^2/s), above 35 °C the diffusion coefficient increased and with 50 °C reached a value (0.41×10^{-10} m^2/s) close to free *Hs*Dod diffusion (about 0.48×10^{-10} m^2/s). One reason for a remaining slow diffusion of the trapped dodecin is the lower stability of the DNA hydrogel built with shorter building blocks, since the shorter sticky ends cause a temperature dependent dehybridization and rehybridization (basically no true gel). At high enough temperatures the DNA hydrogel fully falls apart and *Hs*Dod can diffuse freely. The actual results and the quality of the DNA hydrogel will not further be discussed here, since they are not related to dodecins.

Dodecamer stability can be measured by flavin binding

To measure the temperature dependent diffusion, the probe must also tolerate the evaluated temperatures during the measurement (here 50 °C). Arguable this is the biggest problem of protein nanoparticle, since they can be temperature-sensitive. In the case for dodecins, the stability can be measured via an assay, created and tested in the publications Bourdeaux *et al.* 2018[55] and Bourdeaux *et al.* 2019[53], that measures the ability of dodecins to rebind flavins after a phase of heightened temperatures (heating phase). After each heat phase the sample is cooled down to 5 °C to allow to rebind flavin released from the binding pocket during the heating phase. Since only the assembled dodecamer contains the six binding pockets, it can be assumed that levels of not rebound flavins (measured by their fluorescence) directly correlate to the amount of denatured dodecamer/dodecin. After each heating and cooling phase (cycle),

the heating phase temperature is increased, while the cooling temperature is kept constant. This allows to obtain melting curves for dodecins, which can be used to analyse the dodecin stability under various conditions. Based on the temperature cycles, the assay was called thermocyclic fluorescence assay. The thermocyclic fluorescence assay was inspired by the "ThermoFAD" method (as the affinity measurement introduced in Chapter 5.1.3), which measures the stability of flavoenzymes via the increase in fluorescence upon release of FAD during melting/denaturing of the enzyme.[274] Since dodecins release flavins without melting, the cooling phase was included. With an assay adapted for *Hs*Dod (using RbF instead of FMN) it was shown that the thermostability of *Hs*Dod highly depends on the salt concentration in the buffer solution. While *Hs*Dod dodecin is still relatively stable at the used salt concentration (200 mM NaCl), a to 4 h prolonged heating phase assay (to measure long time stability during the whole diffusion measurement process) showed that after 4 h at 50 °C already about 9% of *Hs*Dod denatured and after 4 h at 60 °C about 39%. The dependency on high ion concentrations is likely based on the high number of acidic amino acid residues (16 negatively charged compared to 4 positively charge residues) to tolerate the high KCl concentrations (about 4 M depending on external salt concentration) in the cytosol of *H. salinarum*.[284–287] While *Hs*Dod was still acceptable for the diffusion measurements under these conditions, other dodecins might have been a better choice.

Stability of bacterial dodecins

Bacterial dodecins are in general highly stable proteins (**Table 4**), but archaeal dodecins require higher salt concentrations, because of their high number of negatively charged amino acid residues (Nöll *et al.* 2018[237]).

Table 4: Overview of the stability of bacterial dodecins. TcFA: Thermocyclic fluorescence assay; temperatures depicted at which a substantial increase of fluorescence was observed (knee of the curves), method is thereby not as precise as other methods and should be seen more as an estimate. For SDS-PAGE data SDS and pH-value of the sample preparation are given (heat treatment step), if no conditions were stated in the reference, Laemmli conditions are assumed (2% SDS, pH 6.8). Sample preparation was in general the same (95 °C, 5 min), no conditions were stated in ref. [14]. M: Monomer. D: Dodecamer (in ref. [14] called hexamer) DSC: Differential scanning calorimetry. CD: Circular dichroism.

Dodecin	Reference	Method	Condition (pH, NaCl, additives) (SDS content, pH)	Transition temperature / Observed species
*Mt*Dod	Bourdeaux *et al.* 2018[55] Bourdeaux *et al.* 2019[53]	TcFA	pH 5.0-7.5, 300 mM NaCl	> 95°C
	Bourdeaux *et al.* 2018[55]		pH 4.0, 300 mM NaCl	87-89 °C
	Bourdeaux *et al.* 2020[236]		PBS (pH 7.4)	> 95°C
			pH 4.2, 150 mM NaCl	75-80 °C
	Bourdeaux *et al.* 2020[236]	SDS-PAGE	2.5% SDS, pH 6.8	D[a] (stable)
			3.3% SDS, pH <4.0	M (denatures)
	ref. [14]	DSC	pH 7.0, no salt, −FMN	118 °C
			pH 7.0, no salt, +FMN	123 °C
		SDS-PAGE	not stated, likely (2% SDS, pH 6.8)	D (stable)
*Tt*Dod	ref. [13]	CD	pH 8.0, no salt	> 95°C
		SDS-PAGE	4% SDS, likely pH 6.8 (not stated)	D + M (stable)
*Sc*Dod	Bourdeaux *et al.* 2019[53]	TcFA	pH 5.0-7.5 300 mM NaCl	>95 °C
		SDS-PAGE	2% SDS, pH 6.8	D (stable)
*Sd*Dod	Bourdeaux *et al.* 2019[53]	TcFA	pH 5.0-7.5 300 mM NaCl	88-91 °C
		SDS-PAGE	2% SDS, pH 6.8	M[b] (denatures)

a: prolonging the heat treatment duration (30 min) appearance of weak monomer bands (*Mt*Dod with C-terminal fused peptides ;Bourdeaux *et al.* 2020[236])
b: Excessive amounts of flavin stabilize SdDod and dodecamer is observable in SDS-PAGE

So far, all studied bacterial dodecins, except *Sd*Dod, are stable above 95 °C at pH 6.8-8.0 (even at low ion strength conditions, for *Mt*Dod not explicitly tested) and can tolerate 2-4% SDS depending on the dodecin. The lower stability of *Sd*Dod can be increased by introducing the mutation D10E, which strengthens the salt bridge to K62 (E10 and K62 are highly conserved in bacterial dodecins >90% (Bourdeaux *et al.* 2018[55])) (Bourdeaux *et al.* 2019[53]). The *Sd*Dod(D10E) mutant showed similar stabilities as the other dodecins (SDS-PAGE: only dodecamer; TcFA: >95 °C (Bourdeaux *et al.* 2019[53])). In line with the importance of this salt bridge for stability, the *Sc*Dod(K62A) mutant was substantially destabilized (SDS-PAGE: only monomer; TcFA: 51-52 °C (Bourdeaux *et al.* 2019[53])). Interestingly the *Sd*Dod(K62A) mutant was less affected (TcFA: 75-77 °C (Bourdeaux *et al.* 2019[53]), indicating that *Sd*Dod evolved to compensate the weakened salt bridge.

The high stability of *Mt*Dod is also not or at least not substantially affected by modifications at its termini, shown by SDS-PAGE of diverse fusion constructs (ranging from peptides to full proteins like msfGFP; Bourdeaux *et al.* 2020[236]). For constructs with thermostable fusions, like peptides, this could also be shown with the TcFA (for most constructs fluorescence increase was observed at 75-80 °C), but thermolabile fusions affected the FMN rebinding, likely through their denaturing and aggregation while the dodecamer stayed intact (Bourdeaux *et al.* 2020[236]). For such constructs the *Mt*Dod dodecamer was still observable in SDS-PAGE (Bourdeaux *et al.* 2020[236]). For other dodecins the effect of fusions is not known, but it can be assumed, that they would behave similarly, since the termini are solvent exposed (based on their crystal structures, see Chapter 2.2.1 **Figure 5** for all available structures). For *Hh*Dod no stability data is available, but based on the high similarity to other bacterial dodecins (see **Figure 23**) it can be assumed that it has a similar stability. A strong salt dependency of the thermal stability of *Hh*Dod seems unlikely, since it has a relatively balanced amount of acidic and basic amino acid residues ($N_{Asp}+N_{Glu} = 9$; $N_{Lys}+N_{Arg} = 7$). A reason for this could be that *H. halophila* does not require high cytosolic KCl concentrations and only uses them as osmoprotectant if needed, alternatively it uses betaine and glycine (in low potassium environments) as osmoprotectants.[288,289]

Concluding dodecin as nanoparticle

Bacterial dodecins can be seen as spheres with a diameter of 7 nm, which tolerate acidic environments, solvents (e.g. about 40-50% DMSO, above *Mt*Dod can precipitate; Bourdeaux *et al.* 2020[236]) detergents (SDS) and heat (>95 °C). In addition to its core stability, the exposed termini on the dodecamer's surface allow easy modification at the gene level to produce functionalized nanoparticles. Those modifications could likely also be used to increase the dodecamer's particle size, although the spherical character will be affected (spiked/brambled sphere or even spherical cluster). Proteins that are comparable to dodecins are proteins of the ferritin superfamily (here considered: ferritin (24-mers) and DNA binding proteins from nutrient starved cells (Dps; 12-mers).[192] Similar to dodecin, these are oligomeric protein complexes that form a hollow sphere-like structure with cubic symmetry, but owing to their larger inner cavity they are rather treated as nanocages (**Table 5**).[192] Although encapsulins and virus capsid/hull proteins (forming VLPs) are technically nanoparticles (based on size definition), they will not be considered here, because the comparison to dodecin is less fitting (much higher oligomeric state and larger size) and they are also more seen as nanocages.[150,290,291]

Table 5: Comparisons of ferritin, Dsp and dodecin nanoparticle/nanocages. Ferritin and Dsp data based on ref. [192]. *Mt*Dod data measured from the crystal structure of *Mt*Dod (PDB ID: 2YIZ)[51] and based on ref. [1] and ref. [51].

Protein type	Oligomeric state	Diameter outer/ inner	Shape
Ferritin	24-mer	12 nm / 8 nm	hollow sphere
Dps	12-mer	9 nm / 5 nm	hollow sphere
Dodecin	12-mer	7 nm / 2-3 nm	hollow sphere

The stability of ferritins and Dsp seems to vastly differ dependent on the source organisms, but in general they can be regarded as stable proteins and there are also highly stable ferritin/Dsp variants, e.g. the ferritin of *Pyrococcus furiosus* (*Pf*Ferritin; tolerates 120 °C about 1 h).[192,292–294] Still based on a rough comparison dodecins seem to be overall, e.g. pH and SDS tolerance, more stable than Dps and ferritins.[192] Because of the smaller inner cavity of dodecins, their potential as nanocage is limited, but small compounds can be inserted, e.g. N-(2-ethyl-iodo-acetamide)-dansyl was attached to a defolded *Mt*Dod(T59C) mutant and incorporated in the

dodecamer by refolding.[51] Overall, bacterial dodecins seem to be a contribution to the existing pool of available nanoparticles.

5.2.2 The Dodecin of *M. tuberculosis* as a Carrier or Scaffold

In this chapter the use of dodecins as carrier or scaffold will be discussed on the aspect of applicability based on the presented data by Bourdeaux *et al.* 2020[236]. The aspect of the stability of the dodecamer was discussed in the chapter above and will not be included here. Beforehand, the use of dodecin as scaffold in the regard to recruiting enzymes (enzyme hubs) was not studied by Bourdeaux *et al.* 2020[236] and therefore cannot be discussed here. While the data might suggest that dodecins could also be used as scaffolds to form enzyme hubs, only the use of *Mt*Dod as antigen carrier was actually tested, and here dodecins will not be discussed beyond the role as a carrier.

Production of *Mt*Dod carrier constructs: The highs and lows of solubility

For *Mt*Dod as carrier two general options of attaching of the "cargo", e.g. an antigen, or rather functionalizing of *Mt*Dod were tested by Bourdeaux *et al.* 2020[236]: 1.) producing the carrier directly with the cargo fused on gene level, here called direct charging. 2.) fusing attachment sites, like the SpyCatcher or SpyTag (small domain that binds and forms a covalent bond with its tag (small peptide)),[295,296] to dodecin on gene level to allow a later attachment of the cargo, here called post-production charging. For simplification, post-production refers to everything after the protein translation, e.g. by enzymes in the cytosol. While direct charging is preferable, since it does not require an extra charging step for production, it requires that *Mt*Dod and also the cargo can be produced as fusion proteins and it is limited to peptides and proteins as cargo. Therefore, the tolerance of *Mt*Dod for terminal fusions was tested by Bourdeaux *et al.* 2020[236]. By expressing diverse *Mt*Dod (7.5 kDa) fusions constructs ranging from small peptides (12-29 amino acids long; referred to as *Mt*Dod peptide constructs) to full proteins like msfGFP[297] (msfGFP: 26.3 kDa; *Mt*Dod msfGFP construct: 36.4kDa), it was shown that *Mt*Dod tolerates even about 4-times larger attachments and assembles still to the dodecamer. In general, it did not make a notable difference, if the C-terminus, the N-terminus or both (two peptides fused to *Mt*Dod) were used for fusion and will not be further discussed here, but additional studies are needed to fully analyse possible effects. While the dodecamer assembly was not problematic,

a few constructs formed non-classical inclusion bodies (inclusion bodies with correctly folded protein, here the dodecamer),[298] as observed in the presence of dodecamers in SDS-PAGE and by the flavin-binding ability of the non-classical inclusion bodies (yellowish colour).

Constructs forming non-classical inclusion bodies contained SpyCatcher[295,296] (C-terminal *Mt*Dod SpyCatcher fusion protein directly aggregated after French press), SYNZIPs[194] (SZ1 or SZ3; about 40-50 amino acid long helices, moderate hydrophobicity), *Se*ACP (*Saccharopolyspora erythraea* ACP, gene *chlB2*), or the peptide "ALMVYRCAPPRSSQF" (called Pep7, contains a cysteine) as fusions. While for *Mt*Dod fusion proteins containing SpyCatcher or *Se*ACP misfolding of the respective fused domain/protein could be the reason for the aggregation, this explanation cannot be used for SYNZIPs (simple helix) and Pep7 (unstructured). The SpyCatcher domain should fold normally in *E. coli*, since it is functional in the cytosol when fused to proteins.[299,300] *Mt*Dod SpyCatcher constructs were successfully refolded, but were aggregation prone, indicating solubility problems (sudden aggregation during concentration with ultra centrifugal filters and membrane filtration). The SpyCatcher:SpyTag complex was reported to be highly soluble (about 110 mg/ml (7 mM) in 50 mM NaCl and 20 mM Tris-HCl pH 8.0),[296] but the SpyCatcher-IMX313-SnoopCatcher fusion protein formed inclusion bodies and was refolded and stored in an alkaline and low ionic strength buffer (50 mM sodium borate, pH 9.5).[190] This could indicate that the SpyCatcher without SpyTag might have aggregation tendencies (although thermally stable)[190] and further buffer optimization (beyond the few tested conditions) might be needed to prevent the severe aggregation of *Mt*Dod SpyCatcher constructs. Low solubility and thereby aggregation would also explain the non-classical inclusion bodies of *Mt*Dod SYNZIP constructs, which were also obtained in soluble form by refolding and were aggregation prone afterwards. The *Mt*Dod *Se*ACP construct was not obtainable in soluble form, which might indicate folding problems, but this was not further investigated. The formation of non-classical inclusion bodies of *Mt*Dod-Pep7 (contains a cysteine) might be caused by disulfide bond formation between different dodecamers forming insoluble clusters/aggregates or by low solubility, but the Pep7 is only moderately hydrophobic. Since the cytosol of *E. coli* is reducing, disulfide bond formation during the expression is unlikely and therefore it seems more plausible that disulphide bonds were formed during cell lysis since the buffers contained no reducing agent (also higher flavin content, because of flavin sequestering might enhance disulfide bond formation).[301] In agreement with this theory, the *Mt*Dod Pep3 (ECYPNEKNSVNMDLD) construct was obtainable in soluble form although slowly formed aggregates, indicating that

the disulfide bond formation happened *in vitro*. For all other constructs, no low solubility was observed.

For nearly all constructs, linkers based on proline, alanine and serine (PAS-linker, e.g. SPAAPAPASPAS; based on ref. [302]) were used, which are suggested to be rather rigid compared to glycine and serine based linkers.[303] While PAS-linkers are not hydrophobic they are far from hydrophilic and some of the solubility problems might be solved via rigid hydrophilic linkers.[303]

Purification and yields

For *Mt*Dod peptide constructs yields of several 100 mg per litre expression culture (yield calculated since only a small fraction of the expression was purified; usually amounts of 20-50 mg) were obtained by using a simple heat denaturation step (75 °C, 15 min; denatures most *E. coli* proteins, subsequently removed by centrifugation), an organic solvent (e.g. DMSO) precipitation protocol after cell lysis and a subsequent size exclusion chromatography (SEC) purification step. This purification strategy also removes enough endotoxins (30 ± 23 EU/mg of purified protein; EU: endotoxin unit) to be directly useable for immunisations in rabbits (above 5-10 EU/kg of rabbit per injection (in general 100 µg))[304,305] For *Mt*Dod constructs with fused proteins, this protocol was also usable by adapting the heat denaturation temperature (down to 55 °C, 15 min; depending on the thermal stability of the fused protein), but resulted in substantial aggregation (peak representing aggregates observable during SEC). For those constructs, affinity purification was the better option. With both methods similar high purity was obtained. For *Mt*Dod protein constructs, no yield was calculated, but obtained amounts were more than sufficient for all experiments. For the *Mt*Dod SpyCatcher and *Mt*Dod SYNZIP constructs only a few milligrams were purified, because of the aggregation problems. These protocols were not optimized since the yields were sufficient for the experiments. In general, *Mt*Dod fusion proteins can be obtained in high purity and high yield as long as the fused cargo does not cause aggregation.

While the purification strategy for *Mt*Dod seems practical and the yields are quite high, also *Pf*Ferritin can be purified simply by incubation of *E. coli* cytosol (*Pf*Ferritin overexpressing cells) at 100 °C for 30 min, centrifugation (removal of denatured proteins) and buffer exchange by ultra centrifugation filters (100 kDa cut-off, which likely removes remaining impurities) with similar yields.[292] This protocol might be an time-saving alternative compared to the

protocol used for *Mt*Dod so far, as it does not rely on a SEC step, which would make the *Mt*Dod production also even more accessible, as only a centrifuge would be required.

Accessibility on the dodecamer surface

For post-production charging, the accessibility of the attachment sites at the surface of the *Mt*Dod dodecamer is critical for the degree of charging. The accessibility of attachments sites linked via PAS-linkers to the 12 (N- or C-terminus) or 24 (both) termini at the surface of the dodecamer were tested with two approaches: covalently attaching cargo proteins via the SpyTag-SpyCatcher system and/or SnoopTag-SnoopCatcher system (similar to SpyCatcher system)[306] and the labelling of *Mus musculus* ACP (*Mm*ACP, gene *Fasn*) with a 4'-phosphopantetheine CoA fluorophore (ATTO-TEC dye: ATTO 488) by the 4'-phosphopantetheine transferase from *Bacillus subtilis* (Sfp).

The covalent attachment of cargo to *Mt*Dod by the Tag-Catcher systems was analysed via SDS-PAGE (shift to higher masses). *Mt*Dod Tag constructs (single terminus (SpyTag) or both termini (SpyTag and SnoopTag)) reacted with excesses of cargo Catcher constructs (*Se*ACP-SpyCatcher (23.8 kDa), mClover3-SnoopCatcher[307] (41.3 kDa)) to a high degree: In none of the reactions, unreacted *Mt*Dod Tag construct was observable, but in the reactions with both cargo Catcher constructs, small amounts of single reaction product (only one of the two cargo constructs was attached) in addition to the double reaction product (both cargo constructs were attached) were found.

The absence of unreacted *Mt*Dod Tag construct shows that the accessibility of tags on the surface is not hindered and freely available for the reaction. However, upon reaching a high degree of charging in the dual Tag constructs (nearly all Tags reacted), the increasing steric hindrance seems to slow down or even prevent the complete saturation. Because of the long reaction time of 20 h and the fast reaction speed of the Tag-Catcher reactions,[295,296,306] it seems more like a saturation limit by steric hindrance than just a slowdown of the reactions. Considering that a fully dual charged dodecamer (dual Tag *Mt*Dod: Monomer: 13.1 kDa; Dodecamer: 157.7 kDa) would reach a mass of 939.0 kDa, even a not fully complete saturation is impressive.

In the case of the *Mt*Dod SpyCatcher constructs (Monomer: 22.4 kDa; Dodecamer: 269.0 kDa), which is the preferred and in general used setup,[220,186,190] since normally the cargo is seen as less stable than the carrier and thereby should be only modified with the SpyTag (or SnoopTag if both Catchers are used), the reaction with SpyTag-*Se*ACP (14.0 kDa) was less successful. While overall the reaction worked, clearly visible bands for the unreacted carrier were observable. As this cannot be caused by steric hindrance, since only twelve attachments sites were used (compared to the 24 above), it is more likely that the aggregation problems of the *Mt*Dod SpyCatcher construct caused this result. In comparison the SpyCatcher-IMX313-SnoopCatcher, which forms heptamers, was reported to be fully saturated without problems.[190] This shows clearly that the *Mt*Dod/SpyCatcher constructs and their production needs to be further optimized to be comparable to the SpyCatcher-IMX313-SnoopCatcher.[190]

In the case of labelling *Mm*ACP fused to *Mt*Dod with Sfp, measured by in-gel fluorescence of the attached fluorophore, only a degree of labelling of 34% was obtainable (assuming 100% degree of labelling of free *Mm*ACP as a reference). The reason for this low degree of labelling is likely caused by the lower accessibility of the *Mm*ACP fused to *Mt*Dod compared to free *Mm*ACP, which slows down the reaction. Surprisingly, prolonging the reaction time from 1 h to 2 h did only slightly increase the fluorescence intensity of *Mm*ACP fused to *Mt*Dod representing bands (about 19%, degree of labelling was not determined, but the increase would translate to maximal about 41% degree of labelling). The reason for this low degree of labelling is not clear. The *Mt*Dod *Mm*ACP constructs were purified by heat denaturation (untagged construct) or Ni-chelating chromatography (His-tagged construct) and both constructs delivered similar results in CoA labelling (His-Tag variant: 36% ± 8%; tag-less variant: 31% ± 8%). It might be possible that *Mt*Dod bound the CoA fluorophore and thereby lowered the amount of available CoA fluorophores for the reaction (CoA fluorophore was used equimolar and not in excess), but that the CoA fluorophore could fit in the flavin binding pocket (ATTO 488 moiety) or the CoA binding site (CoA moiety) seems unlikely but not impossible. The addition of *Mt*Dod to the reaction of free *Mm*ACP was not tested and so the reason for the low degree of labelling remains unclear.

Overall, peptides and proteins fused to the termini of *Mt*Dod are accessible. In case all 24 termini are supposed to be charged, longer linker might be necessary to reduce the cargo density and thereby preventing steric hindrance at high levels of charging.

Heterododecamer or chimera dodecamer

The observed low solubility of some *Mt*Dod fusion constructs (e.g. *Mt*Dod SpyCatcher) and the desire to use more than two different attachment sites per dodecamer, gave birth to the idea of generating heterododecamers. Accordingly, heterododecamers should be produced by letting differently modified *Mt*Dod monomers (including unmodified monomers) jointly assemble to the dodecamer, yielding a mixture of different heterododecamer species. The formation of heterododecamers was tested with the jointly refolding of *Mt*Dod wild type (*Mt*Dod(WT)) monomers and *Mt*Dod monomers modified with a PAS-linker and a StrepII-Tag (*Mt*Dod-PAS-Strep), or the coexpression of both constructs. Because of the abnormal running behaviour of the *Mt*Dod dodecamer (far below its actual molecular weight, which lead to the idea that it runs as a hexamer)[14] in SDS-PAGE and the limited SDS binding to the PAS-linker,[302] which results in a slower migration of the PAS-linker modified constructs in SDS-PAGE, allowed to separate the different heterododecamer species by their *Mt*Dod-PAS-Strep content in SDS-PAGE experiments. Since the separation mainly depends on the PAS-linker content (increase of mass and particle size, also play a role) of the heterododecamer species bands representing all 13 possible composition-species (ratio of different monomer in the dodecamer) were observed.

The position of PAS-Strep peptides did not seem to make a substantial difference in the migration behaviour. If it would do so, more than 13 bands should have appeared on the SDS-PAGE gel, because then the actual number of species would be beyond 13. Interestingly, different to the Tris-Glycine SDS-PAGE (Laemmli)[308], in Tris-Tricine SDS-PAGE (Schägger)[309] only a single large blurred/smeared band was observable, which might indicate unresolved separation of species differing in the relative orientation of the PAS-Strep attachment. This is not caused by unstable migration of the dodecamer, since in Tris-Glycine SDS-PAGE *Mt*Dod homododecamers run as a sharp/normal band.

By jointly refolding different concentration ratios of *Mt*Dod(WT) and *Mt*Dod-PAS-Strep it was shown that the assembly of heterododecamers is concentration controlled and the spectrum of heterododecamer species can be shifted by varying the portion of one of the components. The same effect was observed when coexpressing the respective genes using bicistronic and tricistronic expression vectors. By altering the gene order in the vector system, the pattern of heterododecamer species representing bands shifted, likely because of the gene order dependent expression strength/rate.[310]

The heterododecamer formation was only tested in this proof of concept approach and was not used for actually increasing the solubility or rather creating less aggregation prone dodecamers by thinning out the fused cargo on the surface. The ability to separate *Mt*Dod heterododecamer species by SDS-PAGE is interesting as it could be used as an easy tool to analyse the design dependent expression strength of polycistronic vector systems. For the formation of multifunctional (more than two) heterododecamers, this method can technically be used, but without controlling the actual species formation, this seems rather pointless as only a vast mixture of species with different orientations of attachments will be formed.

The formation of hetero-assemblies (sometimes called chimeras) like the heterododecamer can also be done with ferritins,[150,192] and likely with all self-assembling protein structures, and is thereby not a special feature of dodecins, but the ability to separate them via SDS-PAGE makes them interesting as it could be used to purify the distinct species by, e.g., preparative SDS-PAGE.

Case study Antibody production

To test *Mt*Dod as an actual antigen carrier for the production of custom ABs, 11 peptides were fused to *Mt*Dod with a PAS-linker (called *Mt*Dod-PAS-Pep constructs). The peptides were based on the sequence of different proteins (mostly human heat shock proteins (HSP) and the heat shock cognate 70 interacting protein (CHIP)) without any consideration of the properties of the resulting peptides, like solubility or cysteine content. Basically, the only reasoning for the selection of the peptides was to obtain ABs that recognize the selected peptide/epitope. The *Mt*Dod-PAS-Pep constructs (*Mt*Dod-PAS-Pep1 to *Mt*Dod-PAS-Pep11) were purified with the methods described above and for all constructs 20-50 mg of purified proteins were obtained, except of *Mt*Dod-Pep7, which was not purified because of inclusion body formation and therefore dropped from the study/experiment. All purified *Mt*Dod-PAS-Pep constructs were used to produce ABs in rabbits by Davids Biotechnologie GmbH, Germany. For the immunizations, the adjuvant MF59/AddaVax was used and five injections of 100 µg *Mt*Dod-PAS-Pep construct over 63 days were conducted. The ABs were purified from the sera by affinity chromatography with the specific *Mt*Dod-PAS-Pep construct bound to the column matrix (by Davids Biotechnologie GmbH, Germany), thereby purifying all ABs which recognize the peptide itself and/or *Mt*Dod-PAS. The idea behind this was to remove undesired ABs, although it was likely not needed and can be done if interfering ABs are actually observed.

For all ten *Mt*Dod-PAS-Pep constructs, ABs were obtained recognizing at least the *Mt*Dod-PAS-Pep construct. This indicates clearly that the *Mt*Dod-PAS carrier contains MHC II-epitopes and fulfils the needed requirement for a carrier (required to activate helper T-cells to activate B-cells, see Chapter 2.4.5). Since an adjuvant was used, the potential of the *Mt*Dod-PAS carrier to activate APCs (see Chapter 2.4.5) is not clear, but even if not doing so this does not hinder the use as carrier because adjuvants are anyway used typically for AB production. Eight of the ten *Mt*Dod-PAS-Pep construct derived ABs recognized the respective target proteins (origin of the peptide sequence) to some degree, which indicates that the *Mt*Dod-PAS presents the peptides in a way that the B-cell receptors can bind it (see Chapter 2.4.5). The two remaining *Mt*Dod-PAS-Pep construct derived ABs only recognized their respective *Mt*Dod-PAS-Pep construct, but not the *Mt*Dod-PAS construct, which could be caused by ABs that only recognize epitopes containing parts of the PAS-linker and the peptide, but without further testing this is pure speculation. Nevertheless, in eight from ten cases, ABs were produced with the *Mt*Dod-PAS-Pep constructs that could recognize their target proteins showing the *Mt*Dod potential as an antigen carrier.

The fact that *Mt*Dod-PAS-Pep construct derived ABs did not cause any systematic appearing false positive bands or backgrounds in lysates of HEK293 cells and that *Mt*Dod-PAS has no significant similarity with any human protein (no hits found with BLAST[311] under standard conditions), makes *Mt*Dod-PAS a carrier suited for production of ABs used for human sample analysis.

Concluding *Mt*Dod as antigen carrier for AB production

The small case study with ten selected antigens, shows that *Mt*Dod can be used as a carrier for AB production, but without in-depth analysis of these ABs, e.g., exact epitopes, affinities and serum concentrations, the "quality" of *Mt*Dod as a carrier for AB production cannot be judged.

For the production of custom anti-peptide ABs, the main goal is mostly to just obtain target binding/recognizing ABs and since the broadly available route of crosslinking synthetic peptides to KLH, BSA or OVA (to a lesser extend MAPs) can generally accomplish this, there is not a big incentive to use alternative routes (method stayed basically the same since the early- to mid-1980s)[201,209,211]. This said, for laboratories focussing on studying ABs and their

generation, which are proficient in the basics of heterologous protein production (including construct design and cloning) and purification, the *Mt*Dod approach can be interesting, since it allows more freedom in the design (peptides and the carrier can be altered, other linkers or other dodecins could be used, e.g. *Sc*Dod).

Dodecins might be most interesting in the field of peptide vaccine design/research, which comprises a vast spectrum of applications, like e.g. anti-cancer, anti-Alzheimer and anti-pathogen vaccines.[198,312] Here, dodecins are likely a good addition to the pool of carriers: Ferritins (Dsp included, but are not explicitly mentioned in literature), VLPs and self-assembling peptides.[218] So far, VLPs seem to be the most dominant type likely because of their good immunogenicity (size and structure facilitate APC uptake),[208,218,313] which might be the biggest hurdle for dodecins to overcome.

In the end only time will tell, which carriers are superior or even if a single carrier type can be the general answer.

6 Summary

6.1 Biological Function of Bacterial Dodecins

In this thesis, the dodecins of *Mycobacterium tuberculosis* (*Mt*Dod), *Streptomyces coelicolor* (*Sc*Dod) and *Streptomyces davaonensis* (*Sd*Dod) were studied. Kinetic measurements of the flavin binding of *Mt*Dod revealed that the dodecin binding pocket is filled in two distinct steps, for which a kinetic model then was established and verified by experimental data. The analysis with the two-step model showed that the unique binding pocket of dodecins allows them to bind excessive amounts of flavins, while at low flavin concentrations, flavin is released and only weakly bound. This function of flavin buffering prevents accumulation of free oxidised flavins and therefore helps to keep the redox balance of the cell and prevents potential cell damage caused by excessive free flavins. To further gain insights into the role of bacterial dodecins, the effect of knocking out the dodecin encoding gene in *S. davaonensis* was analysed. The knockout strain showed increased concentrations of various stress related metabolites, indicating that without dodecin the cellular balance is disrupted, which supports the role of dodecins as a flavin homeostasis factor.

With a self-designed affinity measurement method based on the temperature dependent dissociation of the dodecin:flavin complex, which allowed parallel screening of multiple conditions, it was shown that *Mt*Dod, *Sc*Dod and *Sd*Dod have much higher affinities towards FMN and FAD under acidic conditions. Under these conditions, the three dodecins might function as a FMN storage. *M. tuberculosis* encounters multiple acidic environments during its infection cycle of humans and can adopt a state of dormancy. During recovery from the dormant state, a flavin storage might be beneficial. For some *Streptomyces* species it was reported that the formed spores are slightly acidic and therefore *Sc*Dod and *Sd*Dod could function as flavin storages for the spores. Further details on the flavin binding mechanism of *Mt*Dod were revealed by a mutagenesis study, identifying the importance of a histidine residue at the fourth position of the protein sequence for flavin binding, but contrary to expectations, this residue seems only to be partly involved in the pH related affinity shift.

The data, reported in this thesis, demonstrates that bacterial dodecins likely function as flavin homeostasis factors, which allow overall higher flavin pools in the cell without disrupting the cellular balance. Further, the reported acid-dependent increase in binding affinity suggests that under certain conditions bacterial dodecins can also function as a flavin storage system.

6.2 Application of the Dodecin of *M. tuberculosis*

In this thesis, the stability of *Mt*Dod, *Sc*Dod *Sd*Dod and *Hs*Dod was analysed to find a suitable dodecin for the use as a carrier/scaffold. Therefore, a method to easily measure the stability of dodecins was designed, which measures the ability of the dodecamer to rebind flavins after a heating phase with stepwise increasing temperatures. Using this assay and testing the stability against detergents by SDS PAGE, showed that the dodecamer of *Mt*Dod possesses an excellent stability against a vast array of conditions, like temperatures above 95 °C, low pH and about 2% SDS. By solving the crystal structure of *Sc*Dod and *Sd*Dod, the latter forming a less stable dodecamer, combined with a mutagenesis study, the importance of a specific salt bridge for dodecamer stability was revealed and might be helpful to find further highly stable dodecins.

In addition to the intrinsic high stability of the *Mt*Dod dodecamer, also the robustness of the fold was tested by creating diverse *Mt*Dod fusion constructs and producing them in *Escherichia coli*. Here it was shown that *Mt*Dod easily tolerates the attachment of proteins up to 4-times of its own size and that both termini can be modified without affecting the dodecamer noticeably. Further, it was shown that *Mt*Dod and many *Mt*Dod fusion constructs could be purified in high yields via a protocol based on the removal of *E. coli* proteins through heat denaturation and subsequent centrifugation. In a case study, by fusing diverse antigens from mostly human proteins to *Mt*Dod and using these constructs to produce antibodies in rabbits, it was demonstrated that *Mt*Dod is immunogenic and presents the attached antigens to the immune system.

The here reported properties of *Mt*Dod and to a lesser degree of other bacterial dodecins, show that bacterial dodecins are a valuable addition to the pool of scaffold and carrier proteins and have great potential as antigen carriers.

7 Outlook

7.1.1 Missing Proof of the Biological Role of Bacterial Dodecins

While the role of bacterial dodecins to function as a flavin buffer seems plausible and all reported data is in line with this role, it is so far still only a hypothesis. For proofing that bacterial dodecins actually buffer the amount of free flavins in the cell, the ratio of non-dodecin bound flavins, dodecin bound flavins and free flavins in combination with the absolute flavin content and dodecin concentration would be needed. Obtaining this data might be quite difficult, since as soon as the cellular equilibrium is disturbed, for example by cell disruption, dodecins will start to bind or release flavins. Alternatively, a combination of transcriptomic or proteomic and metabolomic studies need to be conducted, which in detail analyse the effects of the knockout of the dodecin gene on the cell compared to the wildtype. This should be combined with studies of binding pocket mutants, which either don't bind flavins or don't bind CoA, to distinguish flavin or CoA related effects. To then relate those found effects to actually altered flavin homeostasis, the introduction of a flavin uptake system, like ribM from *Corynebacterium glutamicum*,[32] could be used to cause a "flavin shock" and then the cellular response can be compared to dodecins knockouts.

7.1.2 The Dodecin of *M. tuberculosis* as a Drug Target?

While the unique binding of flavin dimers by dodecins would likely allow to design specific binders by mimicking the flavin dimers, so far, the importance of *Mt*Dod during the infection is not known. If a *Mt*Dod gene knockout study would show that *Mt*Dod is important for the virulence of *M. tuberculosis*, which might be not too unlikely since the potential flavin sequestering protein Acg (Gene-ID: *Rv2032*) was shown to be important for the virulence of *M. tuberculosis,*[247–249] saturating the *Mt*Dod with a flavin dimer mimetic might help to treat TB. Even without *Mt*Dod being directly important for the virulence, disrupting the flavin buffering of *Mt*Dod might weaken *M. tuberculosis* bacilli and make them more susceptible for other drugs. An indication for this is the upregulation of the isocitrate lyase (Gene-ID: *TTHA1836*) in the *Tt*Dod gene disruption strain of *T. thermophilus*,[238] which is in *M. tuberculosis* related to the oxidative stress response and survival in macrophages.[258,259] Facing the global threat by TB and the gained understanding of dodecin, presented in this thesis, a research focus on disclosing the role of dodecin as drug target seems justified.

7.1.3 Bacterial Dodecins as Carriers and Scaffolds

The here reported results show the great potential of bacterial dodecins as carriers or scaffolds. Because *Mt*Dod is easy to manufacture and immunogenic, it qualifies as an ideal candidate for new vaccine designs. The first aspect that needs to be tested, is how other animals and humans react on the administration of *Mt*Dod antigen conjugates. If *Mt*Dod does not cause any complications, it could directly be used as a carrier for peptide vaccines, which are not limited to the protection against infectious diseases and are currently developed to treat for example cancer and the Alzheimer diseases.[312,314] A problem of peptide vaccines is often the low immunogenicity or effeteness. Here, *Mt*Dod peptide fusions could be beneficial.[198,312]

While the use as a carrier for vaccines is likely the most interesting application of bacterial dodecins, the high stability of the dodecamer and robustness of the fold also makes it interesting for basic research. The high tolerance of diverse conditions makes *Mt*Dod a versatile carrier, which can be included in most experimental setting without the need to adapt parameters. Also, it can easily be produced in *E. coli* without complicated protocols, which makes it accessible for laboratories that are not highly experienced in protein purification. Especially in the field of immunology, a universally used carrier would be beneficial, as so far, the different used carrier systems require careful evaluation of the acquired data, as effects caused by the various carriers might overshadow other aspects.

.

8 Supplementary Information

8.1 Antibodies and their Generation by the Body

Additional information regarding the immune system with a focus on AB production. Information is based on ref. [202] and ref. [200] and gives a short introduction into immunology. In general, these processes might be considered common knowledge.

Description of ABs

ABs are soluble immunoglobins (Ig) and part of the adaptive immune system. They are divided into several different isotypes/classes and subclasses with specific functions, while the basic function to bind foreign material is shared by all Igs.[315] In addition to ABs, Igs are also membrane bound mainly as receptors for B-cell activation and differentiation (called B-cell receptors (BCR)). The Ig superfamily includes many more receptor types, like T-cell receptors (TCR) and the major histocompatibility complex (MHC) proteins (here relevant classes MHC I and MHC II). The general structure of Igs is a homodimer of two complexes containing a heavy chain (HC) and a light chain (LC) (overall two HCs and two LCs). The complete Ig is divided into two regions the "Fragment, crystallizable" (Fc) and the "Fragment, antigen-binding" (Fab). Each Ig contains one Fc, which is only formed by the two HCs, and two identical Fabs (artificial bi-specific ABs can be produced)[316], which are formed by the LCs and HCs. In addition to building the base for dimerization of Igs, the Fc region is also recognized by Fc receptors and proteins of the complement system (part of the innate immune system) deciding the fate of bound antigens, like ingestion by phagocytes. The Fab region has two types of domains, a constant and a variable type, the variable domains contain the actual antigen binding site.[315] The Fc region contains two constant domains and no variable domains. The LC and the HC part of the Fab arm of an Ig each consists a variable domain (V_{LC} and V_{HC}) and a constant domain (C_{LC} and C_{HC}).[315] So, in principle to obtain ABs specific for a chosen antigen, the variable domains need to recognize parts of this antigen. How ABs are produced is briefly described in the next paragraph.

Production of ABs

ABs are part of the adaptive immune system, which can be divided in to the humoral and cellular immunity. The humoral part of the immune system is responsible for the defence against extracellular foreign components and ABs are a major part of this system. In contrast, the cellular part of the immune system defends the body against pathogen-infected cells by apoptosis. While AB production is part of the humoral immune system and will be focussed on here, the cellular system is an important factor for vaccine design, where also carrier proteins can be used (e.g. conjugate vaccines).[197] The most important cells involved in the AB production are "professional" antigen presenting cells (APCs, macrophages and here most important dendritic cells (DC)), B-cells (B lymphocytes formed in the bone marrow from hematopoietic stem cells (HSC)) and helper T-cells (called T_H cells or $CD4^+$ cells, lymphocytes that developed in the thymus, their origin is also in the bone marrow from HSCs). For clarity and conciseness, the subtypes of the involved cells will not be discussed. B-cells or more specific the plasma cells in which B-cells differentiate, secrete ABs after maturation through antigen contact and often T_H-cell involvement. The variable domain sequences of the later produced AB, are created during the B-cell formation in the bone marrow in a randomized process. Since this process is random, several selection steps are involved to prevent the production of ABs, which recognise parts of the body itself (the process will not be explained here). This means that even before any antigen contact was involved, the possibly later produced AB is already defined, but it needs to be noted that further optimisations in germinal centres to improve the antigen binding often occur. To ensure that only relevant ABs are produced, B-cells have a receptor (B-cell receptor; BCR), which is a membrane bound Ig with the exact variable domains as the ABs later produced by the B-cell (other involved proteins here ignored). Basically, only when an antigen binds to the BCR, the B-cell can undergo differentiation into a plasma cell and later produce ABs, but to fully differentiate also T_H-cells and APCs need to be involved. For full differentiation into plasma cells the TCR of an activated T_H cell needs to recognize the peptides bound to the MHC II proteins on the surface of B-cells.

The peptides presented on the MHC II proteins are fragments of foreign proteins ingested and processed by APCs or B-cells (BCR mediated endocytosis). This process is called the MHC II pathway and is mainly for extracellular antigens, although through the cross-presenting of peptides of the MHC I pathway (cellular antigens, e.g. viruses) also fragments of internal antigens can be presented on MHC II proteins and *vice versa*. These fragments/peptides are called T-cell epitopes or also MHC-epitopes, whereas the specific site of the antigen, which is

recognized by the AB, is called B-cell epitope or just epitope (the BCR and the produced ABs recognize the same epitope). For clearance, the term T-cell epitopes refers to MHC I-epitopes and MHC II-epitopes, which have distinct effects/roles. Since for AB production MHC II-epitopes are relevant MHC I-epitopes will not be discussed. Basically, an antigen can have several different B-cell epitopes, which can be linear (continuous sequences) or conformational (depend on the structure of the antigen and are not limited to continuous sequences). In contrast, since T-cell epitopes are fragments of processed proteins, they are continuous sequences (modifications endure the processing are still present). Similar to B-cell epitopes, proteins/antigens can also contain several T-cell epitopes. An important factor for BCR mediated endocytosis is BCR clustering (BCRs in the membrane migrate towards each other), which can be enhanced by multivalent antigens (antigen containing the same B-cell epitope multiple times, for example repetitive structures like virus hulls). Also, monovalent antigens cause BCR clustering.

T-cells develop in the thymus and undergo similar processes like B-cells, their randomised binding region of the TCR is tested against MHC proteins loaded with and without non-foreign peptides (self-antigens) to prevent false activation, while ensuring basic functionality. It is estimated that over 95% of the developing T-cells are removed during this process. The matured T-cells (although still not activated) leave the thymus and enter the blood stream and lymphoid system. T-cells and B-cells are like APCs highly mobile and move through the whole body via the blood stream and the lymphoid system. In lymph nodes and other secondary lymphoid organs (e. g. the spleen), T-cells are exposed to MHC II-epitopes presented by APCs (also other signals are involved) and start to proliferate and differentiate into effector cells (here called activated T_H-cells). In the special microenvironment of activated T_H-cells and antigen presence, B-cells proliferate and differentiate into plasma cells (also other cell types, which will be ignored here) and start to secrete ABs. B-cells that bound an antigen via their BCR get partly activated, engulf the particle, present MHC II-epitopes and migrate into T-cell rich regions of the lymph nodes. While antigens can enter the lymph nodes directly APCs are important to activate the T-cells. This is an important factor for AB production since APCs need to get activated by signals similar to pathogen invasion (basically, APCs need to recognize something as foreign). General factors are size and solvent exposed structures of the particle/compound. An important role here plays a group of receptors that recognize pathogen-associated molecular patterns (PAMPs), called pattern recognition receptors (PRRs). Here to mention are Toll-like receptors (TLRs) that recognize several diverse PAMPs, like bacterial

flagellin, double-stranded RNA of viruses or heat shock proteins (host and bacteria). Depending on what PRRs are activated, the APC (also other immune cells can be involved) produces different sets of signals (cytokines), which affect the development of T-cells (subclass/subset differentiation, e.g. T_H1 or T_H2). In addition to PAMPs, also signals that indicate damaged cells can attract APCs, which are called damage-associated molecular patterns (DAMPs). After "immature" APCs (especially DC fulfil this role) engulf a foreign particle, they activate/mature and migrate from the tissue via the blood stream or the lymph system into the lymph nodes. During this maturation they lose most of their antigen-capturing role and their ability to present antigens (T-cell epitopes) significantly increases. Only when all three cell types (APCs, T-cells and B-cells) are activated and interact with each other AB generation by the body is initiated.

9 Acronyms and Abbreviations

Dodecins

Dod	Dodecin
*Hh*Dod	The dodecin of *Halorhodospira halophila*
*Hs*Dod	The dodecin of *Halobacterium salinarum*
*Mt*CaDod	The dodecin of *Mycobacterium tuberculosis*
*Mt*Dod	The calcium-dodecin of *Mycobacterium tuberculosis*
*Mt*Dod(WT)	Wild type of the dodecin of *Mycobacterium tuberculosis*
*Sc*Dod	The dodecin of *Streptomyces coelicolor*
*Sd*Dod	The dodecin of *Streptomyces davaonensis*
*Tt*Dod	The dodecin of *Thermus thermophilus*

Strains

E. coli	*Escherichia coli*
H. halophila	*Halorhodospira halophila*
H. salinarum	*Halobacterium salinarum*
M. tuberculosis	*Mycobacterium tuberculosis*
S. coelicolor	*Streptomyces coelicolor*
S. davaonensis	*Streptomyces davaonensis*
S. meliloti	*Sinorhizobium meliloti*
T. thermophilus	*Thermus thermophilus*

General

AB	Antibody
ACP	Acyl carrier protein
APC	Antigen presenting cell
B-cell	Type of lymphocyte formed in the bone marrow, also called B lymphocyte
BCG	*Bacillus Calmette-Guérin* strain
BCR	B-cell receptor
BSA	Bovine serum albumin
Catcher	Small protein fold (SpyCatcher or SnoopCatcher) that binds and reacts with Tag
CD	Circular dichroism
CHIP	C-terminus of heat shock cognate 70 interacting
C_{LC} and C_{HC}	Constant domain light chain and heavy chain
CoA	Coenzyme A
DAMP	Damage-associated molecular patterns
DC	Dendritic cells
DHFR-TS	Dihydrofolate reductase thymidylate synthase adduct
DMSO	Dimethyl sulfoxide
DOSY	Diffusion ordered spectroscopy
Dps	DNA binding protein from nutrient starved cells
DSC	Differential scanning calorimetry
Fab	“Fragment, antigen-binding”
FAD	Flavin adenine dinucleotide

FC	“Fragment, crystallizable”
FCA	“Freund’s Complete Adjuvant”
FMN	Riboflavin-5′-phosphate, flavin mononucleotide
GBD	GTPase binding domain
Gene-ID	Identifier used for genes without a gene symbol or name, likely the locus tag
GFP	Green fluorescent protein
GTP	Guanosine triphosphate
HC	Heavy chain (antibody)
His-TAG	Polyhistidine peptide tag
HIV/AIDS	Human immunodeficiency virus infection and acquired immune deficiency syndrome
HSP	Heat shock protein
Ig	Immunoglobin
ITC	Isothermal titration calorimetry
K_D	Dissociation constant
KLH	Keyhole limpet hemocyanin
LC	Light chain (antibody)
LmF	Lumiflavin
LTBI	Latent tuberculosis infection
MAP	Multiple antigen peptides
MBS	*m*-maleimidobenzoyl-*N*-hydroxysuccinimide
MDH-CS	malate dehydrogenase citrate synthase adduct
MHC	Major histocompatibility complex
MHC I	Major histocompatibility complex class I
MHC II	Major histocompatibility complex class II
*Mm*ACP	*Mus musculus* acyl carrier protein
msfGFP	Monomeric superfolder green fluorescent protein
MTC	*Mycobacterium tuberculosis* complex
NEC	Necrosis-associated extracellular cluster
NMR	Nuclear magnetic resonance
OVA	Chicken ovalbumin
PAMPs	Pathogen-associated molecular patterns
PAS-linker	Peptide linker containing proline, alanine and serine
PDB ID	Protein data bank identifier
PDIM	Phthiocerol-dimycocerosate
*Pf*Ferritin	*Pyrococcus furiosus* ferritin
PRR	Pattern recognition receptor
RbF	Riboflavin, vitamin B2
RFN-element	FMN binding riboswitch
SAPN	Self-assembled peptide nanoparticles
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
*Se*ACP	*Saccharopolyspora erythraea* acyl carrier protein
SEC	Size exclusion chromatography
Sfp	4′-Phosphopantetheine transferase from *Bacillus subtilis*
SZ	SYNZIP domain

Tag	Small peptide sequence that interacts with Catcher's (SpyTag or SnoopTag) or affinity peptide tags
TB	Tuberculosis (diseases)
T-cell	Type of lymphocyte that develops in the thymus
TcFA	Thermocyclic fluorescence assay
TCR	T-cell receptors
T_H-cell	Helper T-cell
V_{LC} and V_{HC}	Variable domain light chain and heavy chain
VLP	Virus-like particle

10 List of Figures

11 List of Tables

12 References

[1] Bieger, B.; Essen, L.-O.; Oesterhelt, D. Crystal Structure of Halophilic Dodecin: A Novel, Dodecameric Flavin Binding Protein from *Halobacterium salinarum*. *Structure* **2003**, *11* (4), 375–385. https://doi.org/10.1016/S0969-2126(03)00048-0.

[2] Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S. V.; Eiglmeier, K.; Gas, S.; Barry, C. E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, K.; Osborne, J.; Quail, M. A.; Rajandream, M.-A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, R.; Squares, S.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G. Deciphering the Biology of *Mycobacterium tuberculosis* from the Complete Genome Sequence. *Nature* **1998**, *393* (6685), 537–544. https://doi.org/10.1038/31159.

[3] Camus, J.-C.; Pryor, M. J.; Médigue, C.; Cole, S. T. Re-Annotation of the Genome Sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiol. Read. Engl.* **2002**, *148* (Pt 10), 2967–2973. https://doi.org/10.1099/00221287-148-10-2967.

[4] White, H. B.; Merrill, A. H. Riboflavin-Binding Proteins. *Annu. Rev. Nutr.* **1988**, *8* (1), 279–299. https://doi.org/10.1146/annurev.nu.08.070188.001431.

[5] Grininger, M.; Zeth, K.; Oesterhelt, D. Dodecins: A Family of Lumichrome Binding Proteins. *J. Mol. Biol.* **2006**, *357* (3), 842–857. https://doi.org/10.1016/j.jmb.2005.12.072.

[6] Monaco, H. L. Crystal Structure of Chicken Riboflavin-Binding Protein. *EMBO J.* **1997**, *16* (7), 1475–1483. https://doi.org/10.1093/emboj/16.7.1475.

[7] Grininger, M.; Staudt, H.; Johansson, P.; Wachtveitl, J.; Oesterhelt, D. Dodecin Is the Key Player in Flavin Homeostasis of Archaea. *J. Biol. Chem.* **2009**, *284* (19), 13068–13076. https://doi.org/10.1074/jbc.M808063200.

[8] Gonzalez, O.; Gronau, S.; Pfeiffer, F.; Mendoza, E.; Zimmer, R.; Oesterhelt, D. Systems Analysis of Bioenergetics and Growth of the Extreme Halophile *Halobacterium salinarum*. *PLoS Comput. Biol.* **2009**, *5* (4). https://doi.org/10.1371/journal.pcbi.1000332.

[9] Vauclare, P.; Natali, F.; Kleman, J. P.; Zaccai, G.; Franzetti, B. Surviving Salt Fluctuations: Stress and Recovery in *Halobacterium salinarum* , an Extreme Halophilic Archaeon. *Sci. Rep.* **2020**, *10* (1), 3298. https://doi.org/10.1038/s41598-020-59681-1.

[10] Tarhay, S. L.; Buss, E. G.; Clagett, C. O. Avian Riboflavinuria: 10. Quantitative Changes of Riboflavin-Binding Protein in Individual Egg Tissues during Incubation. *Poult. Sci.* **1975**, *54* (2), 562–571. https://doi.org/10.3382/ps.0540562.

[11] Lee, C. M.; White, H. B. Riboflavin-Binding Protein Induces Early Death of Chicken Embryos. *J. Nutr.* **1996**, *126* (2), 523–528. https://doi.org/10.1093/jn/126.2.523.

[12] Shbailat, S. J.; Qanadilo, S.; Al-Soubani, F. A. Protease Activity in the Egg Yolk during the Development of Meleagris Gallopavo (Galliformes: Phasianidae) Embryos. *Ital. J. Zool.* **2016**, *83* (3), 291–297. https://doi.org/10.1080/11250003.2016.1206976.

[13] Meissner, B.; Schleicher, E.; Weber, S.; Essen, L. O. The Dodecin from *Thermus thermophilus*, a Bifunctional Cofactor Storage Protein. *J. Biol. Chem.* **2007**, *282* (45), 33142–33154. https://doi.org/10.1074/jbc.M704951200.

[14] Liu, F.; Xiong, J.; Kumar, S.; Yang, C.; Ge, S.; Li, S.; Xia, N.; Swaminathan, K. Structural and Biophysical Characterization of *Mycobacterium tuberculosis* Dodecin Rv1498A. *J. Struct. Biol.* **2011**, *175* (1), 31–38. https://doi.org/10.1016/j.jsb.2011.04.013.

[15] Nelson-Sathi, S.; Dagan, T.; Landan, G.; Janssen, A.; Steel, M.; McInerney, J. O.; Deppenmeier, U.; Martin, W. F. Acquisition of 1,000 Eubacterial Genes Physiologically Transformed a Methanogen at the Origin of Haloarchaea. *Proc. Natl. Acad. Sci.* **2012**, *109* (50), 20537–20542. https://doi.org/10.1073/pnas.1209119109.

[16] Müller, F. *Chemistry and Biochemistry of Flavoenzymes : Volume I*; CRC Press, 2018. https://doi.org/10.1201/9781351070577.

[17] Joosten, V.; van Berkel, W. J. Flavoenzymes. *Curr. Opin. Chem. Biol.* **2007**, *11* (2), 195–202. https://doi.org/10.1016/j.cbpa.2007.01.010.

[18] Massey, V. The Chemical and Biological Versatility of Riboflavin. *Biochem. Soc. Trans.* **2000**, *28* (4), 283–296. https://doi.org/10.1042/bst0280283.

[19] Hefti, M. H.; Vervoort, J.; Berkel, W. J. H. van. Deflavination and Reconstitution of Flavoproteins. *Eur. J. Biochem.* **2003**, *270* (21), 4227–4242. https://doi.org/10.1046/j.1432-1033.2003.03802.x.

[20] Holzer, W.; Shirdel, J.; Zirak, P.; Penzkofer, A.; Hegemann, P.; Deutzmann, R.; Hochmuth, E. Photo-Induced Degradation of Some Flavins in Aqueous Solution. *Chem. Phys.* **2005**, *308* (1), 69–78. https://doi.org/10.1016/j.chemphys.2004.08.006.

[21] Massey, V.; Strickland, S.; Mayhew, S. G.; Howell, L. G.; Engel, P. C.; Matthews, R. G.; Schuman, M.; Sullivan, P. A. The Production of Superoxide Anion Radicals in the Reaction of Reduced Flavins and Flavoproteins with Molecular Oxygen. *Biochem. Biophys. Res. Commun.* **1969**, *36* (6), 891–897. https://doi.org/10.1016/0006-291X(69)90287-3.

[22] Ghisla, S.; Massey, V.; Lhoste, J.-M.; Mayhew, S. G. Fluorescence and Optical Characteristics of Reduced Flavines and Flavoproteins. *Biochemistry* **1974**, *13* (3), 589–597. https://doi.org/10.1021/bi00700a029.

[23] Valle, L.; Morán Vieyra, F. E.; Borsarelli, C. D. Hydrogen-Bonding Modulation of Excited-State Properties of Flavins in a Model of Aqueous Confined Environment. *Photochem. Photobiol. Sci.* **2012**, *11* (6), 1051. https://doi.org/10.1039/c2pp05385c.

[24] Surre, J.; Saint-Ruf, C.; Collin, V.; Orenga, S.; Ramjeet, M.; Matic, I. Strong Increase in the Autofluorescence of Cells Signals Struggle for Survival. *Sci. Rep.* **2018**, *8* (1), 12088. https://doi.org/10.1038/s41598-018-30623-2.

[25] WILSON, A. C.; PARDEE, A. B. Regulation of Flavin Synthesis by *Escherichia coli*. *Microbiology,* **1962**, *28* (2), 283–303. https://doi.org/10.1099/00221287-28-2-283.

[26] Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **2009**, *5* (8), 593–599. https://doi.org/10.1038/nchembio.186.

[27] Yurgel, S. N.; Rice, J.; Domreis, E.; Lynch, J.; Sa, N.; Qamar, Z.; Rajamani, S.; Gao, M.; Roje, S.; Bauer, W. D. *Sinorhizobium meliloti* Flavin Secretion and Bacteria-Host Interaction: Role of the Bifunctional RibBA Protein. *Mol. Plant. Microbe Interact.* **2014**, *27* (5), 437–445. https://doi.org/10.1094/mpmi-11-13-0338-r.

[28] Yang, Y.; Wu, Y.; Hu, Y.; Wang, H.; Guo, L.; Fredrickson, J. K.; Cao, B. Harnessing the Periplasm of Bacterial Cells To Develop Biocatalysts for the Biosynthesis of Highly Pure Chemicals. *Appl. Environ. Microbiol.* **2017**, *84* (1). https://doi.org/10.1128/AEM.01693-17.

[29] Kimata, S.; Mochizuki, D.; Satoh, J.; Kitano, K.; Kanesaki, Y.; Takeda, K.; Abe, A.; Kawasaki, S.; Niimura, Y. Intracellular Free Flavin and Its Associated Enzymes Participate in Oxygen and Iron Metabolism in *Amphibacillus xylanus* Lacking a Respiratory Chain. *FEBS Open Bio* **2018**, *8* (6), 947–961. https://doi.org/10.1002/2211-5463.12425.

[30] Eirich, L. D.; Vogels, G. D.; Wolfe, R. S. Proposed Structure for Coenzyme F420 from *Methanobacterium*. *Biochemistry* **1978**, *17* (22), 4583–4593. https://doi.org/10.1021/bi00615a002.

[31] Otani, S.; Takatsu, M.; Nakano, M.; Kasai, S.; Miura, R.; Matsui, K. Roseoflavin, a New Antimicrobial Pigment from *Streptomyces*. *J. Antibiot. (Tokyo)* **1974**, *27* (1), 88–89. https://doi.org/10.7164/antibiotics.27.88.

[32] Langer, S.; Hashimoto, M.; Hobl, B.; Mathes, T.; Mack, M. Flavoproteins Are Potential Targets for the Antibiotic Roseoflavin in *Escherichia coli*. *J. Bacteriol.* **2013**, *195* (18), 4037–4045. https://doi.org/10.1128/JB.00646-13.

[33] Greening, C.; Ahmed, F. H.; Mohamed, A. E.; Lee, B. M.; Pandey, G.; Warden, A. C.; Scott, C.; Oakeshott, J. G.; Taylor, M. C.; Jackson, C. J. Physiology, Biochemistry, and Applications of F420- and Fo-Dependent Redox Reactions. *Microbiol. Mol. Biol. Rev. MMBR* **2016**, *80* (2), 451–493. https://doi.org/10.1128/MMBR.00070-15.

[34] Pedrolli, D. B.; Matern, A.; Wang, J.; Ester, M.; Siedler, K.; Breaker, R.; Mack, M. A Highly Specialized Flavin Mononucleotide Riboswitch Responds Differently to Similar Ligands and Confers Roseoflavin Resistance to *Streptomyces davawensis*. *Nucleic Acids Res.* **2012**, *40* (17), 8662–8673. https://doi.org/10.1093/nar/gks616.

[35] Bacher, A.; Eberhardt, S.; Fischer, M.; Kis, K.; Richter, G. Biosynthesis of Vitamin B2 (Riboflavin). *Annu. Rev. Nutr.* **2000**, *20* (1), 153–167. https://doi.org/10.1146/annurev.nutr.20.1.153.

[36] Lienhart, W.-D.; Gudipati, V.; Macheroux, P. The Human Flavoproteome. *Arch. Biochem. Biophys.* **2013**, *535* (2), 150–162. https://doi.org/10.1016/j.abb.2013.02.015.

[37] Fischer, M.; Bacher, A. Biosynthesis of Vitamin B2: Structure and Mechanism of Riboflavin Synthase. *Arch. Biochem. Biophys.* **2008**, *474* (2), 252–265. https://doi.org/10.1016/j.abb.2008.02.008.

[38] Karp, P. D.; Billington, R.; Caspi, R.; Fulcher, C. A.; Latendresse, M.; Kothari, A.; Keseler, I. M.; Krummenacker, M.; Midford, P. E.; Ong, Q.; Ong, W. K.; Paley, S. M.; Subhraveti, P. The BioCyc Collection of Microbial Genomes and Metabolic Pathways. *Brief. Bioinform.* **2019**, *20* (4), 1085–1093. https://doi.org/10.1093/bib/bbx085.

[39] Massey, V. Activation of Molecular Oxygen by Flavins and Flavoproteins. *J. Biol. Chem.* **1994**, *269* (36), 22459–22462.

[40] García-Angulo, V. A. Overlapping Riboflavin Supply Pathways in Bacteria. *Crit. Rev. Microbiol.* **2017**, *43* (2), 196–209. https://doi.org/10.1080/1040841X.2016.1192578.

[41] Mack, M.; van Loon, A. P. G. M.; Hohmann, H.-P. Regulation of Riboflavin Biosynthesis in *Bacillus subtilis* Is Affected by the Activity of the Flavokinase/Flavin Adenine Dinucleotide Synthetase Encoded by *rib*C. *J. Bacteriol.* **1998**, *180* (4), 950–955. https://doi.org/10.1128/JB.180.4.950-955.1998.

[42] Vitreschak, A. G.; Rodionov, D. A.; Mironov, A. A.; Gelfand, M. S. Regulation of Riboflavin Biosynthesis and Transport Genes in Bacteria by Transcriptional and Translational Attenuation. *Nucleic Acids Res.* **2002**, *30* (14), 3141–3151. https://doi.org/10.1093/nar/gkf433.

[43] Fraaije, M. W.; Mattevi, A. Flavoenzymes: Diverse Catalysts with Recurrent Features. *Trends Biochem. Sci.* **2000**, *25* (3), 126–132. https://doi.org/10.1016/S0968-0004(99)01533-9.

[44] Ghisla, S.; Massey, V. Mechanisms of Flavoprotein-Catalyzed Reactions. *Eur. J. Biochem.* **1989**, *181* (1), 1–17. https://doi.org/10.1111/j.1432-1033.1989.tb14688.x.

[45] Lostao, A.; Harrous, M. E.; Daoudi, F.; Romero, A.; Parody-Morreale, A.; Sancho, J. Dissecting the Energetics of the Apoflavodoxin-FMN Complex. *J. Biol. Chem.* **2000**, *275* (13), 9518–9526. https://doi.org/10.1074/jbc.275.13.9518.

[46] Carlson, R.; Langerman, N. The Thermodynamics of Flavin Binding to the Apoflavodoxin from *Azotobacter Vinelandii*. *Arch. Biochem. Biophys.* **1984**, *229* (1), 440–447. https://doi.org/10.1016/0003-9861(84)90173-5.

[47] Becvar, J.; Palmer, G. The Binding of Flavin Derivatives to the Riboflavin-Binding Protein of Egg White. A Kinetic and Thermodynamic Study. *J. Biol. Chem.* **1982**, *257* (10), 5607–5617.

[48] Abrams, V. A. M.; Bush, L.; Kennedy, T.; Schreiber, R. W.; Sherwood, T. A.; White, H. B. Vitamin-Transport Proteins in Alligator Eggs. *Comp. Biochem. Physiol. Part B Comp. Biochem.* **1989**, *93* (2), 291–297. https://doi.org/10.1016/0305-0491(89)90084-9.

[49] Walsh, C. T.; Wencewicz, T. A. Flavoenzymes: Versatile Catalysts in Biosynthetic Pathways. *Nat. Prod. Rep.* **2013**, *30* (1), 175–200. https://doi.org/10.1039/C2NP20069D.

[50] Heuvel, R. H. H. van den; Westphal, A. H.; Heck, A. J. R.; Walsh, M. A.; Rovida, S.; Berkel, W. J. H. van; Mattevi, A. Structural Studies on Flavin Reductase PheA2 Reveal Binding of NAD in an Unusual Folded Conformation and Support Novel Mechanism of Action. *J. Biol. Chem.* **2004**, *279* (13), 12860–12867. https://doi.org/10.1074/jbc.M313765200.

[51] Vinzenz, X.; Grosse, W.; Linne, U.; Meissner, B.; Essen, L.-O. Chemical Engineering of *Mycobacterium tuberculosis* Dodecin Hybrids. *Chem. Commun.* **2011**, *47* (39), 11071–11073. https://doi.org/10.1039/c1cc12929e.

[52] Arockiasamy, A.; Aggarwal, A.; Savva, C. G.; Holzenburg, A.; Sacchettini, J. C. Crystal Structure of Calcium Dodecin (Rv0379), from *Mycobacterium tuberculosis* with a Unique Calcium-Binding Site. *Protein Sci. Publ. Protein Soc.* **2011**, *20* (5), 827–833. https://doi.org/10.1002/pro.607.

[53] Bourdeaux, F.; Ludwig, P.; Paithankar, K.; Sander, B.; Essen, L.-O.; Grininger, M.; Mack, M. Comparative Biochemical and Structural Analysis of the Flavin-Binding Dodecins from *Streptomyces davaonensis* and *Streptomyces coelicolor* Reveals Striking Differences with Regard to Multimerization. *Microbiology* **2019**. https://doi.org/10.1099/mic.0.000835.

[54] UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506–D515. https://doi.org/10.1093/nar/gky1049.

[55] Bourdeaux, F.; Hammer, C. A.; Vogt, S.; Schweighöfer, F.; Nöll, G.; Wachtveitl, J.; Grininger, M. Flavin Storage and Sequestration by *Mycobacterium tuberculosis* Dodecin. *ACS Infect. Dis.* **2018**, *4* (7), 1082–1092. https://doi.org/10.1021/acsinfecdis.7b00237.

[56] Saviola, B.; Bishai, W. The Genus *Mycobacterium*--Medical. In *The Prokaryotes: Volume 3: Archaea. Bacteria: Firmicutes, Actinomycetes*; Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E., Eds.; Springer: New York, NY, 2006; pp 919–933. https://doi.org/10.1007/0-387-30743-5_34.

[57] Hartmans, S.; de Bont, J. A. M.; Stackebrandt, E. The Genus *Mycobacterium*--Nonmedical. In *The Prokaryotes*; Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E., Eds.; Springer New York: New York, NY, 2006; pp 889–918. https://doi.org/10.1007/0-387-30743-5_33.

[58] Magee, J. G.; Ward, A. C. *Mycobacterium*. In *Bergey's Manual of Systematics of Archaea and Bacteria*; Whitman, W. B., Rainey, F., Kämpfer, P., Trujillo, M., Chun, J., DeVos, P., Hedlund, B., Dedysh, S., Eds.; John Wiley & Sons, Ltd: Chichester, UK, 2015; pp 1–84. https://doi.org/10.1002/9781118960608.gbm00029.

[59] Gupta, R. S.; Lo, B.; Son, J. Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. *Front. Microbiol.* **2018**, *9*. https://doi.org/10.3389/fmicb.2018.00067.

[60] Runyon, E. H. Anonymous Mycobacteria in Pulmonary Disease. *Med. Clin. North Am.* **1959**, *43* (1), 273–290. https://doi.org/10.1016/S0025-7125(16)34193-1.

[61] Kim, C.-J.; Kim, N.-H.; Song, K.-H.; Choe, P. G.; Kim, E. S.; Park, S. W.; Kim, H.-B.; Kim, N.-J.; Kim, E.-C.; Park, W. B.; Oh, M. Differentiating Rapid- and Slow-Growing Mycobacteria by Difference in Time to Growth Detection in Liquid Media. *Diagn. Microbiol. Infect. Dis.* **2013**, *75* (1), 73–76. https://doi.org/10.1016/j.diagmicrobio.2012.09.019.

[62] De Groote, M. A.; Huitt, G. Infections Due to Rapidly Growing Mycobacteria. *Clin. Infect. Dis.* **2006**, *42* (12), 1756–1763. https://doi.org/10.1086/504381.

[63] Bachmann, N. L.; Salamzade, R.; Manson, A. L.; Whittington, R.; Sintchenko, V.; Earl, A. M.; Marais, B. J. Key Transitions in the Evolution of Rapid and Slow Growing Mycobacteria Identified by Comparative Genomics. *Front. Microbiol.* **2020**, *10*. https://doi.org/10.3389/fmicb.2019.03019.

[64] Wee, W. Y.; Dutta, A.; Choo, S. W. Comparative Genome Analyses of Mycobacteria Give Better Insights into Their Evolution. *PLOS ONE* **2017**, *12* (3), e0172831. https://doi.org/10.1371/journal.pone.0172831.

[65] Kazda, J.; Pavlik, I.; III, J. O. F.; Hruska, K. *The Ecology of Mycobacteria: Impact on Animal's and Human's Health*; Springer Science & Business Media, 2010.

[66] Falkinham, J. O. Factors Influencing the Chlorine Susceptibility of *Mycobacterium Avium*, *Mycobacterium intracellulare*, and *Mycobacterium scrofulaceum*. *Appl. Environ. Microbiol.* **2003**, *69* (9), 5685–5689. https://doi.org/10.1128/AEM.69.9.5685-5689.2003.

[67] Brennan, P. J.; Nikaido, H. The Envelope of Mycobacteria. *Annu. Rev. Biochem.* **1995**, *64* (1), 29–63. https://doi.org/10.1146/annurev.bi.64.070195.000333.

[68] Bhatt, A.; Fujiwara, N.; Bhatt, K.; Gurcha, S. S.; Kremer, L.; Chen, B.; Chan, J.; Porcelli, S. A.; Kobayashi, K.; Besra, G. S.; Jacobs, W. R. Deletion of KasB in *Mycobacterium tuberculosis* Causes Loss of Acid-Fastness and Subclinical Latent Tuberculosis in Immunocompetent Mice. *Proc. Natl. Acad. Sci.* **2007**, *104* (12), 5157–5162. https://doi.org/10.1073/pnas.0608654104.

[69] World Health Organization. *Global Tuberculosis Report 2019.*; WORLD HEALTH ORGANIZATION: S.l., 2019.

[70] Golden, M. P.; Vikram, H. R. Extrapulmonary Tuberculosis: An Overview. *Am. Fam. Physician* **2005**, *72* (9), 1761–1768.

[71] Tiemersma, E. W.; van der Werf, M. J.; Borgdorff, M. W.; Williams, B. G.; Nagelkerke, N. J. D. Natural History of Tuberculosis: Duration and Fatality of Untreated Pulmonary Tuberculosis in HIV Negative Patients: A Systematic Review. *PLoS ONE* **2011**, *6* (4). https://doi.org/10.1371/journal.pone.0017601.

[72] Esmail, H.; Barry, C. E.; Young, D. B.; Wilkinson, R. J. The Ongoing Challenge of Latent Tuberculosis. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369* (1645), 20130437. https://doi.org/10.1098/rstb.2013.0437.

[73] Huard, R. C.; Fabre, M.; Haas, P. de; Lazzarini, L. C. O.; Soolingen, D. van; Cousins, D.; Ho, J. L. Novel Genetic Polymorphisms That Further Delineate the Phylogeny of the *Mycobacterium tuberculosis* Complex. *J. Bacteriol.* **2006**, *188* (12), 4271–4287. https://doi.org/10.1128/JB.01783-05.

[74] van Ingen, J.; Rahim, Z.; Mulder, A.; Boeree, M. J.; Simeone, R.; Brosch, R.; van Soolingen, D. Characterization of *Mycobacterium Orygis* as *M. tuberculosis* Complex Subspecies. *Emerg. Infect. Dis.* **2012**, *18* (4), 653–655. https://doi.org/10.3201/eid1804.110888.

[75] Alexander, K. A.; Laver, P. N.; Michel, A. L.; Williams, M.; van Helden, P. D.; Warren, R. M.; Gey van Pittius, N. C. Novel *Mycobacterium tuberculosis* Complex Pathogen, *M. Mungi*. *Emerg. Infect. Dis.* **2010**, *16* (8), 1296–1299. https://doi.org/10.3201/eid1608.100314.

[76] Parsons, S. D. C.; Drewe, J. A.; Gey van Pittius, N. C.; Warren, R. M.; van Helden, P. D. Novel Cause of Tuberculosis in Meerkats, South Africa. *Emerg. Infect. Dis.* **2013**, *19* (12), 2004–2007. https://doi.org/10.3201/eid1912.130268.

[77] Brosch, R.; Gordon, S. V.; Marmiesse, M.; Brodin, P.; Buchrieser, C.; Eiglmeier, K.; Garnier, T.; Gutierrez, C.; Hewinson, G.; Kremer, K.; Parsons, L. M.; Pym, A. S.; Samper, S.; van Soolingen, D.; Cole, S. T. A New Evolutionary Scenario for the *Mycobacterium tuberculosis* Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (6), 3684–3689. https://doi.org/10.1073/pnas.052548299.

[78] King, H. C.; Khera-Butler, T.; James, P.; Oakley, B. B.; Erenso, G.; Aseffa, A.; Knight, R.; Wellington, E. M.; Courtenay, O. Environmental Reservoirs of Pathogenic Mycobacteria across the Ethiopian Biogeographical Landscape. *PLOS ONE* **2017**, *12* (3), e0173811. https://doi.org/10.1371/journal.pone.0173811.

[79] Santos, N.; Santos, C.; Valente, T.; Gortázar, C.; Almeida, V.; Correia-Neves, M. Widespread Environmental Contamination with *Mycobacterium tuberculosis* Complex Revealed by a Molecular Detection Protocol. *PLOS ONE* **2015**, *10* (11), e0142079. https://doi.org/10.1371/journal.pone.0142079.

[80] Velayati, A. A.; Farnia, P.; Mozafari, M.; Malekshahian, D.; Farahbod, A. M.; Seif, S.; Rahideh, S.; Mirsaeidi, M. Identification and Genotyping of *Mycobacterium tuberculosis* Isolated From Water and Soil Samples of a Metropolitan City. *Chest* **2015**, *147* (4), 1094–1102. https://doi.org/10.1378/chest.14-0960.

[81] Russell, D. G.; Cardona, P.-J.; Kim, M.-J.; Allain, S.; Altare, F. Foamy Macrophages and the Progression of the Human TB Granuloma. *Nat. Immunol.* **2009**, *10* (9), 943–948. https://doi.org/10.1038/ni.1781.

[82] Gengenbacher, M.; Kaufmann, S. H. E. *Mycobacterium tuberculosis*: Success through Dormancy. *FEMS Microbiol. Rev.* **2012**, *36* (3), 514–532. https://doi.org/10.1111/j.1574-6976.2012.00331.x.

[83] Getahun, H.; Matteelli, A.; Chaisson, R. E.; Raviglione, M. Latent *Mycobacterium tuberculosis* Infection. *N. Engl. J. Med.* **2015**, *372* (22), 2127–2135. https://doi.org/10.1056/NEJMra1405427.

[84] Cardona, P.-J. What We Have Learned and What We Have Missed in Tuberculosis Pathophysiology for a New Vaccine Design: Searching for the "Pink Swan." *Front. Immunol.* **2017**, *8*. https://doi.org/10.3389/fimmu.2017.00556.

[85] Orme, I. M. A New Unifying Theory of the Pathogenesis of Tuberculosis. *Tuberculosis* **2014**, *94* (1), 8–14. https://doi.org/10.1016/j.tube.2013.07.004.

[86] Peddireddy, V.; Doddam, S. N.; Ahmed, N. Mycobacterial Dormancy Systems and Host Responses in Tuberculosis. *Front. Immunol.* **2017**, *8*. https://doi.org/10.3389/fimmu.2017.00084.

[87] Sasindran, S. J.; Torrelles, J. B. *Mycobacterium tuberculosis* Infection and Inflammation: What Is Beneficial for the Host and for the Bacterium? *Front. Microbiol.* **2011**, *2*. https://doi.org/10.3389/fmicb.2011.00002.

[88] Allard, B.; Panariti, A.; Martin, J. G. Alveolar Macrophages in the Resolution of Inflammation, Tissue Repair, and Tolerance to Infection. *Front. Immunol.* **2018**, *9*. https://doi.org/10.3389/fimmu.2018.01777.

[89] Martin, T. R.; Frevert, C. W. Innate Immunity in the Lungs. *Proc. Am. Thorac. Soc.* **2005**, *2* (5), 403–411. https://doi.org/10.1513/pats.200508-090JS.

[90] Wong, K.-W.; Jacobs, W. R. Postprimary Tuberculosis and Macrophage Necrosis: Is There a Big ConNECtion? *mBio* **2016**, *7* (1). https://doi.org/10.1128/mBio.01589-15.

[91] Rohde, K.; Yates, R. M.; Purdy, G. E.; Russell, D. G. *Mycobacterium tuberculosis* and the Environment within the Phagosome. *Immunol. Rev.* **2007**, *219* (1), 37–54. https://doi.org/10.1111/j.1600-065X.2007.00547.x.

[92] Welin, A.; Eklund, D.; Stendahl, O.; Lerm, M. Human Macrophages Infected with a High Burden of ESAT-6-Expressing *M. tuberculosis* Undergo Caspase-1- and Cathepsin B-Independent Necrosis. *PLOS ONE* **2011**, *6* (5), e20302. https://doi.org/10.1371/journal.pone.0020302.

[93] Pieters, J. *Mycobacterium tuberculosis* and the Macrophage: Maintaining a Balance. *Cell Host Microbe* **2008**, *3* (6), 399–407. https://doi.org/10.1016/j.chom.2008.05.006.

[94] Deb, C.; Lee, C.-M.; Dubey, V. S.; Daniel, J.; Abomoelak, B.; Sirakova, T. D.; Pawar, S.; Rogers, L.; Kolattukudy, P. E. A Novel In Vitro Multiple-Stress Dormancy Model for *Mycobacterium tuberculosis* Generates a Lipid-Loaded, Drug-Tolerant, Dormant Pathogen. *PLoS One* **2009**, *4* (6), e6077. https://doi.org/10.1371/journal.pone.0006077.

[95] Kapoor, N.; Pawar, S.; Sirakova, T. D.; Deb, C.; Warren, W. L.; Kolattukudy, P. E. Human Granuloma In Vitro Model, for TB Dormancy and Resuscitation. *PLoS ONE* **2013**, *8* (1). https://doi.org/10.1371/journal.pone.0053657.

[96] Lipworth, S.; Hammond, R. J. H.; Baron, V. O.; Hu, Yanmin.; Coates, A.; Gillespie, S. H. Defining Dormancy in Mycobacterial Disease. *Tuberculosis* **2016**, *99*, 131–142. https://doi.org/10.1016/j.tube.2016.05.006.

[97] Parrish, N. M.; Dick, J. D.; Bishai, W. R. Mechanisms of Latency in *Mycobacterium tuberculosis*. *Trends Microbiol.* **1998**, *6* (3), 107–112. https://doi.org/10.1016/S0966-842X(98)01216-5.

[98] Scanga, C. A.; Mohan, V. P.; Joseph, H.; Yu, K.; Chan, J.; Flynn, J. L. Reactivation of Latent Tuberculosis: Variations on the Cornell Murine Model. *Infect. Immun.* **1999**, *67* (9), 4531–4538.

[99] McCune, R. M.; Tompsett, R. Fate of *Mycobacterium tuberculosis* in Mouse Tissues as Determined by the Microbial Enumeration Technique I. the Persistence of Drug-Susceptible Tubercle Bacilli in the Tissues Despite Prolonged Antimicrobial Therapy. *J. Exp. Med.* **1956**, *104* (5), 737–762. https://doi.org/10.1084/jem.104.5.737.

[100] McCune, R. M.; Tompsett, R.; McDermott, W. The Fate of *Mycobacterium tuberculosis* in Mouse Tissues as Determined by the Microbial Enumeration Technique Ii. the Conversion of Tuberculous Infection to the Latent State by the Administration of Pyrazinamide and a Companion Drug. *J. Exp. Med.* **1956**, *104* (5), 763–802. https://doi.org/10.1084/jem.104.5.763.

[101] Cardona, P.-J. A Dynamic Reinfection Hypothesis of Latent Tuberculosis Infection. *Infection* **2009**, *37* (2), 80–86. https://doi.org/10.1007/s15010-008-8087-y.

[102] Hunter, R. L. Tuberculosis as a Three-Act Play: A New Paradigm for the Pathogenesis of Pulmonary Tuberculosis. *Tuberculosis* **2016**, *97*, 8–17. https://doi.org/10.1016/j.tube.2015.11.010.

[103] Remmerie, A.; Scott, C. L. Macrophages and Lipid Metabolism. *Cell. Immunol.* **2018**, *330*, 27–42. https://doi.org/10.1016/j.cellimm.2018.01.020.

[104] Berney, M.; Berney-Meyer, L. *Mycobacterium tuberculosis* in the Face of Host-Imposed Nutrient Limitation. *Microbiol. Spectr.* **2017**, *5* (3). https://doi.org/10.1128/microbiolspec.TBTB2-0030-2016.

[105] Macheroux, P.; Kappes, B.; Ealick, S. E. Flavogenomics – a Genomic and Structural View of Flavin-Dependent Proteins. *FEBS J.* **2011**, *278* (15), 2625–2634. https://doi.org/10.1111/j.1742-4658.2011.08202.x.

[106] Griffin, J. E.; Gawronski, J. D.; Dejesus, M. A.; Ioerger, T. R.; Akerley, B. J.; Sassetti, C. M. High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism. *PLoS Pathog.* **2011**, *7* (9), e1002251. https://doi.org/10.1371/journal.ppat.1002251.

[107] DeJesus, M. A.; Gerrick, E. R.; Xu, W.; Park, S. W.; Long, J. E.; Boutte, C. C.; Rubin, E. J.; Schnappinger, D.; Ehrt, S.; Fortune, S. M.; Sassetti, C. M.; Ioerger, T. R. Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. *mBio* **2017**, *8* (1). https://doi.org/10.1128/mBio.02133-16.

[108] Blount, K. F.; Breaker, R. R. Riboswitches as Antibacterial Drug Targets. *Nat. Biotechnol.* **2006**, *24* (12), 1558–1564. https://doi.org/10.1038/nbt1268.

[109] Long, Q.; Ji, L.; Wang, H.; Xie, J. Riboflavin Biosynthetic and Regulatory Factors as Potential Novel Anti-Infective Drug Targets. *Chem. Biol. Drug Des.* **2010**, *75* (4), 339–347. https://doi.org/10.1111/j.1747-0285.2010.00946.x.

[110] Lovewell, R. R.; Sassetti, C. M.; VanderVen, B. C. Chewing the Fat: Lipid Metabolism and Homeostasis during *M. tuberculosis* Infection. *Curr. Opin. Microbiol.* **2016**, *29*, 30–36. https://doi.org/10.1016/j.mib.2015.10.002.

[111] Russell, D. G.; VanderVen, B. C.; Lee, W.; Abramovitch, R. B.; Kim, M.; Homolka, S.; Niemann, S.; Rohde, K. H. *Mycobacterium tuberculosis* Wears What It Eats. *Cell Host Microbe* **2010**, *8* (1), 68–76. https://doi.org/10.1016/j.chom.2010.06.002.

[112] Pandey, A. K.; Sassetti, C. M. Mycobacterial Persistence Requires the Utilization of Host Cholesterol. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (11), 4376–4380. https://doi.org/10.1073/pnas.0711159105.

[113] Soto-Ramirez, M. D.; Aguilar-Ayala, D. A.; Garcia-Morales, L.; Rodriguez-Peredo, S. M.; Badillo-Lopez, C.; Rios-Muñiz, D. E.; Meza-Segura, M. A.; Rivera-Morales, G. Y.; Leon-Solis, L.; Cerna-Cortes, J. F.; Rivera-Gutierrez, S.; Helguera-Repetto, A. C.; Gonzalez-y-Merchand, J. A. Cholesterol Plays a Larger Role during *Mycobacterium tuberculosis* in Vitro Dormancy and Reactivation than Previously Suspected. *Tuberculosis* **2017**, *103*, 1–9. https://doi.org/10.1016/j.tube.2016.12.004.

[114] Binstock, J. F.; Schulz, H. [49] Fatty Acid Oxidation Complex from *Escherichia coli*. In *Methods in Enzymology*; Lipids Part C; Academic Press, 1981; Vol. 71, pp 403–411. https://doi.org/10.1016/0076-6879(81)71051-6.

[115] Houten, S. M.; Wanders, R. J. A. A General Introduction to the Biochemistry of Mitochondrial Fatty Acid β-Oxidation. *J. Inherit. Metab. Dis.* **2010**, *33* (5), 469–477. https://doi.org/10.1007/s10545-010-9061-2.

[116] Yang, M.; Lu, R.; Guja, K. E.; Wipperman, M. F.; St. Clair, J. R.; Bonds, A. C.; Garcia-Diaz, M.; Sampson, N. S. Unraveling Cholesterol Catabolism in *Mycobacterium tuberculosis*: ChsE4-ChsE5 A2β2 Acyl-CoA Dehydrogenase Initiates β-Oxidation of 3-Oxo-Cholest-4-En-26-Oyl CoA. *ACS Infect. Dis.* **2015**, *1* (2), 110–125. https://doi.org/10.1021/id500033m.

[117] Abuhammad, A. Cholesterol Metabolism: A Potential Therapeutic Target in Mycobacteria. *Br. J. Pharmacol.* **2017**, *174* (14), 2194–2208. https://doi.org/10.1111/bph.13694.

[118] Wilburn, K. M.; Fieweger, R. A.; VanderVen, B. C. Cholesterol and Fatty Acids Grease the Wheels of *Mycobacterium tuberculosis* Pathogenesis. *Pathog. Dis.* **2018**, *76* (2). https://doi.org/10.1093/femspd/fty021.

[119] Ghisla, S.; Thorpe, C. Acyl-CoA Dehydrogenases. A Mechanistic Overview. *Eur. J. Biochem.* **2004**, *271* (3), 494–508. https://doi.org/10.1046/j.1432-1033.2003.03946.x.

[120] J. Henriques, B.; K. Olsen, R.; Bross, P.; M. Gomes, C. Emerging Roles for Riboflavin in Functional Rescue of Mitochondrial β-Oxidation Flavoenzymes. *Curr. Med. Chem.* **2010**, *17* (32), 3842–3854. https://doi.org/10.2174/092986710793205462.

[121] Riboflavin Deficiency in Rats and Fatty Acid Oxidation. *Nutr. Rev.* **2009**, *38* (2), 90–92. https://doi.org/10.1111/j.1753-4887.1980.tb05850.x.

[122] Thomas, S. T.; Sampson, N. S. *Mycobacterium tuberculosis* Utilizes a Unique Heterotetrameric Structure for Dehydrogenation of the Cholesterol Side Chain. *Biochemistry* **2013**, *52* (17), 2895–2904. https://doi.org/10.1021/bi4002979.

[123] Dresen, C.; Lin, L. Y.-C.; D'Angelo, I.; Tocheva, E. I.; Strynadka, N.; Eltis, L. D. A Flavin-Dependent Monooxygenase from *Mycobacterium tuberculosis* Involved in Cholesterol Catabolism. *J. Biol. Chem.* **2010**, *285* (29), 22264–22275. https://doi.org/10.1074/jbc.M109.099028.

[124] VanderVen, B. C.; Fahey, R. J.; Lee, W.; Liu, Y.; Abramovitch, R. B.; Memmott, C.; Crowe, A. M.; Eltis, L. D.; Perola, E.; Deininger, D. D.; Wang, T.; Locher, C. P.; Russell, D. G. Novel Inhibitors of Cholesterol Degradation in *Mycobacterium tuberculosis* Reveal How the Bacterium's Metabolism Is Constrained by the Intracellular Environment. *PLoS Pathog.* **2015**, *11* (2). https://doi.org/10.1371/journal.ppat.1004679.

[125] Gopinath, K.; Moosa, A.; Mizrahi, V.; Warner, D. F. Vitamin B12 Metabolism in *Mycobacterium tuberculosis*. *Future Microbiol.* **2013**, *8* (11), 1405–1418. https://doi.org/10.2217/fmb.13.113.

[126] Sritharan, M. Iron Homeostasis in *Mycobacterium tuberculosis*: Mechanistic Insights into Siderophore-Mediated Iron Uptake. *J. Bacteriol.* **2016**, *198* (18), 2399–2409. https://doi.org/10.1128/JB.00359-16.

[127] Ryndak, M. B.; Wang, S. S.; Smith, I.; Rodriguez, G. M. The *Mycobacterium tuberculosis* High-Affinity Iron Importer, IrtA, Contains an FAD-Binding Domain. *J. Bacteriol.* **2010**, *192* (3), 861–869. https://doi.org/10.1128/jb.00223-09.

[128] Sepúlveda Cisternas, I.; Salazar, J. C.; García-Angulo, V. A. Overview on the Bacterial Iron-Riboflavin Metabolic Axis. *Front. Microbiol.* **2018**, *9*. https://doi.org/10.3389/fmicb.2018.01478.

[129] Conrado, R. J.; Varner, J. D.; DeLisa, M. P. Engineering the Spatial Organization of Metabolic Enzymes: Mimicking Nature's Synergy. *Curr. Opin. Biotechnol.* **2008**, *19* (5), 492–499. https://doi.org/10.1016/j.copbio.2008.07.006.

[130] Good, M.; Tang, G.; Singleton, J.; Remenyi, A.; Lim, W. A. Scaffold-Assisted Catalysis: A Novel Domain In the Ste5 Scaffold Protein Is Required to Unlock the MAPK Fus3 for Phosphorylation by the MAPKK Ste7. *Cell* **2009**, *136* (6), 1085–1097. https://doi.org/10.1016/j.cell.2009.01.049.

[131] Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D. Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux. *Nat. Biotechnol.* **2009**, *27* (8), 753–759. https://doi.org/10.1038/nbt.1557.

[132] Shaw, A. S.; Filbert, E. L. Scaffold Proteins and Immune-Cell Signalling. *Nat. Rev. Immunol.* **2009**, *9* (1), 47–56. https://doi.org/10.1038/nri2473.

[133] Good, M. C.; Zalatan, J. G.; Lim, W. A. Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science* **2011**, *332* (6030), 680–686. https://doi.org/10.1126/science.1198701.

[134] Lee, H.; DeLoache, W. C.; Dueber, J. E. Spatial Organization of Enzymes for Metabolic Engineering. *Metab. Eng.* **2012**, *14* (3), 242–251. https://doi.org/10.1016/j.ymben.2011.09.003.

[135] Idan, O.; Hess, H. Engineering Enzymatic Cascades on Nanoscale Scaffolds. *Curr. Opin. Biotechnol.* **2013**, *24* (4), 606–611. https://doi.org/10.1016/j.copbio.2013.01.003.

[136] Anderson, G. P.; Shriver-Lake, L. C.; Liu, J. L.; Goldman, E. R. Orthogonal Synthetic Zippers as Protein Scaffolds. *ACS Omega* **2018**, *3* (5), 4810–4815. https://doi.org/10.1021/acsomega.8b00156.

[137] Backstrom, J. R.; Sanders-Bush, E. Generation of Anti-Peptide Antibodies against Serotonin 5-HT2A and 5-HT2C Receptors. *J. Neurosci. Methods* **1997**, *77* (1), 109–117. https://doi.org/10.1016/S0165-0270(97)00102-7.

[138] Bujacz, A. Structures of Bovine, Equine and Leporine Serum Albumin. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (10), 1278–1289. https://doi.org/10.1107/S0907444912027047.

[139] Kim, D.-M.; Umetsu, M.; Takai, K.; Matsuyama, T.; Ishida, N.; Takahashi, H.; Asano, R.; Kumagai, I. Enhancement of Cellulolytic Enzyme Activity by Clustering Cellulose Binding Domains on Nanoscaffolds. *Small* **2011**, *7* (5), 656–664. https://doi.org/10.1002/smll.201002114.

[140] Siu, K.-H.; Chen, R. P.; Sun, Q.; Chen, L.; Tsai, S.-L.; Chen, W. Synthetic Scaffolds for Pathway Enhancement. *Curr. Opin. Biotechnol.* **2015**, *36*, 98–106. https://doi.org/10.1016/j.copbio.2015.08.009.

[141] Brady, D.; Jordaan, J. Advances in Enzyme Immobilisation. *Biotechnol. Lett.* **2009**, *31* (11), 1639. https://doi.org/10.1007/s10529-009-0076-4.

[142] Dwevedi, A. Basics of Enzyme Immobilization. In *Enzyme Immobilization: Advances in Industry, Agriculture, Medicine, and the Environment*; Dwevedi, A., Ed.; Springer International Publishing: Cham, 2016; pp 21–44. https://doi.org/10.1007/978-3-319-41418-8_2.

[143] DiCosimo, R.; McAuliffe, J.; Poulose, A. J.; Bohlmann, G. Industrial Use of Immobilized Enzymes. *Chem. Soc. Rev.* **2013**, *42* (15), 6437–6474. https://doi.org/10.1039/C3CS35506C.

[144] Giessen, T. W.; Silver, P. A. Encapsulation as a Strategy for the Design of Biological Compartmentalization. *J. Mol. Biol.* **2016**, *428* (5, Part B), 916–927. https://doi.org/10.1016/j.jmb.2015.09.009.

[145] Linko, V.; Nummelin, S.; Aarnos, L.; Tapio, K.; Toppari, J. J.; Kostiainen, M. A. DNA-Based Enzyme Reactors and Systems. *Nanomaterials* **2016**, *6* (8). https://doi.org/10.3390/nano6080139.

[146] Boholm, M.; Arvidsson, R. A Definition Framework for the Terms Nanomaterial and Nanoparticle. *NanoEthics* **2016**, *10* (1), 25–40. https://doi.org/10.1007/s11569-015-0249-7.

[147] Lee, B.-R.; Ko, H. K.; Ryu, J. H.; Ahn, K. Y.; Lee, Y.-H.; Oh, S. J.; Na, J. H.; Kim, T. W.; Byun, Y.; Kwon, I. C.; Kim, K.; Lee, J. Engineered Human Ferritin Nanoparticles for Direct Delivery of Tumor Antigens to Lymph Node and Cancer Immunotherapy. *Sci. Rep.* **2016**, *6* (1), 35182. https://doi.org/10.1038/srep35182.

[148] Ueda, G.; Antanasijevic, A.; Fallas, J. A.; Sheffler, W.; Copps, J.; Ellis, D.; Hutchinson, G. B.; Moyer, A.; Yasmeen, A.; Tsybovsky, Y.; Park, Y.-J.; Bick, M. J.; Sankaran, B.; Gillespie, R. A.; Brouwer, P. J.; Zwart, P. H.; Veesler, D.; Kanekiyo, M.; Graham, B. S.; Sanders, R. W.; Moore, J. P.; Klasse, P. J.; Ward, A. B.; King, N. P.; Baker, D. Tailored Design of Protein Nanoparticle Scaffolds for Multivalent Presentation of Viral Glycoprotein Antigens. *eLife* **2020**, *9*, e57659. https://doi.org/10.7554/eLife.57659.

[149] Falvo, E.; Tremante, E.; Arcovito, A.; Papi, M.; Elad, N.; Boffi, A.; Morea, V.; Conti, G.; Toffoli, G.; Fracasso, G.; Giacomini, P.; Ceci, P. Improved Doxorubicin Encapsulation and Pharmacokinetics of Ferritin–Fusion Protein Nanocarriers Bearing Proline, Serine, and Alanine Elements. *Biomacromolecules* **2016**, *17* (2), 514–522. https://doi.org/10.1021/acs.biomac.5b01446.

[150] Bhaskar, S.; Lim, S. Engineering Protein Nanocages as Carriers for Biomedical Applications. *NPG Asia Mater.* **2017**, *9* (4), e371–e371. https://doi.org/10.1038/am.2016.128.

[151] Giessen, T. W.; Silver, P. A. A Catalytic Nanoreactor Based on in Vivo Encapsulation of Multiple Enzymes in an Engineered Protein Nanocompartment. *ChemBioChem* **2016**, *17* (20), 1931–1935. https://doi.org/10.1002/cbic.201600431.

[152] Linko, V.; Eerikäinen, M.; Kostiainen, M. A. A Modular DNA Origami-Based Enzyme Cascade Nanoreactor. *Chem. Commun.* **2015**, *51* (25), 5351–5354. https://doi.org/10.1039/C4CC08472A.

[153] Minton, A. P. A Molecular Model for the Dependence of the Osmotic Pressure of Bovine Serum Albumin upon Concentration and PH. *Biophys. Chem.* **1995**, *57* (1), 65–70. https://doi.org/10.1016/0301-4622(95)00046-Z.

[154] Zhang, D.-L.; Wu, L.-J.; Chen, J.; Liang, Y. Effects of Macromolecular Crowding on the Structural Stability of Human α-Lactalbumin. *Acta Biochim. Biophys. Sin.* **2012**, *44* (8), 703–711. https://doi.org/10.1093/abbs/gms052.

[155] Cheung, M. S.; Klimov, D.; Thirumalai, D. Molecular Crowding Enhances Native State Stability and Refolding Rates of Globular Proteins. *Proc. Natl. Acad. Sci.* **2005**, *102* (13), 4753–4758. https://doi.org/10.1073/pnas.0409630102.

[156] Eggers, D. K.; Valentine, J. S. Molecular Confinement Influences Protein Structure and Enhances Thermal Protein Stability. *Protein Sci.* **2001**, *10* (2), 250–261. https://doi.org/10.1110/ps.36201.

[157] Politou, A.; Temussi, P. A. Revisiting a Dogma: The Effect of Volume Exclusion in Molecular Crowding. *Curr. Opin. Struct. Biol.* **2015**, *30*, 1–6. https://doi.org/10.1016/j.sbi.2014.10.005.

[158] Cheng, K.; Wu, Q.; Zhang, Z.; Pielak, G. J.; Liu, M.; Li, C. Crowding and Confinement Can Oppositely Affect Protein Stability. *ChemPhysChem* **2018**, *19* (24), 3350–3355. https://doi.org/10.1002/cphc.201800857.

[159] Chien, L.-J.; Lee, C.-K. Biosilicification of Dual-Fusion Enzyme Immobilized on Magnetic Nanoparticle. *Biotechnol. Bioeng.* **2008**, *100* (2), 223–230. https://doi.org/10.1002/bit.21750.

[160] Datta, S.; Christena, L. R.; Rajaram, Y. R. S. Enzyme Immobilization: An Overview on Techniques and Support Materials. *3 Biotech* **2013**, *3* (1), 1–9. https://doi.org/10.1007/s13205-012-0071-7.

[161] Han, G. H.; Seong, W.; Fu, Y.; Yoon, P. K.; Kim, S. K.; Yeom, S.-J.; Lee, D.-H.; Lee, S.-G. Leucine Zipper-Mediated Targeting of Multi-Enzyme Cascade Reactions to Inclusion Bodies in *Escherichia coli* for Enhanced Production of 1-Butanol. *Metab. Eng.* **2017**, *40*, 41–49. https://doi.org/10.1016/j.ymben.2016.12.012.

[162] Yang, Z.; Wang, H.; Wang, Y.; Ren, Y.; Wei, D. Manufacturing Multienzymatic Complex Reactors In Vivo by Self-Assembly To Improve the Biosynthesis of Itaconic Acid in *Escherichia coli*. *ACS Synth. Biol.* **2018**, *7* (5), 1244–1250. https://doi.org/10.1021/acssynbio.8b00086.

[163] Sweetlove, L. J.; Fernie, A. R. The Role of Dynamic Enzyme Assemblies and Substrate Channelling in Metabolic Regulation. *Nat. Commun.* **2018**, *9* (1), 1–12. https://doi.org/10.1038/s41467-018-04543-8.

[164] Miles, E. W.; Rhee, S.; Davies, D. R. The Molecular Basis of Substrate Channeling. *J. Biol. Chem.* **1999**, *274* (18), 12193–12196. https://doi.org/10.1074/jbc.274.18.12193.

[165] Castellana, M.; Wilson, M. Z.; Xu, Y.; Joshi, P.; Cristea, I. M.; Rabinowitz, J. D.; Gitai, Z.; Wingreen, N. S. Enzyme Clustering Accelerates Processing of Intermediates through Metabolic Channeling. *Nat. Biotechnol.* **2014**, *32* (10), 1011–1018. https://doi.org/10.1038/nbt.3018.

[166] Dunn, M. F.; Aguilar, V.; Brzovic, P.; Drewe, W. F.; Houben, K. F.; Leja, C. A.; Roy, M. The Tryptophan Synthase Bienzyme Complex Transfers Indole between the .Alpha.- and .Beta.-Sites via a 25-30 .ANG. Long Tunnel. *Biochemistry* **1990**, *29* (37), 8598–8607. https://doi.org/10.1021/bi00489a015.

[167] Stroud, R. M. An Electrostatic Highway. *Nat. Struct. Biol.* **1994**, *1* (3), 131–134. https://doi.org/10.1038/nsb0394-131.

[168] Wang, N.; McCammon, J. A. Substrate Channeling between the Human Dihydrofolate Reductase and Thymidylate Synthase. *Protein Sci.* **2016**, *25* (1), 79–86. https://doi.org/10.1002/pro.2720.

[169] Knighton, D. R.; Kan, C.-C.; Howland, E.; Janson, C. A.; Hostomska, Z.; Welsh, K. M.; Matthews, D. A. Structure of and Kinetic Channelling in Bifunctional Dihydrofolate Reductase–Thymidylate Synthase. *Nat. Struct. Biol.* **1994**, *1* (3), 186–194. https://doi.org/10.1038/nsb0394-186.

[170] Wu, F.; Minteer, S. Krebs Cycle Metabolon: Structural Evidence of Substrate Channeling Revealed by Cross-Linking and Mass Spectrometry. *Angew. Chem. Int. Ed.* **2015**, *54* (6), 1851–1854. https://doi.org/10.1002/anie.201409336.

[171] Bulutoglu, B.; Garcia, K. E.; Wu, F.; Minteer, S. D.; Banta, S. Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes. *ACS Chem. Biol.* **2016**, *11* (10), 2847–2853. https://doi.org/10.1021/acschembio.6b00523.

[172] Huang, Y. M.; Huber, G. A.; Wang, N.; Minteer, S. D.; McCammon, J. A. Brownian Dynamic Study of an Enzyme Metabolon in the TCA Cycle: Substrate Kinetics and Channeling. *Protein Sci.* **2018**, *27* (2), 463–471. https://doi.org/10.1002/pro.3338.

[173] Poshyvailo, L.; Lieres, E. von; Kondrat, S. Does Metabolite Channeling Accelerate Enzyme-Catalyzed Cascade Reactions? *PLOS ONE* **2017**, *12* (2), e0172673. https://doi.org/10.1371/journal.pone.0172673.

[174] Idan, O.; Hess, H. Origins of Activity Enhancement in Enzyme Cascades on Scaffolds. *ACS Nano* **2013**, *7* (10), 8658–8665. https://doi.org/10.1021/nn402823k.

[175] Hwang, E. T.; Lee, S. Multienzymatic Cascade Reactions via Enzyme Complex by Immobilization. *ACS Catal.* **2019**, *9* (5), 4402–4425. https://doi.org/10.1021/acscatal.8b04921.

[176] Chan, C. Y.; Zhao, H.; Pugh, R. J.; Pedley, A. M.; French, J.; Jones, S. A.; Zhuang, X.; Jinnah, H.; Huang, T. J.; Benkovic, S. J. Purinosome Formation as a Function of the Cell Cycle. *Proc. Natl. Acad. Sci.* **2015**, *112* (5), 1368–1373. https://doi.org/10.1073/pnas.1423009112.

[177] Pedley, A. M.; Benkovic, S. J. A New View into the Regulation of Purine Metabolism: The Purinosome. *Trends Biochem. Sci.* **2017**, *42* (2), 141–154. https://doi.org/10.1016/j.tibs.2016.09.009.

[178] Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. Principles of Membrane Transport. *Mol. Biol. Cell 4th Ed.* **2002**.

[179] Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. Carrier Proteins and Active Membrane Transport. *Mol. Biol. Cell 4th Ed.* **2002**.

[180] Byers, D. M.; Gong, H. Acyl Carrier Protein: Structure–Function Relationships in a Conserved Multifunctional Protein Family. *Biochem. Cell Biol.* **2007**, *85* (6), 649–662. https://doi.org/10.1139/O07-109.

[181] Chen, A.; Re, R. N.; Burkart, M. D. Type II Fatty Acid and Polyketide Synthases: Deciphering Protein-Protein and Protein-Substrate Interactions. *Nat. Prod. Rep.* **2018**, *35* (10), 1029–1045. https://doi.org/10.1039/c8np00040a.

[182] Huber, M. C.; Schreiber, A.; von Olshausen, P.; Varga, B. R.; Kretz, O.; Joch, B.; Barnert, S.; Schubert, R.; Eimer, S.; Kele, P.; Schiller, S. M. Designer Amphiphilic Proteins as Building Blocks for the Intracellular Formation of Organelle-like Compartments. *Nat. Mater.* **2014**, *14* (1), 125–132. https://doi.org/10.1038/nmat4118.

[183] Wang, D.; Chai, Y.; Yuan, Y.; Yuan, R. Lattice-Like DNA Tetrahedron Nanostructure as Scaffold to Locate GOx and HRP Enzymes for Highly Efficient Enzyme Cascade Reaction. *ACS Appl. Mater. Interfaces* **2020**, *12* (2), 2871–2877. https://doi.org/10.1021/acsami.9b18702.

[184] Fu, J.; Liu, M.; Liu, Y.; Woodbury, N. W.; Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **2012**, *134* (12), 5516–5519. https://doi.org/10.1021/ja300897h.

[185] Grossi, G.; Jepsen, M. D. E.; Kjems, J.; Andersen, E. S. Control of Enzyme Reactions by a Reconfigurable DNA Nanovault. *Nat. Commun.* **2017**, *8* (1), 1–8. https://doi.org/10.1038/s41467-017-01072-8.

[186] Jia Lili; Minamihata Kosuke; Ichinose Hirofumi; Tsumoto Kouhei; Kamiya Noriho. Polymeric SpyCatcher Scaffold Enables Bioconjugation in a Ratio-Controllable Manner. *Biotechnol. J.* **2017**, *12* (12), 1700195. https://doi.org/10.1002/biot.201700195.

[187] Zhang, Y.; Yong, Y.; Ge, J.; Liu, Z. Lectin Agglutinated Multienzyme Catalyst with Enhanced Substrate Affinity and Activity. *ACS Catal.* **2016**, *6* (6), 3789–3795. https://doi.org/10.1021/acscatal.6b01047.

[188] Fairhead, M.; Veggiani, G.; Lever, M.; Yan, J.; Mesner, D.; Robinson, C. V.; Dushek, O.; van der Merwe, P. A.; Howarth, M. SpyAvidin Hubs Enable Precise and Ultrastable Orthogonal Nanoassembly. *J. Am. Chem. Soc.* **2014**, *136* (35), 12355–12363. https://doi.org/10.1021/ja505584f.

[189] Zhang, G.; Quin, M. B.; Schmidt-Dannert, C. Self-Assembling Protein Scaffold System for Easy in Vitro Coimmobilization of Biocatalytic Cascade Enzymes. *ACS Catal.* **2018**, *8* (6), 5611–5620. https://doi.org/10.1021/acscatal.8b00986.

[190] Brune, K. D.; Buldun, C. M.; Li, Y.; Taylor, I. J.; Brod, F.; Biswas, S.; Howarth, M. Dual Plug-and-Display Synthetic Assembly Using Orthogonal Reactive Proteins for Twin Antigen Immunization. *Bioconjug. Chem.* **2017**, *28* (5), 1544–1551. https://doi.org/10.1021/acs.bioconjchem.7b00174.

[191] Gao, X.; Yang, S.; Zhao, C.; Ren, Y.; Wei, D. Artificial Multienzyme Supramolecular Device: Highly Ordered Self-Assembly of Oligomeric Enzymes In Vitro and In Vivo. *Angew. Chem. Int. Ed.* **2014**, *53* (51), 14027–14030. https://doi.org/10.1002/anie.201405016.

[192] Jutz, G.; van Rijn, P.; Santos Miranda, B.; Böker, A. Ferritin: A Versatile Building Block for Bionanotechnology. *Chem. Rev.* **2015**, *115* (4), 1653–1701. https://doi.org/10.1021/cr400011b.

[193] Horn, A. H. C.; Sticht, H. Synthetic Protein Scaffolds Based on Peptide Motifs and Cognate Adaptor Domains for Improving Metabolic Productivity. *Front. Bioeng. Biotechnol.* **2015**, *3*. https://doi.org/10.3389/fbioe.2015.00191.

[194] Thompson, K. E.; Bashor, C. J.; Lim, W. A.; Keating, A. E. SYNZIP Protein Interaction Toolbox: *in Vitro* and *in Vivo* Specifications of Heterospecific Coiled-Coil Interaction Domains. *ACS Synth Biol* **2012**, *1* (4), 118–129. https://doi.org/10.1021/sb200015u.

[195] Sakarellos-Daitsiotis, M.; Krikorian, D.; Panou-Pomonis, E.; Sakarellos, C. Artificial Carriers: A Strategy for Constructing Antigenic/Immunogenic Conjugates. *Curr. Top. Med. Chem.* **2006**, *6* (16), 1715–1735. https://doi.org/10.2174/156802606778194190.

[196] Tao Liang, M.; Davies, N.; Blanchfield, J.; Toth, I. Particulate Systems as Adjuvants and Carriers for Peptide and Protein Antigens. *Curr. Drug Deliv.* **2006**, *3* (4), 379–388. https://doi.org/10.2174/156720106778559029.

[197] Pichichero, M. E. Protein Carriers of Conjugate Vaccines. *Hum. Vaccines Immunother.* **2013**, *9* (12), 2505–2523. https://doi.org/10.4161/hv.26109.

[198] Li, W.; Joshi, M. D.; Singhania, S.; Ramsey, K. H.; Murthy, A. K. Peptide Vaccine: Progress and Challenges. *Vaccines* **2014**, *2* (3), 515–536. https://doi.org/10.3390/vaccines2030515.

[199] Moreno-Mendieta, S. A.; Guillén, D.; Espitia, C.; Hernández-Pando, R.; Sanchez, S.; Rodríguez-Sanoja, R. A Novel Antigen-Carrier System: The *Mycobacterium tuberculosis* Acr Protein Carried by Raw Starch Microparticles. *Int. J. Pharm.* **2014**, *474* (1), 241–248. https://doi.org/10.1016/j.ijpharm.2014.07.041.

[200] Wherry, E. J.; Masopust, D. Chapter 5 - Adaptive Immunity: Neutralizing, Eliminating, and Remembering for the Next Time. In *Viral Pathogenesis (Third Edition)*; Katze, M. G., Korth, M. J., Law, G. L., Nathanson, N., Eds.; Academic Press: Boston, 2016; pp 57–69. https://doi.org/10.1016/B978-0-12-800964-2.00005-7.

[201] Lee, B.-S.; Huang, J.-S.; Jayathilaka, L. P.; Lee, J.; Gupta, S. Antibody Production with Synthetic Peptides. In *High-Resolution Imaging of Cellular Proteins: Methods and Protocols*; Schwartzbach, S. D., Skalli, O., Schikorski, T., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2016; pp 25–47. https://doi.org/10.1007/978-1-4939-6352-2_2.

[202] Owen, J. A.; Punt, J.; Stranford, S. A.; Jones, P. P.; Kuby, J. *Kuby Immunology*, 7th ed.; W.H. Freeman: New York, 2013.

[203] Kaba, S. A.; Brando, C.; Guo, Q.; Mittelholzer, C.; Raman, S.; Tropel, D.; Aebi, U.; Burkhard, P.; Lanar, D. E. A Nonadjuvanted Polypeptide Nanoparticle Vaccine Confers Long-Lasting Protection against Rodent Malaria. *J. Immunol.* **2009**, *183* (11), 7268–7277. https://doi.org/10.4049/jimmunol.0901957.

[204] Stils, H. F. Adjuvants and Antibody Production: Dispelling the Myths Associated with Freund's Complete and Other Adjuvants. *ILAR J.* **2005**, *46* (3), 280–293. https://doi.org/10.1093/ilar.46.3.280.

[205] Awate, S.; Babiuk, L. A. B.; Mutwiri, G. Mechanisms of Action of Adjuvants. *Front. Immunol.* **2013**, *4*. https://doi.org/10.3389/fimmu.2013.00114.

[206] Freund, J.; Casals, J.; Hosmer, E. P. Sensitization and Antibody Formation after Injection of Tubercle Bacilli and Paraffin Oil. *Proc. Soc. Exp. Biol. Med.* **1937**, *37* (3), 509–513. https://doi.org/10.3181/00379727-37-9625.

[207] Freund, J.; McDermott, K. Sensitization to Horse Serum by Means of Adjuvants. *Proc. Soc. Exp. Biol. Med.* **1942**, *49* (4), 548–553. https://doi.org/10.3181/00379727-49-13625.

[208] Bachmann, M. F.; Jennings, G. T. Vaccine Delivery: A Matter of Size, Geometry, Kinetics and Molecular Patterns. *Nat. Rev. Immunol.* **2010**, *10* (11), 787–796. https://doi.org/10.1038/nri2868.

[209] Briand, J. P.; Muller, S.; Van Regenmortel, M. H. V. Synthetic Peptides as Antigens: Pitfalls of Conjugation Methods. *J. Immunol. Methods* **1985**, *78* (1), 59–69. https://doi.org/10.1016/0022-1759(85)90329-1.

[210] Grant, G. A. Synthetic Peptides for Production of Antibodies That Recognize Intact Proteins. *Curr. Protoc. Mol. Biol.* **2002**, *59* (1), 11.16.1-11.16.19. https://doi.org/10.1002/0471142727.mb1116s59.

[211] Trier, N. H.; Hansen, P. R.; Houen, G. Production and Characterization of Peptide Antibodies. *Methods* **2012**, *56* (2), 136–144. https://doi.org/10.1016/j.ymeth.2011.12.001.

[212] Kitagawa, T.; Aikawa, T. Enzyme Coupled Immunoassay of Insulin Using a Novel Coupling Reagent. *J. Biochem. (Tokyo)* **1976**, *79* (1), 233–236.

[213] Peeters, J. M.; Hazendonk, T. G.; Beuvery, E. C.; Tesser, G. I. Comparison of Four Bifunctional Reagents for Coupling Peptides to Proteins and the Effect of the Three Moieties on the Immunogenicity of the Conjugates. *J. Immunol. Methods* **1989**, *120* (1), 133–143. https://doi.org/10.1016/0022-1759(89)90298-6.

[214] Boeckler, C.; Frisch, B.; Muller, S.; Schuber, F. Immunogenicity of New Heterobifunctional Cross-Linking Reagents Used in the Conjugation of Synthetic Peptides to Liposomes. *J. Immunol. Methods* **1996**, *191* (1), 1–10. https://doi.org/10.1016/0022-1759(95)00284-7.

[215] Harris, J. R.; Markl, J. Keyhole Limpet Hemocyanin (KLH): A Biomedical Review. *Micron* **1999**, *30* (6), 597–623. https://doi.org/10.1016/S0968-4328(99)00036-0.

[216] Swaminathan, A.; Lucas, R. M.; Dear, K.; McMichael, A. J. Keyhole Limpet Haemocyanin – a Model Antigen for Human Immunotoxicological Studies. *Br. J. Clin. Pharmacol.* **2014**, *78* (5), 1135–1142. https://doi.org/10.1111/bcp.12422.

[217] Zeigler, D. F.; Gage, E.; Roque, R.; Clegg, C. H. Epitope Targeting with Self-Assembled Peptide Vaccines. *Npj Vaccines* **2019**, *4* (1), 1–8. https://doi.org/10.1038/s41541-019-0125-5.

[218] Negahdaripour, M.; Golkar, N.; Hajighahramani, N.; Kianpour, S.; Nezafat, N.; Ghasemi, Y. Harnessing Self-Assembled Peptide Nanoparticles in Epitope Vaccine Design. *Biotechnol. Adv.* **2017**, *35* (5), 575–596. https://doi.org/10.1016/j.biotechadv.2017.05.002.

[219] Chackerian, B. Virus-like Particles: Flexible Platforms for Vaccine Development. *Expert Rev. Vaccines* **2007**, *6* (3), 381–390. https://doi.org/10.1586/14760584.6.3.381.

[220] Brune, K. D.; Leneghan, D. B.; Brian, I. J.; Ishizuka, A. S.; Bachmann, M. F.; Draper, S. J.; Biswas, S.; Howarth, M. Plug-and-Display: Decoration of Virus-Like Particles via Isopeptide Bonds for Modular Immunization. *Sci. Rep.* **2016**, *6*. https://doi.org/10.1038/srep19234.

[221] Munesinghe, D. Y.; Clavijo, P.; Calle, M. C.; Nussenzweig, R. S.; Nardin, E. Immunogenicity of Multiple Antigen Peptides (MAP) Containing T and B Cell Epitopes of the Repeat Region of the *P. falciparum* Circumsporozoite Protein. *Eur. J. Immunol.* **1991**, *21* (12), 3015–3020. https://doi.org/10.1002/eji.1830211217.

[222] Bennett, N. R.; Zwick, D. B.; Courtney, A. H.; Kiessling, L. L. Multivalent Antigens for Promoting B and T Cell Activation. *ACS Chem. Biol.* **2015**, *10* (8), 1817–1824. https://doi.org/10.1021/acschembio.5b00239.

[223] Schubert, B.; Kohlbacher, O. Designing String-of-Beads Vaccines with Optimal Spacers. *Genome Med.* **2016**, *8*. https://doi.org/10.1186/s13073-016-0263-6.

[224] Liu, L. M.; MacPherson, G. G. Antigen Processing: Cultured Lymph-Borne Dendritic Cells Can Process and Present Native Protein Antigens. *Immunology* **1995**, *84* (2), 241–246.

[225] Reece, J. C.; Geysen, H. M.; Rodda, S. J. Mapping the Major Human T Helper Epitopes of Tetanus Toxin. The Emerging Picture. *J. Immunol.* **1993**, *151* (11), 6175–6184.

[226] Tanabe, S.; Kobayashi, Y.; Takahata, Y.; Morimatsu, F.; Shibata, R.; Nishimura, T. Some Human B and T Cell Epitopes of Bovine Serum Albumin, the Major Beef Allergen. *Biochem. Biophys. Res. Commun.* **2002**, *293* (5), 1348–1353. https://doi.org/10.1016/S0006-291X(02)00381-9.

[227] Hirata, A. A.; Sussdorf, D. H. Immunogenicity of Insolubilized Bovine Serum Albumin. *J. Immunol.* **1966**, *96* (4), 611–613.

[228] Bachmann, M. F.; Zinkernagel, R. M. Neutralizing Antiviral B Cell Responses. *Annu. Rev. Immunol.* **1997**, *15* (1), 235–270. https://doi.org/10.1146/annurev.immunol.15.1.235.

[229] Tolar, P.; Hanna, J.; Krueger, P. D.; Pierce, S. K. The Constant Region of the Membrane Immunoglobulin Mediates B Cell-Receptor Clustering and Signaling in Response to Membrane Antigens. *Immunity* **2009**, *30* (1), 44–55. https://doi.org/10.1016/j.immuni.2008.11.007.

[230] Liu, W.; Sohn, H. W.; Tolar, P.; Pierce, S. K. It's All About Change: The Antigen-Driven Initiation of B-Cell Receptor Signaling. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (7), a002295. https://doi.org/10.1101/cshperspect.a002295.

[231] Adamczyk, M.; Buko, A.; Chen, Y.-Y.; Fishpaugh, J. R.; Gebler, J. C.; Johnson, D. D. Characterization of Protein-Hapten Conjugates. 1. Matrix-Assisted Laser Desorption Ionization Mass Spectrometry of Immuno BSA-Hapten Conjugates and Comparison with Other Characterization Methods. *Bioconjug. Chem.* **1994**, *5* (6), 631–635. https://doi.org/10.1021/bc00030a019.

[232] Singh, K. V.; Kaur, J.; Varshney, G. C.; Raje, M.; Suri, C. R. Synthesis and Characterization of Hapten−Protein Conjugates for Antibody Production against Small Molecules. *Bioconjug. Chem.* **2004**, *15* (1), 168–173. https://doi.org/10.1021/bc034158v.

[233] Pedersen, M. K.; Sorensen, N. S.; Heegaard, P. M. H.; Beyer, N. H.; Bruun, L. Effect of Different Hapten-Carrier Conjugation Ratios and Molecular Orientations on Antibody Affinity against a Peptide Antigen. *J. Immunol. Methods* **2006**, *311* (1), 198–206. https://doi.org/10.1016/j.jim.2006.02.008.

[234] Li, Q.; Rodriguez, L. G.; Farnsworth, D. F.; Gildersleeve, J. C. Effects of Hapten Density on the Induced Antibody Repertoire. *Chembiochem Eur. J. Chem. Biol.* **2010**, *11* (12), 1686–1691. https://doi.org/10.1002/cbic.201000235.

[235] Ludwig, P.; Sévin, D. C.; Busche, T.; Kalinowski, J.; Bourdeaux, F.; Grininger, M.; Mack, M. Characterization of the Small Flavin-Binding Dodecin in the Roseoflavin Producer *Streptomyces davawensis*. *Microbiology,* **2018**, *164* (6), 908–919. https://doi.org/10.1099/mic.0.000662.

[236] Bourdeaux, F.; Kopp, Y.; Lautenschläger, J.; Gößner, I.; Besir, H.; Vabulas, R. M.; Grininger, M. Dodecin as Carrier Protein for Immunizations and Bioengineering Applications. *Sci. Rep.* **2020**, *10* (1), 13297. https://doi.org/10.1038/s41598-020-69990-0.

[237] Nöll, T.; Wenderhold-Reeb, S.; Bourdeaux, F.; Paululat, T.; Nöll, G. Diffusion-Ordered NMR Spectroscopy of Guest Molecules in DNA Hydrogels and Related Matrices. *ChemistrySelect* **2018**, *3* (37), 10287–10297. https://doi.org/10.1002/slct.201802364.

[238] Agari, Y.; Kuramitsu, S.; Shinkai, A. Identification of Novel Genes Regulated by the Oxidative Stress-Responsive Transcriptional Activator SdrP in *Thermus thermophilus* HB8. *FEMS Microbiol. Lett.* **2010**, *313* (2), 127–134. https://doi.org/10.1111/j.1574-6968.2010.02133.x.

[239] Pedrolli, D. B.; Kühm, C.; Sévin, D. C.; Vockenhuber, M. P.; Sauer, U.; Suess, B.; Mack, M. A Dual Control Mechanism Synchronizes Riboflavin and Sulphur Metabolism in *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* **2015**, *112* (45), 14054–14059. https://doi.org/10.1073/pnas.1515024112.

[240] Hemberger, S.; Pedrolli, D. B.; Stolz, J.; Vogl, C.; Lehmann, M.; Mack, M. RibM from *Streptomyces davawensis* Is a Riboflavin/Roseoflavin Transporter and May Be Useful for the Optimization of Riboflavin Production Strains. *BMC Biotechnol.* **2011**, *11* (1), 119. https://doi.org/10.1186/1472-6750-11-119.

[241] Fuentes Flores, A.; Sepúlveda Cisternas, I.; Vásquez Solis de Ovando, J. I.; Torres, A.; García-Angulo, V. A. Contribution of Riboflavin Supply Pathways to Vibrio Cholerae in Different Environments. *Gut Pathog.* **2017**, *9* (1), 64. https://doi.org/10.1186/s13099-017-0214-9.

[242] McAnulty, M. J.; Wood, T. K. YeeO from *Escherichia coli* Exports Flavins. *Bioengineered* **2014**, *5* (6), 386–392. https://doi.org/10.4161/21655979.2014.969173.

[243] Yang, G.; Bhuvaneswari, T. V.; Joseph, C. M.; King, M. D.; Phillips, D. A. Roles for Riboflavin in the *Sinorhizobium*-Alfalfa Association. *Mol. Plant-Microbe Interactions®* **2002**, *15* (5), 456–462. https://doi.org/10.1094/MPMI.2002.15.5.456.

[244] von Canstein, H.; Ogawa, J.; Shimizu, S.; Lloyd, J. R. Secretion of Flavins by *Shewanella* Species and Their Role in Extracellular Electron Transfer. *Appl. Environ. Microbiol.* **2008**, *74* (3), 615–623. https://doi.org/10.1128/aem.01387-07.

[245] Marsili, E.; Baron, D. B.; Shikhare, I. D.; Coursolle, D.; Gralnick, J. A.; Bond, D. R. *Shewanella* Secretes Flavins That Mediate Extracellular Electron Transfer. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (10), 3968–3973. https://doi.org/10.1073/pnas.0710525105.

[246] Park, J. O.; Rubin, S. A.; Xu, Y.-F.; Amador-Noguez, D.; Fan, J.; Shlomi, T.; Rabinowitz, J. D. Metabolite Concentrations, Fluxes, and Free Energies Imply Efficient Enzyme Usage. *Nat. Chem. Biol.* **2016**, *12* (7), 482–489. https://doi.org/10.1038/nchembio.2077.

[247] Harold, L. K.; Antoney, J.; Ahmed, F. H.; Hards, K.; Carr, P. D.; Rapson, T.; Greening, C.; Jackson, C. J.; Cook, G. M. FAD-Sequestering Proteins Protect Mycobacteria against Hypoxic and Oxidative Stress. *J. Biol. Chem.* **2019**, *294* (8), 2903–2912. https://doi.org/10.1074/jbc.RA118.006237.

[248] Chauviac, F.-X.; Bommer, M.; Yan, J.; Parkin, G.; Daviter, T.; Lowden, P.; Raven, E. L.; Thalassinos, K.; Keep, N. H. Crystal Structure of Reduced MsAcg, a Putative Nitroreductase from *Mycobacterium Smegmatis* and a Close Homologue of *Mycobacterium tuberculosis* Acg. *J. Biol. Chem.* **2012**, *287* (53), 44372–44383. https://doi.org/10.1074/jbc.M112.406264.

[249] Hu, Y.; Coates, A. R. M. *Mycobacterium tuberculosis acg* Gene Is Required for Growth and Virulence In Vivo. *PLOS ONE* **2011**, *6* (6), e20958. https://doi.org/10.1371/journal.pone.0020958.

[250] Schubert, O. T.; Ludwig, C.; Kogadeeva, M.; Zimmermann, M.; Rosenberger, G.; Gengenbacher, M.; Gillet, L. C.; Collins, B. C.; Röst, H. L.; Kaufmann, S. H. E.; Sauer, U.; Aebersold, R. Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* **2015**, *18* (1), 96–108. https://doi.org/10.1016/j.chom.2015.06.001.

[251] Farr, S. B.; Natvig, D. O.; Kogoma, T. Toxicity and Mutagenicity of Plumbagin and the Induction of a Possible New DNA Repair Pathway in *Escherichia coli*. *J. Bacteriol.* **1985**, *164* (3), 1309–1316.

[252] Hayes, J. D.; McLellan, L. I. Glutathione and Glutathione-Dependent Enzymes Represent a Co-Ordinately Regulated Defence against Oxidative Stress. *Free Radic. Res.* **1999**, *31* (4), 273–300. https://doi.org/10.1080/10715769900300851.

[253] Deponte, M. Glutathione Catalysis and the Reaction Mechanisms of Glutathione-Dependent Enzymes. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2013**, *1830* (5), 3217–3266. https://doi.org/10.1016/j.bbagen.2012.09.018.

[254] Barrett, T.; Wilhite, S. E.; Ledoux, P.; Evangelista, C.; Kim, I. F.; Tomashevsky, M.; Marshall, K. A.; Phillippy, K. H.; Sherman, P. M.; Holko, M.; Yefanov, A.; Lee, H.; Zhang, N.; Robertson, C. L.; Serova, N.; Davis, S.; Soboleva, A. NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Res.* **2013**, *41* (D1), D991–D995. https://doi.org/10.1093/nar/gks1193.

[255] Hutchins, A. M.; Holden, J. F.; Adams, M. W. W. Phosphoenolpyruvate Synthetase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **2001**, *183* (2), 709–715. https://doi.org/10.1128/JB.183.2.709-715.2001.

[256] Zhang, C.; Anderson, A. J. The Gluconeogenic Pathway in a Soil *Mycobacterium* Isolate with Bioremediation Ability. *Curr. Microbiol.* **2013**, *66* (2), 122–131. https://doi.org/10.1007/s00284-012-0248-7.

[257] Shimizu, K. Regulation Systems of Bacteria Such as *Escherichia coli* in Response to Nutrient Limitation and Environmental Stresses. *Metabolites* **2014**, *4* (1), 1–35. https://doi.org/10.3390/metabo4010001.

[258] Ahn, S.; Jung, J.; Jang, I.-A.; Madsen, E. L.; Park, W. Role of Glyoxylate Shunt in Oxidative Stress Response. *J. Biol. Chem.* **2016**, *291* (22), 11928–11938. https://doi.org/10.1074/jbc.M115.708149.

[259] Lorenz, M. C.; Fink, G. R. Life and Death in a Macrophage: Role of the Glyoxylate Cycle in Virulence. *Eukaryot. Cell* **2002**, *1* (5), 657–662. https://doi.org/10.1128/EC.1.5.657-662.2002.

[260] Wong, K. S.; Houry, W. A. Novel Structural and Functional Insights into the MoxR Family of AAA+ ATPases. *J. Struct. Biol.* **2012**, *179* (2), 211–221. https://doi.org/10.1016/j.jsb.2012.03.010.

[261] Dannenberg, A. M. *Pathogenesis of Human Pulmonary tuberculosis: Insights from the Rabbit Model*; ASM Press, 2006.

[262] Baker, J. J.; Johnson, B. K.; Abramovitch, R. B. Slow Growth of *Mycobacterium tuberculosis* at Acidic pH Is Regulated by *phoPR* and Host-Associated Carbon Sources. *Mol. Microbiol.* **2014**, *94* (1), 56–69. https://doi.org/10.1111/mmi.12688.

[263] Vandal, O. H.; Pierini, L. M.; Schnappinger, D.; Nathan, C. F.; Ehrt, S. A Membrane Protein Preserves Intrabacterial pH in Intraphagosomal *Mycobacterium tuberculosis*. *Nat. Med.* **2008**, *14* (8), 849–854. https://doi.org/10.1038/nm.1795.

[264] Vandal, O. H.; Nathan, C. F.; Ehrt, S. Acid Resistance in *Mycobacterium tuberculosis*. *J. Bacteriol.* **2009**, *191* (15), 4714–4721. https://doi.org/10.1128/jb.00305-09.

[265] Velayati, A. A.; Abeel, T.; Shea, T.; Konstantinovich Zhavnerko, G.; Birren, B.; Cassell, G. H.; Earl, A. M.; Hoffner, S.; Farnia, P. Populations of Latent *Mycobacterium tuberculosis* Lack a Cell Wall: Isolation, Visualization, and Whole-Genome Characterization. *Int. J. Mycobacteriology* **2016**, *5* (1), 66–73. https://doi.org/10.1016/j.ijmyco.2015.12.001.

[266] Rohde, K. H.; Abramovitch, R. B.; Russell, D. G. *Mycobacterium tuberculosis* Invasion of Macrophages: Linking Bacterial Gene Expression to Environmental Cues. *Cell Host Microbe* **2007**, *2* (5), 352–364. https://doi.org/10.1016/j.chom.2007.09.006.

[267] Magill, N. G.; Cowan, A. E.; Koppel, D. E.; Setlow, P. The Internal pH of the Forespore Compartment of *Bacillus Megaterium* Decreases by about 1 pH Unit during Sporulation. *J. Bacteriol.* **1994**, *176* (8), 2252–2258.

[268] Wilks, J. C.; Slonczewski, J. L. pH of the Cytoplasm and Periplasm of *Escherichia coli*: Rapid Measurement by Green Fluorescent Protein Fluorimetry. *J. Bacteriol.* **2007**, *189* (15), 5601–5607. https://doi.org/10.1128/JB.00615-07.

[269] Quirós, L. M.; Salas, J. Intracellular Water Volume and Internal pH of *Streptomyces Antibioticus* Spores. *FEMS Microbiol. Lett.* **1996**, *141* (2–3), 245–249. https://doi.org/10.1111/j.1574-6968.1996.tb08392.x.

[270] Chae, W.-B.; Kim, Y.-B.; Choi, S.-W.; Lee, H.-B.; Kim, E.-K. Enhancing the Sporulation of *Streptomyces Kasugaensis* by Culture Optimization. *Korean J. Chem. Eng.* **2009**, *26* (2), 438–443. https://doi.org/10.1007/s11814-009-0074-1.

[271] Seckbach, J. *Journey to Diverse Microbial Worlds: Adaptation to Exotic Environments*; Springer Science & Business Media, 2012.

[272] Ericsson, U. B.; Hallberg, B. M.; Detitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-Based High-Throughput Stability Optimization of Proteins for Structural Studies. *Anal. Biochem.* **2006**, *357* (2), 289–298. https://doi.org/10.1016/j.ab.2006.07.027.

[273] Huynh, K.; Partch, C. L. Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay. *Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al* **2015**, *79*, 28.9.1-28.9.14. https://doi.org/10.1002/0471140864.ps2809s79.

[274] Forneris, F.; Orru, R.; Bonivento, D.; Chiarelli, L. R.; Mattevi, A. ThermoFAD, a Thermofluor®-Adapted Flavin Ad Hoc Detection System for Protein Folding and Ligand Binding. *FEBS J.* **2009**, *276* (10), 2833–2840. https://doi.org/10.1111/j.1742-4658.2009.07006.x.

[275] Zhang, X.-D.; Wu, D.; Shen, X.; Liu, P.-X.; Yang, N.; Zhao, B.; Zhang, H.; Sun, Y.-M.; Zhang, L.-A.; Fan, F.-Y. Size-Dependent in Vivo Toxicity of PEG-Coated Gold Nanoparticles. *Int. J. Nanomedicine* **2011**, *6*, 2071–2081. https://doi.org/10.2147/IJN.S21657.

[276] Truong, N. P.; Whittaker, M. R.; Mak, C. W.; Davis, T. P. The Importance of Nanoparticle Shape in Cancer Drug Delivery. *Expert Opin. Drug Deliv.* **2015**, *12* (1), 129–142. https://doi.org/10.1517/17425247.2014.950564.

[277] Nöll, T.; Schönherr, H.; Wesner, D.; Schopferer, M.; Paululat, T.; Nöll, G. Construction of Three-Dimensional DNA Hydrogels from Linear Building Blocks. *Angew. Chem.* **2014**, *126* (32), 8468–8472. https://doi.org/10.1002/ange.201402497.

[278] Fleming, P. J.; Fleming, K. G. HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamic Properties. *Biophys. J.* **2018**, *114* (4), 856–869. https://doi.org/10.1016/j.bpj.2018.01.002.

[279] Zhang, X.; Tran, S.; Gray-Weale, A. Hydrodynamic Drag on Diffusing Nanoparticles for Size Determination. *J. Phys. Chem. C* **2016**, *120* (38), 21888–21896. https://doi.org/10.1021/acs.jpcc.6b03981.

[280] Yu, Y.; Heidel, B.; Parapugna, T. L.; Wenderhold-Reeb, S.; Song, B.; Schönherr, H.; Grininger, M.; Nöll, G. The Flavoprotein Dodecin as a Redox Probe for Electron Transfer through DNA. *Angew. Chem. Int. Ed.* **2013**, *52* (18), 4950–4953. https://doi.org/10.1002/anie.201208987.

[281] Gutiérrez Sánchez, C.; Su, Q.; Schönherr, H.; Grininger, M.; Nöll, G. Multi-Ligand-Binding Flavoprotein Dodecin as a Key Element for Reversible Surface Modification in Nano-Biotechnology. *ACS Nano* **2015**, *9* (4), 3491–3500. https://doi.org/10.1021/nn506993s.

[282] *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press, 2005. https://doi.org/10.1385/1592598900.

[283] Shao, Y.; Jia, H.; Cao, T.; Liu, D. Supramolecular Hydrogels Based on DNA Self-Assembly. *Acc. Chem. Res.* **2017**, *50* (4), 659–668. https://doi.org/10.1021/acs.accounts.6b00524.

[284] Elcock, A. H.; McCammon, J. A. Electrostatic Contributions to the Stability of Halophilic Proteins Edited by B. Honig. *J. Mol. Biol.* **1998**, *280* (4), 731–748. https://doi.org/10.1006/jmbi.1998.1904.

[285] Ng, W. V.; Kennedy, S. P.; Mahairas, G. G.; Berquist, B.; Pan, M.; Shukla, H. D.; Lasky, S. R.; Baliga, N. S.; Thorsson, V.; Sbrogna, J.; Swartzell, S.; Weir, D.; Hall, J.; Dahl, T. A.; Welti, R.; Goo, Y. A.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M. J.; Hough, D. W.; Maddocks, D. G.; Jablonski, P. E.; Krebs, M. P.; Angevine, C. M.; Dale, H.; Isenbarger, T. A.; Peck, R. F.; Pohlschroder, M.; Spudich, J. L.; Jung, K.-H.; Alam, M.; Freitas, T.; Hou, S.; Daniels, C. J.; Dennis, P. P.; Omer, A. D.; Ebhardt, H.; Lowe, T. M.; Liang, P.; Riley, M.; Hood, L.; DasSarma, S. Genome Sequence of *Halobacterium* Species NRC-1. *Proc. Natl. Acad. Sci.* **2000**, *97* (22), 12176–12181. https://doi.org/10.1073/pnas.190337797.

[286] Fukuchi, S.; Yoshimune, K.; Wakayama, M.; Moriguchi, M.; Nishikawa, K. Unique Amino Acid Composition of Proteins in Halophilic Bacteria. *J. Mol. Biol.* **2003**, *327* (2), 347–357. https://doi.org/10.1016/S0022-2836(03)00150-5.

[287] Tadeo, X.; López-Méndez, B.; Trigueros, T.; Laín, A.; Castaño, D.; Millet, O. Structural Basis for the Aminoacid Composition of Proteins from Halophilic Archea. *PLoS Biol.* **2009**, *7* (12). https://doi.org/10.1371/journal.pbio.1000257.

[288] Deole, R.; Challacombe, J.; Raiford, D. W.; Hoff, W. D. An Extremely Halophilic Proteobacterium Combines a Highly Acidic Proteome with a Low Cytoplasmic Potassium Content. *J. Biol. Chem.* **2013**, *288* (1), 581–588. https://doi.org/10.1074/jbc.M112.420505.

[289] Deole, R.; Hoff, W. D. A Potassium Chloride to Glycine Betaine Osmoprotectant Switch in the Extreme Halophile *Halorhodospira halophila*. *Sci. Rep.* **2020**, *10* (1), 3383. https://doi.org/10.1038/s41598-020-59231-9.

[290] Nichols, R. J.; Cassidy-Amstutz, C.; Chaijarasphong, T.; Savage, D. F. Encapsulins: Molecular Biology of the Shell. *Crit. Rev. Biochem. Mol. Biol.* **2017**, *52* (5), 583–594. https://doi.org/10.1080/10409238.2017.1337709.

[291] Ren, H.; Zhu, S.; Zheng, G. Nanoreactor Design Based on Self-Assembling Protein Nanocages. *Int. J. Mol. Sci.* **2019**, *20* (3), 592. https://doi.org/10.3390/ijms20030592.

[292] Tatur, J.; Hagedoorn, P.-L.; Overeijnder, M. L.; Hagen, W. R. A Highly Thermostable Ferritin from the Hyperthermophilic Archaeal Anaerobe *Pyrococcus furiosus*. *Extremophiles* **2006**, *10* (2), 139–148. https://doi.org/10.1007/s00792-005-0484-x.

[293] Minato, T.; Teramoto, T.; Kakuta, Y.; Ogo, S.; Yoon, K.-S. Biochemical and Structural Characterization of a Thermostable Dps Protein with His-Type Ferroxidase Centers and Outer Metal-Binding Sites. *FEBS Open Bio* **2020**, *10* (7), 1219–1229. https://doi.org/10.1002/2211-5463.12837.

[294] Zhang, X.; Zang, J.; Chen, H.; Zhou, K.; Zhang, T.; Lv, C.; Zhao, G. Thermostability of Protein Nanocages: The Effect of Natural Extra Peptide on the Exterior Surface. *RSC Adv.* **2019**, *9* (43), 24777–24782. https://doi.org/10.1039/C9RA04785A.

[295] Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. *Proc. Natl. Acad. Sci.* **2012**, *109* (12), E690–E697. https://doi.org/10.1073/pnas.1115485109.

[296] Li, L.; Fierer, J. O.; Rapoport, T. A.; Howarth, M. Structural Analysis and Optimization of the Covalent Association between SpyCatcher and a Peptide Tag. *J. Mol. Biol.* **2014**, *426* (2), 309–317. https://doi.org/10.1016/j.jmb.2013.10.021.

[297] Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24* (1), 79. https://doi.org/10.1038/nbt1172.

[298] Jevševar, S.; Gaberc-Porekar, V.; Fonda, I.; Podobnik, B.; Grdadolnik, J.; Menart, V. Production of Nonclassical Inclusion Bodies from Which Correctly Folded Protein Can Be Extracted. *Biotechnol. Prog.* **2005**, *21* (2), 632–639. https://doi.org/10.1021/bp0497839.

[299] Schoene, C.; Fierer, J. O.; Bennett, S. P.; Howarth, M. SpyTag/SpyCatcher Cyclization Confers Resilience to Boiling on a Mesophilic Enzyme. *Angew. Chem. Int. Ed Engl.* **2014**, *53* (24), 6101–6104. https://doi.org/10.1002/anie.201402519.

[300] Schoene, C.; Bennett, S. P.; Howarth, M. SpyRing Interrogation: Analyzing How Enzyme Resilience Can Be Achieved with Phytase and Distinct Cyclization Chemistries. *Sci. Rep.* **2016**, *6*, 21151. https://doi.org/10.1038/srep21151.

[301] Inaba, K. Disulfide Bond Formation System in *Escherichia coli*. *J. Biochem. (Tokyo)* **2009**, *146* (5), 591–597. https://doi.org/10.1093/jb/mvp102.

[302] Schlapschy, M.; Binder, U.; Börger, C.; Theobald, I.; Wachinger, K.; Kisling, S.; Haller, D.; Skerra, A. PASylation: A Biological Alternative to PEGylation for Extending the Plasma Half-Life of Pharmaceutically Active Proteins. *Protein Eng. Des. Sel.* **2013**, *26* (8), 489–501. https://doi.org/10.1093/protein/gzt023.

[303] Chen, X.; Zaro, J. L.; Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* **2013**, *65* (10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039.

[304] Wachtel, R. E.; Tsuji, K. Comparison of Limulus Amebocyte Lysates and Correlation with the United States Pharmacopeial Pyrogen Test. *Appl. Environ. Microbiol.* **1977**, *33* (6), 1265–1269.

[305] Pearson, F. C.; Weary, M. E.; Sargent, H. E.; Novitsky, T. J.; Lin, H.; Lindsay, G.; Berzofsky, R. N.; Lane, A. L.; Wilson, J. D.; Cooper, J. F. Comparison of Several Control Standard Endotoxins to the National Reference Standard Endotoxin -an HIMA Collaborative Study. *Appl. Environ. Microbiol.* **1985**, *50* (1), 91–93.

[306] Veggiani, G.; Nakamura, T.; Brenner, M. D.; Gayet, R. V.; Yan, J.; Robinson, C. V.; Howarth, M. Programmable Polyproteams Built Using Twin Peptide Superglues. *Proc. Natl. Acad. Sci.* **2016**, *113* (5), 1202–1207. https://doi.org/10.1073/pnas.1519214113.

[307] Bajar, B. T.; Wang, E. S.; Lam, A. J.; Kim, B. B.; Jacobs, C. L.; Howe, E. S.; Davidson, M. W.; Lin, M. Z.; Chu, J. Improving Brightness and Photostability of Green and Red Fluorescent Proteins for Live Cell Imaging and FRET Reporting. *Sci. Rep.* **2016**, *6*. https://doi.org/10.1038/srep20889.

[308] Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **1970**, *227* (5259), 680–685. https://doi.org/10.1038/227680a0.

[309] Schägger, H. Tricine–SDS-PAGE. *Nat. Protoc.* **2006**, *1* (1), 16–22. https://doi.org/10.1038/nprot.2006.4.

[310] Shieh, Y.-W.; Minguez, P.; Bork, P.; Auburger, J. J.; Guilbride, D. L.; Kramer, G.; Bukau, B. Operon Structure and Cotranslational Subunit Association Direct Protein Assembly in Bacteria. *Science* **2015**, *350* (6261), 678–680. https://doi.org/10.1126/science.aac8171.

[311] Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

[312] Malonis, R. J.; Lai, J. R.; Vergnolle, O. Peptide-Based Vaccines: Current Progress and Future Challenges. *Chem. Rev.* **2020**, *120* (6), 3210–3229. https://doi.org/10.1021/acs.chemrev.9b00472.

[313] Hume, H. K. C.; Vidigal, J.; Carrondo, M. J. T.; Middelberg, A. P. J.; Roldão, A.; Lua, L. H. L. Synthetic Biology for Bioengineering Virus-like Particle Vaccines. *Biotechnol. Bioeng.* **2019**, *116* (4), 919–935. https://doi.org/10.1002/bit.26890.

[314] Vetter, V.; Denizer, G.; Friedland, L. R.; Krishnan, J.; Shapiro, M. Understanding Modern-Day Vaccines: What You Need to Know. *Ann. Med.* **2017**. https://doi.org/10.1080/07853890.2017.1407035.

[315] Chiu, M. L.; Goulet, D. R.; Teplyakov, A.; Gilliland, G. L. Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies* **2019**, *8* (4). https://doi.org/10.3390/antib8040055.

[316] Kontermann, R. E.; Brinkmann, U. Bispecific Antibodies. *Drug Discov. Today* **2015**, *20* (7), 838–847. https://doi.org/10.1016/j.drudis.2015.02.008.

13 Acknowledgment

First of all, I want to thank my supervisor, Prof. Dr. Martin Grininger, who gave me the opportunity to pursue my PhD in his group. Martin, without you I would have never entered the field of biochemistry, for this you have my deepest gratitude. I'm thankful for your guidance and ideas during the PhD, but also for the freedom and trust to let me develop my own concepts. Your open and honest view on science kept my passion for the project burning.

Many thanks to Prof. Dr. Stefan Knapp for taking the time to be my second referee.

I would like to extend my sincere thanks to my collaboration partners outside of the Grininger group: Prof. Dr. Josef Wachtveitl, Dr. Christopher A. Hammer, Dr. Felix Schweighöfer, Dr. Gilbert Nöll and Dr. Stefan Vogt, who helped characterizing the dodecin of *Mycobacterium tuberculosis*. Prof. Dr. R. Martin Vabulas, Dr Hysein Besir and Dr. Yannick Kopp, without whom the dodecin carrier project would have never been possible. Prof. Dr. Matthias Mack and Dr. Petra Ludwig, who introduced me to the dodecins of *Streptomyces coelicolor* and *Streptomyces davaonensis*.

I am also grateful to all members of the Grininger group for supporting me and making my PhD an unforgettable experience. I want to especially thank Alex, Ilka, Ines Karthik and Mirko: Alex, there are no words to fully express my gratitude for your constant support and the many things you taught me, without you, I would not dare to call myself a biochemist. Ilka, thank you for keeping everything in the laboratory running and your positive attitude. Ines, I want to thank you for supporting me during my first experiments. Karthik, I'm extremely grateful for your constructive criticism and for encouraging me to challenge myself to try and learn new things. Mirko, thank you for always offering help and your insightful suggestions.

Thanks also to all students, I had the pleasure to supervise, especially Emily Hensch, Julia Lautenschläger and Dominik Scheliu, who worked hard on their projects and made teaching enjoyable.

Many thanks to the technical staff of the BMLS, especially Elvira and Nicole, who worked relentlessly to keep all the laboratories running.

Finally, I would like to thank my family and friends for their support and reminding me that there is more than a laboratory and a computer.