THÈSE EN COTUTELLE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ET DE L'UNIVERSITÉ DE FRANCFORT

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

SFB807-INTEGRATED RESEARCH TRAINING GROUP 'TRAM'

SPÉCIALITÉ : BIOCHIMIE

Par Narek YOUSEFIAN

La pompe à efflux multidrogue de type MFS à trois composants EmrAB-TolC d'*Escherichia coli* : du clonage à l'analyse structurale

The three-component multidrug MFS-type efflux pump EmrAB-TolC from *Escherichia coli*: from cloning to structural analysis

Sous la direction de : Olivier LAMBERT

et de : Klaas Martinus POS

Soutenue le 20 Juillet 2020

Membres du jury :

Mme VÉNIEN-BRYAN, Catherine	Professeure	Sorbonne Université	Présidente
Mme BROUTIN, Isabelle	Directrice de Recherche	Université Paris Descartes	Rapporteure
M. VAN VEEN, Hendrik	Professeur agrégé (reader)	University of Cambridge	Rapporteur
M. LAMBERT, Olivier	Directeur de Recherche	Université de Bordeaux	Co-directeur
M. POS, Klaas Martinus	Professeur	Goethe-University Frankfurt	Co-directeur
Mme GIRAUD, Marie-France	Chargée de Recherche	Université de Bordeaux	Invitée

Titre

La pompe à efflux multidrogue de type MFS à trois composants EmrAB-TolC d'*Escherichia coli* : du clonage à l'analyse structurale.

Résumé long

Depuis la découverte de la pénicilline à la fin de années 1920, les antibiotiques se sont révélés très efficaces contre les bactéries pathogènes. Nous sommes malheureusement confrontés aujourd'hui à la menace croissante de souches multiresistantes. En effet, la mauvaise utilisation de ces médicaments a conduit à une évolution accélérée d'un phénomène naturel de résistance aux antibiotiques chez les bactéries. Ainsi, les possibilités de traiter efficacement les maladies infectieuses deviennent de plus en plus limitées, notamment en l'absence de développement de nouveaux antibiotiques. Des études détaillées de ces organismes sont donc nécessaires pour mieux comprendre à l'échelle moléculaire les mécanismes de résistance afin de développer de nouvelles molécules thérapeutiques.

Les bactéries sont en mesure de résister aux effets des antibiotiques grâce à cinq mécanismes principaux : (i) l'altération du site cible, (ii) la modification de la voie métabolique de la cible, (iii) la réduction de l'accumulation de médicament par une absorption réduite et/ou par une augmentation de l'efflux actif, (iv) la modification ou l'inactivation du médicament, et (v) la surproduction de la cible.

Dans ce travail de thèse, le mécanisme de résistance par efflux actif sera étudié. En effet, les pompes à efflux contribuent considérablement à la résistance aux antibiotiques grâce à leur capacité de reconnaître et de transporter des antibiotiques de toutes les classes. En utilisant différents critères tels que le mode de transport, le couplage énergétique, la spécificité de substrat et la phylogénie, ces pompes ont été classées en sept grandes classes : (i) ATP-Binding Cassette (ABC) superfamily, (ii) Major Facilitator Superfamily (MFS), (iii) Drug/Metabolite Transporter (DMT) superfamily incluant la famille Small Multi-drug Resistance (SMR), (iv) Multi-drug/Oligosaccharidyl-lipid/Polysaccharide (MOP) superfamily incluant la famille Multi-drug And Toxic compound Extrusion (MATE), (v) Resistance-Nodulation-cell Division (RND) superfamily, (vi) Antimetabolite transporters (AbgT) family, et (vii) Proteobacterial Antimicrobial Compound Efflux transporters (PACE) family. La MFS est une

superfamille très ancienne, grande et diversifiée formée actuellement par 103 familles qui transportent divers composés. Chez les bactéries à Gram négatif, les membres de cette superfamille contribuent considérablement à la résistance aux antibiotiques grâce à une augmentation de l'efflux actif. Ainsi le présent travail de thèse se concentre sur les systèmes d'efflux tripartite de type EmrAB (*E. coli* multidrug resistance) – TolC (Tolerance to Colicin E1) composés d'un transporteur de la membrane interne (EmrB), d'un canal de la membrane externe (TolC) et d'un adaptateur periplasmique (EmrA). Le transporteur de la membrane interne (l'antiporteur utilisant le gradient de protons) reconnaît les composés hydrophobes (CCCP, l'acide nalidixique, la thiolactomycine etc.) dans le cytoplasme et les transporte à travers la membrane interne et à travers EmrA-TolC vers l'extérieur de la cellule.

Contrairement aux systèmes d'efflux de la superfamille des RND (tels que AcrAB-TolC) qui ont été largement étudiés, peu d'informations structurales sont disponibles sur le système EmrAB-TolC. Il y a en effet peu d'information pour EmrA aucune pour EmrB et aucune donnée structurale sur le complexe entier. Ainsi, de nombreuses interrogations persistent concernant entre autre la stœchimétrie de l'assemblage tripartite et son mécanisme de fonctionnement. Par conséquent, des études structurales détaillées sont nécessaires pour commencer à mieux comprendre son rôle global dans la résistance aux antibiotiques médiée par l'efflux actif.

L'objectif de mon travail de thèse a été de produire et catactériser sur un plan structural au moins un complexe d'efflux EmrAB-TolC extrait directement de bactéries surexprimant les trois protéines. Dans un premier temps, l'amplification des gènes *emrA*, *emrB* et *tolC* pour 15 systèmes homologues a été réalisé suivie pour certains inserts du clonage (Fragment eXchange, FX-cloning), ce qui a permis de construire une première librairie pour le système EmrAB-TolC. Seuls les systèmes provenant d'*E. coli* et de *V. cholerae* ont pu être correctement clonés dans les vecteurs d'expression contenant des marqueurs fluorescents pour le suivi de leur expression et pour l'étude de la formation des complexes. Je me suis appuyé sur une méthode de criblage à haut débit développé par Alina Ornik-Cha dans le laboratoire du Prof. Dr. Klaas Martinus Pos. Dans un premier temps, les niveaux d'expression des protéines (EmrB-mRFP1 et EmrA-sfGFP) furent étudiés pour plusieurs souches d'expression d'*E. coli* en mesurant les niveaux de fluorescences rouge et verte et par Western blot (anti-His, Myc, et Strep pour EmrB, EmrA et TolC). La souche d'*E. coli* C41(DE3) était le mieux adapté pour la co-expression d'EmrAB-TolC. Afin d'obtenir des indications sur la formation du complexe, la

méthodologie appelée FSEC (Fluorescence detection Size Exclusion Chromatography) a permis de constater que le complexe EmrAB-TolC d'*E. coli* était produit en plus grande quantité que celui de *V. cholerae*. En parallèle, deux autres stratégies de clonage ont été initiées, l'une créant des chimères EmrA-EmrB avec un peptide polyGS de longeur variable et l'autre en clonant individuellement chaque protéine au cas où la première stratégie ne fournirait pas de résultats satisfaisants.

La co-expression d'EmrAB-TolC d'*E. coli* dans la souche d'*E. coli* C41(DE3) a été optimisée en analysant l'influence de paramètres de culture. Pour la co-purification, le protocole finale consiste à effectuer une lyse douce en utilisant le lysozyme, une solubilisation en présence de détergent (DDM) puis une étape de chromatography d'affinité Ni^{2+}-NTA suivie d'une chromatographie d'exclusion stérique. Ensuite le complexe EmrAB-TolC a été stabilisé après échange du détergent par l'Amphipol A8-35. Le complexe a été soumis à une analyse en microscopie éléctronique.

L'observation par microscopie éléctronique en coloration négative révèle des objets allongés d'une longueur de 33 nm. Une image moyenne d'EmrAB-TolC montre certaines similitudes avec celle du complexe AcrAB-TolC observé dans des conditions similaires. En effet on retrouve les densités caractéristiques de la protéine TolC à une extrémité du complexe. Cependant à l'autre extrémité les densités d'EmrAB apparaissent plus fines que celles d'AcrAB. Les densités visibles au-dessus de l'amphipol correspondent seulement à EmrA. En effet une analyse en microscopie électronique d'EmrB purifié seule et stabilisée en amphipol ne montre pas de densités visibles à l'extérieur du disque d'amphipol. L'architecture générale du complexe MFS EmrAB-TolC est comparable à celle décrite pour les complexes de type RND et ABC avec une disposition en « tip-to-tip » dans laquelle le transporteur n'est pas en contact direct avec le canal de la membrane externe.

Sachant que les images de microscopie électronique ont montré qu'en plus des particules qui correspondent au complexe EmrAB-TolC, d'autres petites particules peuvent également être observées (EmrB seul, TolC seul), il serait nécessaire d'optimiser davantage la purification du complexe, par exemple avec d'autres étapes de chromatographie. Une autre approche serait la reconstitution des composants individuels *in vitro* afin d'augmenter la proportion des complexes tripartites. D'autres substitutions membranaires telles que les nanodisques, SaliPro ou SMALPs pourraient également être utilisées. Le présent travail montre pour la première fois les contours d'un système d'efflux tripartite de type MFS. Il y a des premières indications sur la stœchiométrie des composants individuels du système

et l'image moyenne montre que le canal EmrA-TolC traverse tout le périplasme ainsi que la membrane externe.

Mots clés

Protéines membranaires, antibioresistance, système d'efflux, système tripartite de type MFS, microscopie électronique

Titel

Die dreikomponentige Multidrug-Effluxpumpe vom MFS-Typ EmrAB-TolC aus *Escherichia coli*: von der Klonierung bis zur Strukturanalyse.

Lange Zusammenfassung

Seit der Entdeckung von Penicillin Ende der 1920er Jahre haben sich Antibiotika als sehr wirksam gegen pathogene Bakterien erwiesen. Leider sind wir heute der wachsenden Bedrohung durch multiresistente Stämme ausgesetzt. Tatsächlich hat der Missbrauch dieser Medikamente zu einer beschleunigten Entwicklung eines natürlichen Phänomens der Antibiotikaresistenz bei Bakterien geführt. Infolgedessen werden die Möglichkeiten zur wirksamen Behandlung von Infektionskrankheiten zunehmend eingeschränkt, insbesondere wenn keine neuen Antibiotika entwickelt werden. Detaillierte Untersuchungen dieser Organismen sind daher erforderlich, um die molekularen Resistenzmechanismen besser zu verstehen und neue therapeutische Moleküle zu entwickeln.

Bakterien können den Wirkungen von Antibiotika dank fünf Hauptmechanismen widerstehen: (i) Veränderung der Zielstelle, (ii) Veränderung des Stoffwechselwegs, (iii) Verringerung der Akkumulation von Arzneimittel durch verringerte Absorption und / oder erhöhten aktiven Ausfluss, (iv) Modifikation oder Inaktivierung des Arzneimittels und (v) Überproduktion des Ziels.

In dieser Arbeit wird der Mechanismus der aktiven Ausfluss untersucht. Tatsächlich tragen Effluxpumpen dank ihrer Fähigkeit, Antibiotika aller Klassen zu erkennen und zu transportieren, erheblich zur Antibiotikaresistenz bei. Unter Verwendung verschiedener Kriterien wie Transportart, Energiekopplung, Substratspezifität und Phylogenie wurden diese Pumpen in sieben Hauptklassen eingeteilt: (i) ATP-Binding Cassette (ABC) Superfamilie, (ii) Major Facilitator Superfamily (MFS), (iii) Drug/Metabolite Transporter (DMT) Superfamilie einschließlich der Small Multi-Drug Resistance (SMR) Familie, (iv) Multi-Drug/Oligosaccharidyllipid/Polysaccharide (MOP) Superfamilie einschließlich der Multi-drug And Toxic compound Extrusion (MATE) Familie, (v) Resistance-Nodulation-cell Division (RND) Superfamilie, (vi) Antimetabolite transporters (AbgT) Familie und (vii) Proteobacterial Antimicrobial Compound Efflux transporters (PACE) Familie. Die MFS ist eine sehr alte, große und vielfältige Superfamilie, die derzeit aus 103 Familien besteht, die verschiedene Moleküle transportieren.

Bei gramnegativen Bakterien tragen Mitglieder dieser Superfamilie durch eine Erhöhung des aktiven Ausflusses erheblich zur Antibiotikaresistenz bei. Daher konzentriert sich diese Arbeit auf dreiteilige MFS-basierte Effluxsysteme vom Typ EmrAB (*E. coli* Multidrug Resistance) - TolC (Tolerance to Colicin E1), die aus einem Transporter der inneren Membran (EmrB), einem Kanal der äußeren Membran (TolC) und ein periplasmatischer Adapter (EmrA) bestehen. Der Transporter der inneren Membran ist ein H^+/Substrat Antiporter und erkennt hydrophobe Moleküle (CCCP, Nalidixinsäure, Thiolactomycin usw.) im Zytoplasma und transportiert diese durch die innere Membran und durch EmrA-TolC zur Außenseite der Zelle.

Im Gegensatz zu Effluxsystemen der RND-Superfamilie (wie AcrAB-TolC), die umfassend untersucht wurden, sind für das EmrAB-TolC System nur wenige Strukturinformationen verfügbar. Es gibt in der Tat wenig Informationen für EmrA, keine für EmrB und keine Strukturdaten für den gesamten Komplex. Daher bleiben viele Fragen offen, unter anderem hinsichtlich der Stöchiometrie der dreiteiligen Zusammenbau und ihres Transportmechanismus. Daher sind detaillierte Strukturstudien erforderlich, um die allgemeine Rolle der Pumpe bei der durch aktiven Efflux vermittelten Antibiotikaresistenz besser zu verstehen.

Das Ziel meiner Doktorarbeit war es, mindestens einen EmrAB-TolC Effluxkomplex herzustellen, der direkt aus Bakterien die die drei Proteine überexprimieren herausgenommen wurde um seine struktur zu charakterisieren. Zunächst wurde eine Amplifikation der *emrA*-, *emrB*- und *tolC*-Gene für 15 homologe Systeme durchgeführt, gefolgt vür einige Inserts mit einer Klonierung (Fragment eXchange, FX-Klonierung), die es ermöglichte, eine erste Bibliothek für das EmrAB-TolC System aufzubauen. Nur die Systeme von *E. coli* und *V. cholerae* konnten korrekt in Expressionsvektoren kloniert werden, die fluoreszierende Marker enthielten, um ihre Expression zu testen und die Bildung von Komplexen zu untersuchen. Ich habe eine Hochdurchsatz-Screenings Methode der von Alina Ornik-Cha im Labor von Prof. Dr. Klaas Martinus Pos entwickelt wurde betnutzt. Zunächst wurden die Expressionsniveaus von Proteinen (EmrB-mRFP1 und EmrA-sfGFP) für mehrere *E. coli* Expressionsstämme untersucht durch Messung der roten und grünen Fluoreszenz und durch Western Blot (Anti-His, Myc und Strep für EmrB, EmrA und TolC). Der Stamm von *E. coli* C41(DE3) war am besten für die Koexpression von EmrAB-TolC geeignet. Um Hinweise auf die Bildung des Komplexes zu erhalten, ergab die als FSEC (Fluorescence Detection Size Exclusion Chromatography) bezeichnete

Methode, dass der EmrAB-TolC Komplex von *E. coli* in größerer Menge als die von *V. cholerae* produziert wurde. Parallel dazu wurden zwei weitere Klonierungsstrategien initiiert, von denen eine für die erzugung EmrA-EmrB-Chimären mit einem PolyGS-Peptid von variabler Länge und die andere durch Klonierung jedes Gens einzeln, falls die erste Strategie keine zufriedenstellenden Ergebnisse lieferte.

Koexpression von EmrAB-TolC von *E. coli* in *E. coli* Stamm C41(DE3) wurde durch Analyse des Einflusses von Kulturparametern optimiert. Zur gemeinsamen Reinigung besteht das endgültige Protokoll darin, eine sanfte Lyse unter Verwendung von Lysozym, eine Solubilisierung in Gegenwart eines Detergens (DDM) und anschließend ein Ni^{2+}-NTA-Affinitätschromatographie gefolgt von einer sterischen Ausschlusschromatographie durchzuführen. Dann wurde der EmrAB-TolC Komplex nach Austausch des Detergens durch Amphipol A8-35 stabilisiert. Der Komplex wurde danach elektronenmikroskopisch analysiert.

Die Negativkontrastierungs-(negative-stain) elektronenmikroskopische Beobachtung zeigt längliche Partikel mit einer Länge von 33 nm. Eine durch Mittlung der Partikel erhaltenes Bild von EmrAB-TolC zeigte einige Ähnlichkeiten mit dem des AcrAB-TolC Komplexes, der unter ähnlichen Bedingungen beobachtet wurde. Tatsächlich befinden sich die charakteristischen Dichten des TolC Proteins an einem Ende des Komplexes. Am anderen Ende erscheinen die Dichten von EmrAB jedoch schmaler als die von AcrAB. Die über Amphipol sichtbaren Dichten entsprechen nur EmrA. In der Tat zeigt eine elektronenmikroskopische Analyse von EmrB, das allein gereinigt und in Amphipol stabilisiert wurde, keine sichtbaren Dichten außerhalb der Amphipolscheibe. Die allgemeine Architektur des MFS Komplexes EmrAB-TolC ist vergleichbar mit der für die RND- und ABC- Typ beschriebenen Komplexe mit einer "tip-to-tip" Zusammenbau, bei der der Transporter der inneren Membran nicht in direktem Kontakt mit dem Kanal des äußeren Membran ist.

Da die Elektronenmikroskopische Bilder gezeigt haben, dass neben Partikel die den EmrAB-TolC Gesamtkomplex entsprechen, auch weiterer, kleiner Partikel zu beobachten (EmrB alleine, TolC alleine), wäre es notwendig, die Reinigung des Gesamtkomplexes weiter zu optimieren, z.B. durch weiterer Chromatographie-Schritten. Ein anderer Ansatz wäre die Rekonstitution der einzelnen Komponenten *in vitro*, um somit der Anteil der dreiteilige Komplexe zu erhöhen. Andere Membran-Substitutionen, wie Nanodiscs, SaliPro oder SMALPs könnten ebenfalls eingesetzt warden. Die

vorliegende Arbeit zeigt zum ersten Mal die Konturen eines MFS-basiertes dreiteiliges System. Es gibt erste Hinweise auf die Stöchiometrie der einzelnen Komponenten innerhalb des Systems und es zeigt, dass der EmrA-TolC Kanal das gesamte Periplasma sowie die äusseren Membran durchquert.

Schlüsselwörter

Membranproteine, Antibiotikaresistenz, Effluxsystem, MFS-basiertes dreiteiliges system, Elektronenmikroskopie

Titre

La pompe à efflux multidrogue de type MFS à trois composants EmrAB-TolC d'*Escherichia coli* : du clonage à l'analyse structurale.

Résumé court

A l'heure actuelle, suite à une mauvaise utilisation des antibiotiques, nous faisons face à un problème majeur de santé publique. En effet la résistance aux antibiotiques de certaines souches bactériennes rend le traitement des infections très complexe.

Dans ce contexte, le présent projet de thèse concerne l'étude d'un complexe d'efflux bactérien capable de transporter des antibiotiques du cytoplasme vers l'extérieur de la cellule. Ce complexe est composé d'un transporteur de la membrane interne appartenant à la Major Facilitator Superfamily (MFS) (EmrB, *E. coli* multidrug resistance), d'un canal de la membrane externe TolC (Tolerance to Colicin E1) et d'un adaptateur périplasmique (EmrA, *E. coli* multidrug resistance).

Contrairement aux systèmes d'efflux de type RND (tels que AcrAB-TolC), peu de choses sont connues sur le système EmrAB-TolC de type MFS. Il est donc important d'étudier l'ensemble du complexe sur le plan structurale et fonctionnel afin d'identifier les différences entre ces deux types de systèmes d'efflux.

L'objectif de mon projet de thèse était d'étudier au moins un complexe EmrAB-TolC d'un point de vue structurale. Ainsi durant mes études, le but était d'isoler le complexe directement des bactéries surexprimant les trois partenaires protéiques. Dans un premier temps, 15 systèmes homologues EmrAB-TolC ont été identifiés et leurs gènes correspondants amplifiés à partir de l'ADN génomique de différentes bactéries à Gram négatif. Parmi les gènes des 15 systèmes, les gènes codant pour les systèmes d'*E. coli* et de *V. cholerae* ont été étudiés plus en détail. Les vecteurs d'expression codaient pour des marqueurs fluorescents pour la mesure des niveaux d'expression de différentes protéines et pour l'étude de la formation des complexes. Dans un premier temps, les différents niveaux d'expression des protéines (EmrB-mRFP1 et EmrA-sfGFP) ont été étudiés pour plusieurs souches d'expression d'*E. coli* en mesurant les niveaux de fluorescence rouge et verte et par Western blot (anti-His, Myc et Strep pour EmrB, EmrA et TolC). La souche d'*E. coli* C41(DE3) était la mieux adaptée pour la co-expression

d'EmrAB-TolC. Dans un deuxième temps, la méthodologie FSEC (Fluorescence detection Size Exclusion Chromatography) a été utilisée pour identifier un complexe adapté à l'étude structurale. Ainsi, cette méthode a permis d'observer que le complexe EmrAB-TolC d'*E. coli* était produit en plus grande quantité que celui de *V. cholerae*.

Le protocole final de co-purification consiste à effectuer une lyse douce des bactéries à l'aide du lysozyme, puis après solubilisation avec le DDM, la purification est débutée par une étape de chromatographie d'affinité Ni^{2+}-NTA suivie d'une étape de chromatographie d'exclusion stérique. Enfin, les fractions contenant les trois partenaires protéiques sont utilisées pour l'échange de détergent par l'amphipol A8-35 avant l'étude structurale par microscopie électronique.

Les images de microscopie électronique en coloration négative montrent des objets allongés d'une longueur de 33 nm en vue de côté. Une image moyenne d'EmrAB-TolC montre des similitudes avec celle du complexe AcrAB-TolC observé dans des conditions similaires. Les similitudes concernent les densités caractéristiques de TolC. Des différences ont été trouvées pour la partie inférieure d'EmrAB qui est plus fine que la partie inférieure d'AcrAB. Les densités visibles au-dessus de l'anneau d'amphipol correspondent à EmrA, qui présente une structure en forme de canal comme observé avec AcrA. Le canal semble cependant s'étendre plus loin vers la ceinture d'amphipol. Comme EmrB n'a pas de domaine périplasmique étendu présent dans le cas des protéines RND, ces densités sont donc uniquement attribuées à EmrA. EmrA, de l'autre côté, contacte TolC de manière similaire à l'interaction d'AcrA/MexA avec leurs canaux de la membrane externe respectifs (TolC/OprM) de façon «tip-to-tip».

Mots clés

Protéines membranaires, antibioresistance, système d'efflux, système tripartite de type MFS, microscopie électronique

Titel

Die dreikomponentige Multidrug-Effluxpumpe vom MFS-Typ EmrAB-TolC aus *Escherichia coli*: von der Klonierung bis zur Strukturanalyse.

Kurze Zusammenfassung

Aufgrund des Missbrauchs von Antibiotika stehen wir derzeit vor einem großen Problem der öffentlichen Gesundheit. Die Antibiotikaresistenz bestimmter Bakterienstämme macht die Behandlung von Infektionen sehr komplex.

In diesem Zusammenhang befasst sich diese Arbeit mit der Untersuchung eines bakteriellen Effluxkomplexes, der Antibiotika vom Zytoplasma zur Außenseite der Zelle transportieren kann. Dieser Komplex besteht aus einem Major Facilitator Superfamily (MFS) Transporter der inneren Membran (EmrB, *E. coli* multidrug resistance), einem Kanal der äußeren Membran TolC (Tolerance to Colicin E1) und einem periplasmatischen Adapter (EmrA, *E. coli* multidrug resistance).

Im Gegensatz zu Effluxsystemen vom RND-Typ (wie AcrAB-TolC) ist über das EmrAB-TolC-System vom MFS-Typ wenig bekannt. Es ist daher wichtig, den gesamten Komplex auf struktureller und funktioneller Sicht zu untersuchen, um die deutlichen Unterschiede zwischen diesen beiden Arten von Effluxsystemen zu analysieren.

Ziel meiner Doktorarbeit war es, mindestens einen EmrAB-TolC-Komplex aus struktureller Sicht zu untersuchen. Ziel meiner Studien war es, den Komplex direkt aus Bakterien, die die drei Proteinpartner überexprimieren, zu isolieren. In einem ersten Schritt wurden 15 homologe EmrAB-TolC-Systeme identifiziert und ihre entsprechenden Gene aus der genomischen DNA verschiedener gramnegativer Bakterien amplifiziert. Unter den Genen der 15 Systeme wurden die Gene, die für die *E. coli* und *V. cholerae* Systeme kodieren, weiter untersucht. Die Expressionsvektoren codierten fluoreszierende Marker zur Untersuchung der Expression verschiedener Proteine und zur Untersuchung der Komplexbildung. In einem ersten Schritt wurden die verschiedenen Niveaus der Proteinexpression (EmrB-mRFP1 und EmrA-sfGFP) für mehrere *E. coli* Expressionsstämme untersucht durch Messen der roten und grünen Fluoreszenzniveaus und durch Western Blot (Anti-His, Myc und Strep für EmrB, EmrA und TolC). Der Stamm von *E. coli* C41(DE3) war am besten für die Koexpression von EmrAB-TolC

geeignet. In einem zweiten Schritt wurde die FSEC-Methode (Fluorescence Detection Size Exclusion Chromatography) verwendet, um einen für Strukturuntersuchungen geeigneten Komplex zu identifizieren. Somit konnte mit dieser Methode festgestellt werden, dass der EmrAB-TolC-Komplex von *E. coli* in größerer Menge als der von *V. cholerae* produziert wurde.

Das endgültige Ko-Reinigungsprotokoll besteht darin, eine sanfte Lyse der Bakterien unter Verwendung von Lysozym durchzuführen. Nach der Solubilisierung mit DDM wird die Reinigung durch einen Ni^{2+}-NTA Affinitätschromatographieschritt gefolgt von einem Größenausschlusschromatographieschritt gestartet. Schließlich werden die Fraktionen, die die drei Proteinpartner enthalten, für den Detergensaustausch durch Amphipol A8-35 vor der Strukturuntersuchung durch Elektronenmikroskopie verwendet.

EM-Aufnahmen mit negativer Kontrastierung zeigten längliche Objekte mit einer Länge von 33 nm in Seitenansicht. Ein durch Mittlung der Partikel erhaltenes Bild von EmrAB-TolC zeigt Ähnlichkeiten mit dem des AcrAB-TolC-Komplexes, der unter ähnlichen Bedingungen beobachtet wurde. Ähnlichkeiten schlossen die charakteristischen Dichten von TolC ein. Während im unteren Teil von EmrAB Unterschiede festgestellt wurden, der dünner ist als der untere Teil von AcrAB. Die über dem Amphipolring sichtbaren Dichten entsprechen EmrA, das wie bei AcrA eine kanalartige Struktur aufweist. Der Kanal scheint sich jedoch weiter in Richtung des Amphipolgürtels zu erstrecken. Da EmrB keine erweiterte periplasmatische Domäne aufweist wie die RND-Proteine, werden diese Dichten daher ausschließlich EmrA zugeordnet. Auf der anderen Seite kontaktiert EmrA TolC, ähnlich der Interaktion von AcrA/MexA mit ihren jeweiligen Außenmembrankanälen (TolC/OprM), von “tip-to-tip”.

Schlüsselwörter

Membranproteine, Antibiotikaresistenz, Effluxsystem, MFS-basiertes dreiteiliges system, Elektronenmikroskopie

Title

The three-component multidrug MFS-type efflux pump EmrAB-TolC from *Escherichia coli*: from cloning to structural analysis.

Short abstract

Currently, due to the misuse of antibiotics, we are facing a major public health problem. The resistance to antibiotics of certain bacterial strains makes the treatment of infections very complex.

In this context, the present thesis project concerns the study of a bacterial efflux complex capable of transporting antibiotics from the cytoplasm to the outside of the cell. This complex is composed of an inner-membrane Major Facilitator Superfamily (MFS) transporter (EmrB, *E. coli* multidrug resistance), a channel of the outer membrane TolC (Tolerance to Colicin E1) and a periplasmic adapter (EmrA, *E. coli* multidrug resistance).

Unlike RND-type efflux systems (such as AcrAB-TolC), little is known about the MFS-type EmrAB-TolC system. It is therefore important to study the entire complex on a structural and functional level, to analyse the marked differences between these two types of transport systems.

The goal of my thesis project was to study at least one EmrAB-TolC complex from a structural point of view. For my studies the aim was to isolate the complex directly from bacteria overexpressing the three protein partners. In a first step, 15 homologous EmrAB-TolC systems were identified and their corresponding genes amplified from genomic DNA of different Gram-negative bacteria. Among the genes of the 15 systems, the genes coding for the *E. coli* and *V. cholerae* systems were further studied. The expression vectors encoded fluorescent markers for the monitoring of the expression levels of different proteins and for studying the formation of complexes. In a first step, the different protein expression levels (EmrB-mRFP1 and EmrA-sfGFP) were studied for several expression strains of *E. coli* by measuring the red and green fluorescence levels and by Western blot (anti-His, Myc, and Strep for EmrB, EmrA, and TolC). The *E. coli* strain C41(DE3) was best suited for co-expression of EmrAB-TolC. In a second step, the FSEC (Fluorescence detection Size Exclusion Chromatography) methodology was used to identify a complex suitable for structural study. Thus this method enabled the

observation that the EmrAB-TolC complex of *E. coli* was produced in higher amount than that of *V. cholerae*.

The final co-purification protocol consists in perfoming a gentle lysis of the bacteria using lysozyme, then after solubilization with DDM, the purification is started by a Ni^{2+}-NTA affinity chromatography step followed by a size exclusion chromatography step. Finally, the fractions containing the three protein partners are used for the detergent-exchange by amphipol A8-35 before the structural study by electron microscopy.

Negative stain EM-micrographs displayed elongated objects with a length of 33 nm in side view. An average image of EmrAB-TolC shows similarities to that of the AcrAB-TolC complex observed under similar conditions. Similarities included the characteristic densities of TolC. Whereas differences were found in the lower part of EmrAB which is thinner than the lower part of AcrAB. The densities visible above the amphipol-ring correspond to EmrA, which displays a channel-like structure as in AcrA. The channel however seems to extend further towards the amphipol belt. Since EmrB does not have an extended periplasmic domain as the RND proteins have, these densities are therefore solely assigned to EmrA. EmrA, on the other side, contacts TolC akin to the interaction of AcrA/MexA to their cognate outer membrane channels (TolC/OprM) in a 'tip-to-tip' fashion.

Keywords

Membrane proteins, antibiotic resistance, efflux system, MFS-type tripartite system, electron microscopy

Unités de recherche

[CBMN-Chimie et Biologie des Membranes et Nano-objets, CNRS UMR 5248, 14 Allée Geoffroy Saint Hilaire, 33600 Pessac, France]

[Institute of Biochemistry, Goethe University Frankfurt am Main, Max-von-Laue Str. 9, Biocenter N200 1.20, D-60438 Frankfurt am Main, Germany]

Acknowledgments

Thank you to Dr. Isabelle Broutin and Dr. Hendrik Van Veen for doing me the great honor of reviewing my thesis manuscript. I would also like to thank Dr. Catherine Vénien-Bryan for accepting to be an examiner and the president of my thesis committee. Finally I would like to thank Dr. Marie-France Giraud for participating in my thesis committee as a guest.

I would like to thank my supervisors Prof. Dr. Klaas Martinus Pos and Dr. Olivier Lambert for giving me the opportunity to work on this exciting project and for their welcome in their respective laboratories.

Many thanks to my Master's professor Dr. Alain Brisson for his guidance for choosing my Ph.D. project and for the good advices given to me.

Thanks to Dr. Laetitia Daury for her hospitality in the French laboratory. Thank you for your incredible support from the beginning till the end of my thesis.

Thanks to Dr. Frank Bernhard and Dr. Misha Kudryashev for the nice advices given to me during the TRAM examinations.

Thanks to Beate Braungart and Katharina Dolata for their help for all the paperwork done in Germany. Thank you for the kind support during the first half of my Ph.D thesis.

Thanks to Patricia Dulor and Celia Gauthier for all the paperwork done for the french laboratory and university.

I would like to thank Dr. Oliver Richter for all the orders made in the german laboratory. Thank you for your nice support.

Many thanks to Selena Đorđević-Marquardt for all the help and kind advices from the beginning till the end of my stay in Germany. I wish you all the best for the end of your thesis.

I would also like to thank Dr. Viveka Nand Malviya and Dr. Heng Keat Tam for all their explanations at the theoretical and practical levels in Germany. Thanks for all the advices that you have given to me.

Thanks to Dr. Marion Decossas for your incredible kindness and all the help you have provided for the electron microscopy part in France. Thanks for all the advices you have given to me. Our conversations have brought to me a lot.

Thanks to Dr. Sylvie Poussard for all the help provided in France for the experiments concerning the purification of EmrB.

I would also like to thank Dr. Marie-Christine Gerbod-Giannone for the help provided in France for the small scale experiments concening the EmrAB fusion.

Thank you to Dr. Eric Geertsma for all the help you have provided for the cloning part in Germany both on the theoretical and practical levels. I will not forget the “DNA-Sudoku lessons”.

Many thanks to Céline Gounou and Mélanie Berbon for all the help provided for the membrane protein purification experiments in France. All the advices about the akta devices you have given to me brought me a lot.

I would also like to thank Sisareuth Tan for the electron microscopy training in France and all the good advices given to me.

I thank Andrea Herrmann, Christian Groß, and Jasmin El-Delik for technical assistance for the membrane protein purification in Germany and sample shipment to France.

Many thanks to Alina Ornik-Cha for all the help you have provided for the screening part in Germany with the use of your high-throughput screening pipeline.

Thanks to my Master's internship student Assouane Bouafia for all the help provided for the cloning and small scale experiments in France.

I would also like to thank all my other colleagues from the Pos, Geertsma and Lambert labs, Julia Wilhelm, Reinke Müller, Jenifer Cuesta Bernal, Jessica Kobylka, Ganna Krasnoselka, Wuen Ee Foong, Nita Chang, Benedikt Kuhn, Katharina Holzhüter, Anja Roden, Melanie Engelin, Marie Glavier, Dimitri Salvador, Laurence Dallet, Esther Boyer, and Ghita Touti for the shared "Beer hours" and for all the good moments spent with them.

Thanks to all the other members of both the French and German institutes for the good moments during Christmas parties, seminars, picnics and during the retreat in Austria.

I would like to thank François Reynaert for always being there for my Family and for me. You have been such an important role model to whom I look up to. Thanks for believing in me.

Finally, I would like to thank my parents and my brother for their incredible support during these years. None of these would have been possible without you.

“Put it before them briefly so they will read it, clearly so they will appreciate it, picturesquely so they will remember it, and above all, accurately so they will be guided by its light.”

—Joseph Pulizer

List of abbreviations

A

Å	Angström
Acr	Acridine resistance protein
A.D.	Anno Domini
ABC	ATP Binding Cassette
AMR	Antimicrobial Resistance/Resistant
ANAPOE-$C_{12}E_{10}$	3,6,9,12,15,18,24,27,30-Decaoxadotetracontan-1-ol
AP	Alkaline Phosphatase

B

BCIP	5-Bromo-4-Chloro-3-Indoyl Phosphate
BSA	Bovine Serum Albumin

C

ca.	Circa
CE	Common Era

D

2D	Two-dimension
$\Delta\Psi$	Membrane potential
DDM	*n*-Dodecyl-β-D-Maltopyranoside
DFP	Diisopropyl fluorophosphate
DHA1/2/3	Drug:H^+ Antiporter
DMNG	Decyl Maltose Neopentyl Glycol
DMSO	Dimethyl sulfoxide
DMT	Drug/Metabolite Transporter
DNA	Deoxyribonucleic acid

E

EDTA	Ethylenediaminetetraacetic acid
EEA	European Economic Area
EM	Electron Microscopy
Emr	*Escherichia coli* multi-drug resistance
ESβL	Extended Spectrum β-Lactamase
EU	European Union

F

FSEC	Fluorescence Detection Size Exclusion Chromatography
FX	Fragment eXchange

I

IgG	Immunoglobulin G
IPTG	Isopropylthiogalactopyranoside

L

LB	Lysogeny Broth

M

MATE	Multi-drug And Toxic compound Extrusion
MDR	Multi-drug Resistance/Resistant
MFS	Major Facilitator Superfamily
MOP	Multidrug Oligosaccharidyl lipid Polysaccharide
mRFP1	Monomeric Red Fluorescent Protein
MRSA	Methicillin Resistant *Staphylococcus aureus*

N

NBT	Nitro Blue tetrazolium
NCBI	National Center for Biotechnology Information
NDM	New Delhi Metallo-β-lactamase

O

OD_{600}	Optical density at 600 nm
ORF	Open Reading Frame
OXA	Oxacillin-hydrolyzing

P

PACE	Proteobacterial Antimicrobial Compound Efflux
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PMSF	Phenylmethylsulfonyl fluoride

R

RND	Resistance Nodulation cell Division

S

SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
sfGFP	Superfolder Green Fluorescent Protein
SMALPs	Styrene Maleic Acid copolymer Lipid Particles
SMR	Small Multidrug Resistance

T

TB	Terrific Broth
TBST	Tris Buffered Saline with Tween
TCDB	Transporter Classification DataBase
TMS/TM	Transmembrane α-helix spanner
TolC	Tolerance to Colicin E1
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
Triton X-100	4-(1,1,3,3-Tetramethylbuthyl)phenyl-polyethylene glycol

V

vol/vol	Volume/Volume percent concentration

W

wt/vol	Weight/Volume percent concentration

Y

2xYT	2x Yeast Extract Tryptone

List of figures

Figure 1. Current and estimated number of future deaths per year linked to AMR. 44

Figure 2. Timeline and delay between antibiotic introduction and antibiotic resistance development.. 46

Figure 3. Graphic showing the decreasing number of drugs available for the treatment of severe bacterial infections. 48

Figure 4. Flight travel routes and spread of antibiotic resistance. 49

Figure 5. Multiple, complex and interconnected routes of antibiotic resistance transmission. 50

Figure 6. Percentage of MDR *E. coli* isolates with combined resistance to third-generation cephalosporins, fluoroquinolones and aminoglycosides, within EU/EEA countries in 2017. 54

Figure 7. Proportion of *E. coli* isolates from EU/EEA countries in 2017 corresponding to fully susceptible, resistant to one, two, three, four or all five antibiotic groups (*i.e.* aminopenicillins, third-generation cephalosporins, carbapenems, fluoroquinolones and aminoglycosides). 55

Figure 8. Resistance acquisition pathways. 57

Figure 9. Mechanisms of resistance to antibiotics. 58

Figure 10. Different families of efflux pumps. 59

Figure 11. Representation of the transport modes of members belonging to the MFS. 60

Figure 12. Transport modes of LacY. 68

Figure 13. Global structure of LacY. 70

Figure 14. The substrate binding site of LacY. 71

Figure 15. Organization of critical residues involved H^+ translocation and coupling. 72

Figure 16. Four motifs identified within MdfA homologs. 73

Figure 17. Global structure of MdfA. 74

Figure 18. Protonation sites and important residues for substrate binding. 75

Figure 19. Site coupling protonation status to substrate binding. 76

Figure 20. The cytoplasmic rim of MdfA. 77

Figure 21. Global structural fold of EmrD. 78

Figure 22. The hydrophobic cavity of EmrD. 79

Figure 23. View of the selectivity filter of EmrD. 79

Figure 24. Representation of the overall topology of $PepT_{So}$. 80

Figure 25. Global structure of $PepT_{So}$. 81
Figure 26. Gating regions in $PepT_{So}$. 82
Figure 27. Substrate binding region of $PepT_{So}$. 83
Figure 28. Structure of the human glucose transporter 3 (GLUT3). 84
Figure 29. Illustration of the global MFS fold. 85
Figure 30. Major conformational states of MFS transporters. 86
Figure 31. Symport mechanism of LacY. 87
Figure 32. Suggested antiport mechanism of MdfA. 88
Figure 33. Antiport mechanism of EmrD. 89
Figure 34. Symport mechanism of $PepT_{So}$. 90
Figure 35. The rocker-switch model. 91
Figure 36. The clamp and switch model. 92
Figure 37. Illustration and structural evidence of different types of gates. 94
Figure 38. *In vitro* transcription-translation products of *emrAB* labeled with [^{35}S]methionine separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). 96
Figure 39. SDS-PAGE analysis (with the EZ blue stain) of EmrA and EmrB purified separately. 97
Figure 40. Isolation of the EmrAB complex. 98
Figure 41. Negative staining EM analyses of EmrAB. 99
Figure 42. Models of tripartite MFS-type systems. 101
Figure 43. Overall structure of TolC in its closed conformation. 103
Figure 44. External and internal views of the periplasmic end of the α-helical barrel of TolC in its closed conformation. 104
Figure 45. Detailed view of the periplasmic entrance of TolC in its closed conformation. 105
Figure 46. Global structural fold of EmrA from *Aquifex aeolicus*. 106
Figure 47. Global structural comparison of different periplasmic adaptors. 107
Figure 48. Comparison of the α-helical coiled-coil domains of different adaptors. 108
Figure 49. Sequence alignment of EmrA from *Aquifex aeolicus* (AaEmrA) and from *Escherichia coli* (EcEmrA). 109
Figure 50. Structural overlay of β-barrel domains from different adaptors. 110

Figure 51. Schematic description of the FX cloning methodology. 123
Figure 52. Agarose gel electrophoresis analysis of amplified *emrA*, *emrB* and *tolC* type genes. 134
Figure 53. Plasmid maps of the final expression constructs used for the screening purposes..... 135
Figure 54. Cloning methodology used for the preparation of fusion stabilized EmrAB-TolC systems from *E. coli*. 139
Figure 55. Schematic representations of the constructs encoding stabilized EmrAB-TolC efflux systems..... 143
Figure 56. Plasmid maps of two types of individual expression constructs. 145
Figure 57. Schematic representation of the high-throughput screening pipeline. 148
Figure 58. Fluorescence spectra of mRFP1 and sfGFP. 149
Figure 59. mRFP1 based expression screen of EmrB. 155
Figure 60. sfGFP based expression screen of EmrA. 156
Figure 61. Expression level analyses of the *E. coli* efflux system components. 157
Figure 62. Electrophoretic mobility of mRFP1 and sfGFP fusion proteins expressed in *E. coli* C41(DE3)*ΔacrAB* cells. 159
Figure 63. Comparison between green fluorescence measurements of whole cell samples at different solubilization test steps..... 160
Figure 64. FSEC chromatograms of the EmrAB-TolC system from *E. coli* compared to each negative control. 161
Figure 65. FSEC chromatograms of the EmrAB-TolC system from *E. coli* obtained via two different methods. 162
Figure 66. Comparison of the expression levels of the *E. coli* and *V. cholerae* systems..... 163
Figure 67. Comparison of the FSEC chromatograms of the *E. coli* and *V. cholerae* systems. 164
Figure 68. Schematic representation of the four constructs selected for biochemical analyses. 165
Figure 69. Western blot analyses of the EmrAB fusion chimeras co-expressed with TolC. 167
Figure 70. Western blot analyses of the pull down assay samples (fusion chimeras)..... 167
Figure 71. Scheme describing the purification procedures tested for the isolation of EmrAB-TolC... 172
Figure 72. Comparison of the Ni^{2+} and Co^{2+} based affinity purifications. 174
Figure 73. Large-scale co-purification analysis of EmrAB-TolC. 175

Figure 74. Size-exclusion chromatography profile and electrophoretic mobility analysis of EmrAB-TolC stabilized with Amphipol A8-35. 176
Figure 75. TEM analysis of the tripartite EmrAB-TolC efflux system. 177
Figure 76. Wide field image of negatively stained and gold labeled sample containing EmrAB-TolC complexes 178
Figure 77. A gallery of electron micrographs of negatively stained and gold labeled EmrAB-TolC complexes 179
Figure 78. Western blot analyses of the Ni^{2+}-NTA purification of the 15 GS EmrAB fusion-TolC system. 181
Figure 79. SDS-PAGE analysis of EmrB after mRFP1 cleavage. 186
Figure 80. Reconstitution of EmrB in Amphipol A8-35. 187
Figure 81. Mass spectrometry sequence coverage for EmrB. 187
Figure 82. Negative stain EM analysis of EmrB and Amphipol A8-35. 188
Figure 83. Structures of RND and ABC superfamily tripartite efflux systems. 195
Figure 84. Structural model of ecEmrA and overview of tripartite efflux systems from Gram-negative bacteria. 197

List of tables

Table 1. Crystal structures of bacterial MFS members. 61

Table 2. Bacterial MFS-type multi-drug transporters. 64

Table 3. Various operons selected for subsequent PCR amplifications. 116

Table 4. Identity and similarity values of the homologous proteins compared to the *E. coli* efflux system components. 120

Table 5. Primers used for the insertion of the ORFs into pINITIAL. 124

Table 6. PCR reaction mixture composition for the *E. coli* targets (*emrAB-tolC*). 128

Table 7. PCR reaction mixture composition for the *V. cholerae*, *H. elongata*, *D. tiedjei*, *T. lienii*, *T. albus* and *C. metallidurans* targets..... 128

Table 8. PCR reaction mixture composition for the *E. coli* (*emrKY*), *S. blattae*, *S. fumaroxidans*, *K. pneumoniae*, and *S. enterica* targets. 129

Table 9. PCR cycling program used for all the targets..... 129

Table 10. Reaction mixture composition of the initial FX cloning step. 130

Table 11. Reaction mixture composition of the FX subcloning step. 131

Table 12. Primers used for the traditional cloning of *E. coli tolC* and *V. cholerae vceC*. 132

Table 13. PCR reaction mixture composition for the amplification of *E. coli tolC* and *V. cholerae vceC* for traditional cloning. 132

Table 14. Cloning progress of the ORFs of 15 homologous EmrAB-TolC systems. 136

Table 15. Primers used for the 'modified FX cloning' and traditional clonings for the preparation of genetically engineered EmrAB-TolC systems from *E. coli*. 140

Table 16. Other co-expression conditions tested for the EmrAB-TolC complex from *E. coli*. 170

Table 17. Top 4 covarying residue pairs of EmrA and TolC predicted by GREMLIN. 196

Table 18. Predictions of covarying residue pairs between EmrB and TolC by GREMLIN. 196

Table of contents

INTRODUCTION 41

CHAPTER I: ANTIMICROBIAL RESISTANCE: FROM BACTERIAL CELLS TO TRIPARTITE EFFLUX SYSTEMS 43

1. Introduction 44

1.1. Antimicrobial resistance: a global health concern 44

1.2. Antibiotic introduction and antibiotic resistance 45

1.3. Spread of antibiotic resistance 48

1.4. Antibiotic resistance: an old phenomenon 51

1.5. Notorious antibiotic resistant pathogenic bacteria 51

1.6. Antibiotic resistant *Escherichia coli* 54

1.7. Tackling antibiotic resistance 56

2. Antibiotic resistance at the bacterial cell level 57

2.1. Acquisition of resistance 57

2.2. Mechanisms of resistance 58

3. The major facilitator superfamily (MFS) 60

3.1. Single component MFS members 67

3.1.1. Structural insights of single component MFS members 67

3.1.1.1. LacY 67

3.1.1.2. MdfA 73

3.1.1.3. EmrD 77

3.1.1.4. $PepT_{So}$ 80

3.1.1.5. Generalization of the structural fold of single component MFS members 84

3.1.2. Mechanistic insights of single component MFS members 86

3.1.2.1. LacY 86

3.1.2.2. MdfA 88

3.1.2.3. EmrD 89

3.1.2.4. $PepT_{So}$ 89

3.1.2.5. Generalization of the transport mechanism of single component MFS members 90

3.2. Tripartite MFS members 95

3.2.1. General introduction to the EmrAB-TolC system 95

3.2.2. Structural insights of tripartite MFS members 97

3.2.2.1. Structure of the EmrAB complex 97
3.2.2.2. Postulated models of the EmrAB-TolC and FarAB-MtrE systems 100
3.2.2.3. Structural review of the protein partners forming the EmrAB-TolC system 102
3.2.2.3.1. The inner membrane drug:H^+ antiporter EmrB 102
3.2.2.3.2. The outer membrane exit duct TolC....102
3.2.2.3.3. The periplasmic adaptor protein EmrA....106
3.2.3. Functional insights of tripartite MFS members....111
4. Strategies employed for the isolation of the EmrAB-TolC system 111
METHODS AND RESULTS 113
CHAPTER II: IDENTIFICATION OF '*emrAB*' AND '*emrAB-tolC*' OPERONS AND SUBSEQUENT CLONING 115
1. Identification of emrAB-tolC *ORFs from* Escherichia coli *and other gram-negative bacteria.. 116*
2. Fragment eXchange (FX) cloning 122
3. First cloning strategy for the subsequent use of the properties of fluorescent labels and affinity tags for screening purposes 124
3.1. Cloning methodology for the production of proteins fused to fluorescent labels and affinity tags 124
3.2. Results of the first cloning strategy employed for screening purposes....133
4. Alternative cloning strategy for the production of genetically engineered EmrAB-TolC systems from E. coli*.... 138*
4.1. Cloning methodology for the co-production of fusion stabilized EmrAB-TolC systems from *E. coli* ..138
4.2. Results of the alternative cloning strategy used for the stabilization of the EmrAB-TolC system from *E. coli*....143
5. Cloning strategy for the individual expression of the emrA*,* emrB *and* tolC *genes..... 144*
5.1. Cloning methodology for the preparation of individual expression constructs for the *emrAB-tolC* inserts from *E. coli*....144
5.2. Results of the third cloning strategy representing an alternative method for the isolation of the EmrAB-TolC system from *E. coli*....144
CHAPTER III: SMALL-SCALE ANALYSES OF EmrAB-TolC FROM *E. coli* AND VceABC FROM *V. cholerae* 147
1. A high-throughput screening pipeline for initial analyses of EmrAB-TolC systems 148
1.1. Presentation of the high-throughput screening pipeline 148
1.2. Experimental procedures for screening purposes....150
1.2.1. Co-expressions....150

1.2.1.1. Co-expression level analyses of the EmrAB-TolC from *E. coli* in 96 deep-well blocks.....150
1.2.1.2. Co-expressions of EmrAB-TolC from *E. coli* and VceABC from *V. cholerae* for FSEC experiments.....151
1.2.1.2.1. Membrane preparation of cells co-expressing EmrAB-TolC from *E. coli*.....151
1.2.1.2.2. Small-scale co-expressions of EmrAB-TolC from *E. coli* and VceABC from *V. cholerae* 151
1.2.2. Small-scale preliminary verifications 152
1.2.2.1. *In gel* mobility controls of EmrAB-TolC from *E. coli* 152
1.2.2.1.1. *In gel* fluorescence.....152
1.2.2.1.2. Immunodetection 152
1.2.2.2. Initial *E. coli* EmrA integrity verification 153
1.2.3. FSEC 154
1.2.3.1. FSEC starting from cellular membranes 154
1.2.3.2. FSEC starting from whole cells 154
1.3. Results of the screening procedures.....155
1.3.1. Co-expressions.....155
1.3.1.1. *E. coli* C41 cells are best suited for the co-expression of EmrB and EmrA from *E. coli* 155
1.3.1.2. Comparison of the expression behaviors of the three partners forming the *E. coli* efflux system 157
1.3.2. Small-scale preliminary verifications 158
1.3.2.1. The two different fluorescent labels and three different affinity tags are correctly located on the three *E. coli* protein partners.....158
1.3.2.2. *E. coli* EmrA fused to sfGFP is correctly folded 160
1.3.3. EmrAB-TolC from *E. coli* is a tripartite complex 161
1.4. Comparison of the expression and complex formation behaviors of the *E. coli* and *V. cholerae* efflux systems 163
2. Initial analyses of fusion stabilized EmrAB-TolC systems from E. coli..... *165*
2.1. Experimental procedures 165
2.1.1. Small-scale co-expressions.....165
2.1.1.1. Co-expressions for Western blot analyses.....165
2.1.1.2. Co-expressions for a pull down assay 166
2.1.2. Pull down assay.....166

2.2. Results....166
2.2.1. The EmrAB fusion chimeras are co-expressed with TolC....166
2.2.2. The 15GS EmrAB fusion can be isolated with TolC....167
CHAPTER IV: LARGE-SCALE CO-EXPRESSION, CO-PURIFICATION AND EM ANALYSIS OF THE EmrAB-TolC COMPLEX FROM *E. coli*.... 169
1. First strategy making use of the red and green fluorescent labels 170
1.1. Experimental procedures....170
1.1.1. Co-expression of EmrAB-TolC....170
1.1.2. Co-purification of EmrAB-TolC....171
1.1.3. Negative staining EM analysis of EmrAB-TolC....173
1.1.4. Identification of pumps by labelling with nitrilotriacetic acid-nanogold....173
1.2. Results....174
1.2.1. The entire EmrAB-TolC complex can be isolated directly from bacterial cells....174
1.2.2. The entire EmrAB-TolC complex exhibits an elongated structure....177
1.2.3. The identity of the entire EmrAB-TolC efflux system could be confirmed by gold labelling....178
2. Second strategy making use of the EmrAB fusion chimera 179
2.1. Experimental procedures....180
2.1.1. Co-expression of the 15 GS EmrAB fusion-TolC system....180
2.1.2. Co-purification of the 15 GS EmrAB fusion-TolC system....180
2.2. Results....181
2.2.1. The EmrAB-fusion and TolC protein partners did not seem to form stable complexes....181
CHAPTER V: LARGE-SCALE EXPRESSION, PURIFICATION AND EM ANALYSIS OF EmrB FROM *E. coli*.... 183
1. Experimental procedures.... 184
1.1. Expression of EmrB....184
1.2. Purification of EmrB....184
1.3. Proteomic analysis....185
1.4. Negative staining EM analysis of EmrB....185
2. Results.... 185
2.1. Biochemical characterizations of EmrB....185
2.2. Structural characterization of EmrB....188
DISCUSSION AND CONCLUSION....189

CHAPTER VI: GLOBAL DISCUSSION AND CONCLUSION 191

1. Global discussion and conclusion 192

1.1. Generation of different DNA constructs..... 192

1.2. Utilization of a high-throughput screening pipeline..... 193

1.3. Co-expression, co-purification and EM analysis of EmrAB-TolC 194

1.4. Expression, purification and EM analysis of EmrB..... 198

1.5. Conclusion 198

REFERENCES 201

ANNEXES 227

INTRODUCTION

Chapter I: Antimicrobial resistance: from bacterial cells to tripartite efflux systems

The present chapter will briefly describe the global health crisis attributed to antimicrobial resistance (AMR) with its multiple and variable aspects.

The different aspects of resistance acquisition and of the resistance mechanisms at the bacterial cell level will be mentioned.

Active efflux is a prominent resistance mechanism in Gram negative bacteria. Efflux is catalyzed by efflux systems which span both inner and outer membranes. The inner membrane proteins belong to different superfamilies. Here, members of the major facilitator superfamily (MFS) of transporters will be introduced with a special focus on the tripartite EmrAB-TolC system from *Escherichia coli*.

Finally, the strategies employed during the present Ph.D. project for the isolation and subsequent structural characterization of the tripartite EmrAB-TolC system will be described.

1. Introduction

In this first section an overview will be given about the major health concern, it's main causes, how it spreads worldwide, and its ancient origin. The main pathogenic organisms will be presented with a special focus on *Escherichia coli*. Different approaches proposed to win the battle against antimicrobial resistance that are currently under investigation will be discussed.

1.1. Antimicrobial resistance: a global health concern

Currently, about 700,000 patients die each year from antimicrobial resistant infections worldwide[1,2]. With the actual rhythm of excessive usage of antibiotics this number could reach 10 million deaths per year by 2050[1,2]. Thus, antimicrobial resistance could become one of the leading causes of death in the world (Figure 1)[2]. In Europe, the number of deaths attributed to antimicrobial resistance is about 25,000 per year with an associated economic burden of about €1.5 billion annually[3,4].

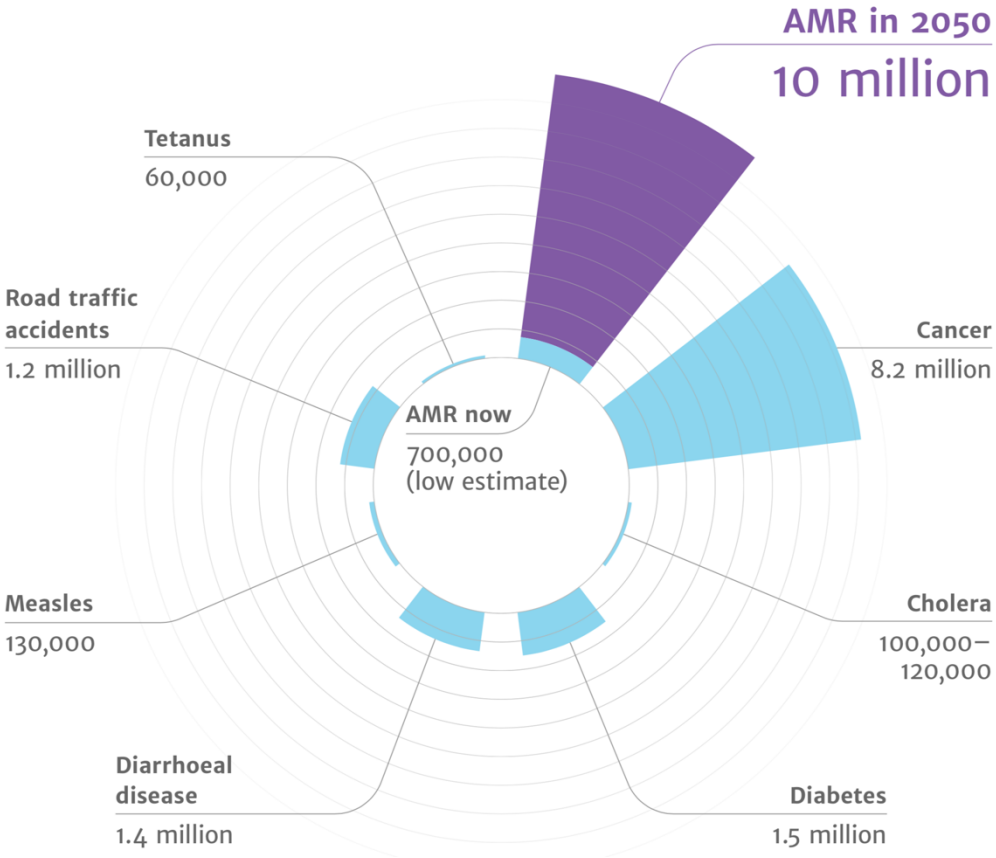

Figure 1. Current and estimated number of future deaths per year linked to AMR.

Diagram showing the estimated global health impact of antimicrobial resistance (terminology including antibiotic resistance phenomenon (bacteria) but also the health burden caused by viruses, fungi and parasites). Adapted from reference 2.

Antimicrobial resistance can be defined as the ability of microorganisms and infectious agents (*e.g.* bacteria) to resist to drugs that would usually either kill them (*e.g.* bactericidal) or stop their proliferation (*e.g.* bacteriostatic)[2,5]. Henceforth, the present description of the research project context will mainly focus on bacteria and therefore the terminology antimicrobial resistance will be replaced by antibiotic resistance.

The excessive use of antibiotics has led to a proliferation of this phenomenon leading to the development of strains able to resist to multiple antibiotics (MDR). For example, the appearance of methicillin resistant *Staphylococcus aureus* (MRSA) represents an unprecedented threat to human health because of the difficulty or impossibility of an available treatment[2].

Previously, resistant infections were mostly associated to hospitals with the so-called nosocomial infections. Nevertheless, studies show that antibiotic resistant infections can be found in other reservoirs as well[2,3].

Without reliable alternatives to antibiotics, we could head to a post-antibiotic era where numerous medical procedures (*e.g.* cancer chemotherapy, dialysis treatment, organ transplant, orthopaedic surgery, caesarian section) would present a very high risk and common infections could kill once again[2,5].

1.2. Antibiotic introduction and antibiotic resistance

The modern era of antibiotics started with the discovery of penicillin by Sir Alexander Fleming in 1928 changing completely the practice of medicine. In the 1940s penicillin was highly used for the treatment of serious infections and to control bacterial infections among the casualties from World War II. However, Sir Alexander Fleming already predicted that the incorrect use of penicillin could lead to penicillin resistance which would become a major clinical problem when he received the Nobel Prize in Physiology or Medicine in 1945[6]. Nevertheless, the success of penicillin led to the development of numerous classes of antibiotics. Unfortunately, the overuse and especially misuse of these 'miracle-drugs' induced the appearance of resistance to nearly all classes of antibiotics (Figure 2)[7,5,8].

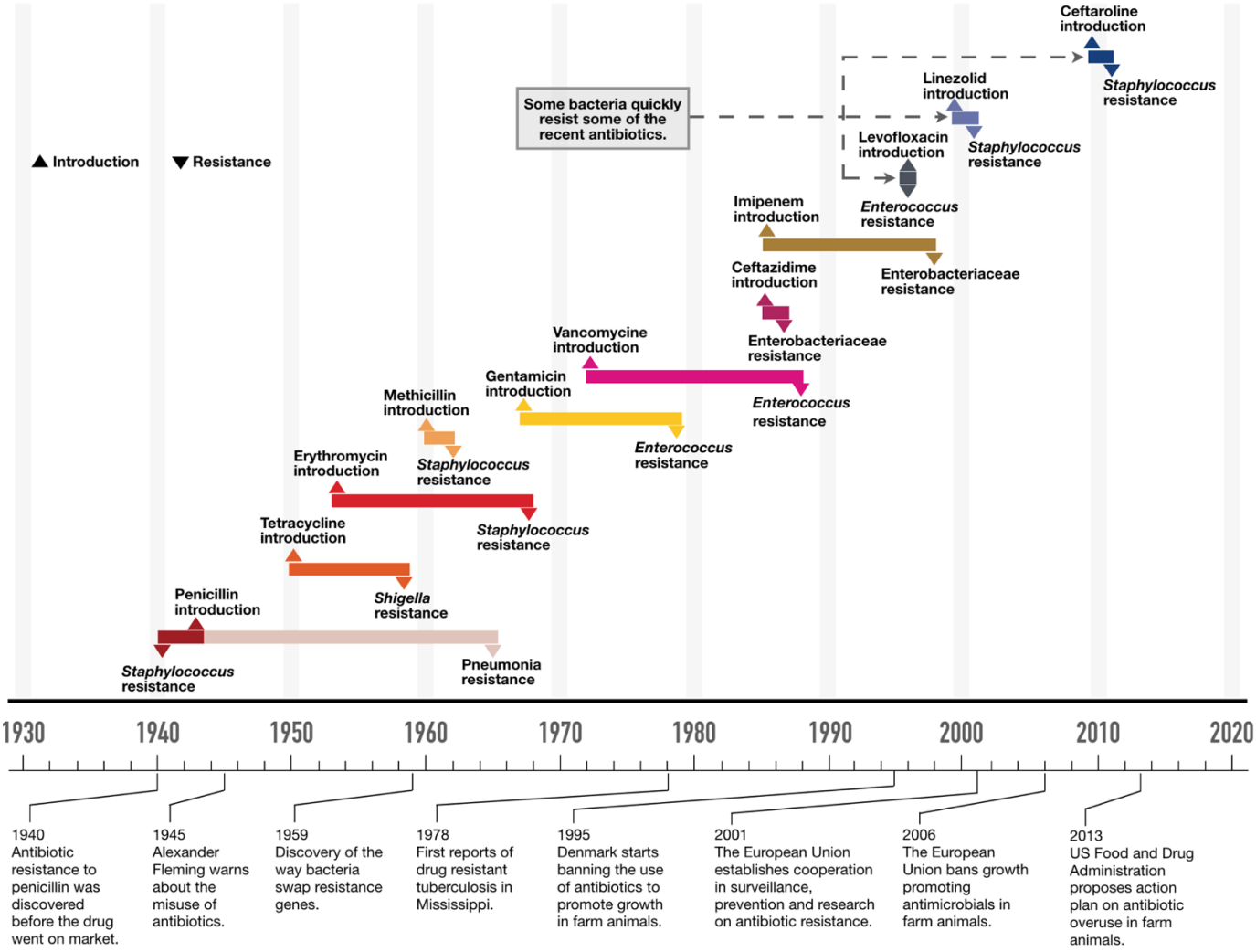

Figure 2. Timeline and delay between antibiotic introduction and antibiotic resistance development.

Shortly after the introduction of antibiotics on the market (upward arrows), antibiotic resistance is detected (downward arrows). However, their usage continues. More and more antibiotics are introduced to the market over time, but bacteria progressively learn to resist to the different drugs. Adapted from reference 8.

The emergence of clinically relevant antibiotic resistance is considered to be mainly due to human activity. The increased selection pressure has dramatically contributed to the appearance and development of resistant pathogens. Since their approval, antibiotics have been widely prescribed by physicians worldwide. It is considered that 50% of the antibiotics prescribed in human medicine were unnecessary[9]. Therefore, a general guidance is needed for the appropriate prescription of antibiotics with a careful estimation of the duration of the treatment for each patient. Moreover, special care with last resort antibiotics must be taken and these drugs should be kept for infections with multi-drug resistance profiles. More importantly, in the USA and in Europe, antibiotics have been used four times more in the food industry than in human medicine. Estimates described the presence of 4 mg to 400 mg of antibiotics per kg of meat that had been produced in European countries[10–12]. It should be mentioned that antibiotics used in the food industry were closely related or even identical to those used for human medicine therefore contributing to the global health crisis[13,14]. It was only in 2006 that the European Union banned the use of antibiotics given as growth factors to food producing animals[15,16]. The US Food and Drug Administration finally banned the use of antibiotics as growth promoters in 2017[17]. In countries such as Vietnam and Thailand farmers are still using antibiotics in food producing animals even if their use as growth promoters has been prohibited since 2018[18].

Unfortunately, whereas the sales of antibiotics was on the increase, the admission of new antibiotics from different classes to the market have declined. Between 1980 and 1984, 19 new drug applications were approved by the US Food and Drug Administration, whereas between 2005 and 2009 only 3 new antibiotics were approved (Figure 3)[8]. This decline in approvals correlates with the diminished attractivity of antibiotics for pharmaceutical companies. In fact, with the rapid development of resistance to different classes of antibiotics, the return on investment was low and cannot be considered as economically relevant for the industry. For example, one of the leading companies in antibiotics development, namely Pfizer closed its antibiotics research and development department in 2011. Similarly, Roche closed its antibiotics facility in 1999. Fortunately, other companies such as Merck, and Cubist pharmaceuticals remained in this business. Moreover, in order to encourage the development of new drugs, public-private partnerships such as the Innovative Medicines Initiative in Europe with its ‘New Drugs for Bad Bugs’ program were created. Such collaborations including

academic researchers, actors from public health and small/large industries could be one of the best solutions to promote the development of efficient and economically viable antibiotics[19].

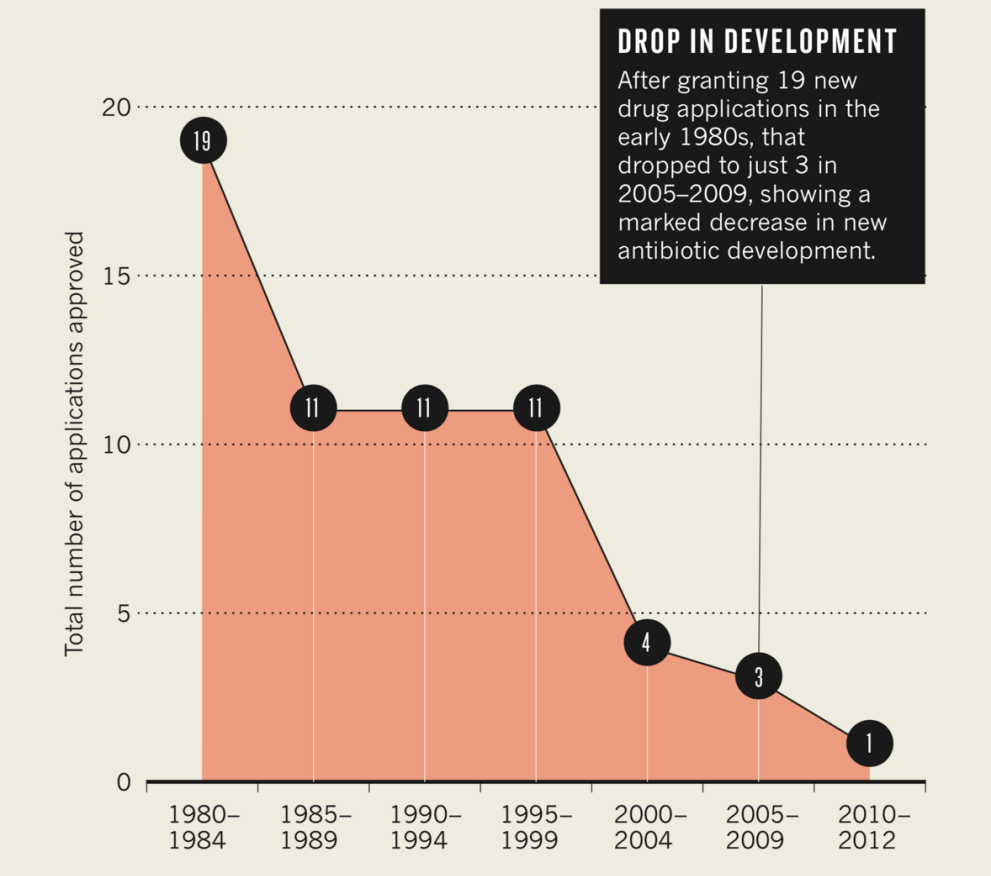

Figure 3. Graphic showing the decreasing number of drugs available for the treatment of severe bacterial infections.

Since the 1980s fewer new antibiotics were approved by the US Food and Drug Administration. Adapted from reference 8.

1.3. Spread of antibiotic resistance

The spread of antibiotic resistance for both gram-positive and gram-negative bacteria can occur through two main categories of transmissions: transmissions between humans and by complex animal-human-environment transfer routes.

One of the most striking pathways of antibiotic resistance occurs through modern travel. For instance, the level of gut colonization of extended spectrum β-lactamase (ESβL) positive Enterobacteriaceae, and the transfer of carbapenem resistance mechanisms with the New Delhi metallo-β-lactamase (NDM), *Klebsiella pneumoniae* carbapenemase, and carbapenem resistant OXA

β-lactamases (oxacillin-hydrolyzing, OXA-48) enzymes has been linked to travel (Figure 4). A second thoroughly studied transfer pathway between humans concerns the hospital and health-care acquired (nosocomial) infections. Amongst other organisms, infection with MRSA was linked to the duration of the stay of patients in hospitals as well as the hand-contamination of health-care professionals. Faecal-oral transmissions can occur as well, in cases of poor sanitation. Last, sexual transmitted diseases can also occur, in the case of for example *Neisseria gonorrhoeae* [16].

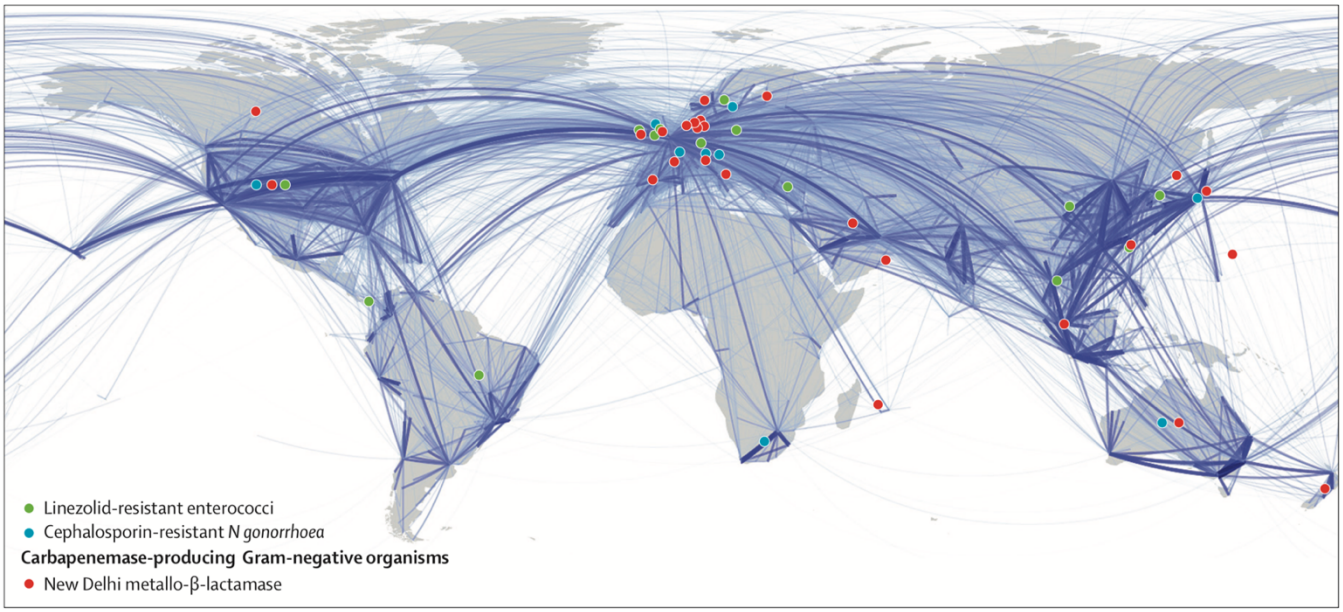

Figure 4. Flight travel routes and spread of antibiotic resistance.

Data shown for NDM-positive bacteria form patients epidemiologically linked to the Indian subcontinent, linezolid resistant enterococci, and cefixime/ceftriaxone resistant *Neisseria gonorrhoea*. Adapted from reference 16.

Antibiotic resistance is actually present everywhere in the environment. Multiple complex interconnections between different ecological niches make it easy for antibiotic resistant bacteria and antibiotic resistance genes to move from one reservoir to another (Figure 5)[10]. First, as stated earlier, antibiotic resistance transfer has been identified between food producing animals and humans. Evidences show the transmission of resistant zoonotic pathogens such as *Salmonella serovars* and *Campylobacter spp.* but also of resistant opportunistic pathogens such as enterococci and *Staphylococcus aureus* [16]. Antibiotic resistance is also present in aquaculture and transmission routes between aquaculture and livestock but also between aquaculture and humans has been evidenced.

Obviously, there is a clear link between wastewater treatment from households, farms and industries and the transfer of resistance to aquatic ecosystems[20,21]. Finally, resistance is also present in cultivation farms because of the use of manure and nitrogen fertilizers but also with the use of metals as bactericides and fungicides[16].

The antibiotic resistance phenomenon presents therefore an unprecedent global challenge comparable to climate change which must be considered at all levels within low-income, moderate-income and high-income countries worldwide.

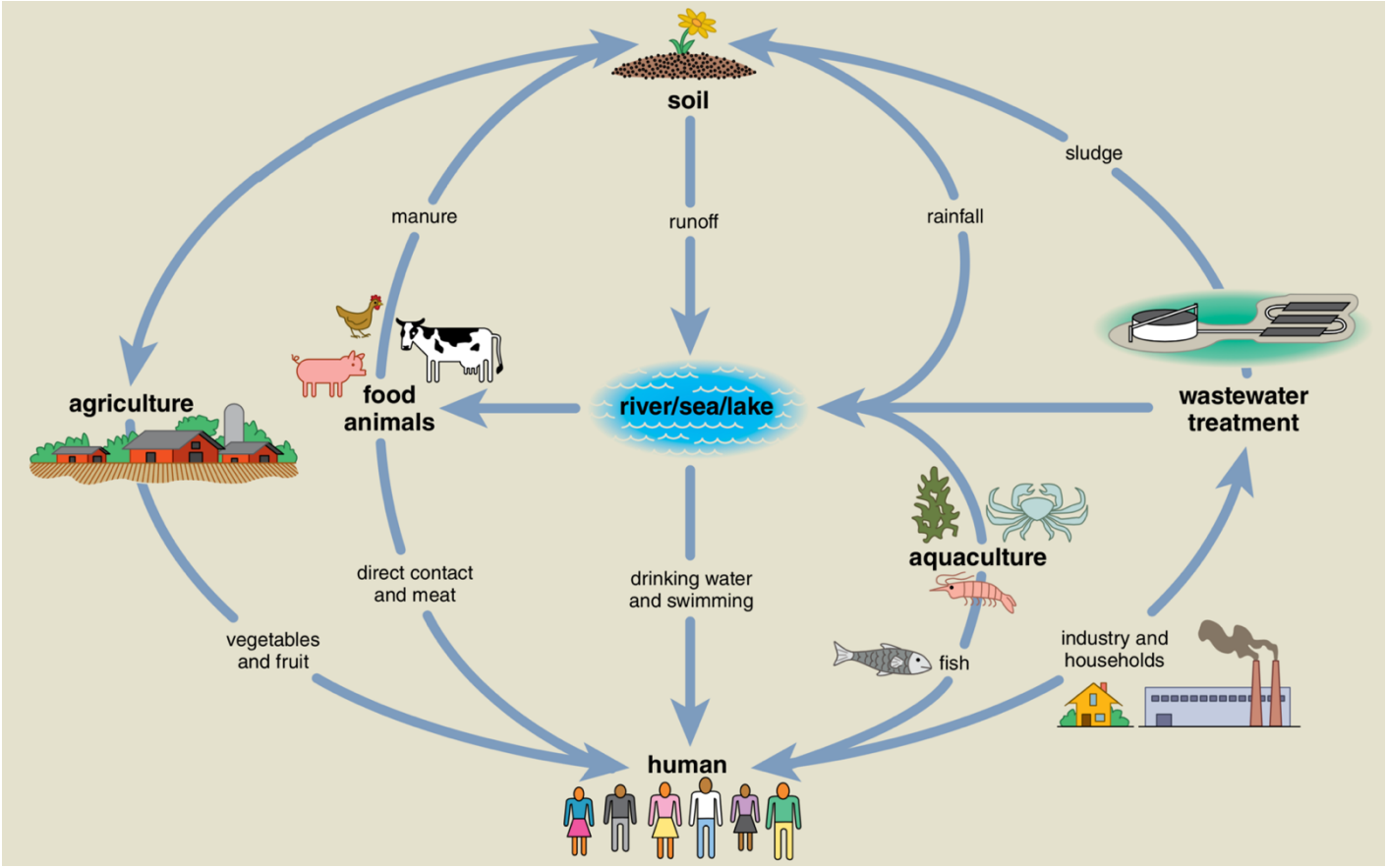

Figure 5. Multiple, complex and interconnected routes of antibiotic resistance transmission.

Multiple routes of exchange promote the development and spread of antibiotic resistance between humans, animals and the environment. For example, an antibiotic resistant bacterium occurring in livestock animals could travel through manure to the soil and be transferred to humans via consumption of vegetables and fruit. Adapted from reference 10.

1.4. Antibiotic resistance: an old phenomenon

Antibiotics employed in medicine and agriculture are mainly derivatives or natural compounds produced by Actinomycetes (*e.g. Streptomyces*). These organisms have different antibiotic resistance mechanisms to survive to their own toxic compounds conferring these an immense advantage for instance for nutrient competition. Therefore, it is considered that such antibiotic producing organisms constitute the origin of multiple antibiotic resistance genes present in various bacteria[22,23].

Given the vast presence of antibiotic resistance in different environmental reservoirs, an evolutionary link must exist between different bacteria.

Examples in the literature include the presence of genes encoding for resistance to β-lactam, tetracycline and glycopeptide antibiotics in 30,000 year old Beringian permafrost samples[24]. Analyses of ancient human samples (gut microbiome of an 11th Century A.D. pre-Columbian Andean mummy and oral microbiome of human skeletons from the medieval monastic site of Dalheim, Germany ca. 950-1200 CE) showed the presence of genes conferring resistance to β-lactams, aminoglycosides, and macrolides amongst other antibiotics[25,26]. Finally, the study of the gut flora of a current but remote human community of Chayahuita Indians from Angaiza situated in the Alto Amazonas province of Peru and therefore isolated from modern civilization showed the presence of resistance to ampicillin, tetracycline, trimethoprim, streptomycin and chloramphenicol in commensal *Escherichia coli* [27].

Even if modern use of antibiotics has greatly influenced the selection and dissemination of antibiotic resistance, it is very important to study the ancient genetic history of these microorganisms in order to predict the future development of antibiotic resistance in the environment and establish strategies to limit this propagation and possibly win the battle against microbes.

1.5. Notorious antibiotic resistant pathogenic bacteria

The main MDR pathogenic bacteria causing significant mortality around the globe include both gram-positive and gram-negative organisms and are currently known as the 'ESKAPE' pathogens. This group is composed of *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species[28–32].

Enterococci are gram-positive bacteria. Amongst these organisms, *Enterococcus faecium* and *Enterococcus faecalis* constitute major nosocomial pathogens. Enterococci can be responsible for

variable diseases including bloodstream, urinary tract, skin and soft-tissue infections[33]. These organisms are particularly problematic because of high survival rates for long periods in hospitals. For instance, they can remain present on medical material despite cleaning procedures with alcohol disinfectants. Enterococci present high genome plasticity and developed therefore multiple resistances to several antibiotic groups. Examples include resistance to glycopeptides (*e.g.* vancomycin), β-lactams (*e.g.* ampicillin), and aminoglycosides (*e.g.* gentamycin)[34–36]. To treat enterococcal infections antibiotics such as daptomycin and linezolid are employed even if resistance to these antibiotics of some strains is known from the literature[33].

A second gram-positive opportunistic pathogen, namely *Staphylococcus aureus*, causes life threatening illnesses including infective endocarditis, and necrotizing pneumonia[37]. *Staphylococcus aureus* presents great antibiotic resistance acquisition capacities and resistances have been well described for numerous antibiotic groups. Examples include resistance to β-lactams (*e.g.* methicillin), aminoglycosides, macrolides, and tetracyclines[38–41]. Amongst different strains of *Staphylococcus aureus*, MRSA represents a major health concern. Alternative treatments for infections caused by *Staphylococcus aureus* include the use of daptomycin, linezolid, and vancomycin[37]. Nevertheless, resistance of some *Staphylococcus aureus* strains is known for linezolid[42] and studies in the literature indicated in some cases reduced susceptibilities for daptomycin and vancomycin as well[43–45].

Klebsiella pneumoniae is a gram-negative opportunistic pathogen that can cause diseases including urinary tract infection, cystitis, pneumonia, surgical wound infections, endocarditis and septicemia[46]. It also presents a great adaptability and resistance acquisition capacities towards multiple antibiotic groups. Examples include resistance to β-lactams (*e.g.* third-generation cephalosporins and carbapenems), aminoglycosides, and fluoroquinolones[46–49]. The prevalence of Extended Spectrum β-Lactamase (ESβL) and New Delhi Metallo-β-lactamase (NDM-1) producing strains is of particular concern. With this type of resistance range, establishing effective treatments is particularly challenging with few antibiotics available including colistin, and tigecycline[50,51]. Nevertheless, also in this case resistance to both antibiotics of some strains has been reported[52,53].

Acinetobacter baumannii, a major, gram-negative, opportunistic, nosocomial pathogen is responsible for different illnesses including bloodstream, urinary tract, skin and soft tissue infections[54]. It presents important survival rates for long periods in hospitals and on human surfaces. With important

resistance acquisition capacities, it is able to resist to multiple antibiotic groups. Examples include resistance to β-lactams (*e.g.* cephalosporins and carbapenems), fluoroquinolones (*e.g.* ciprofloxacin), and aminoglycosides[55–58]. Therefore, only few treatment options remain including colistin, tigecycline and rifampin[50,59,60]. However, resistance towards these three compounds of some strains is also known in the literature[61–63].

Pseudomonas aeruginosa is a gram-negative, common, opportunistic, nosocomial pathogen responsible for a variety of diseases including urinary tract, respiratory, soft tissue, corneal, and catheter associated infections[64]. This organism is particularly problematic because of its intrinsic tolerance towards many disinfectants. It also has a great ability to acquire novel resistance mechanisms from other organisms and can develop drug resistance during therapy. Resistance examples towards numerous antibiotic groups include β-lactams (*e.g.* carbapenems and cephalosporins), fluoroquinolones (*e.g.* ciprofloxacin), and aminoglycosides[65–68]. Against such infections, colistin is usually employed as a last treatment option[50]. However, resistance for this antibiotic by some strains has been reported as well[69,70].

Enterobacter species are gram-negative organisms. The main opportunistic, nosocomial pathogens are represented by *Enterobacter aerogenes* and *Enterobacter cloacae*. These organisms can cause diverse illnesses including infections of the urinary tract, the lower respiratory tract, wounds, and the central nervous system[71]. Because of their great resistance acquisition capacities and their ability to develop resistance during infection, they pose a major threat to human health. Resistance occurrences towards numerous antibiotic groups are known in the literature. Well known examples include resistance towards β-lactams (*e.g.* carbapenems, cephalosporins and penicillins)[72–75]. Indeed, strains producing ESβLs and carbapenemases are particularly problematic. Limited treatment options for such infections include the use of colistin and tigecycline[50,51]. However, resistance to both antibiotics of some strains is also known in the literature[76,77].

MDR pathogen associated infections are challenging illnesses and require therefore special treatments including combinatorial therapies with last resort antibiotics. New solutions must be developed in order to restore our set of medicines to combat such dreadful diseases.

1.6. Antibiotic resistant Escherichia coli

Escherichia coli, a gram-negative bacterium, is mainly a beneficial organism present in the human/animal gastrointestinal flora. Indeed, it is involved in the production of vitamin K (menaquinone) and prevents colonization of bacterial pathogens[78,79]. However, pathogenic strains of this organism exist as well and represent common causes of illnesses which include neonatal meningitidis, bloodstream, urinary tract and intra-abdominal infections[80]. Similar to the ESKAPE pathogens, *Escherichia coli* presents effective antibiotic resistance acquisition capacities. Examples include resistances towards aminopenicillins, third-generation cephalosporins, carbapenems, fluoroquinolones, and aminoglycosides[81–85]. According to the 2017 AMR surveillance report from the European Antimicrobial Resistance Surveillance Network, in total 58.2% of *Escherichia coli* clinical isolates obtained from the EU/EEA zone were resistant to at least one of the antibiotic groups mentioned previously. The EU/EEA population-weighted mean resistance percentages for the different antibiotic groups with a top down ranking corresponded to 58.7% for aminopenicillins, 25.7% for fluoroquinolones, 14.9% for cephalosporins, 11.4% for aminoglycosides and 0.1% for carbapenems. In general, resistance percentages for the different antibiotic groups mentioned varied among the EU/EEA countries with higher percentages for the southern and eastern countries (Figure 6)[81].

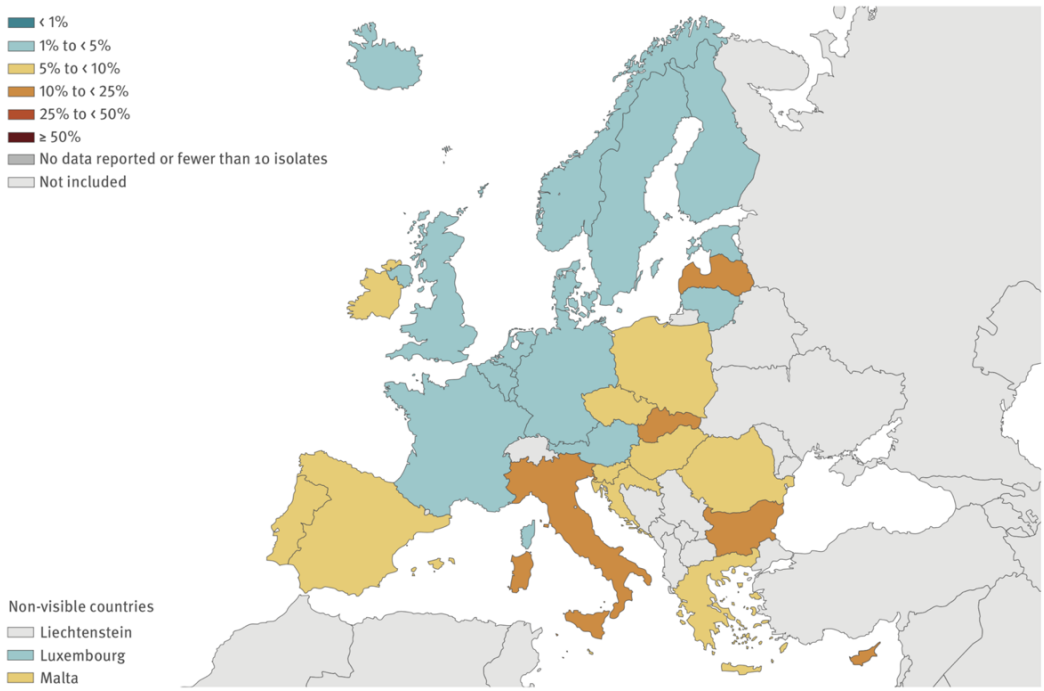

Figure 6. Percentage of MDR *E. coli* isolates with combined resistance to third-generation cephalosporins, fluoroquinolones and aminoglycosides, within EU/EEA countries in 2017.

MDR *E. coli* was present in all the countries included in the surveillance project with important occurrences in the southern and eastern nations. Adapted from reference 81.

Susceptibility differences were also identified between the different countries with a nearly similar fashion (*i.e.* higher susceptibility proportions for northern countries) (Figure 7)[81].

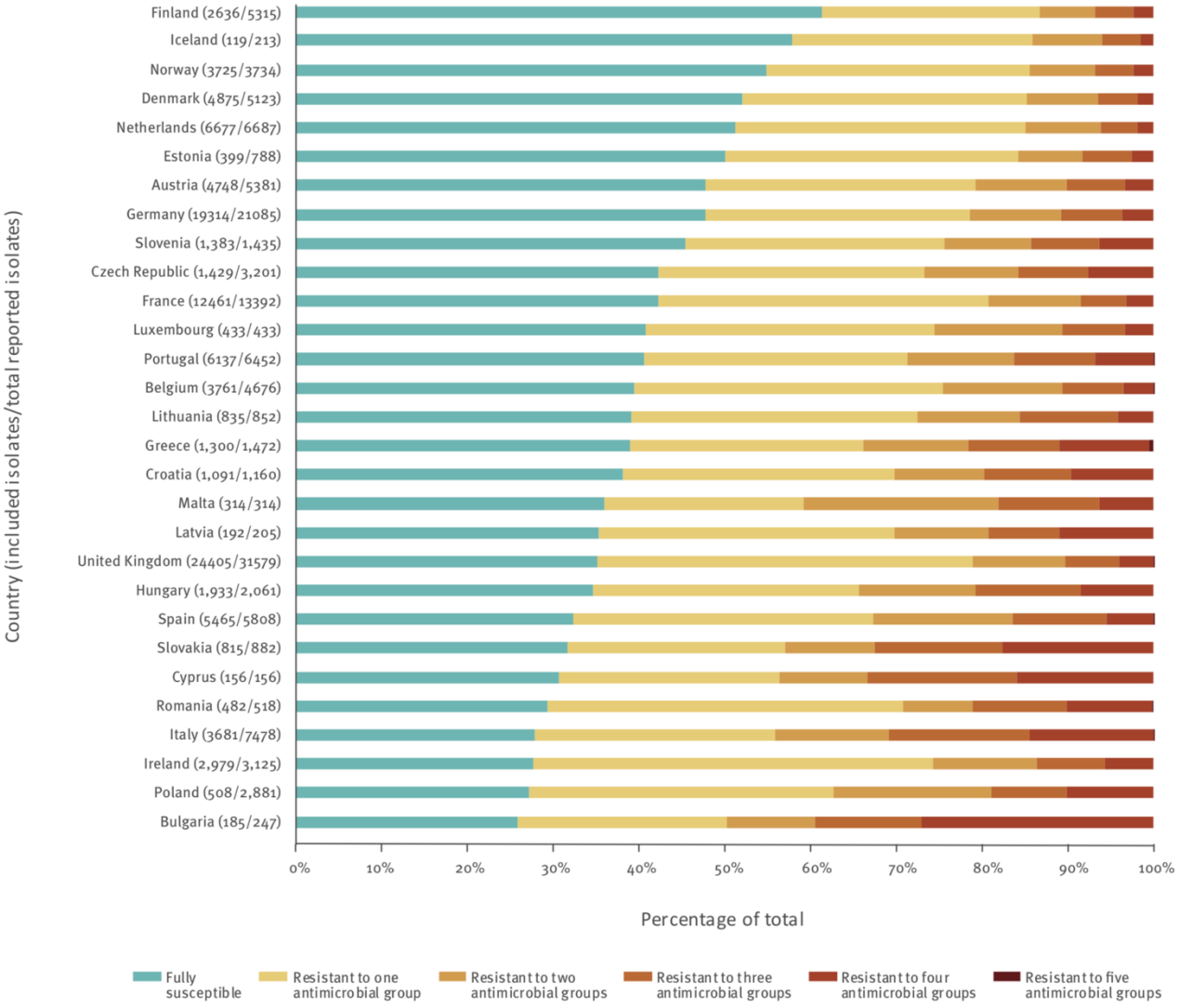

Figure 7. Proportion of *E. coli* isolates from EU/EEA countries in 2017 corresponding to fully susceptible, resistant to one, two, three, four or all five antibiotic groups (*i.e.* aminopenicillins, third-generation cephalosporins, carbapenems, fluoroquinolones and aminoglycosides).

Proportion differences for the susceptible and MDR *E. coli* within the isolates from the different EU/EEA countries can be distinguished. Adapted from reference 81.

Few treatment options remain against MDR *Escherichia coli* infections including the use for instance of colistin and tigecycline[50,51]. Nevertheless resistance to both antibiotics of some strains is known[86,87].

Therefore, it is important to understand how bacteria are able to resist almost all antibacterial molecules currently available. Beyond the clinical importance of *Escherichia coli,* it also represents a great model organism used in most studies conducted in laboratories.

1.7. Tackling antibiotic resistance

Beyond the behavior change of our usage of antibiotics in general, new alternatives are being developed in order to restore our capacity to treat bacterial infections.

First new antibiotics are being made in order to gain more time in the battle against MDR bacteria. For instance, a new platform was recently developed for the synthesis of new macrolide drugs. Using the convergent assembly of chemical building blocks, 300 new macrolide antibiotics were synthesized amongst which the approved drug telithromycin and the clinical candidate solithromycin[88].

Therapy using bacteriophages has also regained interest because these agents are able to target specifically a given strain of a bacterium and decay during therapy once the given bacterium is dead. However, phages are not assumed to replace entirely antibacterial drugs but could be used more in a synergistic manner together with antibiotics to combat MDR[89].

Antimicrobial peptides, which are short and generally positively charged peptides found in various forms of life are also being studied. Natural and synthetic variants of antimicrobial peptides have the ability to either directly disrupt bacterial cell membrane or reach to intracellular targets to kill specifically pathogenic bacteria[90].

Pathoblockers and antivirulence agents, also constitute new forms of therapies that focus on disabling the virulence capacities of a bacterial pathogen rather than killing it directly. It is believed that such therapies can diminish damage to the host and permit the clearance of the pathogen by the immune system. Moreover, as selection pressures are believed to be low in these cases, resistance should not appear as quickly as for antibiotics[91].

Finally, vaccination against MDR pathogenic bacteria is also investigated. For instance, Th17 (T helper 17 cell) mediated adaptive immunity confers broad protection against MDR pathogens. Potential immunogenic agents include highly conserved outer membrane proteins and virulence factors. Vaccination could have a global positive effect on antibiotic resistance as reduction of the number of infections would also decrease the amount of antibiotics used to treat such infections[92].

2. Antibiotic resistance at the bacterial cell level

In this section, the different resistance acquisition pathways will be mentioned as well as the various resistance mechanisms conferring bacteria the ability to survive despite the presence of antibiotics.

2.1. Acquisition of resistance

Bacteria can become resistant to antibiotics through four main pathways. *De novo* mutations giving the ability to resist to antibacterial drugs can occur in the genome and be passed to the following generations through vertical transmission. Moreover, resistance acquisition can also occur through horizontal transmission either by conjugation (*i.e.* transmission of genes from another cell through the pilus), transformation (*i.e.* acquisition of genetic material from dead cells present within the environment), or transduction (*i.e.* transmission mediated by bacteriophages infecting cells) (Figure 8)[93,94,10].

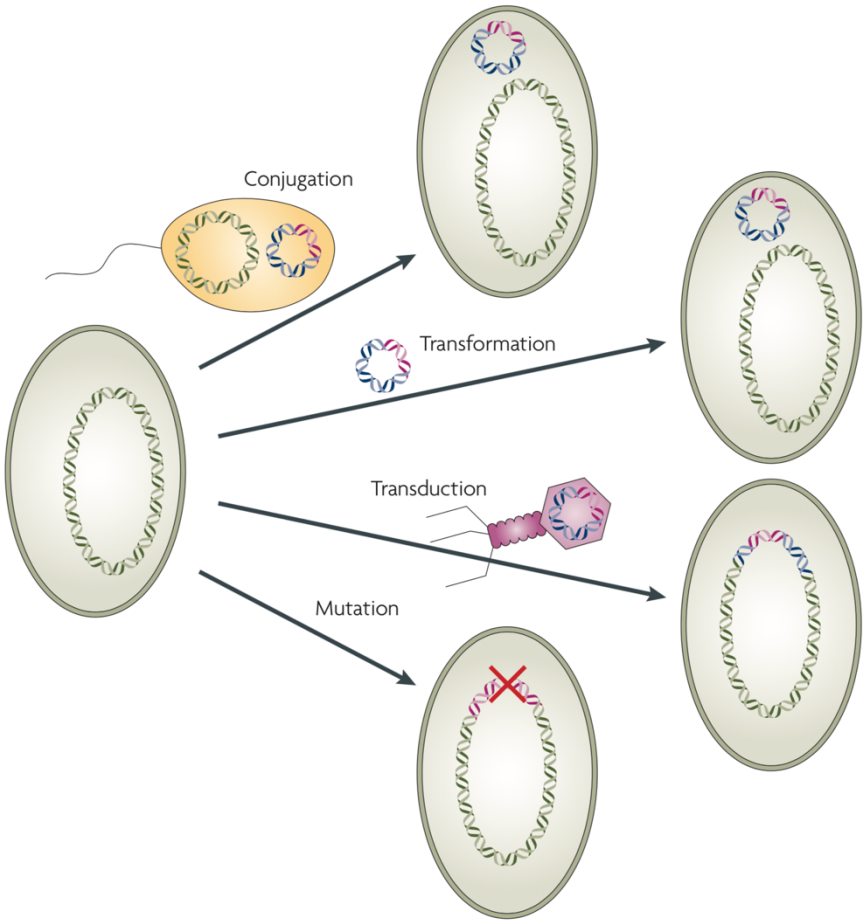

Figure 8. Resistance acquisition pathways.

The gene conferring resistance is colored in pink within the blue colored plasmid. The *de novo* mutation is marked as a red cross. Adapted from reference 93.

2.2. Mechanisms of resistance

Bacteria have developed numerous mechanisms to resist to antibiotics. The main clinically relevant mechanisms are well described in the literature and include, (I) decreased uptake of antibiotics through reduced membrane permeability or/and active efflux, (II) alteration of the antibiotic target, (III) enzymatic modification or inactivation of the antibiotic, (IV) bypass of pathways targeted by antibiotics and (V) overproduction of the antibiotic target (Figure 9)[95–100].

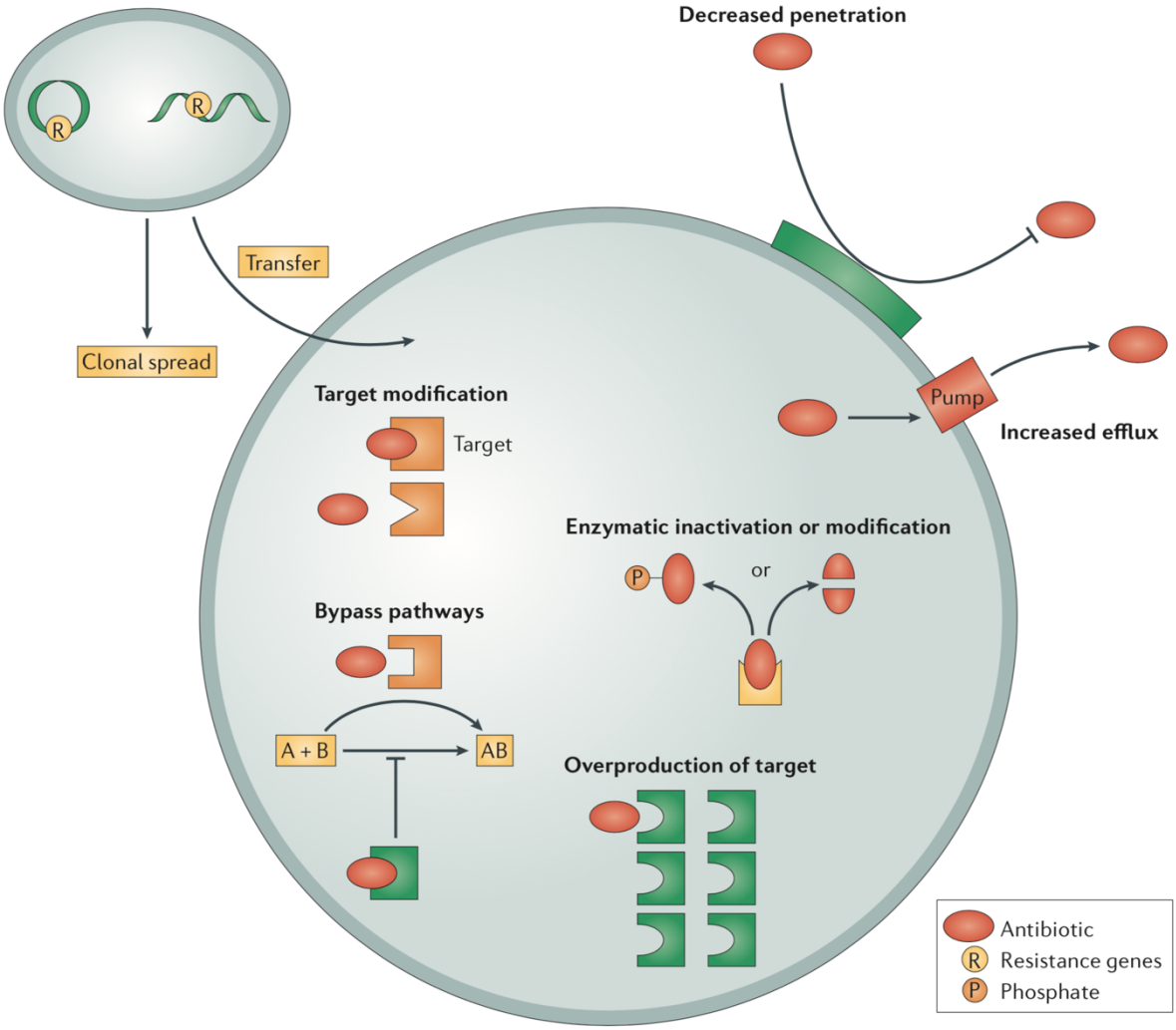

Figure 9. Mechanisms of resistance to antibiotics.

The five main clinically relevant mechanisms developed by bacteria are shown. Adapted from reference 95.

Amongst all these mechanisms, multi-drug efflux pumps contribute substantially to antibiotic resistance with a wide range of substrate polyspecificity (*i.e.* ability to transport structurally different

molecules). Thus, it is important to study their structure/function in order to elaborate specific pump inhibition strategies and restore the clinical efficacy of antibiotics.

Henceforth, given the focus of the present Ph.D. project description context, only efflux mediated antibiotic resistance will be mentioned. Moreover, special attention will be given to gram-negative bacterial multi-drug efflux systems.

Currently 7 families of multi-drug efflux pumps are known comprising: (I) the ATP-binding cassette (ABC) superfamily[101], (II) the major facilitator superfamily (MFS)[102], (III) the drug/metabolite transporter (DMT) superfamily (containing the small multi-drug resistance (SMR) family)[103], (IV) the multi-drug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily (containing the multi-drug and toxic compound extrusion (MATE) family[104], (V) the resistance-nodulation-cell division (RND) superfamily[105], (VI) the antimetabolite transporters (AbgT family)[106], and (VII) the proteobacterial antimicrobial compound efflux transporters (PACE family)[107]. Tripartite efflux systems in gram-negative bacteria, comprised of an inner membrane transporter, a periplasmic adaptor and an outer membrane exit duct are found in the ABC, MFS and RND families (Figure 10)[108].

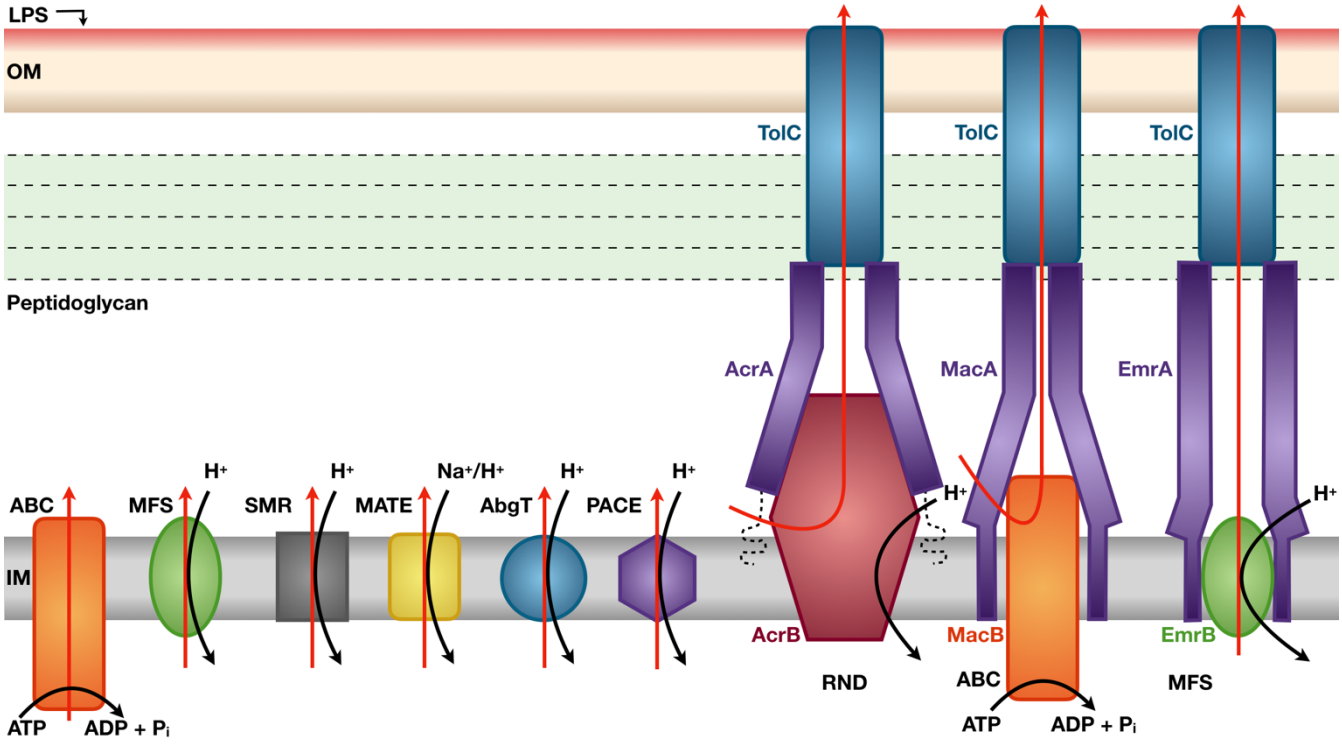

Figure 10. Different families of efflux pumps.

Schematic showing the diversity of efflux pump families comprising both single component members as well as tripartite system forming representatives. The red arrows indicate the efflux pathway of drugs. LPS, lipopolysaccharide (colored in red); OM, outer membrane (colored in light pink); IM, inner membrane (colored in light grey).

3. The major facilitator superfamily (MFS)

The present section aims to give a global overview about the major facilitator superfamily which includes both single component members as well as tripartite efflux system forming representatives.

The major facilitator superfamily is a very large, diverse and old family of transporters including uniporters, symporters and antiporters. Uniporters transport a single substrate downhill of its concentration gradient. Symporters (cotransporters) and antiporters (exchangers) use the energy from the concentration gradient of one substrate to transport a second substrate either in the same direction or in opposite directions respectively. Moreover, in the case of antiporters the transport of one substrate is dependent on the prior release of the other (Figure 11)[109,110].

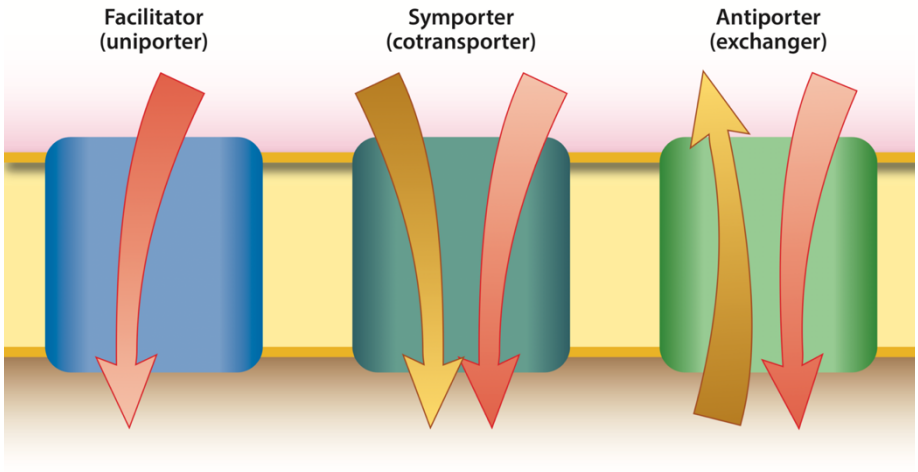

Figure 11. Representation of the transport modes of members belonging to the MFS.

The electrochemical gradients of the substrates are indicated by the respective color gradients of the arrows. Adapted from reference 109.

Transporters belonging to this superfamily are present in all kingdoms of life. Comprising currently 103 subfamilies of transporters, the MFS is actually the largest superfamily of secondary carriers[111].

Substrates transported by the different MFS representatives are diverse and include amongst others: antibiotics, monosaccharides, oligosaccharides, amino acids, peptides, vitamins, enzyme

cofactors, chromophores, nucleobases, nucleosides, nucleotides, iron chelates, and inorganic/organic ions[112–115].

The different members are usually composed of 400-600 amino acids. In most cases the transporters present a general topology with 12 transmembrane α-helix spanners (TMSs) connected by hydrophilic loops and with both the N- and C-termini located in the cytoplasm[111,112,116,117]. However, there are also exceptions to the general rule with some families containing members with 6, 13, 14, 15, 16, 18, and 24 TMSs[111,102,118]. From an evolutionary context, it is considered that the 12 TMSs arose from two subsequent duplication events of a 3 TMS repeat unit, with a 2 TMS containing protein being the precursor of the fundamental 3 TMS repeat unit. Moreover, the additional 2 TMSs present in the 14 TMS containing representatives and usually inserted in between the two 6 TMS helix bundles possibly arose from an additional duplication of an adjacent 2 TMS containing hairpin element[119].

The most extensively studied bacterial member of the MFS is the lactose:H^+ symporter LacY from *E. coli*, serving as a general model to understand the transport mechanism of the MFS members. Over the years numerous X-ray crystal structures have been obtained for many MFS representatives (examples of bacterial MFS members with known X-ray crystal structures are listed in Table 1)[109,110].

Table 1. Crystal structures of bacterial MFS members.

Adapted from references 109 and 110.

Transporter	Function	TCDB	Organism	Conformation (PDB ID)	Resolution limit (Å)	Year of first structure
$GlcP_{Se}$	Glucose:H^+ symporter	2.A.1.1	*Staphylococcus epidermidis*	Inward open (4LDS)	3.2	2013
XylE	Xylose:H^+ symporter	2.A.1.1	*Escherichia coli*	Outward-facing occluded (4GBY, 4GBZ, 4GC0, 6N3I) Inward-facing partially occluded (4JA3) Inward open (4JA4, 4QIQ)	2.6	2012

Transporter	Function	TCDB	Organism	Conformation (PDB ID)	Resolution limit (Å)	Year of first structure
EmrD	drug:H^+ antiporter	2.A.1.2	*Escherichia coli*	Inward-facing occluded (2GFP)	3.5	2006
MdfA	drug:H^+ antiporter	2.A.1.2	*Escherichia coli*	Inward open (4ZOW, 4ZP0, 4ZP2, 6EUQ, 6OOM, 6OOP, 6OOQ) Outward open (6GV1)	2	2015
YajR	Unknown	2.A.1.2	*Escherichia coli*	Outward open (3WDO)	3.15	2013
GlpT	Glycerol-3-phosphate antiporter	2.A.1.4	*Escherichia coli*	Inward open (1PW4)	3.3	2003
LacY	Lactose:H^+ symporter	2.A.1.5	*Escherichia coli*	Inward open (1PV6, 1PV7, 2CFP, 2CFQ, 2V8N, 2Y5Y) Outward-facing partially occluded (4OAA, 4ZYR, 5GXB, 6C9W)	2.95	2003
FucP	Fucose:H^+ symporter	2.A.1.7	*Escherichia coli*	Outward open (3O7P, 3O7Q)	3.14	2010
NarK	Nitrate/nitrite antiporter	2.A.1.8	*Escherichia coli*	Inward open (4JR9, 4JRE, 4U4T, 4U4V) Inward-facing occluded (4U4W)	2.35	2013
NarU	Nitrate or nitrite sym- or antiporter	2.A.1.8	*Escherichia coli*	Inward-facing occluded, partially inward open (4IU8, 4IU9)	3.01	2013
MelB	Melibiose/Na^+ or Li^+ symporter	2.A.2	*Salmonella typhimurium*	Outward open and outward-facing partially occluded (4M64)	3.35	2014

Transporter	Function	TCDB	Organism	Conformation (PDB ID)	Resolution limit (Å)	Year of first structure
GkPOT	Peptide:H^+ symporter	2.A.17	*Geobacillus kaustophilus*	Inward open (4IKV, 4IKW, 4IKX, 4IKY, 4IKZ)	1.9	2013
$PepT_{So}$	Peptide:H^+ symporter	2.A.17	*Shewanella oneidensis*	Inward-facing occluded (2XUT) Inward open (4UVM)	3	2011
$PepT_{So2}$	Peptide:H^+ symporter	2.A.17	*Shewanella oneidensis*	Inward open (4LEP, 4TPG, 4TPH, 4TPJ, 6JKC)	3.16	2013
$PepT_{St}$	Peptide:H^+ symporter	2.A.17	*Streptococcus thermophilus*	Inward open (4APS, 4D2B, 4D2C, 4D2D, 4XNI, 4XNJ, 5MMT, 5OXM, 5OXN, 5OXO, 6GHJ) Inward-facing partially occluded (5OXK, 5OXL)	1.9	2010
YbgH	Peptide:H^+ symporter	2.A.17	*Escherichia coli*	Inward open (4Q65)	3.4	2014
YePEPT	Peptide:H^+ symporter	2.A.17	*Yersinia enterolitica*	Inward open (4W6V)	3.02	2015
BbFPN	Divalent metal ion transporter	2.A.100	*Bdellovibrio bacteriovorus*	Inward open (5AYO, 6BTX) Outward open (5AYM, 5AYN)	2.2	2015

Within the MFS, multi-drug transporters are grouped into different subfamilies (drug:H^+ antiporter (DHA1/2/3) subfamilies) mostly based on the number of TMSs. DHA1 and DHA3 members contain 12 TMSs whereas DHA2 subfamily members possess 14 TMSs[102,118,111]. DHA1 and DHA2 subfamily representatives are present both amongst prokaryotes and eukaryotes. DHA3 subfamily

members are only found within prokaryotes. Representatives of the DHA1 subfamily present a large polyspecificity transporting various substrates including amongst others: antibiotics, sugars, polyamines, uncouplers, monoamines, acetylcholine, paraquat and methyl glyoxal. DHA2 and DHA3 subfamily members having a narrower substrate specificity, are known to transport antibiotics amongst other substrates[120,121]. Examples of well-known bacterial MFS-type multi-drug transporters in the literature are listed in Table 2[122,111].

Table 2. Bacterial MFS-type multi-drug transporters.

Type	Transporter (system)	Family	Organism	Examples of substrates	Reference
Single component	Bcr	DHA1	*Escherichia coli*	BCM, FOF, KAN, L-CYS, TET	123,124
Single component	Dep	*nd*	*Escherichia coli*	DHCP	125
Single component	EmrD	DHA1	*Escherichia coli*	ARA, CCCP, TCS	126–128
Single component	FloR	DHA1	*Escherichia coli*	FLO	129,130
Single component	MdfA/Cmr/CmlA	DHA1	*Escherichia coli*	ACO, AG, ARA, BAC, CHL, DAU, EB, ERY, FQ, IPTG, PUR, R6G, RIF, TET, TPP	131,128,132–134
Single component	MdtG/YceE	DHA1	*Escherichia coli*	DOC, FOF	135,136
Single component	MdtM	DHA1	*Escherichia coli*	CHL, CHO, DOC, EB	137–140
Single component	Mef	*nd*	*Escherichia coli*	ML	141
Single component	QepA/QepA2	DHA2	*Escherichia coli*	FQ	142
Single component	TetA	DHA1	*Escherichia coli*	TET	143

Type	Transporter (system)	Family	Organism	Examples of substrates	Reference
Multi-component	EmrB (EmrAB-TolC)	DHA2	*Escherichia coli*	CA, CCCP, CHH, DOC, EST, MV, NAL, PMA, PRG, R6G, SDS, TCA, TCS, TLM, TRX	143–147,135,148
Multi-component	EmrY (EmrKY-TolC)	DHA2	*Escherichia coli*	DOC, HP, MIT, NAL	127,149–152
Single component	NorA	DHA1	*Staphylococcus aureus*	ACO, BAC, CET, CHL, CIP, EB, ENX, FLO, NOR, OFX, PUR, R6G, TPP	153–156
Single component	QacA	DHA2	*Staphylococcus aureus*	ACL, ACR, ACY, BAC, CTA, CV, DAP, DAPI, DAZ, DBP, EB, HED, PAD, PPD, PRO, PTD, PY, QR, R6G, SO, STD, TMA-DPH, TPA, TPP	157–160
Single component	LmrS	DHA2	*Staphylococcus aureus*	CHL, EB, ERY, FLO, FUA, GAT, KAN, LIN, LZD, OXY, SDS, STR, TMP, TPP	161
Multi-component	KpnH (KpnGH)	*nd*	*Klebsiella pneumoniae*	ACR, AZI, BAC, CAZ, CHX, CIP, DOC, EB, ERT, ERY, GEN, IMI, NOR, PIP, PMB, SDS, SPE, STR, TIC, TOB, TRI	162

Type	Transporter (system)	Family	Organism	Examples of substrates	Reference
Single component	CraA	DHA1	*Acinetobacter baumannii*	CHL	163
Single component	TetA	DHA1	*Acinetobacter baumannii*	TET	164
Single component	AmvA	DHA2	*Acinetobacter baumannii*	ACO, ACR, BAC, CHX, DAPI, DOC, EB, ERY, MV, NOR, NOV, SDS, TPP	165
Single component	Cml	DHA1	*Pseudomonas aeruginosa*	CHL	166
Single component	TetA	*nd*	*Pseudomonas aeruginosa*	TET	166
Single component	EmeA	*nd*	*Enterococcus faecalis*	ACR, CIP, CLI, EB, ERY, FQ, NOR, NOV	167
Single component	CmlB	*nd*	*Enterobacter aerogenes*	CHL	168
Single component	Bmr	DHA1	*Bacillus subtilis*	ACD, EB, DOR, FQ, R6G, TPP	169,170
Single component	LmrP	DHA1	*Lactococcus lactis*	DAU, EB, TPP	171
Single component	MdfA	*nd*	*Salmonella enterica*	CHL, DOR, NOR, TET	172
Multi-component	EmrB (EmrAB)	*nd*	*Salmonella enterica*	NAL, NOV, R6G, TRI	172,173
Single component	EmrD-3	DHA1	*Vibrio cholerae*	CHL, EB, ERY, LZD, MIN, R6G, RIF, TMP, TPP	174
Multi-component	VceB (VceCAB)	DHA2	*Vibrio cholerae*	CCCP, DOC, NAL, PCP, PMA	175,176
Single component	SmtcrA	*nd*	*Stenotrophomonas maltophilia*	TET	177

Type	Transporter (system)	Family	Organism	Examples of substrates	Reference
Multi-component	EmrB (EmrCAB)	*nd*	*Stenotrophomonas maltophilia*	CCCP, ERY, NAL, TCS	178
Multi-component	EmrB (EmrAB)	*nd*	*Sinorhizobium meliloti*	-	179
Multi-component	EmrB (EmrAB-TolC)	*nd*	*Erwinia chrysanthemi*	ACR, BER, CAR, CCCP, CHL, LOA, NAR, NOR, NOV, OLA, OXA, PRT, QUE, R6G, TET	180,181
Multi-component	FarB (FarAB-MtrE)	DHA2	*Neisseria gonorrhoeae*	OLA, LOA, PA	182
Single component	MefE	DHA3	*Streptococcus pneumoniae*	ML	183
Multi-component	HmrAB	*nd*	*Haemophilus influenzae*	ERY, NAL	184

nd not described.

ACD acridine dyes, ACL amicarbalide, ACO acridine orange, ACR acriflavine, ACY acridine yellow, AG aminoglycosides, ARA arabinose, AZI azithromycin, BAC benzalkonium chloride, BCM bicyclomycin, BER berberine, CA cholic acid, CAR carbenicillin, CAZ ceftazidime, CCCP carbonyl cyanide *m*-chlorophenyl hydrazone, CET cetrimide, CHH 2-chlorophenylhydrazine hydrochloride, CHL chloramphenicol, CHO cholate, CHX chlorhexidine, CIP ciprofloxacin, CLI clindamycin, CTA cetyltrimethylammonium, CV crystal violet, DAP diamidinodiphenylamine, DAPI 4',6-diamidino-2-phenylindole, DAU daunomycin, DAZ diminazene, DBP dibromopropamidine, DHCP 4,5-dihydroxy-2-cyclopentan-1-one, DOC deoxycholate, DOR doxorubicin, EB ethidium bromide, ENX enoxacin, ERT ertapenem, ERY erythromycin, EST estradiol, FLO florfenicol, FOF fosfomycin, FQ fluoroquinolones, FUA fusidic acid, GAT gatifloxacin, GEN gentamicin, HED hexamidine, HP hydrogen peroxide, IMI imipenem, IPTG isopropyl-β-D-thiogalactopyranoside, KAN kanamycin, L-CYS L-cysteine, LIN lincomycin, LOA linoleic acid, LZD linezolid, MIN minocycline, MIT mitomycin, ML macrolides, MV methyl viologen (paraquat), NAL nalidixic acid, NAR naringenin, NOR norfloxacin, NOV novobiocin, OFX ofloxacin, OLA oleic acid, OXA oxacillin, OXY oxytetracycline, PA palmitic acid, PAD phenamidine, PCP pentachlorophenol, PIP piperacillin, PMA phenylmercuric acetate, PMB polymyxin-B, PPD propamidine, PRG progesterone, PRO proflavine, PRT protamine, PTD pentamidine, PUR puromycin, PY pyronin Y, QR quinaldine red, QUE quercetin, R6G rhodamine 6G, RIF rifampin, SDS sodium dodecyl sulphate, SO safranin O, SPE spectinomycin, STD stilbamidine, STR streptomycin, TCA taurocholic acid, TCS tetrachlorosalicylanilide, TET tetracycline, TIC ticarcillin, TLM thiolactomycin, TMA-DPH 1-(4-trimethylammoniumphenyl)-6-phenyl-1,3,5-hexatriene, TMP trimethoprim, TOB tobramycin, TPA tetraphenylarsonium, TPP tetraphenylphosphonium, TRI triclosan, TRX triton X-100.

3.1. Single component MFS members

3.1.1. Structural insights of single component MFS members

3.1.1.1. LacY

The LacY transporter from *E. coli* is the most thoroughly studied bacterial MFS member. The protein is encoded by the *lacY* gene located within the *lac* operon of *E. coli*. It is involved in the symport

of D-galactose and D-galactopyranosides together with the transfer of H^+ towards the cytoplasm. The coupling reaction occurs stoichiometrically (with the translocation of a D-galactopyranoside and an H^+). In fact, the symporter makes use of the energy from the electrochemical gradient of H^+. Therefore, the downhill transport of H^+ (in the direction of the concentration gradient) is coupled to the uphill transport of sugars (against their concentration gradient). In the absence of an electrochemical gradient of H^+, LacY catalyzes also the downhill transfer of sugars coupled to the uphill transfer of H^+ generating an electrochemical gradient of H^+ which's polarity depends on the direction of sugar translocation (Figure 12). However, in the absence of sugars LacY cannot translocate H^+. Moreover, LacY catalyzes exchange or counterflow of sugar without H^+ translocation. Therefore, it is believed that the primary driving force for the global symport reaction is linked to the binding and dissociation of sugar on either side of the membrane[185].

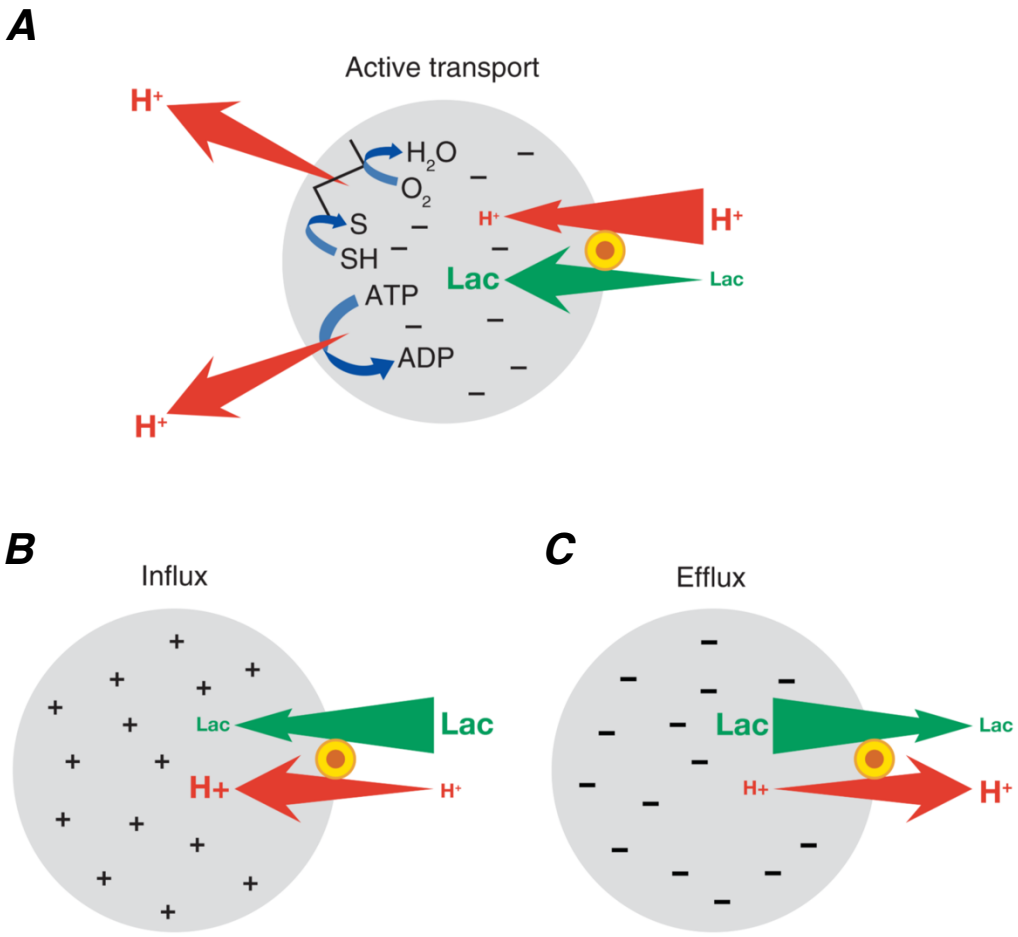

Figure 12. Transport modes of LacY.

(A) In the presence of lactose and an H^+ electrochemical gradient generated by the hydrolysis of ATP by the F_1F_o ATPase and the electron transfer chain, LacY co-transports lactose and H^+. (B, C) In the presence of lactose, electrochemical gradients of H^+ of either polarity depending on the sugar gradient can be generated. Adapted from reference 185.

The first crystal structures of LacY were obtained in 2003[186] and represent actually structures of a mutant LacY (C154G) stabilized and constrained in an inward open conformation. Numerous other crystal structures have been obtained since 2003: of the same mutant but with better resolutions in 2006[187], of the WT LacY in 2007[188] but with a similar resolution and of another mutant (A122C) in 2011[189] also with a better resolution. However LacY in all these structures retained an inward open conformation. Since 2011 other structures were obtained (in 2014, 2015, 2016 and 2018) in a new conformation (outward-facing partly occluded) with the help of a double mutant (G46W/G262W) and nanobodies (Table 1)[190–193].

The following description of the structural fold of LacY mainly focuses on the 2003 crystal structure solved with a high affinity homolog of lactose named TDG (β-D-galactopyranosyl-1-thio-β-D-galactopyranoside) at a resolution of 3.6 Å.

Within the asymmetric unit of the crystal, an artificial dimer composed of two molecules oriented in opposite directions was observed. Thus, LacY is likely to be functional as a monomer.

Globally, LacY contains 12 TMSs which are grouped in two six-helix bundles (helices I to VI and VII to XII) called the N- and C-terminal domains with a pseudo two-fold symmetry along an axis perpendicular to the plane of the membrane (Figure 13, A and B). The two six-helix bundles are connected via a long and flexible cytoplasmic loop between helices VI and VII (Figure 13C).

Each six-helix bundle can further be divided in two three-helix bundles with pseudo symmetry. Helices III, VI, IX, and XII are completely embedded in the membrane and not exposed to the external medium. A large hydrophilic cavity open on the cytoplasmic side is visible in the global structure, formed in between helices I, II, IV and V of the N-terminal domain and helices VII, VIII, X and XI of the C-terminal domain. The overall dimensions of the hydrophilic cavity are of 25 Å by 15 Å.

The TDG molecule is bound at the apex of the hydrophilic cavity at an equivalent distance from both sides of the membrane. Therefore, the substrate binding site within the hydrophilic cavity in LacY, is located at the center of the protein in between the N- and C-terminal domains.

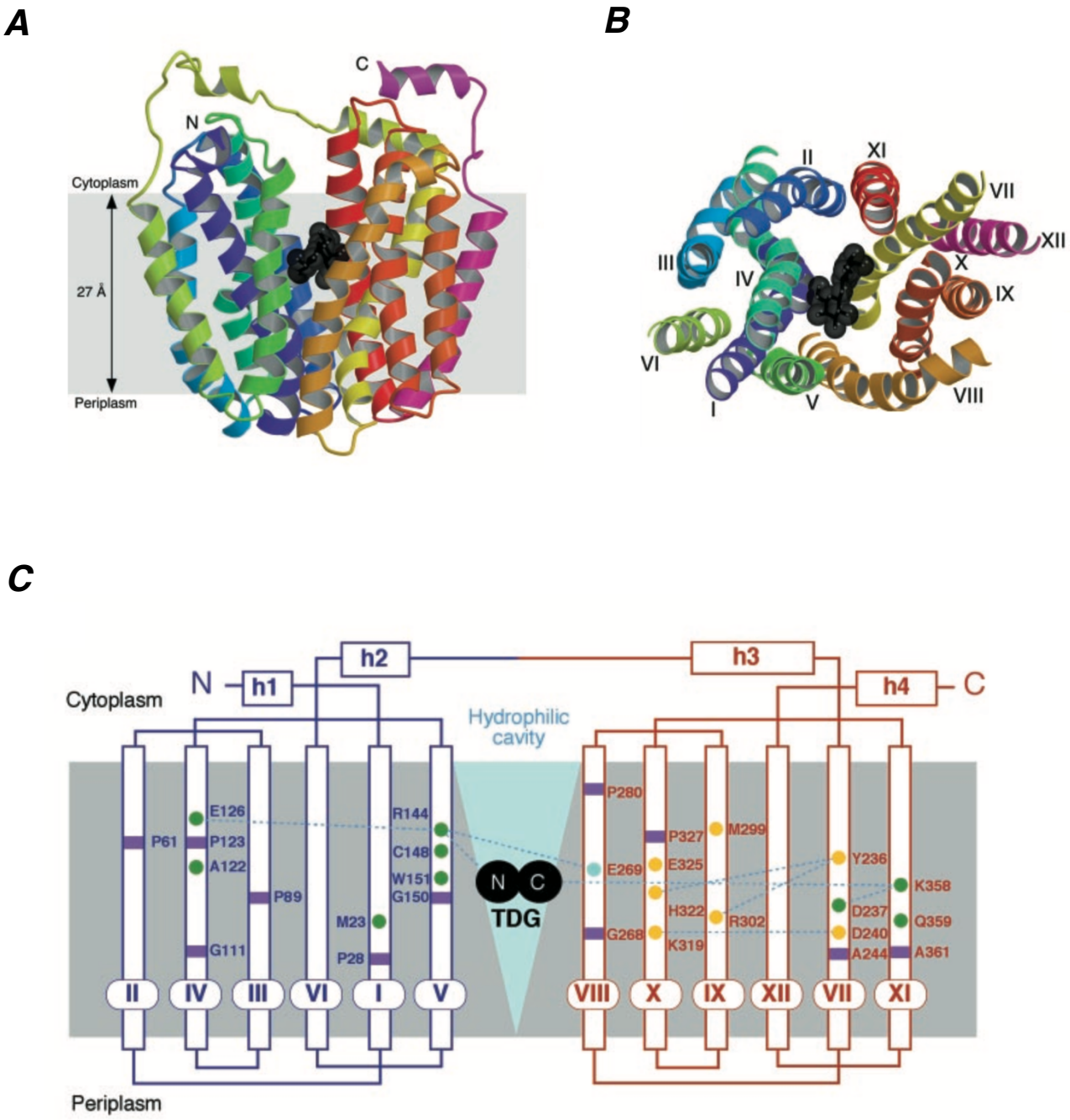

Figure 13. Global structure of LacY.

(A) Ribbon representation of the inward open structure of TDG bound LacY (C154G) viewed parallel to the membrane. All 12 TMSs and cytoplasmic helices are shown in different colors with the N- and C-terminal helices represented in purple and pink respectively. The TDG bound in the substrate binding site is shown as black spheres. (B) Cytoplasmic view of the global structure without the loop regions for more clarity. The TMSs are numbered from one to twelve in roman numbers. (C) Overall topology of LacY with the secondary structures. The N- and C-terminal domains are colored in blue and red respectively. The cytoplasmic helices are shown and numbered from h1 to h4. A light blue triangle indicates the position of the hydrophilic cavity with the TDG molecule marked as 2 black circles. Residues involved in substrate binding and proton translocation are marked with green and orange circles respectively. E269 important both for substrate binding and proton translocation is indicated as a light blue circle. Other residues present at the kinks of the helices are shown as purple rectangles. Adapted from reference 186.

A representation of the substrate binding site of LacY is shown in Figure 14. Half of the site is formed by residues from helices I, IV, and V in the N-terminal domain and the other half by residues from helices VII and XI in the C-terminal domain. The essential residue R144 interacts with TDG through hydrogen bonding with the O3 and O4 atoms of the galactopyranosyl ring. A second essential residue, E126 is likely to interact with the O4, O5 and O6 atoms of the galactopyranosyl ring through water molecules. Hydrophobic effects exist as well: between the indole ring of W151 and the galactopyranosyl ring but also between the M23 and the C6 of the galactopyranosyl ring. A third essential residue namely E269 in the C-terminal domain forming a salt bridge with R144 and a possible hydrogen bond with W151 constitutes an energetic link between the C- and N-terminal domains and might also interact with TDG. However, few other residues interact with the sugar in the C-terminal domain. K358 interacts with TDG through hydrogen bonding with the O4' atom of the galactopyranosyl ring and D237 might also interact with the O4' atom of the galactopyranosyl ring through a water molecule. These residues provide additional affinity for disaccharide molecules and therefore only play a supporting role compared to the N-terminal interaction site. Thus, substrate specificity is mainly provided by residues in the N-terminal domain.

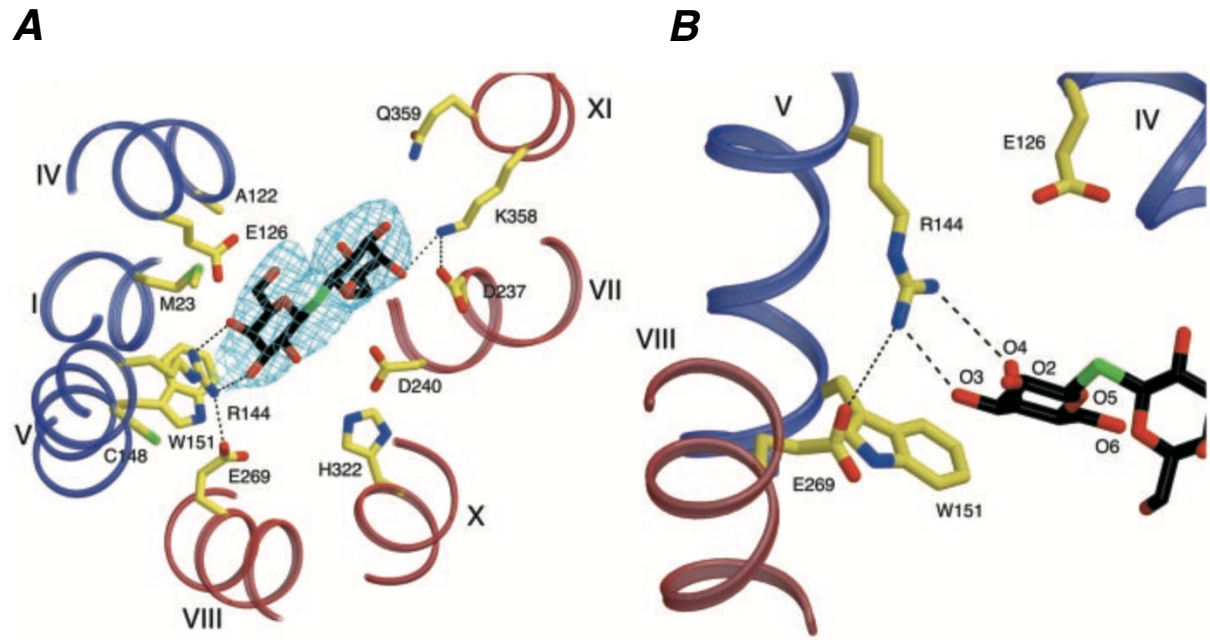

Figure 14. The substrate binding site of LacY.

TMSs of the N- and C-terminal domains are colored in blue and red respectively. The color code for the different atoms shown is: yellow for carbon of side chains, black for carbon of TDG, red for oxygen, blue for nitrogen and green for sulfur. Possible hydrogen bonds and salt bridges are shown as dashed lines. (A) Cytoplasmic view along the membrane normal of the substrate binding site with the electron density map indicated for TDG. (B) Enlarged view of the substrate binding site in the N-terminal domain. Adapted from reference 186.

Within LacY, residues involved in H^+ translocation form a complex salt-bridge and hydrogen bond network in the C-terminal domain (Figure 15). The network is composed of Y236, D240, K319, H322, E325 and R302. Mutational studies indicate that H322, E325 and R302 are directly involved in H^+ translocation[185]. E325 being located in a hydrophobic environment is considered to remain protonated in the substrate bound inward facing conformation. It is suggested that H322 could possibly be the H^+ donor to E325. The D240 K319 salt bridge might not be directly involved in the H^+ translocation pathway as mutations of these residues do not impair transport activity[185]. As the distance between substrate binding site and the salt-bridge and hydrogen network is about 6 Å, there is no direct interaction between both sites. The only residue of the C-terminal domain interacting with the N-terminal substrate binding site is E269. Therefore, E269 might contact the H322 (located at a distance of 5.8 Å) with an interaction involving a water molecule. Hence, E269 is critical for coupling of sugar and H^+ translocation.

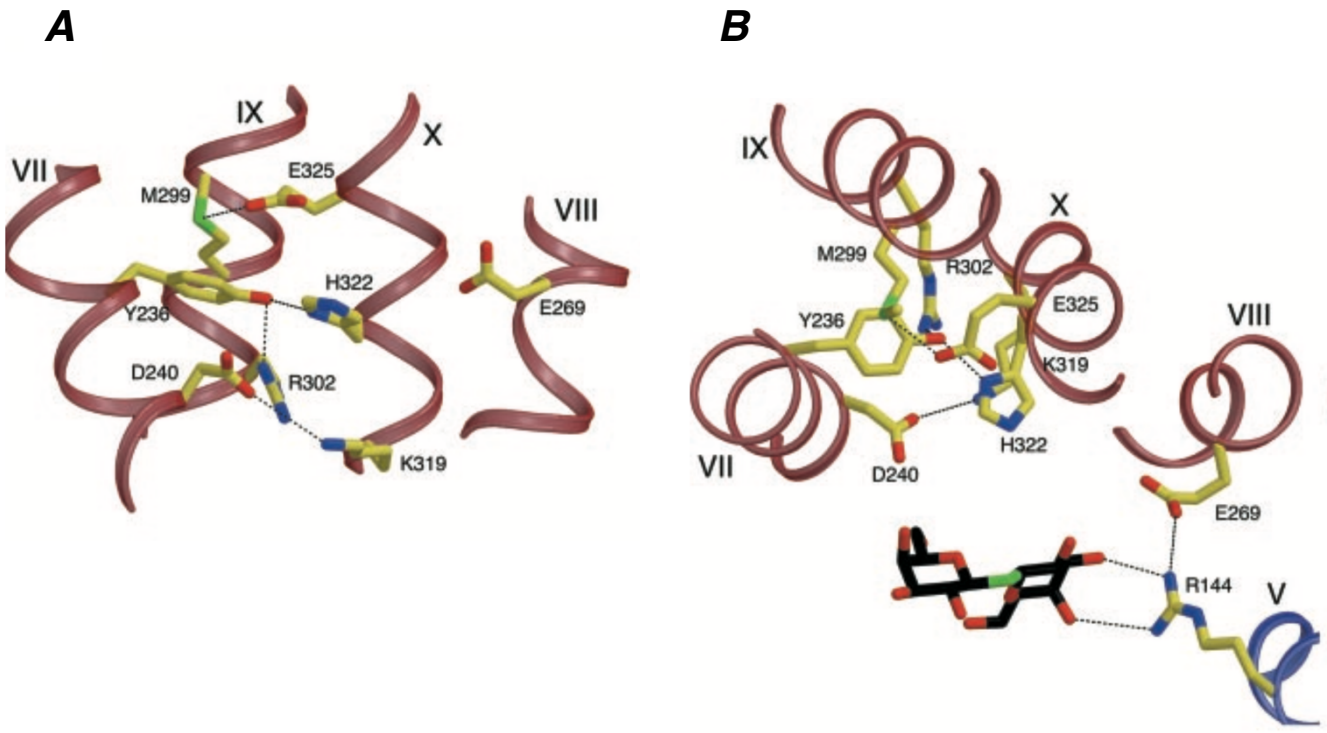

Figure 15. Organization of critical residues involved H^+ translocation and coupling.

TMSs of the N- and C-terminal domains are colored in blue and red respectively. The color code for the atoms is identical to the color code used in Figure 14. Hydrogen bonds are indicated as dashed lines. (A) View of the salt-bridge and hydrogen bond network parallel to the membrane. (B) Cytoplasmic view along the membrane normal of the different residues involved in H^+ translocation and coupling. Adapted from reference 186.

3.1.1.2. MdfA

MdfA from *E. coli* is one of the most studied bacterial multi-drug MFS transporters. This antiporter is involved in the translocation of a large variety of substrates including antibiotics (Table 2) together with the downhill translocation of H^+. The structural diversity of its substrates appears to coincide with the observed large hydrophobic pocket in the MdfA structure[194].

Sequence alignments of different MFS antiporters demonstrated the presence of four distinct motifs (named motif A-D). These motifs for different MdfA homologs are shown in Figure 16. Motif A has been suggested to be involved in the stabilization of antiporters in the outward-facing conformation. Motif B could possibly be linked to energy coupling and conformational change induced by substrate binding. Two acidic residues of MdfA homologs present in motif D are important for its function[195,194].

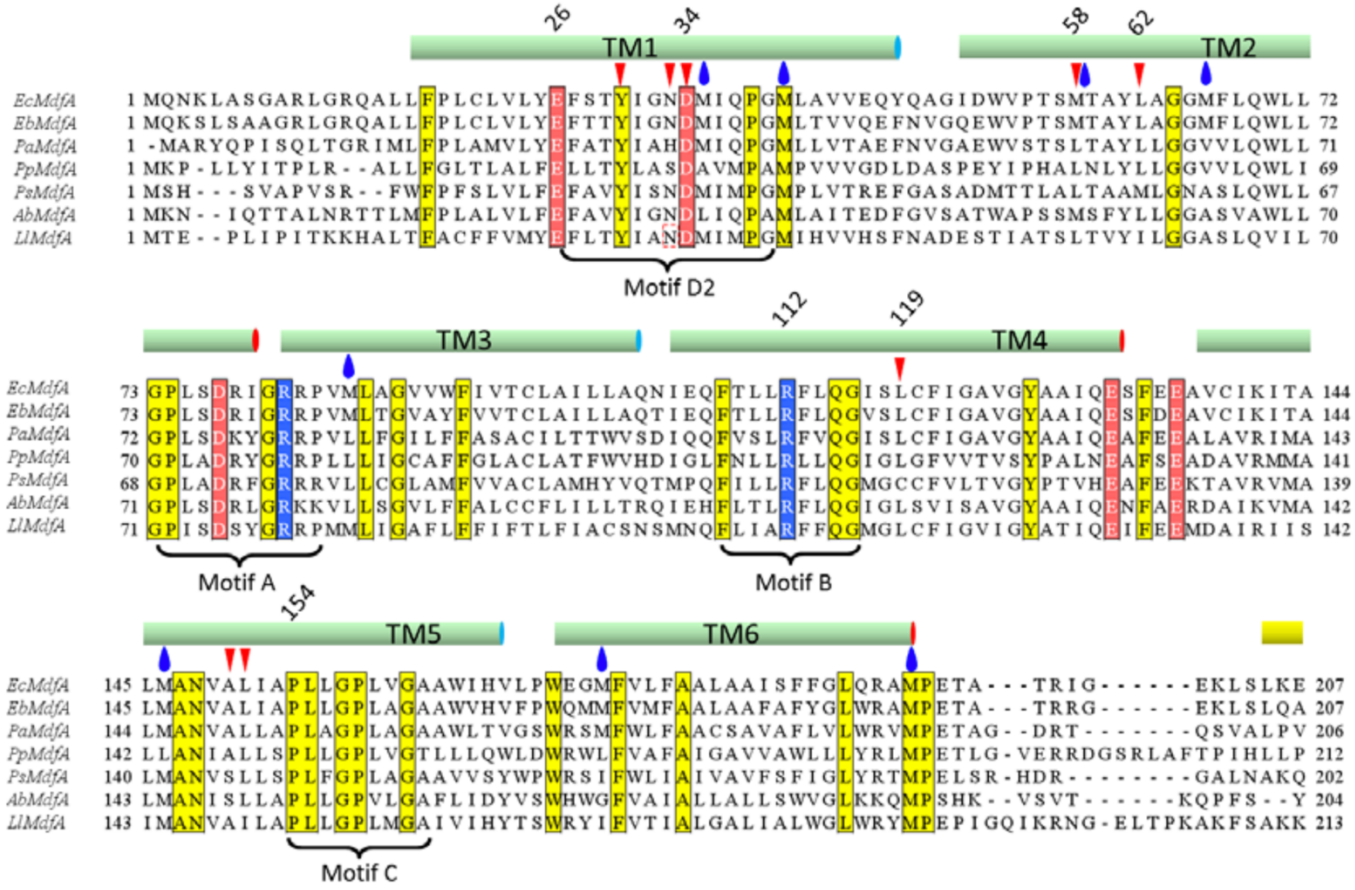

Figure 16. Four motifs identified within MdfA homologs.

Partial sequence alignment (N-terminal half) showing the four distinct motifs present within MdfA homologs. The different MdfA sequences used for the alignment are from *E. coli* (Ec) and six other pathogenic bacteria: *Enterobacteriaceae bacterium* (Eb), *Pantoea ananatis* (Pa), *Pseudomonas putida* (Pp), *Pandoraea sp.* (Ps), *Acinetobacter baumannii* (Ab), and *Legionella longbeachae* (Ll). Secondary structures of EcMdfA are marked above the alignment. Some of the residues involved in chloramphenicol binding are indicated by red triangles within the portion of the alignment. Blue drops show the positions of some of the Se-M residues used in initial phasing. Adapted from reference 194.

The first crystal structures of MdfA were obtained in 2015 at 2.0, 2.2 and 2.45 Å resolutions[194] and represent the structures of a mutant MdfA (Q131R) in an inward open conformation. Since then two additional structures were obtained in 2018. The first one represents a double mutant (MdfA (Q131R/L339)[196]) in the same conformation as the single mutant Q131R and with a comparable resolution. Whereas the second structure shows the wild-type MdfA in an outward open conformation and at a lower resolution (3.4 Å) obtained with the help of Fab fragments[197]. Finally, in 2019 three additional structures of a double mutant (I239/G354E) also in the same conformation as the single mutant Q131R, at 2.2, 2.8 and 3 Å resolutions were obtained[198] (Table 1).

The following structural description of MdfA mainly concerns the structure obtained in 2015 together with the antibiotic chloramphenicol at a resolution of 2.45 Å.

Overall, MdfA belonging to the same superfamily as LacY, it shares similar structural features. It is composed of 12 TMSs divided in two six-helix bundles, called the N- and C-terminal domains (Figure 17). Both regions are connected by a long cytoplasmic loop containing an amphipathic helix located between helices 6 and 7. The structure can also be described as a total of four three-helix groups with pseudo symmetry. Similarly, helices 3, 6, 9, and 12 are completely embedded in the membrane and are not exposed. However, in this case the central inward open cavity (~ 3 000 Å^3) formed in between the

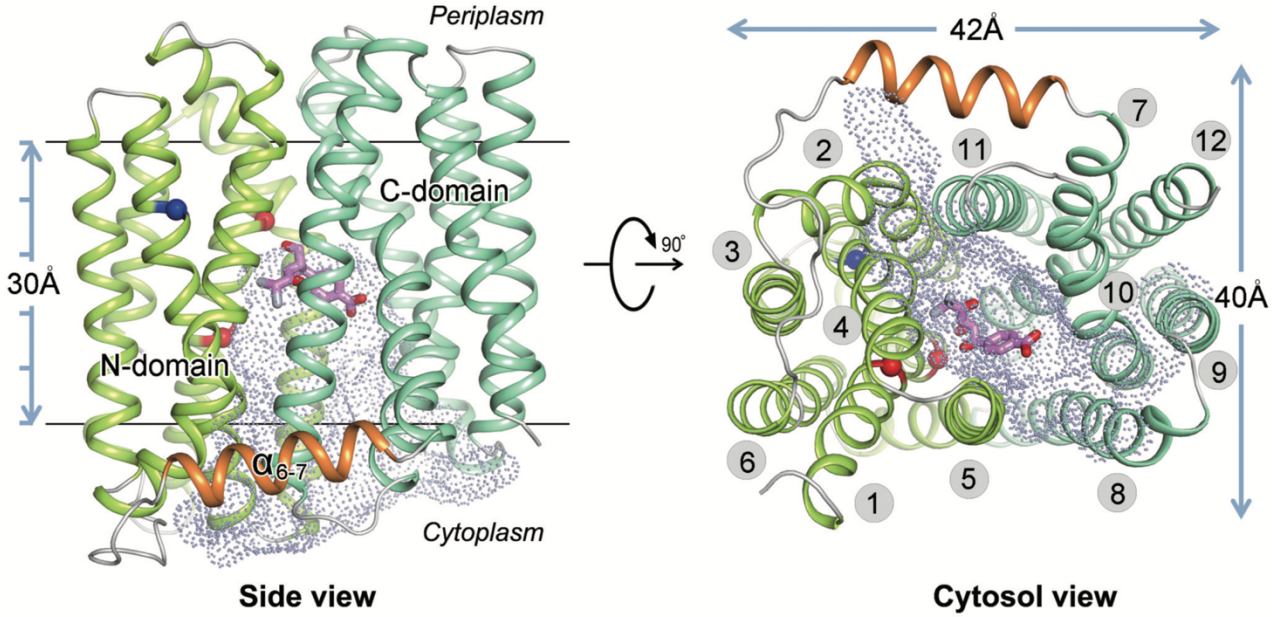

Figure 17. Global structure of MdfA.

The N- and C-terminal domains are colored in green and cyan respectively. The amphipathic helix within the cytoplasmic loop between helices 6 and 7 ($\alpha_{6\text{-}7}$) is colored in orange. The TMSs are numbered from 1 to 12 on the right panel. The central cavity is shown as a dot surface representation. Chloramphenicol located at the apex of the cavity is shown as purple sticks. Residues E26 and D34 are marked as red spheres. R112 is represented as a blue sphere. Adapted from reference 194.

Inside the cavity, two acidic residues from motif D (E26 and D34) are present (Figure 18A). These two residues constitute protonation sites and are important for the transport function of MdfA. Both residues are buried by mostly hydrophobic side chains. E26 is surrounded by Y30, I122, M146, V149, and A150. D34 (located at the apex of the cavity) is buried by the side chains of N33, M58, A153, P154, and I239.

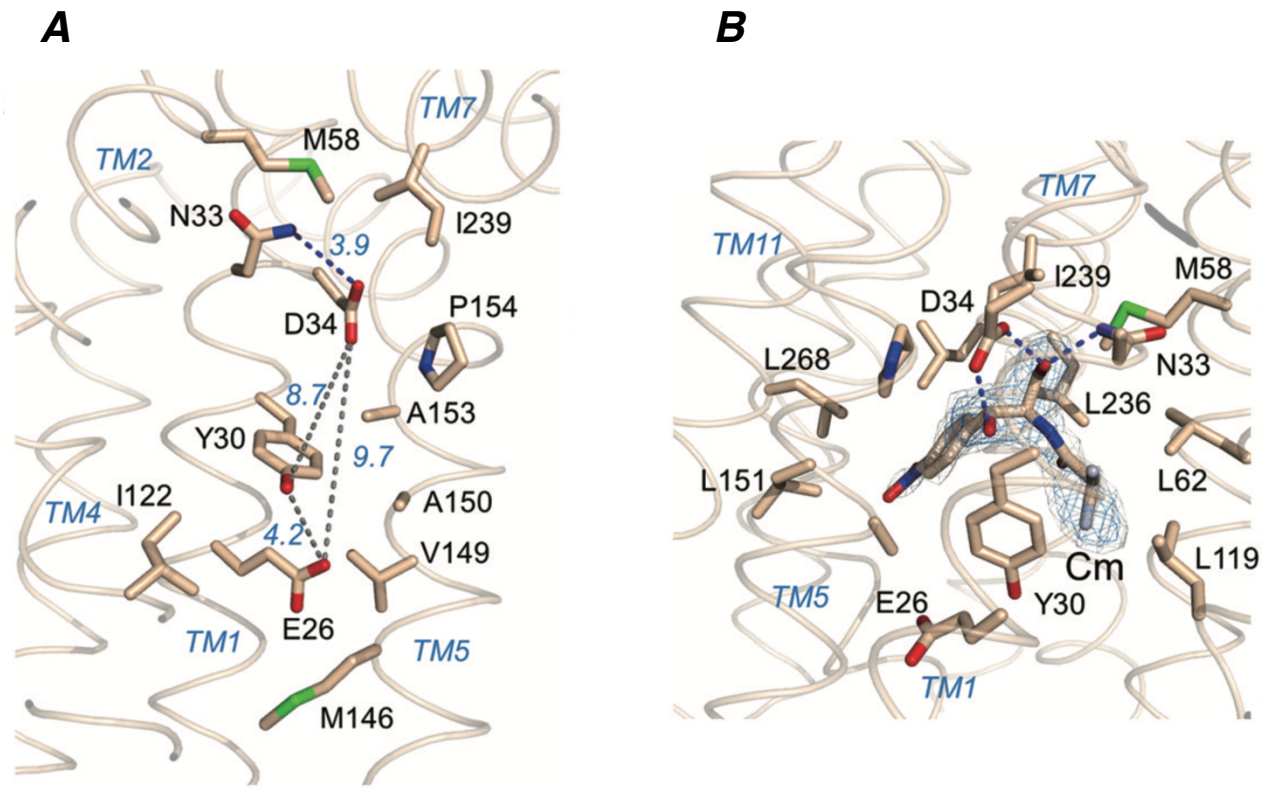

Figure 18. Protonation sites and important residues for substrate binding.

(A) The two protonation sites and their surrounding residues shown as sticks. The TMSs are shown as tubes. Distances between the atoms of different residues are marked as dashed lines and labeled. The substrate chloramphenicol (Cm) was removed for more clarity. (B) Representation of the residues involved in the binding of Cm. Cm is shown as wheat sticks with its electron density maps in blue and wheat. Hydrogen bonds are shown as dashed blue lines. This view is different from (A) by ~ 180°. Adapted from reference 194.

In total, twelve amino acid residues seem to be important for chloramphenicol binding. Besides A150 and L235, all of these residues are represented and labeled in Figure 18B. Most of these residues are hydrophobic. D34 and N33 interact with the oxygen atoms from the hydroxyl groups of chloramphenicol via hydrogen bonding. The location of the nitrile group of chloramphenicol at a solvent exposed site in the cavity might indicate why thiamphenicol is also recognized as a substrate.

Two other crystal structures obtained with deoxycholate (Dxc) and LDAO (n-dodecyl-N, N-dimethyl-amine-N-oxide), show similar interactions, mainly with the side chains of Y30, N33, D34, and L236. Moreover, as the distances between D34 and the negatively or positively charged groups seem to be different, it has been shown that MdfA recognizes structurally diverse substrates.

A hydrogen bond network located close to the hydrophobic cavity, is constituted by the essential residue from motif B (R112), as well as residues C96, Q115, G32, and a water molecule (Figure 19). Because of the close proximity of R112 to D34 (distance of 9 Å), it has been proposed that motif B (with the positive electrostatic field) has a more important effect on D34 than E26 (located at 16 Å). In addition, based on mutational studies it has been postulated that motif B and the surrounding residues couple the protonation status to substrate binding[194].

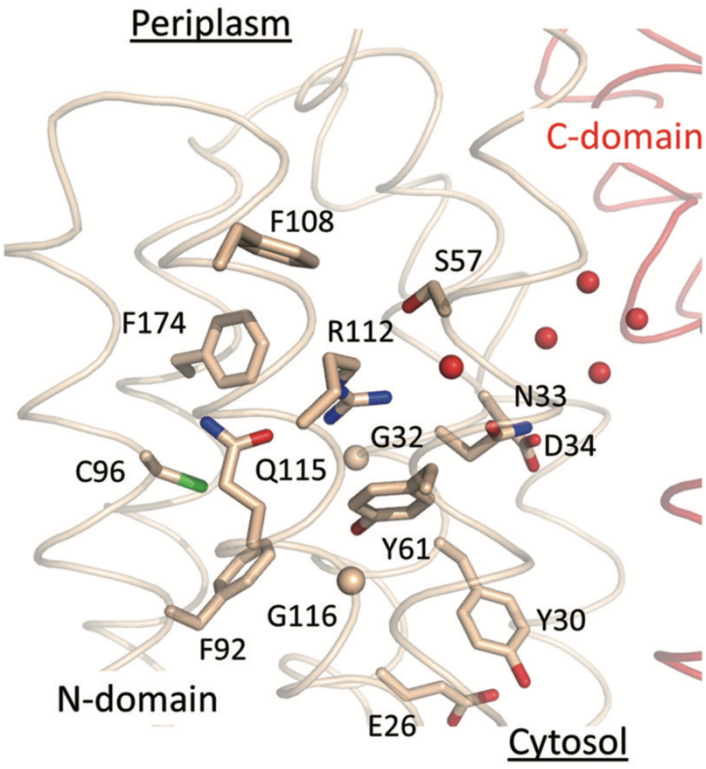

Figure 19. Site coupling protonation status to substrate binding.

Different residues involved in the proposed coupling process are shown. The helices from the N- and C-domain are represented as wheat and red colored tubes respectively. Glycine residues and water molecules are depicted as wheat and red spheres respectively. Adapted from reference 194.

More recently, it has been suggested that a cytoplasmic rim (constituted by residues Q131, E132, K346, R336, E136, and E135) plays an essential role in substrate recognition, translocation function and conformational change of the transporter (Figure 20)[196].

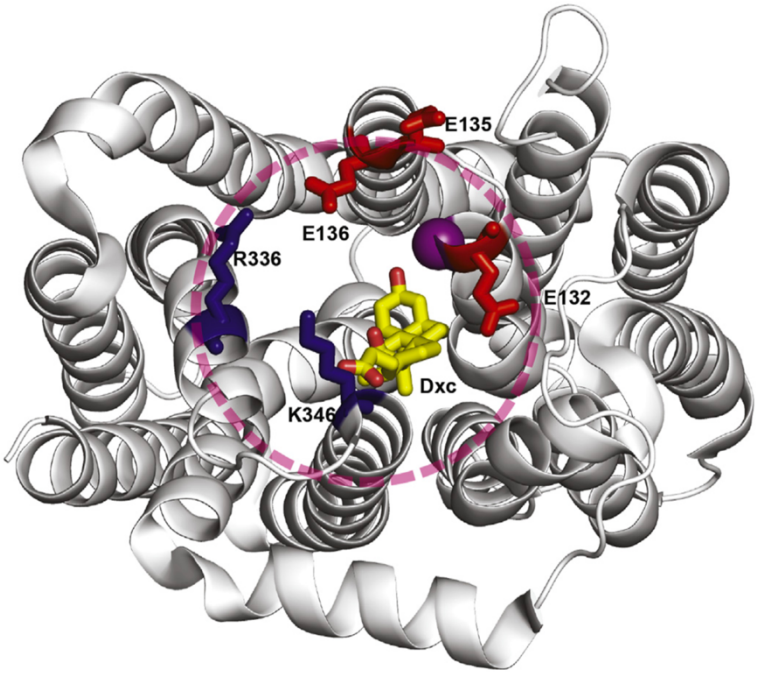

Figure 20. The cytoplasmic rim of MdfA.

The rim in the structure of MdfA (PDB ID: 4ZP0) viewed from the cytoplasmic side is depicted by a pink dashed circle. Negatively and positively charged residues are shown in red and blue respectively. Q131 is marked as a purple sphere. Dxc bound in the center of the cavity is shown as yellow sticks. Adapted from reference 196.

3.1.1.3. EmrD

EmrD from *E. coli* is homologous to the previously described transporter MdfA (with 26% of identity and 39% of similarity). This antiporter is involved in the translocation of structurally diverse substrates including for instance the uncouplers carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) and tetrachlorosalicylanilide (TCS) (Table 2). Similarly, EmrD couples the downhill translocation of H^+ to the uphill transfer of variable substrates[199].

One crystal structure of this antiporter in an inward-facing occluded conformation was obtained in 2006[199] at a resolution of 3.5 Å (Table 1). Since then no other crystal structures of EmrD have been reported thus far.

Within the asymmetric unit of the crystal, two molecules were found. However, it was suggested that the observed dimer was not physiologically relevant. Thus, EmrD is likely to be functional as a monomer.

Globally, the structural fold of EmrD is similar to the previously described structures of LacY and MdfA (Figure 21). Indeed, it is composed of the typical 12 TMSs divided into the N- and C-terminal domains connected via a cytoplasmic loop between helices 6 and 7. In addition, one lateral helix is present on the periplasmic side between helices 3 and 4 and two other lateral helices are located in the cytoplasmic loop region between helices 6 and 7. Moreover each six-helix bundle can further be subdivided in two pseudo symmetric three-helix groups. Likewise, helices 3, 6, 9 and 12 are not exposed to the external medium. Similar to MdfA, a central hydrophobic cavity formed by the remaining helices of the N- and C-terminal domains can be distinguished in this case.

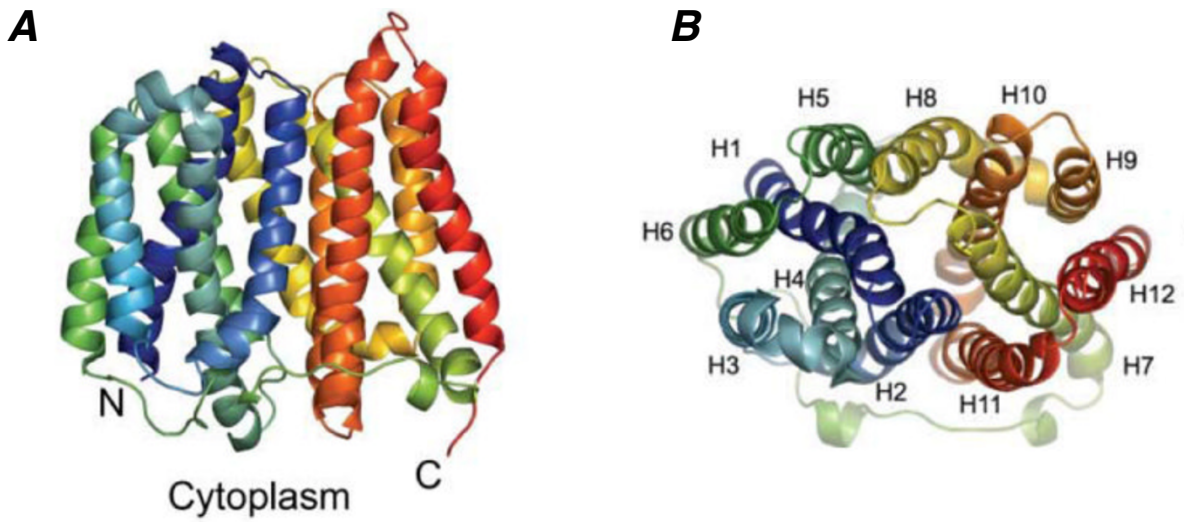

Figure 21. Global structural fold of EmrD.

(A) Structure of EmrD viewed from the side. The N- and C-termini are indicated. (B) Periplasmic view of the global structure. The 12 TMSs are numbered accordingly (H, helix). Adapted from reference 199.

Indeed, mostly hydrophobic residues are found inside the cavity. In particular, I28, I217, I253, Y52, Y56, W300, and F249 are suggested to be important for substrate translocation and might also contribute to the substrate specificity of EmrD. Amongst these residues Y52, Y56, W300 and F249 might interact with numerous aromatic drugs through stacking (Figure 22). T25, D33, and E227 located on the periplasmic side of the cavity are suggested to be involved in H^+ translocation.

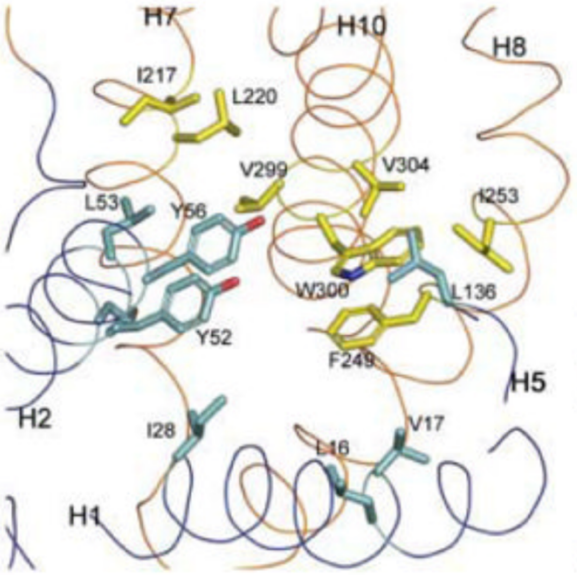

Figure 22. The hydrophobic cavity of EmrD.

Various hydrophobic residues present in the cavity are shown as sticks. The N- and C-terminal domains are colored in blue and orange respectively. Adapted from reference 199.

Two helical regions (from helix 4 to helix 5 and from helix 10 to helix 11) located towards the cavity and pointing in the cytoplasmic direction have been proposed to be important for substrate specificity. Moreover, V17 (at the cytoplasmic end of helix 1) with its additional accessibility from the inner membrane side might play an important role in substrate binding as well. Together these regions could possibly constitute the so-called 'selectivity filter' of EmrD (Figure 23).

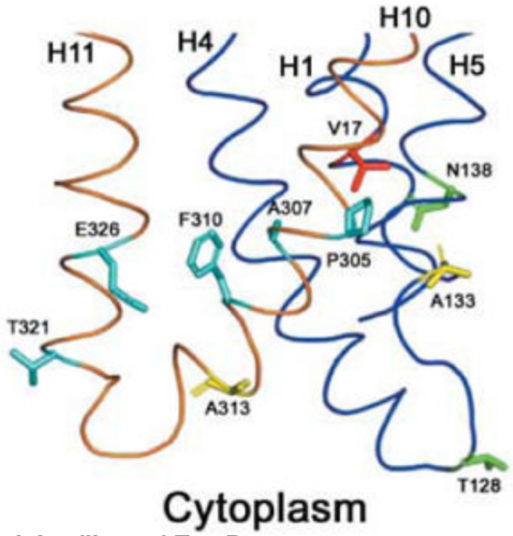

Figure 23. View of the selectivity filter of EmrD.

V17 from helix 1 is depicted as red sticks. The remaining residues shown were suggested to be involved in substrate recognition based on homology studies with other MFS drug:H+ transporters. Adapted from reference 199.

3.1.1.4. $PepT_{So}$

$PepT_{So}$ is a peptide transporter from *Shewanella oneidensis*. It is a symporter which couples the uptake of peptides to the electrochemical gradient of H^+.

The first crystal structure of $PepT_{So}$ was obtained in 2011[200], representing the symporter in an inward-facing occluded conformation. A second structure with higher resolution (3 Å) has been reported in 2015[201] and shows the inward open conformation of the symporter (Table 1).

The following description will mainly focus on the structure obtained in 2011 together with possibly a non-natural ligand or high-affinity inhibitor at 3.62 Å resolution. Three molecules with identical structures were found within the asymmetric unit of the crystal.

Globally, $PepT_{So}$ has a similar topology to the previously described MFS members (Figure 24). However, it is composed of 14 TMSs with two additional TMSs called A and B just in between the N- and C-terminal domains. Therefore, in the present case these two TMSs together with the cytoplasmic loop form the link between the N- and C-terminal domains. Their location within the global structure is more peripheral and their function is still unclear (Figure 25). In addition, two lateral helices have been found with the first one located on the periplasmic side and the second one (which was solved in the most recent structure and is not shown here) located on the cytoplasmic side between helices 6 and A. Here, the global 12 TMS containing core can also be described as a total of four 3-helix groups with pseudo symmetry. Similar to the previous descriptions, helices 3, 6, 9, and 12 are completely embedded in the membrane and are not exposed. Likewise, the remaining helices contribute to the formation of a central occluded cavity (13 × 12 × 11 Å) in between the N- and C-terminal domains. Similar to LacY, this cavity is hydrophilic.

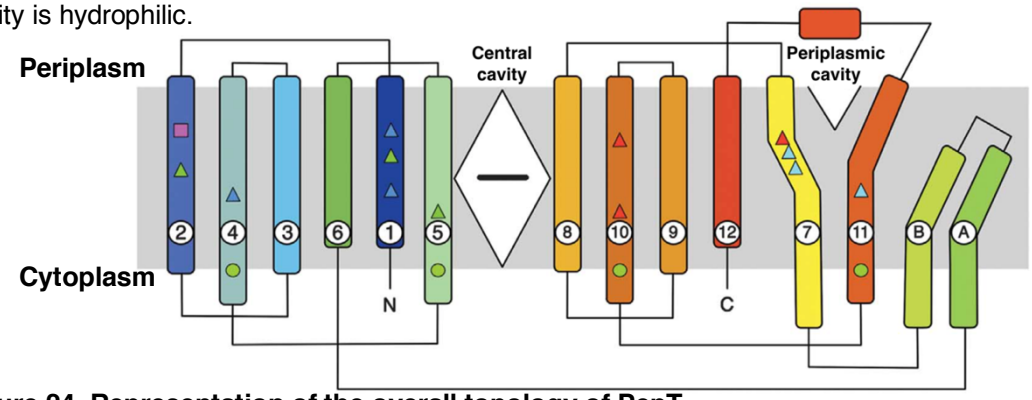

Figure 24. Representation of the overall topology of $PepT_{So}$.

The different TMSs constituting the central MFS core are numbered from 1 to 12. The two additional TMSs are labeled A and B. The periplasmic lateral helix between helices 11 and 12 is colored in red. The central hydrophilic cavity is shown as a white diamond shape with the bound ligand is marked as a black line. An additional periplasmic open hydrophilic cavity is shown as a white triangle. Residues important for the transport function and conserved between different peptide transporters are marked with various shapes. Adapted from reference 200.

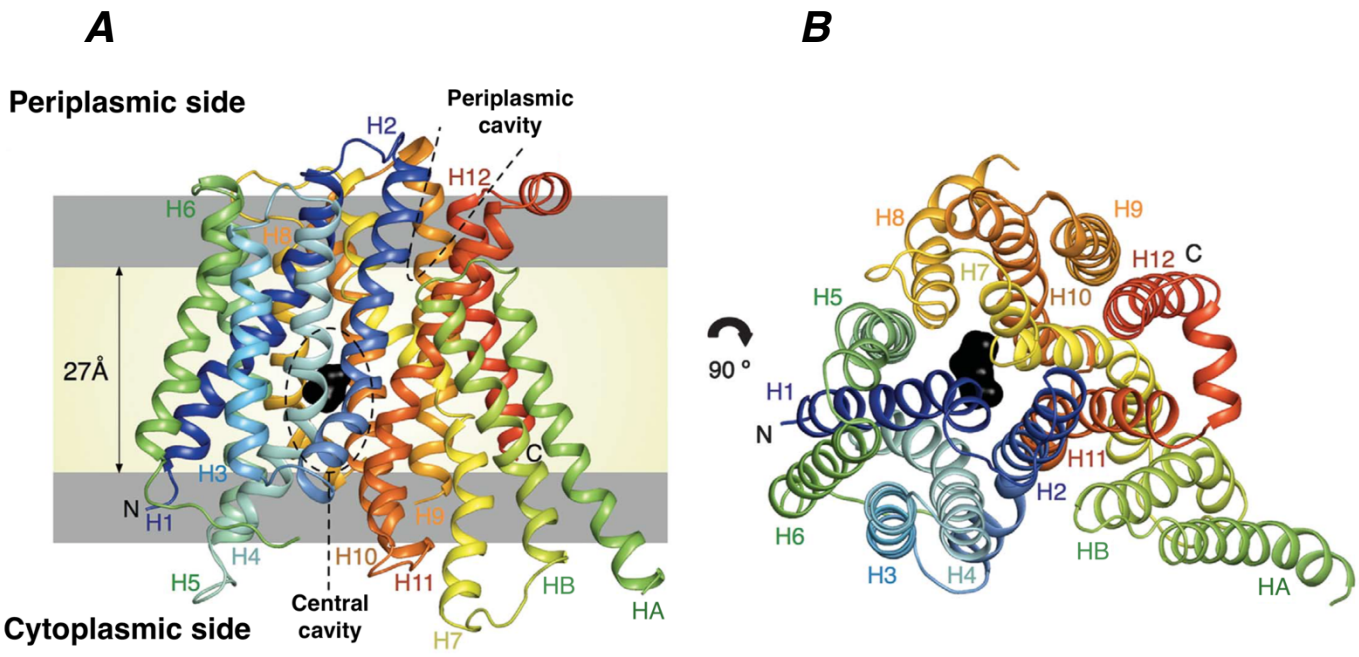

Figure 25. Global structure of $PepT_{So}$.

(A) Side view of $PepT_{So}$. The different TMSs are labeled and colored in a similar fashion compared to the topological representation. The N- and C-termini are indicated. The hydrophobic and interfacial portions of the membrane bilayer are colored in pale yellow and light grey respectively. The two hydrophilic cavities observed are marked by dashed lines. The ligand inside the central cavity is colored in black. (B) Periplasmic view of $PepT_{So}$. Adapted from reference 200.

A periplasmic hydrophilic cavity (16 × 8 × 8 Å) is also evidenced within the structure of $PepT_{So}$ (Figures 24 and 25). This cavity is cone shaped (open to the periplasmic side) and protrudes towards the central cavity. Thus, it has been suggested that it could represent the entry pathway for peptides once the central cavity is outward open.

Within the structure, two gate regions occluding the central cavity from both sides of the membrane are distinguished. First on the periplasmic side, the central cavity is closed by a periplasmic gate formed by the constriction involving two pairs of helices (1 and 2 as well as 7 and 8) from the N- and C-terminal domains (Figure 26A). In a similar fashion, on the cytoplasmic side the access from the central cavity is hindered by the contribution of two pairs of helices (4 and 5 as well as 10 and 11) from the N- and C-terminal domains forming the cytoplasmic gate. Here, the interaction occurs through the side chains of residues such as S131, F150, L427 and M443 (Figure 26B). As will be mentioned later, the most recent considerations refer to both types of gates as the 'thick' and 'thin' gate respectively[201,110].

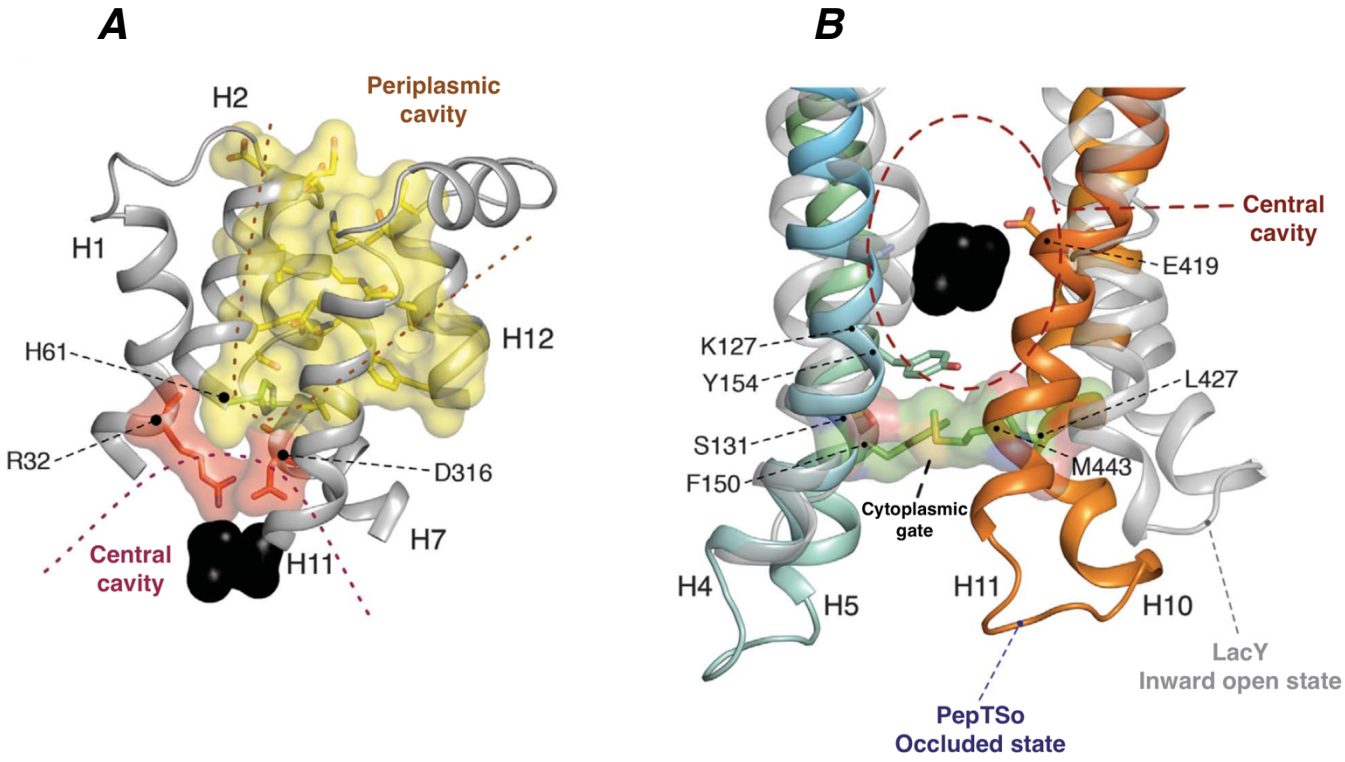

Figure 26. Gating regions in $PepT_{So}$.

(A) Enlarged side view of the periplasmic gate region with the contribution of the helices from the N- and C-terminal domains. Residues from the periplasmic cavity are shown as yellow sticks with a transparent surface. Residues from the central cavity are shown as red sticks with a transparent surface. The bound ligand is colored in black. H61 part of the H^+ substrate coupling system is shown in a stick representation and colored in green. (B) Enlarged side view of the cytoplasmic gate region with the implication of helices from the N- and C-terminal domains. Residues involved in the interaction are shown in a sticks representation in green together with their transparent surfaces. The structure of $PepT_{So}$ is superimposed to the structure of LacY in an inward open state. Adapted from reference 200.

Given the overall dimensions of the central cavity, di- and tri-peptides can be accommodated. Moreover, as single amino acids due to their lower size cannot interact with both the N- and C-terminal domains, they may not be recognized as substrates.

The various residues suggested to be important for the interaction with substrates are located within helices 1, 2, 4 and 5 from the N-terminal domain as well as helices 7, 8, 10 and 11 from the C-terminal domain (Figure 27). Two regions with opposite charges are present within the central cavity. Indeed, a positively charged cluster is constituted by the residues R25, R32, and K127 from the N-terminal domain. Located at the opposite side of K127 a negatively charged residue (E419) is found at the C-terminal domain. These charges may play a role in the recognition and orientation of peptides, as supported by mutational analyses of residues R25 and E419. Y29, Y68 and Y154 could possibly interact with substrates via hydrogen bonding and hydrophobic effects. Finally, several other hydrophobic

residues such as I154, W312, F315, and W446 could be important for hydrophobic effects possibly needed for accommodation of hydrophobic side chains of peptides. This suggestion is supported by mutational studies of W446[200].

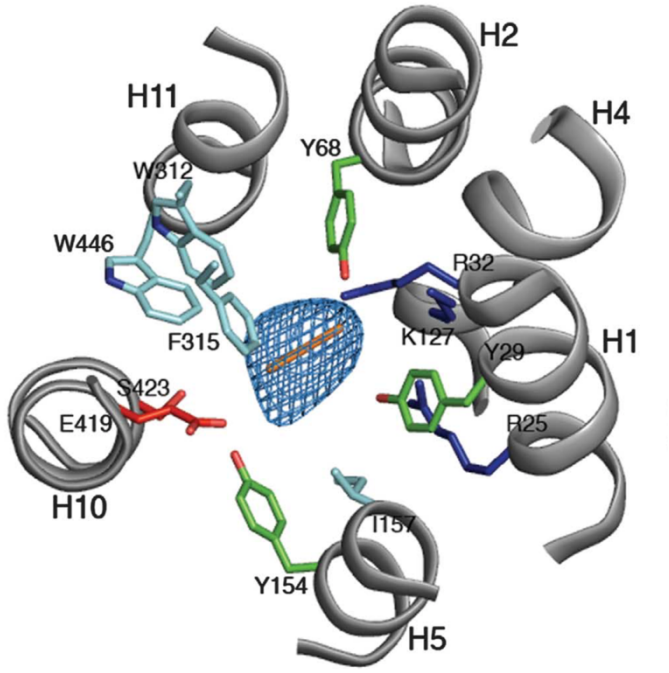

Figure 27. Substrate binding region of $PepT_{So}$.

View of the peptide binding region from the periplasmic side. Conserved residues amongst peptide transporters are viewed as sticks and colored according to their type side chain types: R25, R32, and K127 (blue); Y29, Y68, and Y154 (green); I157, W312, and W446 (cyan); E419, and S423 (red). A di-peptide C_α stick in orange is fitted within the electron density in blue as a size reference. Adapted from reference 200.

Residues constituting possible protonation sites within $PepT_{So}$ include H61, and D316 (Figure 26A). Mutational studies of H61 indicate the importance of its protonation status for the transport activity[200]. Therefore, this region located next to the periplasmic gate might be important for H^+ translocation and also for the periplasmic gate opening.

3.1.1.5. Generalization of the structural fold of single component MFS members

Based on numerous structures obtained over the years for both prokaryotic and eukaryotic single component members of the MFS family it has been possible to decipher the general features of the so-called MFS fold present in all cases.

Indeed, as mentioned previously the overall MFS fold is composed of at least 12 TMSs grouped in two six-helix bundles (helices 1 to 6 and 7 to 12) called the N- and C-terminal domains (Figure 28). These two domains present a two-fold pseudo symmetry according to an axis which is perpendicular to the plane of the membrane. Both domains are connected to each other via a long cytoplasmic loop located in between helices 6 and 7 (not shown in Figure 28). Each of the six-helix bundles can further be subdivided in two three-helix repeat units with pseudo symmetry (Figure 29)[120]. In the most recent considerations, the first, second, and third helices from each repeat unit were named helix A (for helices 1, 4, 7, and 10), helix B (for helices 2, 5, 8, and 11) and helix C (for helices 3, 6, 9, and 12) based on their shape and localization within the overall structure (Figure 28)[110].

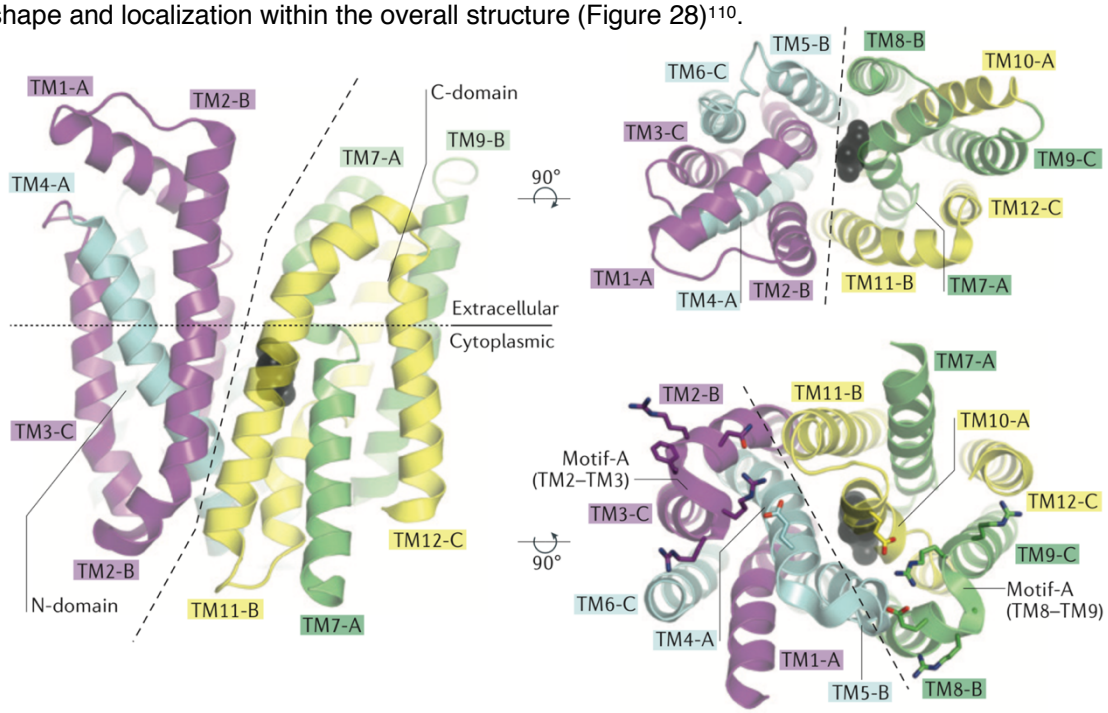

Figure 28. Structure of the human glucose transporter 3 (GLUT3).

The structure of GLUT3 (PDB ID: 4ZW9) in an outward-facing partially occluded conformation was chosen as an example because of its high resolution (1.5 Å). On the left side, GLUT3 is viewed from the side. In the top right panel, the transporter is viewed from the extracellular side. A cytoplasmic view of the structure is shown in the bottom right panel. Overall the 12 helices are numbered from 1 to 12 starting from the N-terminus. The N- and C-terminal domains are labeled and separated from each other by a vertical dashed line. Each of the domains is further divided in two three-helix repeats colored in purple and cyan within the N-terminal domain and in green and yellow within the C-terminal domain. Each of the first, second and third helices within the repeat units are labeled with an additional A, B, and C respectively. The substrate (glucose) bound at the center of the protein is colored in black. The conserved A-motifs are shown as sticks in the right bottom panel. Adapted from reference 110.

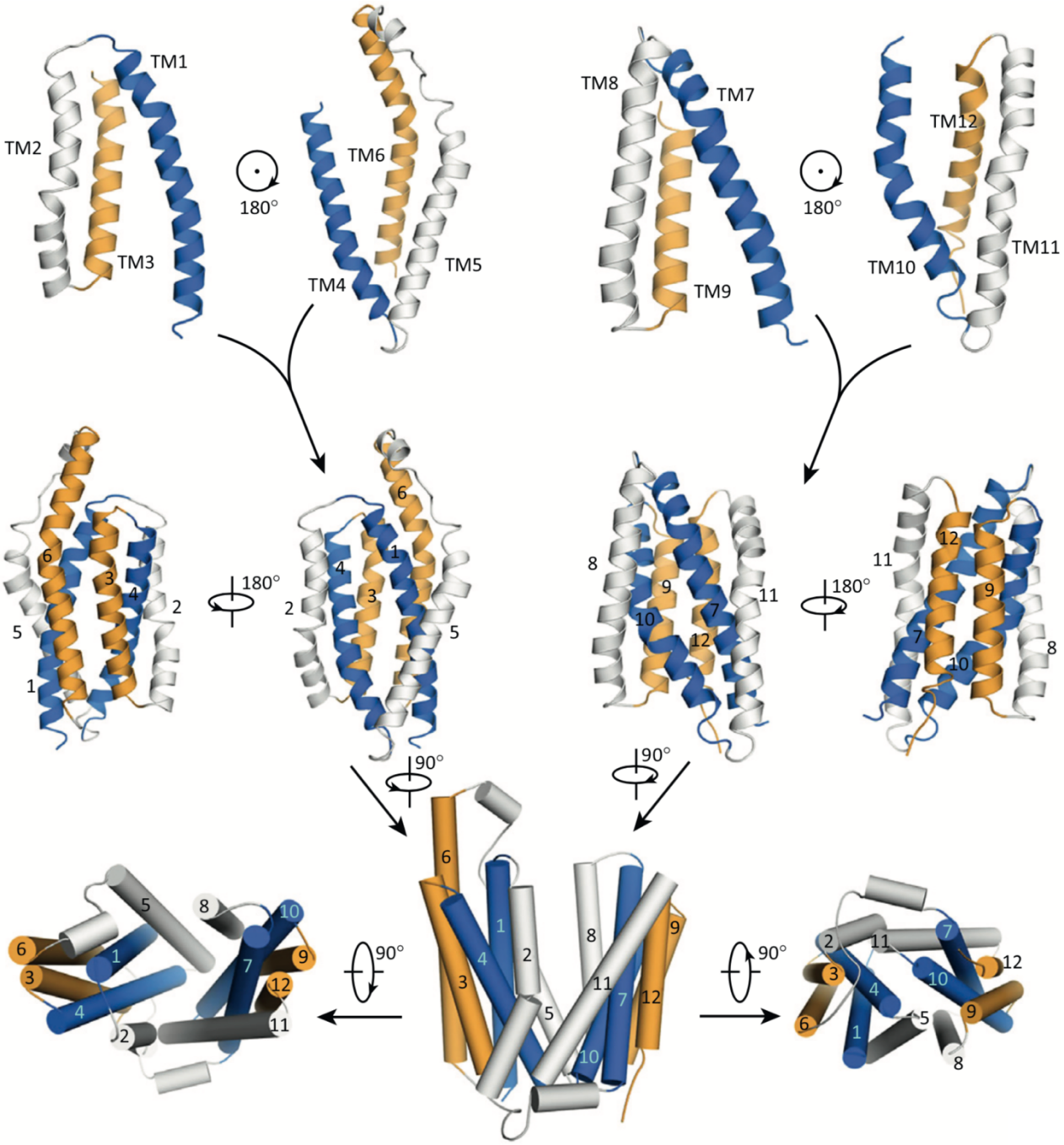

Figure 29. Illustration of the global MFS fold.

The structure of FucP (PDB ID: 3O7Q) in an outward open conformation is used as an example. (Top row) Four repeat units constituted by three helices are present in the global MFS fold. (Middle row) Two inverted repeat units in each case constitute the N- and C-terminal domains. (Bottom row) The N- and C- terminal domains within the overall structure are related to each other by a 180° rotation along an axis perpendicular to the plane of the membrane. Symmetry related helices are shown with the same colors. Adapted from reference 120.

The main conformations identified for MFS transporters include: the inward open, the occluded and the outward open conformations. In addition, a so-called 'cavity' present in between the N- and C-terminal domains is usually described in the different structures. Therefore, in the inward open conformation the cavity is visible on the cytoplasmic side allowing access to the substrate binding site located in the exact center of the protein. Moreover, in the occluded conformation the cavity is buried in the structure and the substrate binding site is inaccessible. Finally, in the outward open conformation the cavity is visible on the periplasmic or extracellular side and therefore the substrate binding site is accessible again (Figure 30)[110].

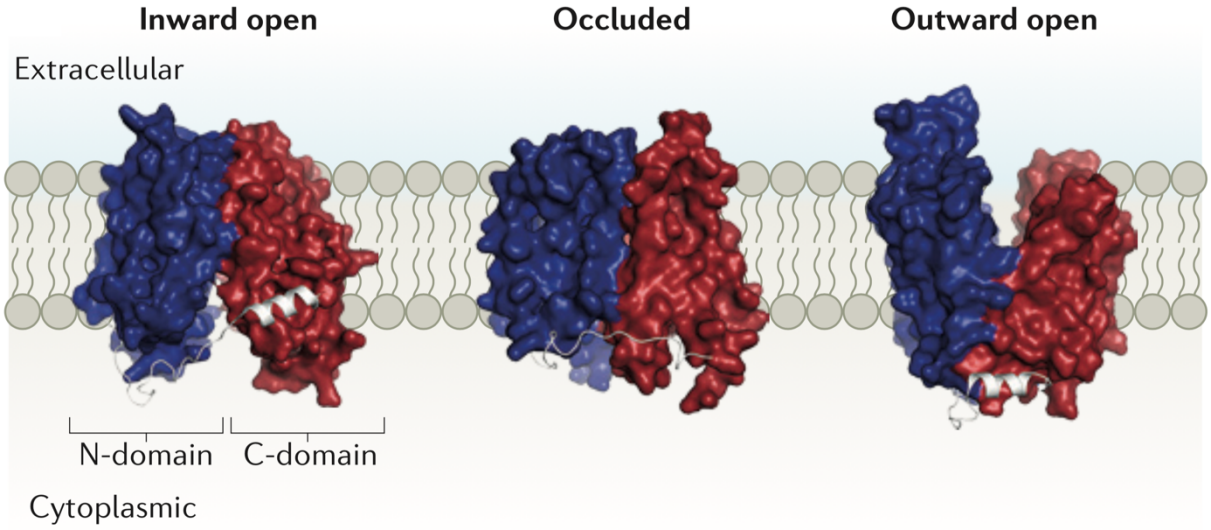

Figure 30. Major conformational states of MFS transporters.

The first crystal structures of MFS members representatives of the different conformational states. LacY (PDB ID: 1PV6) is used for the illustration of the inward open conformation. EmrD (PDB ID: 2GFP) is shown as an example of the occluded conformation. FucP (PDB ID: 3O7Q) is used for the representation of the outward open conformation. The N- and C-terminal domains are colored in blue and red respectively. The cytoplasmic loop is colored in light grey. Adapted from reference 110.

3.1.2. Mechanistic insights of single component MFS members

3.1.2.1. LacY

Based on structural and biochemical information obtained regarding LacY, a mechanistic model has been proposed for the different events that might occur during symport. Here, the description mainly

focuses on the transition between the inward and outward open conformations together with the roles of the different residues involved in the translocation of D-galactopyranosides and H^+ (Figure 31).

Starting from an outward open conformation (Figure 31A), the first event occurring is the protonation of E269 (Figure 31B). Next, the substrate arrives through the hydrophilic cavity and interacts first with W151 and subsequently with R144 as well as E269. The binding of the substrate induces the deprotonation of E269. This event in turn disrupts the salt bridge present between R144 and E126. A new salt bridge is formed instead between R144 and E269. The H^+ released by E269 is transferred to H322 (Figure 31C). Following this first transfer, the H^+ is further conveyed from H322 to E325 (through the salt bridge and hydrogen bond network) and the transition from outward open to inward open conformation occurs (Figure 31D). Upon the release of the substrate, the salt bridge between R144 and E269 is disrupted and another one is formed again between R144 and E126 (Figure 31E). It has been suggested that the deprotonation of E325 might probably occur either because of the proximity to R144, or with the exposure to the cytoplasmic environment and maybe because of both factors simultaneously (Figure 31F). Finally, a transition from the inward open to the outward open conformation occurs to return to the initial state[187].

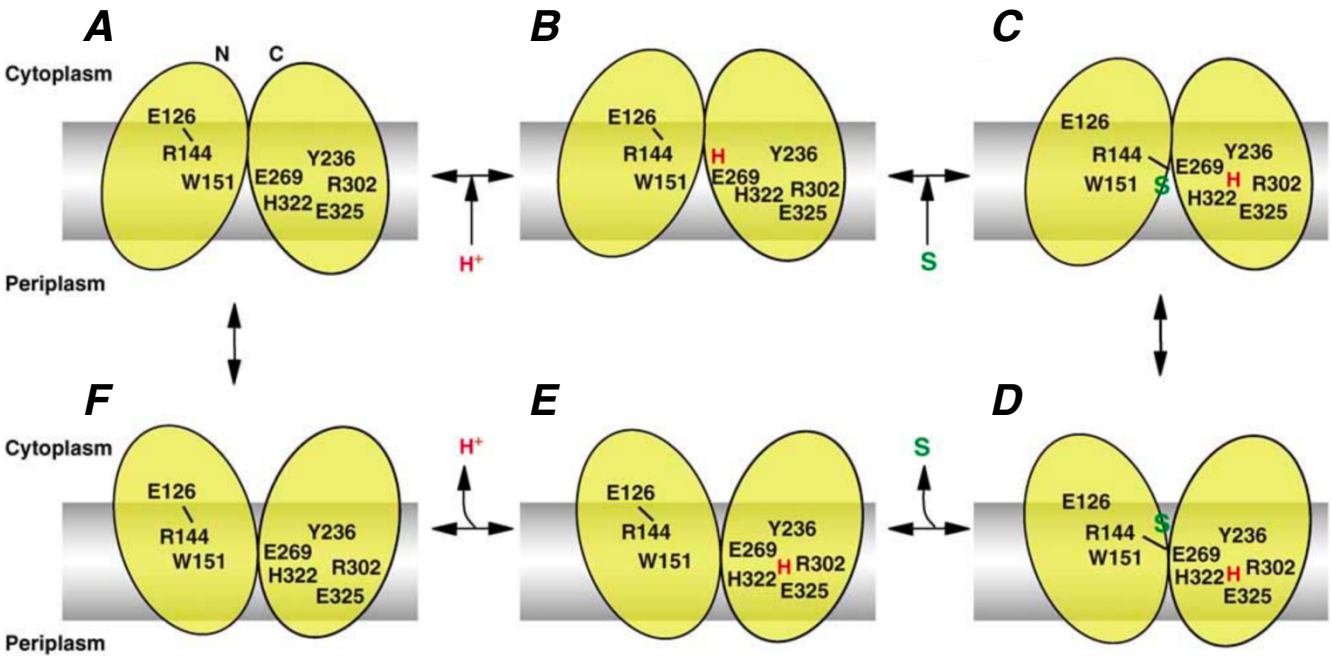

Figure 31. Symport mechanism of LacY.

Role of different residues during the symport cycle of LacY. The N and C-terminal domains are represented as oval shapes colored in yellow and labeled in (A). Salt bridges between residues are indicated by black lines. H^+ and the substrate are shown in red and green respectively. Adapted from reference 187.

3.1.2.2. MdfA

With the help of structural and biochemical information, an antiport mechanism has been proposed for the transport of different molecules including antibiotics with the translocation of H^+ in MdfA. The following description mainly concerns the inward and outward open conformations as well as two important residues for protonation (Figure 32).

Starting from the inward open conformation, D34 located at the apex of the hydrophobic cavity is protonated (Figure 32A). Upon the binding of a substrate (from the cytosol or from the inner leaflet of the inner membrane) and the exclusion of solvent molecules, the dielectric constant inside the cavity is decreased which increases the positive electrostatic field generated by Motif B. This event in turn triggers the deprotonation of D34 by decreasing its pK_a (Figure 32B). The deprotonation of D34 destabilizes the inward open conformation by disrupting the electrostatic interaction with the membrane potential. A so-called 'elastic energy' contained in the inward open state is subsequently released and triggers the transition towards the outward open state (Figure 32C). After the release of the substrate, E26 now located on the apex of the hydrophobic cavity and not exposed to the solvent is protonated (Figure 32D). With the protonation of E26, an electrostatic interaction occurs with the membrane potential (Figure 32E). This interaction together with the simultaneous and direct or indirect H^+ transfer from E26 to D34 triggers the transition towards an inward open conformation stabilized by the new electrostatic interaction of protonated D34 with the membrane potential (Figure 32A). Part of this energy is stored again in the inward open conformation and constitutes the so-called 'elastic energy'[194].

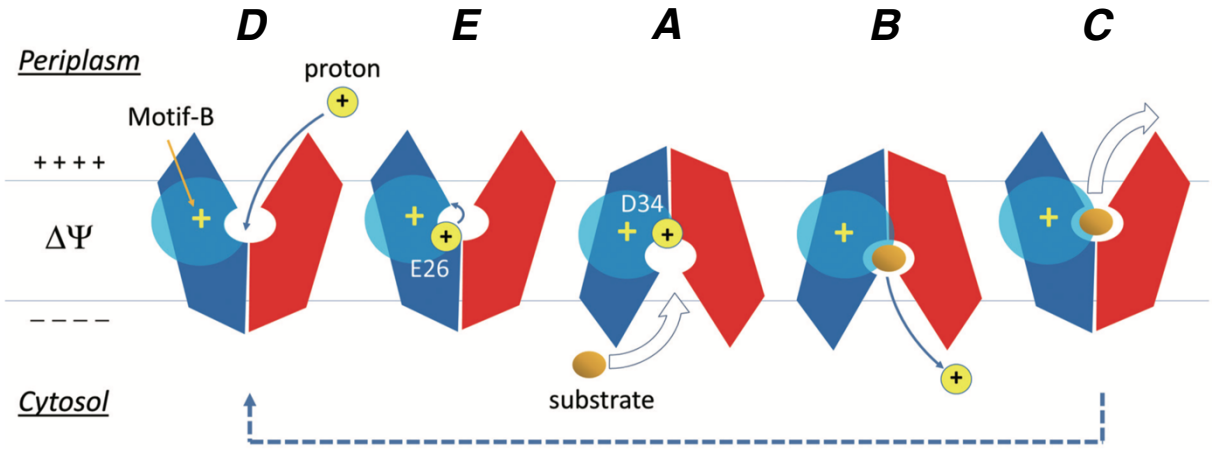

Figure 32. Suggested antiport mechanism of MdfA.

The N- and C-terminal domains are colored in blue and red respectively. Motif B is indicated with a yellow plus sign within a transparent blue circle. The proton is shown as a yellow circle with a black plus sign. The substrate is colored in orange. The two protonation sites essential for the transport cycle are indicated. ΔΨ, membrane potential. Adapted from reference 194.

3.1.2.3. EmrD

Because EmrD still remains poorly characterized, its mechanism is mostly proposed on the basis of its structure and through homology comparison with other MFS antiporters (Figure 33).

Substrates could either be recognized from the cytosol directly or from the inner leaflet of the inner membrane when the transporter is in an inward open conformation (Figure 33A). Subsequently a transition occurs from inward open to occluded conformation (Figure 33B). Finally, substrates could be released on the periplasmic side with a second transition from occluded to outward open conformation (Figure 33C). Obviously, substrate translocation would also in this case be coupled to H^+ transfer[199].

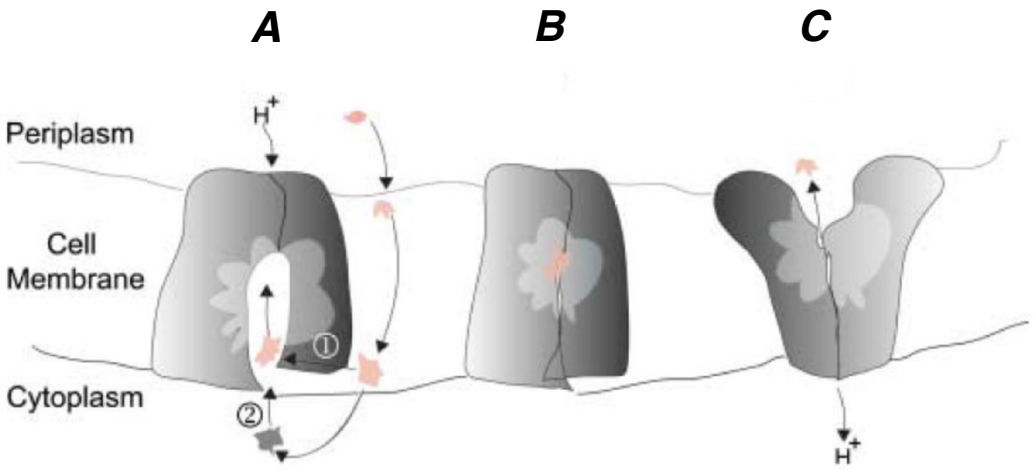

Figure 33. Antiport mechanism of EmrD.

Substrates can be recognized either from the inner leaflet of the inner membrane (pathway 1) or from the cytosol (pathway 2). Adapted from reference 199.

3.1.2.4. $PepT_{So}$

Based on structural and biochemical studies, a symport mechanism has been proposed for the peptide transporter from *Shewanella oneidensis* $PepT_{So}$ (Figure 34).

Starting from an outward open conformation (Figure 34A), a peptide (arriving through the outward open cavity) is bound at the peptide binding site at the interface of the N- and C-terminal domains. However, H^+ (also arriving through the outward open cavity) is bound at a region close to the periplasmic gate at the interface of the N- and C-terminal domains (with H61, and D316 representing possible protonation sites). After the transition from the outward open to the occluded conformation, the central cavity containing the bound peptide is occluded from both sides by the periplasmic and

cytoplasmic gates. On the contrary the H^+ binding site is still open on the periplasmic side through the periplasmic cavity (Figure 34B). The peptide and H^+ are released on the cytoplasmic side (through the inward open cavity) after the transition from the occluded to the inward open conformation. Here, the H^+ binding site is open towards the cytosol (Figure 34C)[200].

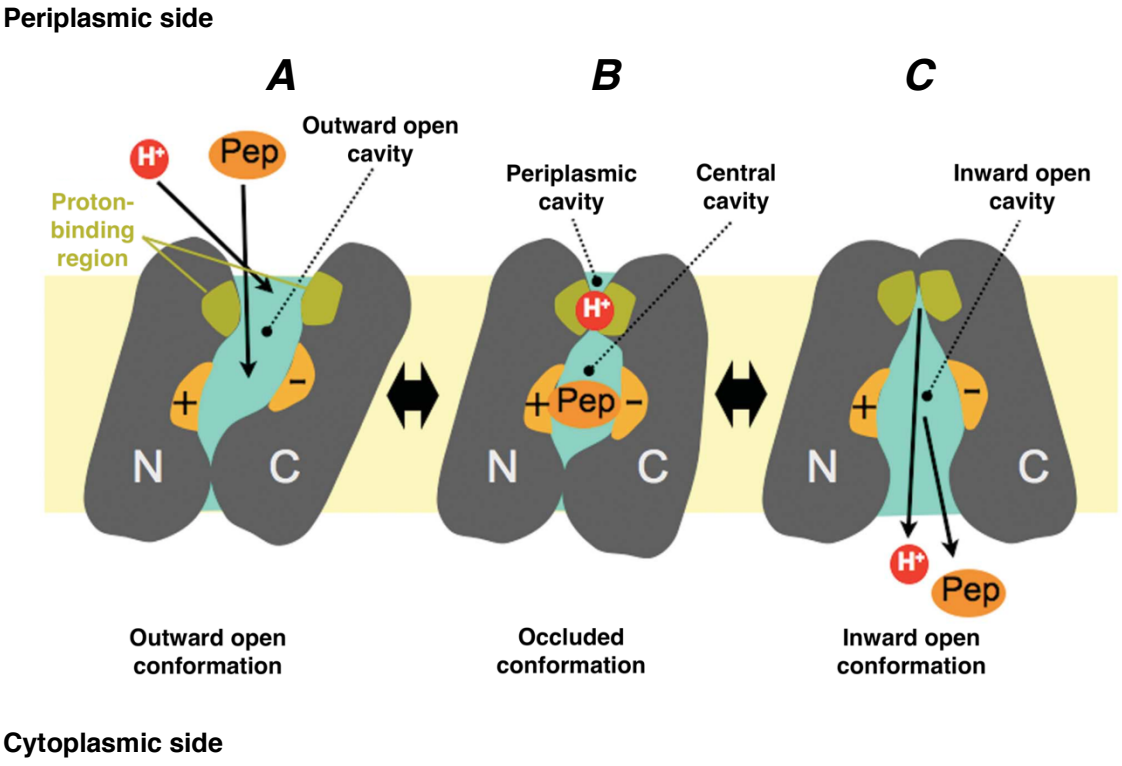

Figure 34. Symport mechanism of $PepT_{So}$.

The peptide (Pep) and proton (H^+) are shown in orange and red respectively. The N- and C-terminal domains of the transporter are colored in dark grey. The peptide binding site is colored in yellow and indicated by + and – signs. The proton binding region is colored in green. Adapted from reference 200.

3.1.2.5. Generalization of the transport mechanism of single component MFS members

Based on various structural features observed within numerous prokaryotic and eukaryotic single component MFS members, increasing proposals of transport mechanisms have been made over the years. Therefore, it is interesting to try to understand the transport mode of the MFS with a global overview of all the structures obtained thus far.

Classically, the first mechanistic proposal made was called an 'alternating access model'. This model considers that the transporter undergoes global conformational changes exposing alternatively a

central substrate binding side to either side of the membrane. Indeed, the inward and outward open conformations of various MFS members are in agreement with this type of transport mechanism[202,110].

Based on the first proposal, a second model called the 'rocker-switch model' has been suggested afterwards (Figure 35). In this case, it is considered that the transporter undergoes conformational changes with a rocker-switch type rotation of the N- and C-terminal domains towards a central axis. This would be favored by the curved shape of the B helices. Moreover, the inward and outward open conformations would be stabilized by interactions between the cytoplasmic and periplasmic/extracellular ends of the A and B helices. However, this model presents a limitation. Indeed, several occluded conformations observed for different MFS members (for examples see Table 1) cannot be explained by this model as the rocker-switch type rotation of the N- and C-terminal domains is not sufficient for the occlusion of the binding site from both sides of the membrane[203,110].

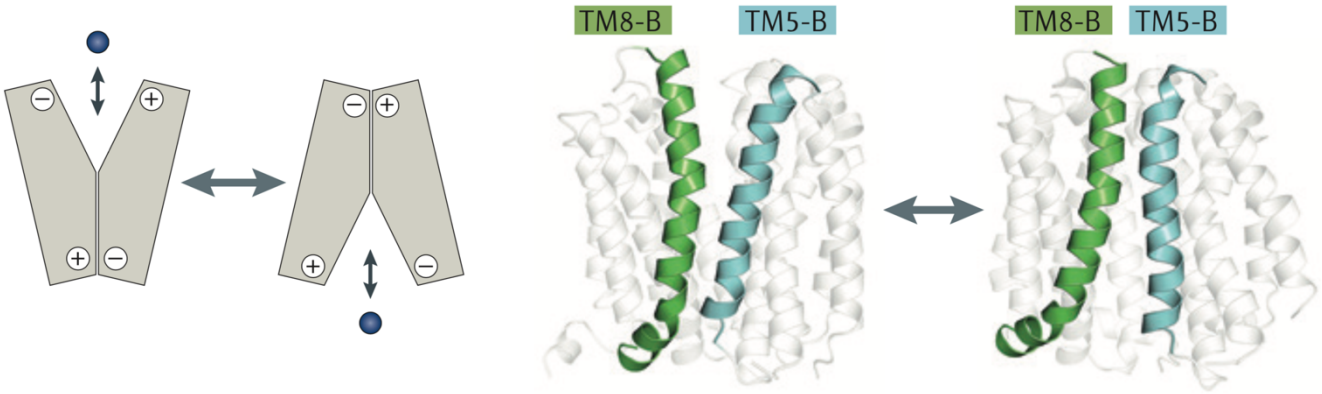

Figure 35. The rocker-switch model.

(Left panel) Illustration of the rocker-switch model explaining the transport mechanism of single component MFS transporters. The N- and C-terminal domains are colored in light grey. The substrate is shown as a blue circle. The plus and minus signs indicate the ionic interactions present in some cases at the tips of helices from both domains. (Right panel) Outward open structure of LacY (PDB ID: 4OAA) on the left and inward open structure of LacY (PDB ID: 2Y5Y) on the right supporting the rocker-switch model. Two curved shape B helices lining the central axis are colored in green and cyan and labeled in order to illustrate their respective interactions on the periplasmic and cytoplasmic sides. Adapted from reference 110.

Recently, an updated model called the 'clamp and switch model' has been proposed instead based on the structures of MFS members in occluded conformations (Figure 36). The different structures obtained in an occluded conformation thus far are described as inward-facing occluded, outward-facing occluded, inward-facing partially occluded, and outward facing partially occluded. The terminologies

inward and outward facing mainly indicate that the global orientations of the N- and C-terminal six-helix bundles are similar to the respective orientations observed in the inward and outward open conformations. Therefore, in addition to the switch movement of both domains, bending of the tips of one or several of the A helices 4 and 10 as well as 1 and 7 and in some cases their respective flaking B helices are involved in the occlusion of the substrate binding site from the remaining cytoplasmic and periplasmic/extracellular sides respectively[110].

Thus, the so-called 'clamp and switch model' is better suited for the overall description of the conformational cycle of the MFS transporters. In fact, (I) in the so-called clamping step, the bending of the tips of the previously mentioned helices closes the substrate binding site from the remaining cytoplasmic or periplasmic/extracellular sides and (II) in the so-called switching step, it is the rotation of the N- and C-terminal domains towards a central axis that exposes the substrate binding site to the opposite side of the membrane (compared to a given clamping direction).

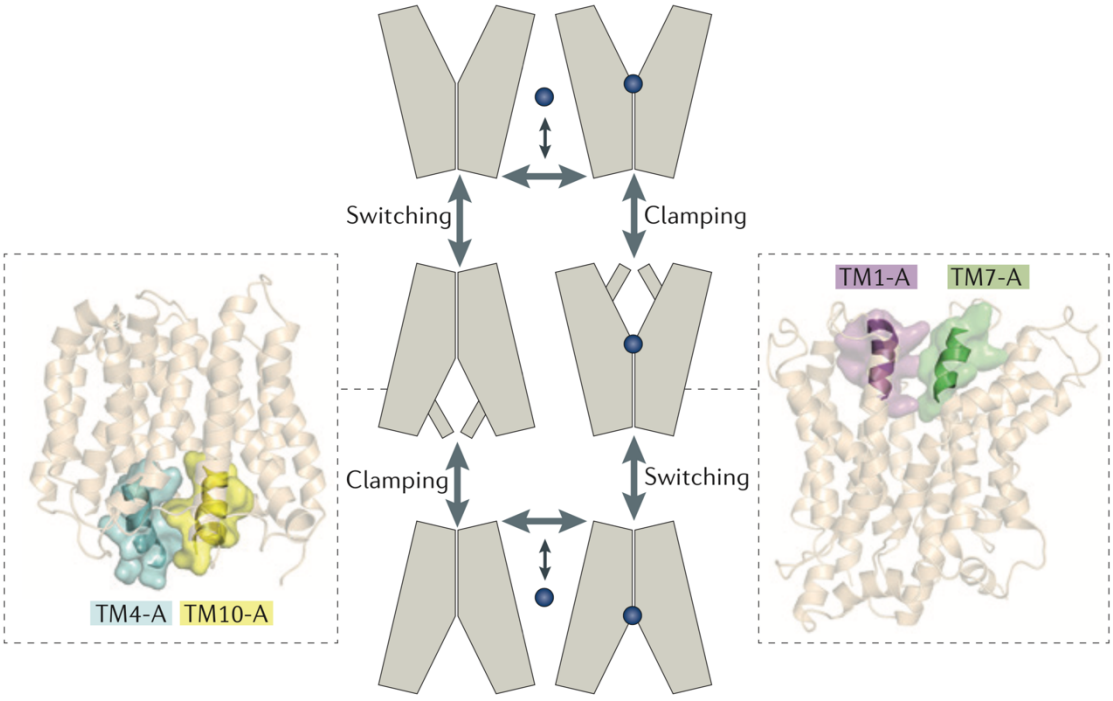

Figure 36. The clamp and switch model.

Illustration of the clamp and switch model taking into account the occluded states of the single component MFS transporters. The N- and C-terminal domains are colored in light grey. The substrate is shown as a blue circle. On the left and right sides, the structures of EmrD in an inward facing occluded conformation (PDB ID: 2GFP) and MelB in an outward facing occluded conformation (PDB ID: 4M64) respectively are shown. The corresponding bent tips of A helices are indicated with their transparent surfaces and colored in the same color code as in Figure 28. Thus, in total, four conformational states (inward open, inward facing occluded, outward open and outward facing occluded conformations) are considered by this model. Adapted from reference 110.

In order to provide a complete description of the transport mechanism of single component MFS members, it is important to mention the involvement of gating residues in the conformational changes mentioned previously. Gates can be defined as transient structural elements formed via interactions between residues (gating residues) from two regions within or in between the N- and C-terminal domains (after movements of the tips of one or several A helices and in some cases flaking B helices as well as after global rotations of both domains). Therefore, a distinction has been made between 'intra-domain', 'thin' and 'thick' gates (Figure 37A). Intra-domain gates are formed via transient interactions between residues located within either the N- or C-terminal domains. However, thin and thick gates are formed via transient interactions between residues from both the N- and C-terminal domains. In addition, the terminologies thin and thick refer to the degrees of solvent accessibility and conformational changes needed for their formation. Thus, thin gates might be accessible to solvent molecules and only need low degree conformational changes (bending of helix tips) for their formation. On the contrary, thick gates are tight seals which need high degree conformational changes (rotation of the N- and C-terminal domains) for their formation. Because of the conservation of gating residues from motif A: GX_3-(D/E)-(R/K)-X-G-[X]-(R/K)-(R/K), located in between helices 2 and 3 of the N-terminal domain and/or helices 8 and 9 of the C-terminal domain in most MFS members (Figure 28), interactions involving this motif of the C-terminal domain in XylE have been used to explain cytoplasmic gating in the context of the clamp and switch model (Figure 37B first row). Starting from the inward open conformation of XylE, R341 (from motif A) and E397 both located within the C-terminal domain interact to form the closed intra-domain gate. In the inward-facing occluded conformation, E397 from the C-terminal domain is interacting with R160 from the N-terminal domain to form the closed thin gate (here the intra-domain gate is open). Finally, in the outward-facing occluded conformation (and by extension in the outward open conformation) additional interactions between the N- and C-terminal domains including the interaction between D337 (from motif A) and the positive charge of the N-terminus of helix 5 form the closed thick gate (here the closed intra-domain gate is formed again via the interaction between R341 (from motif A) and E397). Similar interactions are also indicated for motif A of both the N- and the C-terminal domains of other MFS members (Figure 37B middle and bottom rows). Even if thus far only gating on the cytoplasmic side has been mentioned, it is considered that similar gates may also be formed on the periplasmic/extracellular side[110].

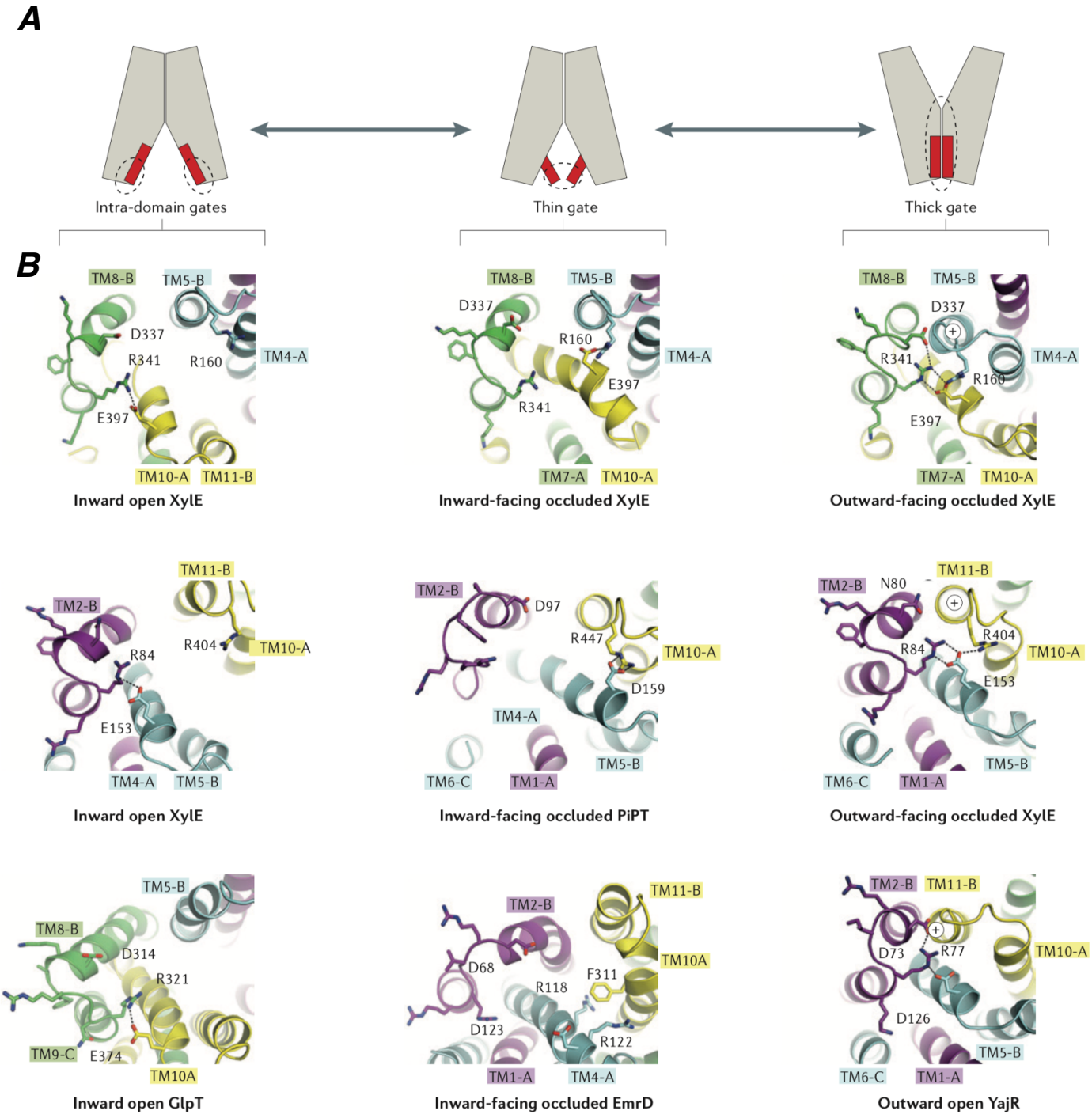

Figure 37. Illustration and structural evidence of different types of gates.

(A) Schematic representation of the intra-domain, thin and tick gates (shown by dashed circles) on the cytoplasmic side in inward open, inward-facing occluded and outward open conformations respectively. The N- and C-terminal domains are colored in light grey. The flexible helix tips are colored in red. (B) The different helices and belonging residues are colored in the same color code as in Figure 28. Top row, illustration of the different interactions of residues from motif A of the C-terminal domain involved in the formation of different types of gates. Middle and bottom rows show similar interactions involving residues from motif A of the N- and C-terminal domains. XylE Xylose:H^+ symporter from *Escherichia coli*; PiPT Phosphate:H^+ symporter from *Piriformospora indica*; GlpT Glycerol-3-phosphate antiporter from *Escherichia coli*; EmrD drug:H^+ antiporter from *Escherichia coli*, YajR Transporter of unidentified function from *Escherichia coli*. Adapted from reference 110.

3.2. Tripartite MFS members

3.2.1. General introduction to the EmrAB-TolC system

As stated in the very beginning, the aim of the present Ph.D. project was to gain structural information about the thus far poorly studied EmrAB-TolC tripartite MFS-type efflux system. As will be described in the following sections, this system is composed of the inner membrane drug:H^+ antiporter (EmrB), the outer membrane exit duct (TolC) and the periplasmic adaptor protein (EmrA) making the connection between the two former components.

The chromosomal *emr* (*Escherichia coli* multi-drug resistance) locus (containing the *emrA* and *emrB* open reading frames (ORFs)) was first described by Lomovskaya and Lewis 28 years ago[144]. It was evidenced that this locus conferred intrinsic resistance towards different compounds including uncouplers of oxidative phosphorylation such as CCCP and TCS as well as the antibiotics Nalidixic acid and Thiolactomycin (Table 2). It is interesting to note that most of these compounds are rather hydrophobic.

Because of the absence of promoter regions upstream of both *emrA* and *emrB*, it was suggested that these ORFs together with *mprA* (microcin production regulation, locus A)[204] having a promoter upstream its coding region could form an operon[144]. Later studies, revealed that *mprA* is a negative regulator of the *emr* locus and was thereafter renamed *emrR* [205]. Overexpression of EmrAB occurs either through mutation in the *emrR* ORF region or through induction[205]. Indeed, this negative regulator similarly to a 'sensor' can bind structurally variable toxic compounds including EmrAB substrates (CCCP, TCS, and Nalidixic acid) and this event releases EmrR bound to its promoter region inducing the *emrRAB* operon[205–207]. EmrR also participates in the regulation of the plasmid encoded *mcb* operon responsible for the production of the peptide antibiotic Microcin B17. Interestingly, substances (including CCCP) inducing the *emr* operon repress the *mcb* operon indicating the existence of variable survival strategies that can be adopted by the bacterium[208]. Finally, within the *emr* locus, an additional promoter (independent of EmrR) for the *emrB* ORF located in the *emrA* ORF has also been identified[205].

The *emrA* ORF codes for a 390 amino acid containing protein with an estimated molecular weight of 42.7 kDa. Hydropathy analyses indicated that EmrA has one hydrophobic domain located at its N-terminus. A chimera of EmrA fused to alkaline phosphatase (only active in the periplasm) at its C-terminus indicated that this C-terminal hydrophilic region is located in the periplasm. Therefore, as EmrA

lacks a signal sequence (found in periplasmic, outer membrane and exported proteins), it was deduced that overall, a short N-terminal end of EmrA was located on the cytoplasmic side followed by a TMS located in the inner membrane and a hydrophilic C-terminal domain located in the periplasm[144]. Based on the first amino acid sequence analysis[144], EmrA was found to be homologous to CyaD[209] (involved in the export of cyclolysin), HlyD[210] (involved in the export of hemolysin), as well as to CvaA[211] (involved in the export of colicin V).

The *emrB* ORF codes for a 513 amino acid containing protein with an estimated molecular weight of 55.6 kDa. Hydropathy analyses indicated that EmrB contains 14 TMSs and is thus an integral membrane protein[144]. Based on the first amino acid sequence analysis[144], EmrB was found to be homologous amongst others to QacA (Table 2).

The first *in vitro* transcription-translation analysis of *emrAB* showed the presence of three bands (at 55, 44 and 36 kDa) (Figure 38). The weak 55 kDa signal was attributed to EmrB. The prominent signals visible at 44 and 36 kDa were both assigned to EmrA. In fact, it was suggested that the presence of the second smaller 36 kDa signal could be due to an alternative translation from another ATG located at position 396 (with a putative ribosome binding site found next to it)[144].

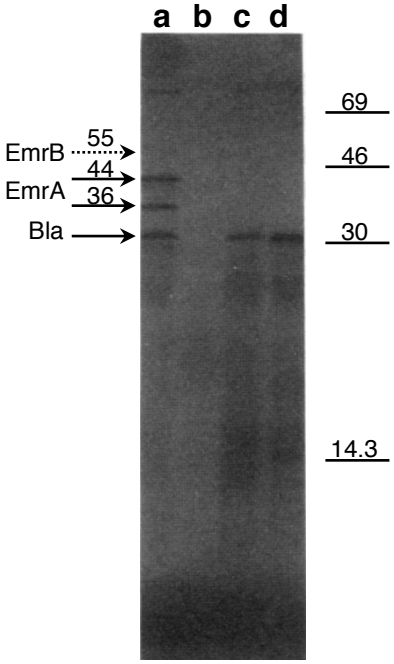

Figure 38. *In vitro* transcription-translation products of *emrAB* labeled with [^{35}S]methionine separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

The molecular weight markers in kDa are indicated on the right. Bla, β-lactamase. Lane a, plasmid pEMR2.1 (*emrAB*); lane b, no plasmid; lane c, plasmid pEMR2.6 (*emrB*); lane d, plasmid pUC18. Bla was used as a control of the *in vitro* expression system. The weak expression of EmrB (dashed arrow) was suggested to be due to the limitation of the *in vitro* expression system. Adapted from reference 144.

Finally, early on based on homology to the HlyBD-TolC system, it was suggested that in a similar manner EmrAB could together with TolC form an MFS-type tripartite efflux system[212,213].

3.2.2. Structural insights of tripartite MFS members

3.2.2.1. Structure of the EmrAB complex

In 2009, Tanabe and co-workers performed the first structural investigations of EmrAB using negative staining electron microscopy (EM)[214] and the following description only focuses on that study.

Interestingly, the results of the SDS-PAGE obtained after the separate purification of EmrA and EmrB were interpreted differently compared to the first *in vitro* transcription-translation analysis mentioned previously (Figure 39). Indeed, only one corresponding signal was seen for EmrA at about 48 kDa and the signal corresponding to EmrB was found at about 38 kDa. However, it is important to mention that in this case *E. coli* BL21(DE3) and C43(DE3) were used as expression hosts for the separate expression of EmrA and EmrB respectively.

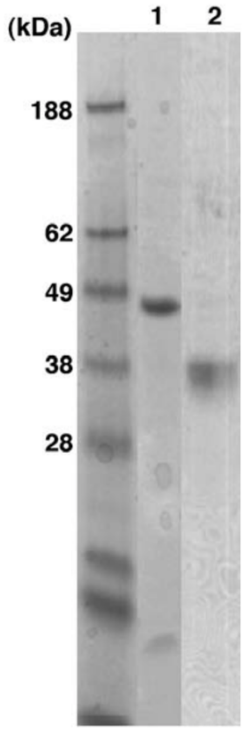

Figure 39. SDS-PAGE analysis (with the EZ blue stain) of EmrA and EmrB purified separately.
(Lane 1) EmrA purified via a His-tag. (Lane 2) EmrB purified via a Strep-tagII. Adapted from reference 214.

Following the purification, both proteins were mixed and reconstituted into liposomes and then re-exchanged to detergent micelles via size exclusion chromatography (SEC) using a Superose 6 10/30 column and the buffer 50 mM Tris/HCl pH 7.5, 200 mM NaCl and 0.1% DDM (Figure 40A and B). Fractions 2 and 3 containing the EmrAB complex were used for subsequent negative staining EM analyses.

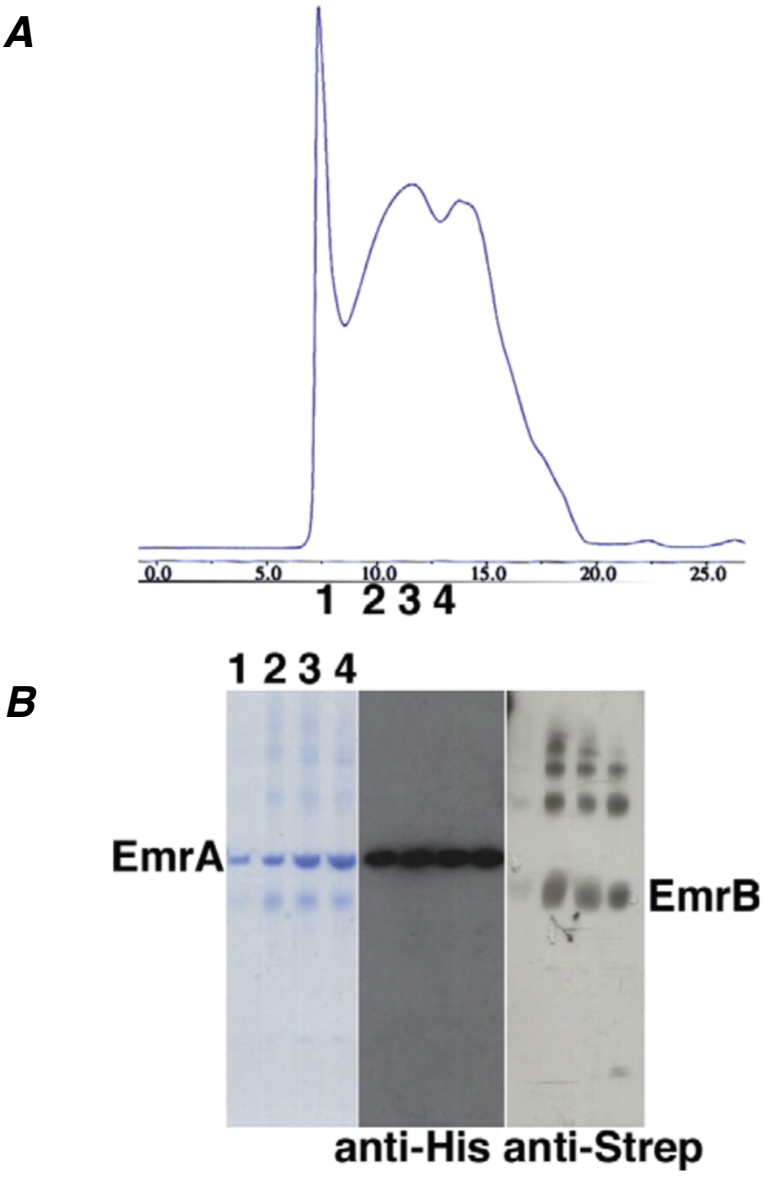

Figure 40. Isolation of the EmrAB complex.

(A) SEC profile of the EmrAB sample reconstituted into liposomes and injected on a Superose 6 10/30 column for the isolation of the EmrAB complex in DDM. Different fractions labeled 1, 2, 3, and 4 were taken for the verification of the presence of both proteins. The first peak corresponded to EmrA aggregates, the second peak was attributed to the EmrAB complex and the third peak corresponded to free EmrA. (B) SDS-PAGE (with the EZ blue stain), anti-His and anti-Strep Western blotting analyses. Adapted from reference 214.

The negative staining EM analyses of the fractions containing the EmrAB complex are shown in Figure 41.

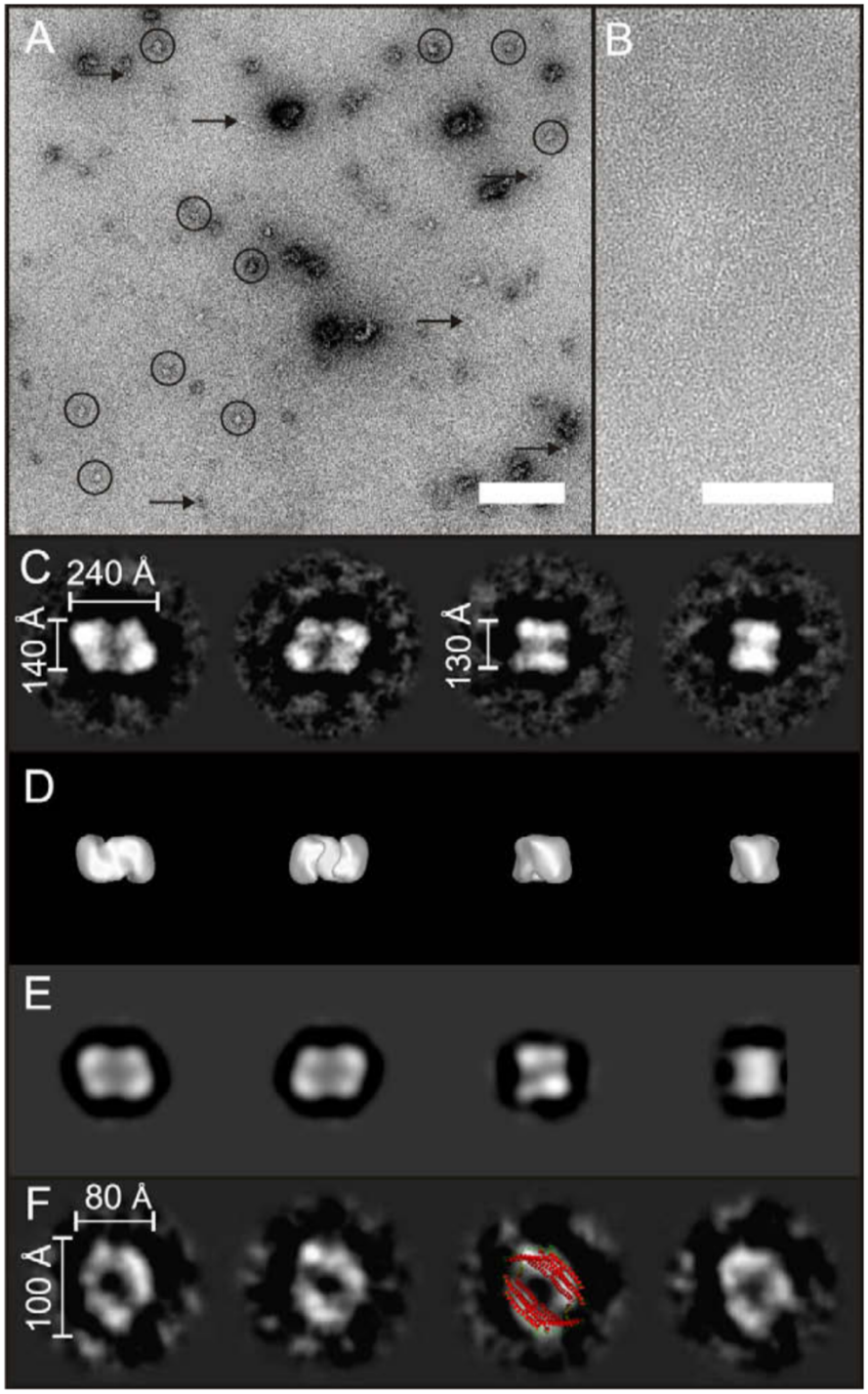

Figure 41. Negative staining EM analyses of EmrAB.

(A) Representative micrograph of fractions containing the EmrAB complex. The larger particles (encircled) were attributed to the EmrAB complex. The smaller particles (indicated by arrows) were assigned to free EmrA. The scale bar represents 100 nm. (B) Micrograph of EmrB used as a negative control. The scale bar represents 100 nm. (C) Averages of EmrAB complexes in different orientations. (D) Three-dimensional (3D) map of the EmrAB complex in the same orientation as in (C) at 30 Å resolution. (E) Reprojections through the 3D map. (F) Averages of the smaller particles attributed to EmrA only. The crystal structure of MexA (PDB ID: 1VF7) modelled as a dimer (colored in red) is superimposed on the third average to indicated the similarity in size and shape compared to a possible EmrA dimer. Adapted from reference 214.

A representative micrograph of the fractions containing the EmrAB complex is shown in Figure 41A. The alignment and classification analyses revealed two distinct populations based on overall size. Larger particles (indicated by circles in Figure 41A) were attributed to the EmrAB complex and smaller particles (indicated by arrows in Figure 41A) were considered to correspond to free EmrA. Averages of the EmrAB complex (Figure 41C) having dimensions of 240 × 140 Å were interpreted as being of a dimeric nature. Moreover, it was considered that the side views with a height of 130 Å had an additional dimeric organization. Figure 41D shows a 3D map of the complex in the same orientation as in Figure 41C with a resolution of 30 Å. Based on the volume of the map, the molecular mass of the complex was estimated to be about 268 kDa. Overall, based on these observations, it was postulated that the EmrAB complex was organized in the following manner: 2 × (2 × (42 kDa EmrA + 56 kDa EmrB)) = 392 kDa (being in the same range as the previously estimated value). Figure 41E represents reprojections through the 3D map of the complex. Averages of the smaller particles attributed to free EmrA were described as having a 'doughnut' shape composed of two cylinders joining at one end (Figure 41F). As a size and shape comparison, the crystal structure of MexA was modeled in a dimeric form (colored in red) on the third average. Therefore, it was hypothesized that the excess EmrA within the preparation was present as a dimer. Finally, the negative staining EM analysis of EmrB alone (Figure 41B) as a negative control also supported that finding.

3.2.2.2. Postulated models of the EmrAB-TolC and FarAB-MtrE systems

With the help of the known crystal structures of various protein partners it has been possible to propose several hypothetical models of tripartite MFS-type systems (Figure 42). Starting from the left, a first model of EmrAB-TolC was proposed by Hinchliffe and co-workers in 2014 (Figure 42A) after the crystallization of EmrA from *Aquifex aeolicus* (modeled as a hexamer)[215] and using the crystal structures of trimeric TolC[216] as well as $PepT_{So}$ from *Shewanella oneidensis* [200] (to model the contours of a hypothesized EmrB monomer). In 2015, Symmons and co-workers proposed a second model of EmrAB-TolC[217] (Figure 42B) based on the first one but replacing the $PepT_{So}$ contour from *Shewanella oneidensis* by the structure of EmrD[199] (modeled as a dimer). Finally, Shafer and co-workers proposed in 2016 a model of FarAB-MtrE from *Neisseria gonorrhoeae* [218] (Figure 42C) using the crystal structures

of EmrA from *Aquifex aeolicus* (modeled as a hexamer), of trimeric MtrE[219] and EmrD (modeled as a dimer).

Finally, the stoichiometry of EmrAB and the homologous FarAB in the different models does not correspond to the previously mentioned dimer of dimers identified by Tanabe and co-workers[214]. In addition, it is interesting to note that in all three models the interaction zones between the periplasmic adaptor and the outer membrane exit duct are quite substantial (*i.e.* the outer membrane exit duct is inserted quite far in the hexametric adaptor in a so-called 'deep-interpenetration model').

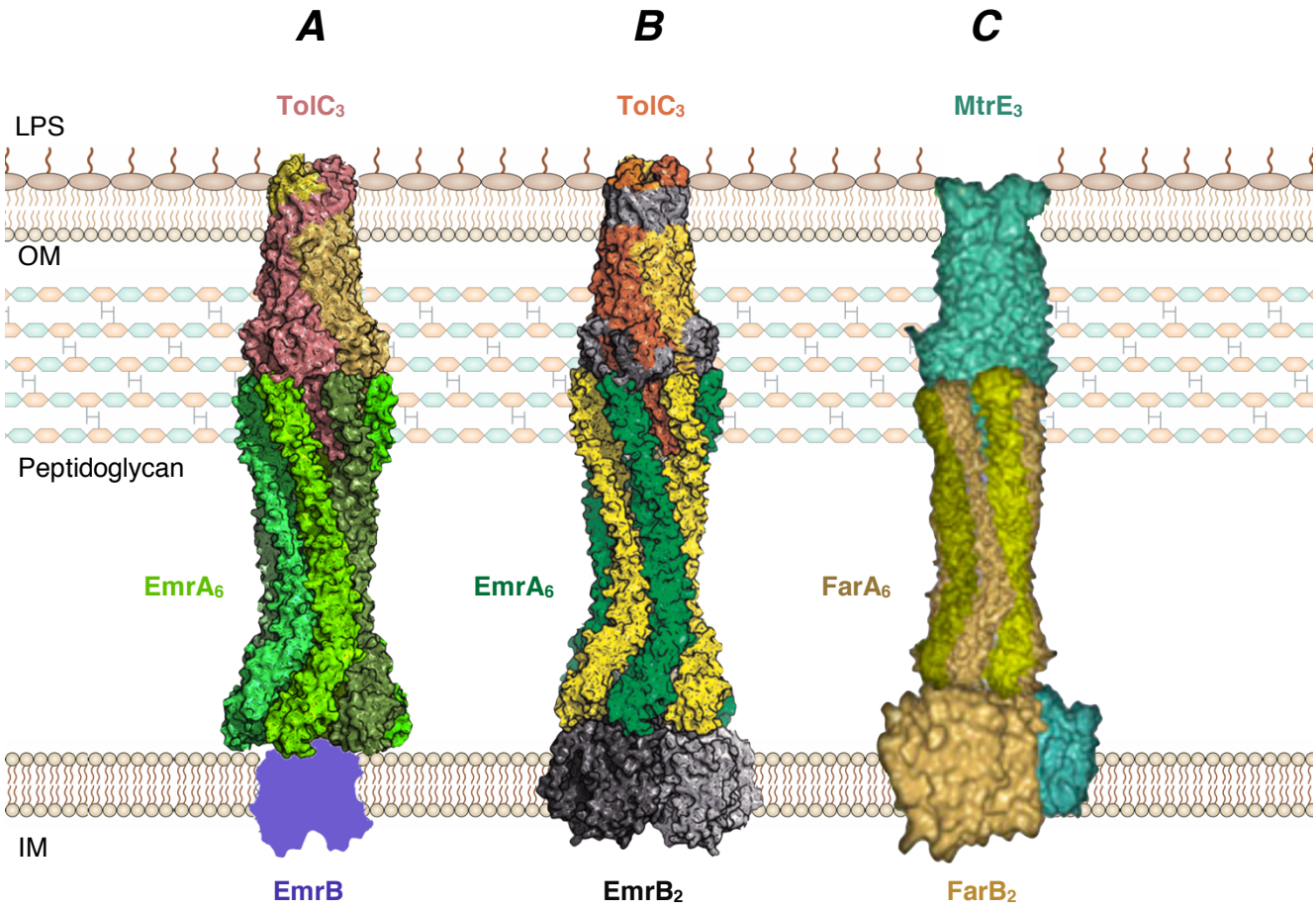

Figure 42. Models of tripartite MFS-type systems.

(A) The first model of the EmrAB-TolC system composed of the trimeric TolC exit duct, a hexameric EmrA and a monomer of EmrB. (B) A second model of the EmrAB-TolC system similar to the first model but containing a dimeric EmrB. (C) A model of the FarAB-MtrE system from *Neisseria gonorrhoeae* composed of the trimeric MtrE exit duct, a hexameric FarA and a dimeric FarB. LPS, lipopolysaccharide; OM, outer membrane; IM, inner membrane. Adapted from references 215, 217 and 218.

3.2.2.3. Structural review of the protein partners forming the EmrAB-TolC system

3.2.2.3.1. The inner membrane drug:H^+ antiporter EmrB

Contrary to the previously mentioned single component MFS members, no structural information is available about EmrB. In fact, as mentioned previously it is only known that the drug:H^+ antiporter is composed of 14 TMSs. However, it presumably presents a typical MFS fold composed of the 12 TMSs with two additional TMSs located in between the N- and C-terminal domains. Therefore, as shown in 3.2.2.2., it can be compared to $PepT_{So}$ from *Shewanella oneidensis* [200] which also contains 14 TMSs even if Symmons and co-workers compared it to EmrD in the previously described second model of EmrAB-TolC[217]. Nevertheless, care must be taken when considering the exact packing and position of the 2 additional TMSs as $PepT_{So}$ from *Shewanella oneidensis* is not a homologue of EmrB.

3.2.2.3.2. The outer membrane exit duct TolC

The first crystal structure of TolC (at a resolution of 2.1 Å) was obtained 19 years ago describing the closed conformation of TolC[220]. Since then other crystal structures were obtained of a ligand blocked TolC in 2004[221], of a partially open TolC double mutant (Y362F and R367E) in 2008[222], and of an early stage open TolC single mutant (R367S) as well as an advanced stage open TolC double mutant (Y362F and R367S) in 2011[216].

The following description mainly focuses on the first crystal structure and describes its main characteristics.

The various structural features of TolC are shown in Figure 43. The outer membrane exit duct is trimeric (Figure 43A). Overall, it has a length of 140 Å, divided into a 40 Å long outer membrane domain and a 100 Å long periplasmic domain. The outer membrane domain is a constitutively open β-barrel (also called the 'channel') whereas the periplasmic domain is a α-helical barrel (also called the 'tunnel') which is constricted near its end. Therefore, TolC is in the present case in a resting closed conformation. The internal diameter for the most part is constant at 35 Å. Finally, the additional external α-helices and β-strands forming a belt at the moiety of the α-helical barrel constitute the so called 'equatorial domain'.

Because of these observations, each of the three protomers can been divided into three separate domains (Figure 43B) defined as the β-, α-helical, and mixed α/β-domains. The topological view of a protomer (Figure 43C) shows that for the most part, the polypeptide backbone moves sequentially up and down four times with a transition from outer membrane β-strands to periplasmic α-helices. Therefore, each protomer provides four antiparallel β-strands (S1, S2, S4, and S5) for the formation of the outer membrane β-barrel. Likewise, together with two long helices (H3 and H7) as well as two pairs of stacked short helices (H2/H4 and H6/H8) forming pseudo continuous long helices, each protomer contributes to the formation of the α-helical barrel. Finally, three additional external α-helices (H1, H5, and H9) as well as two additional external β-strands (S3 and S6) of each protomer form the equatorial domain.

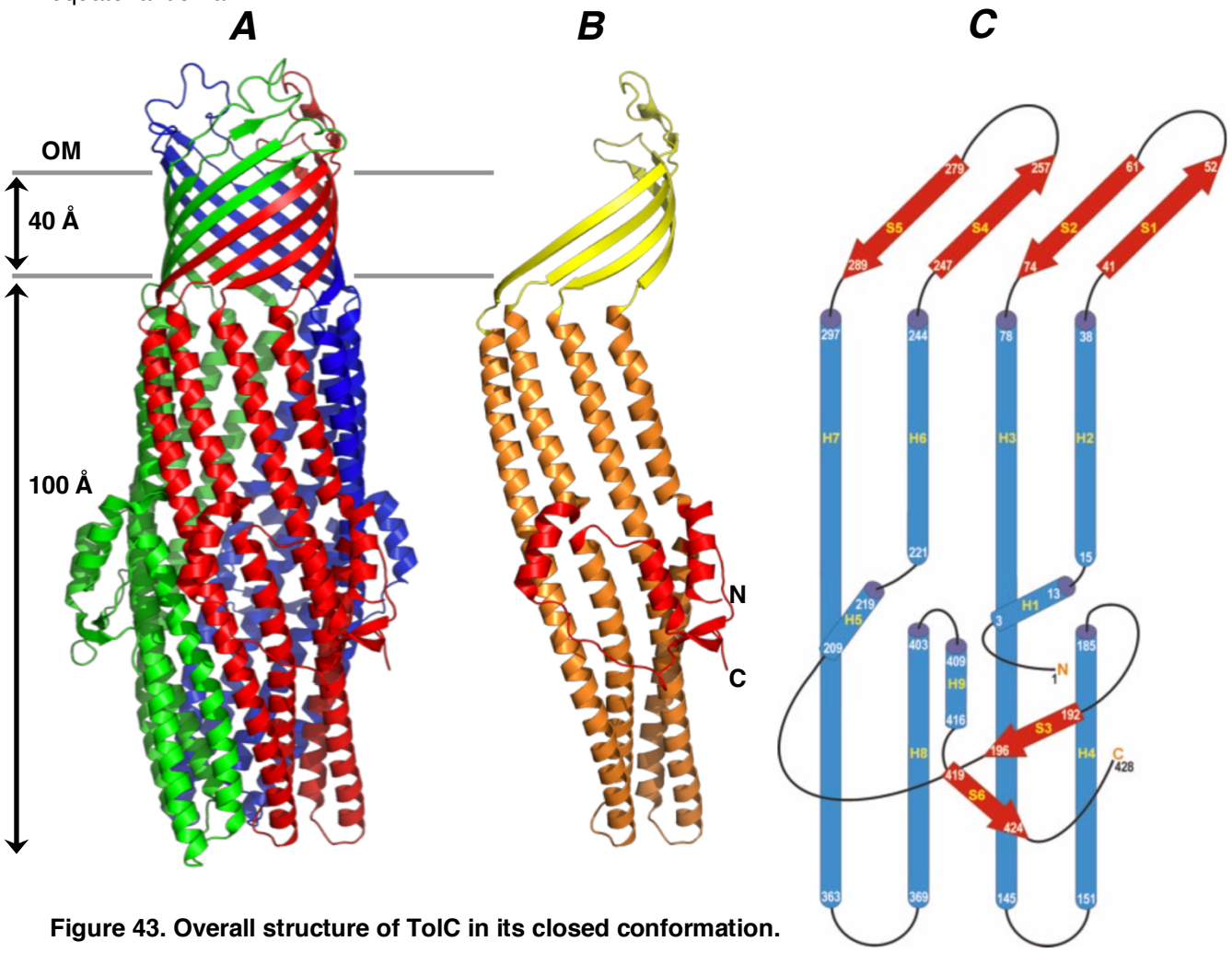

Figure 43. Overall structure of TolC in its closed conformation.

(A) The polypeptide backbone representation of TolC with the three protomers colored in red, blue and green. The outer membrane β-strands form a 12 stranded β-barrel and the periplasmic α-helices also form a 12 stranded barrel. At the mid-section of the α-helical barrel the additional external α-helices and β-strands form the equatorial domain. (B) Structural subdivision of the polypeptide backbone representation of the red protomer from (A) composed of the β-domain (colored in yellow), the α-helical domain (colored in orange) and the mixed α/β-domain (colored in red). The N- and C-termini are also indicated. (C) Topological representation of a single protomer. The β-strands and α-helices are colored in red and blue respectively. OM, outer membrane. Adapted from reference 220.

Considering the α-helical barrel further into detail, it is important to mention that the six pairs of helices involved in its formation are actually coiled-coils. In addition, the helix pairs (H7 and H8) as well as (H3 and H4) of the three protomers have been divided into inner and outer coiled-coils respectively. Therefore, the periplasmic entrance of TolC is closed as the inner coiled-coils fold inwards (Figure 44A). Furthermore, an early on identified aspartate ring with a diameter of 3.9 Å (comprised by D371 and D374 from H8 forming hydrogen bonds) represents the narrowest pore constriction of the α-helical barrel (Figure 44B)[223,224,221].

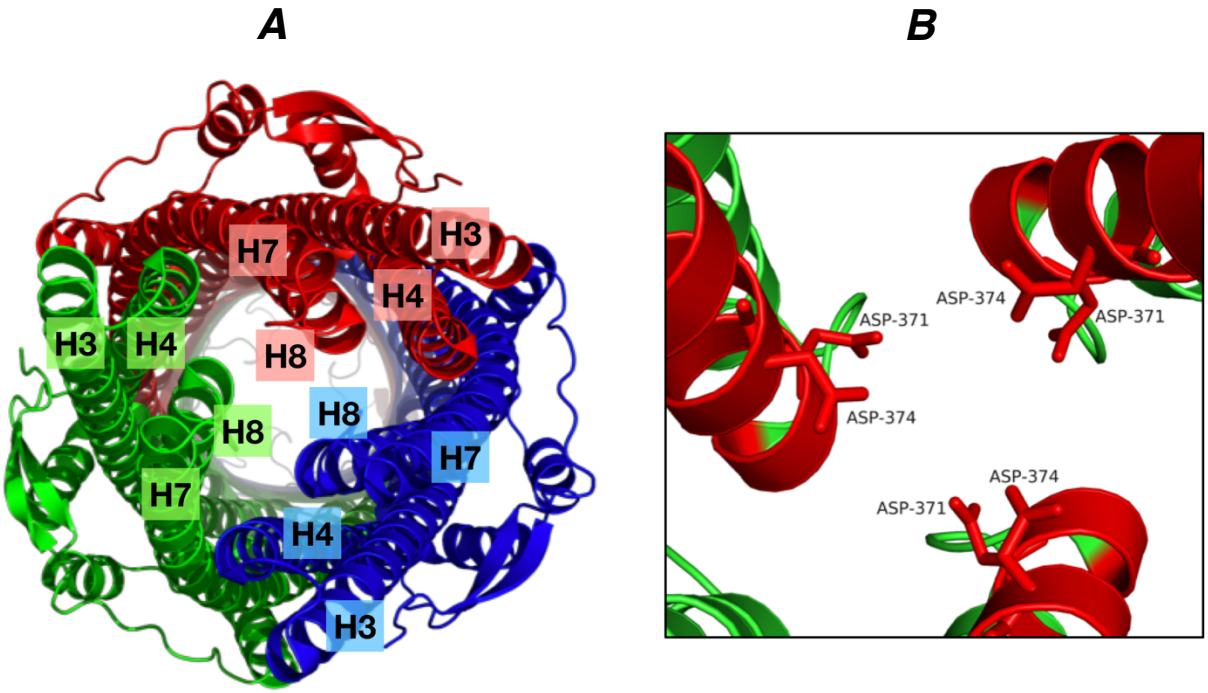

Figure 44. External and internal views of the periplasmic end of the α-helical barrel of TolC in its closed conformation.

(A) View of the periplasmic end of the α-helical barrel from the periplasmic side. The protomers are colored in the same manner as in Figure 43A. The different helices are labeled. The inner coiled-coils are folded towards the center of the pore closing the periplasmic entrance of TolC. (B) The aspartate ring formed by Asp-371 and Asp-374 of the H8 of each protomer in a stick view colored in red viewed from the interior of the α-helical barrel towards the periplasm. The lower ends of the H8 of each protomer are also colored in red. The neighboring part of the structure partially visible is colored in green.

Later on, a second hydrogen bond network located below the aspartate ring was described being composed of R367 forming interprotomer bonds with T152 and D153 linking H8 to H4, as well as Y362 forming an intraprotomer bond with D153 linking the inner coiled coil to the outer coiled coil (Figure 45 (lower center), cross-section 2). In addition, a relatively unconstricted glycine ring (underneath the

second hydrogen bond network) formed by G365 located within the loop between H7 and H8 was also identified (Figure 45 (lower right), cross-section 3). Overall, it is considered that the opening of the TolC α-helical barrel from the periplasmic side occurs with the most notable displacements observed for the inner coiled-coils repacking towards the exterior in a so called 'iris-like movement'. Disruption of the second hydrogen bond network by mutating both R367 as well as Y362 has been demonstrated to be key for the opening of TolC. Finally it is in general considered that the complete opening of the TolC tunnel is most probably triggered by the interaction with an inner membrane transporter/periplasmic adaptor complex[222,216].

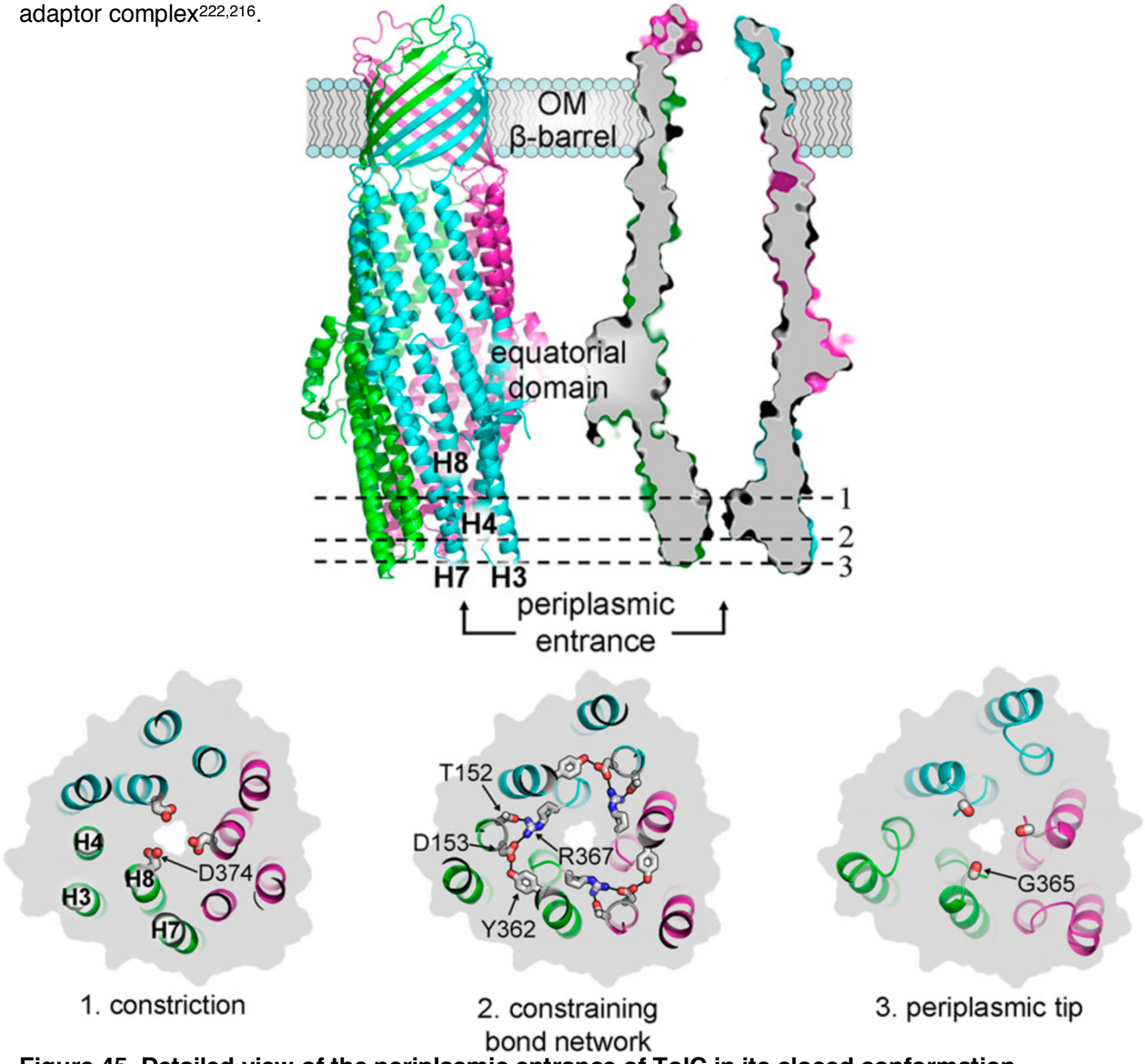

Figure 45. Detailed view of the periplasmic entrance of TolC in its closed conformation.

(Upper left) The overall structure of TolC with the three protomers colored in green, blue and magenta. (Upper right) A sliced view of the surface representation of TolC showing the central continuous pore and the constrictions. The dashed lines labeled 1, 2 and 3 represent different cross sections. (Lower panels) Periplasmic views of the three different cross sections shown in the upper right view. The grey background in each case corresponds to the outline of the surface representation of TolC. (Lower left) Cross section 1 at the level of the aspartate ring constriction. (Lower center) Cross section 2 at the level of the second hydrogen bond network constriction. (Lower right) Cross section 3 at the level of the glycine ring at the periplasmic tip. Adapted from reference 216.

3.2.2.3.3. The periplasmic adaptor protein EmrA

The exact structure of EmrA from *Escherichia coli* is not available yet. Nevertheless, as mentioned previously (in 3.2.2.2.) the crystal structure of the homologous EmrA from *Aquifex aeolicus* is known and was obtained in 2014 (at a resolution of 2.85 Å) describing for the first time the adaptor fold of an MFS-type efflux system[215].

The different structural features of EmrA from *Aquifex aeolicus* are shown in Figure 46. The adaptor was crystallized without the N-terminal TMS region anchoring it into the inner membrane. The periplasmic region of the adaptor can be subdivided into three domains: the β-barrel, lipoyl, and α-helical coiled-coil domains. Overall the periplasmic region has a length of 185 Å.

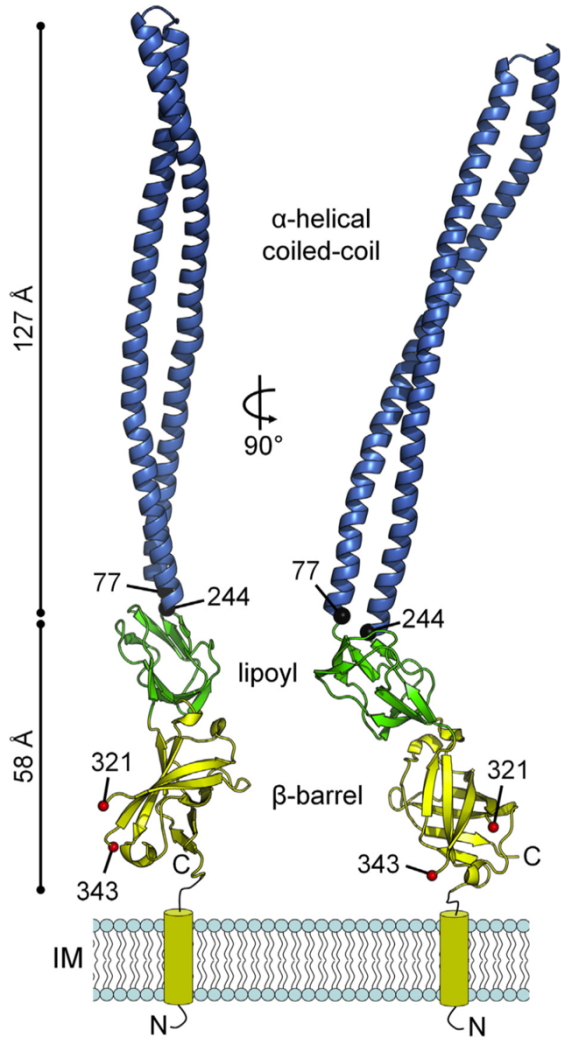

Figure 46. Global structural fold of EmrA from *Aquifex aeolicus*.

The N-terminal TMS is represented as a yellow tube. The β-barrel, lipoyl, and α-helical coiled-coil domains are colored in yellow, green and blue respectively. The red spheres at residues 321 and 343 indicate the beginning and end of the 322-342 loop region which could not be solved due to poorly defined density. The N- and C-termini are indicated. IM, inner membrane. Adapted from reference 215.

The β-barrel domain is composed of 7 antiparallel β-strands and 3 short α-helices. The lipoyl domain is described as a 'β-sandwich of 2 interlocking motifs of 4 antiparallel β-strands' and was found to be structurally homologous to biotinyl/lipoyl carrier domains of dehydrogenase enzymes. Finally, the α-helical coiled coil domain is composed of 2 antiparallel α-helices[217].

It is interesting to note that when the overall structure of EmrA from *Aquifex aeolicus* is compared to the structures of other adaptors, the main difference evidenced is the lack of a membrane proximal domain in EmrA (Figure 47) which correlates well with the absence of a periplasmic domain in EmrB being entirely submerged in the inner membrane. This domain present in other adaptors (interacting with inner membrane RND and ABC transporters having periplasmic protrusions) is a so-called β-roll with a similar topology to the β-barrel domain resulting from a possible domain duplication event[215,217].

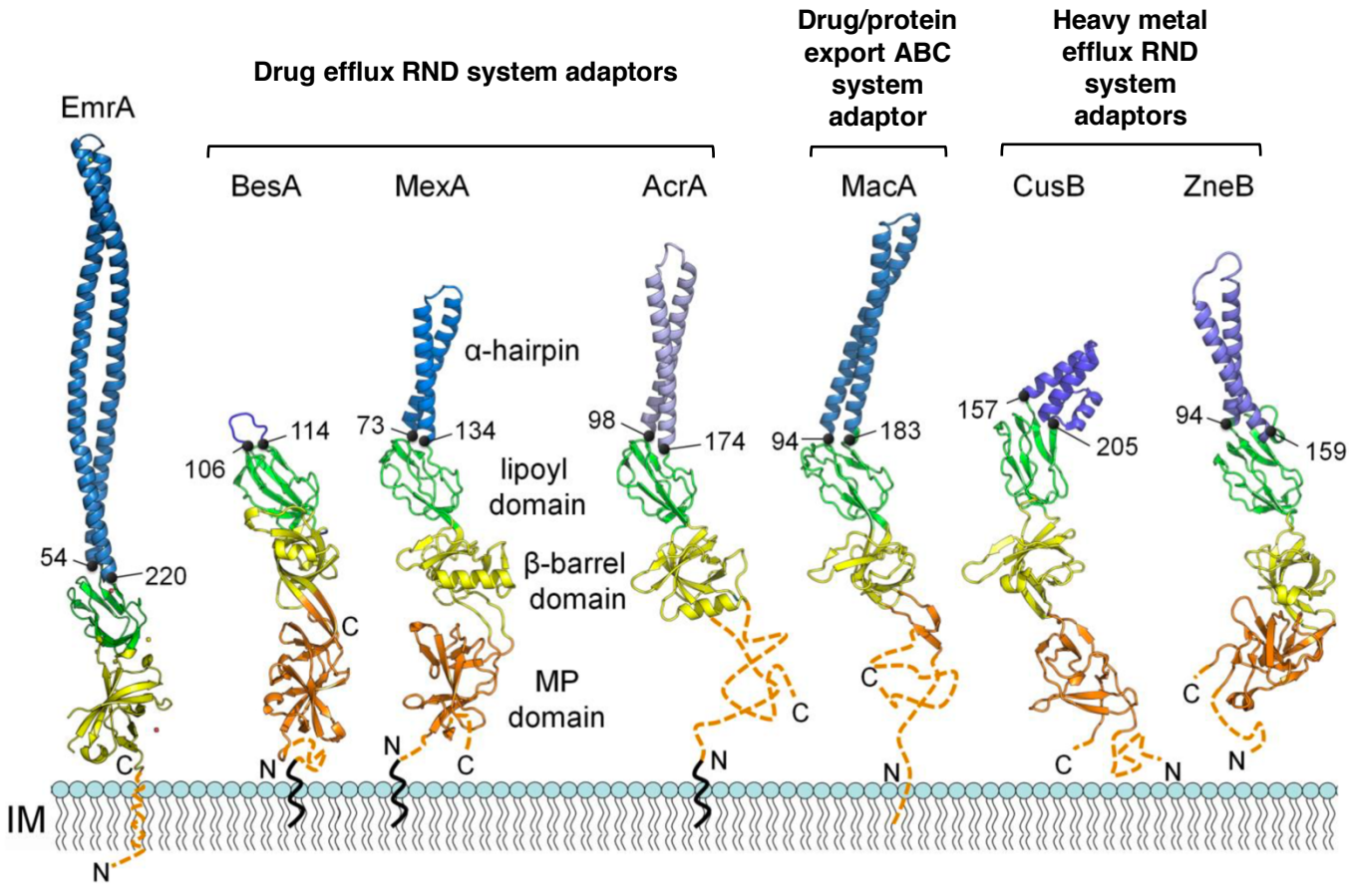

Figure 47. Global structural comparison of different periplasmic adaptors.

The membrane proximal (MP), β-barrel, lipoyl, and α-helical coiled-coil domains are colored in orange, yellow, green, and blue respectively. Dotted orange lines indicate the unobserved N- and C-terminal regions. The compared adaptors with known structures are BesA (*Borrelia burgdorferi*, PDB ID: 4KKS), MexA (*Pseudomonas aeruginosa*, PDB ID: 2V4D), AcrA (*Escherichia coli*, PDB ID: 2F1M), MacA (*Escherichia coli*, PDB ID: 3FPP), CusB (*Escherichia coli*, PDB ID: 3OOC), ZneB (*Cupriavidus metallidurans*, PDB ID: 3LNN). Note the absence of a MP domain in EmrA, the absence of an α-helical coiled-coil domain in BesA and the presence of a third α-helix in the α-helical coiled-coil domain of CusB. IM, inner membrane. Adapted from reference 215.

The α-helical coiled-coil domains from different adaptors can be well aligned structurally (Figure 48) indicating that EmrA might interact with TolC via this region in a similar fashion as the other adaptor types.

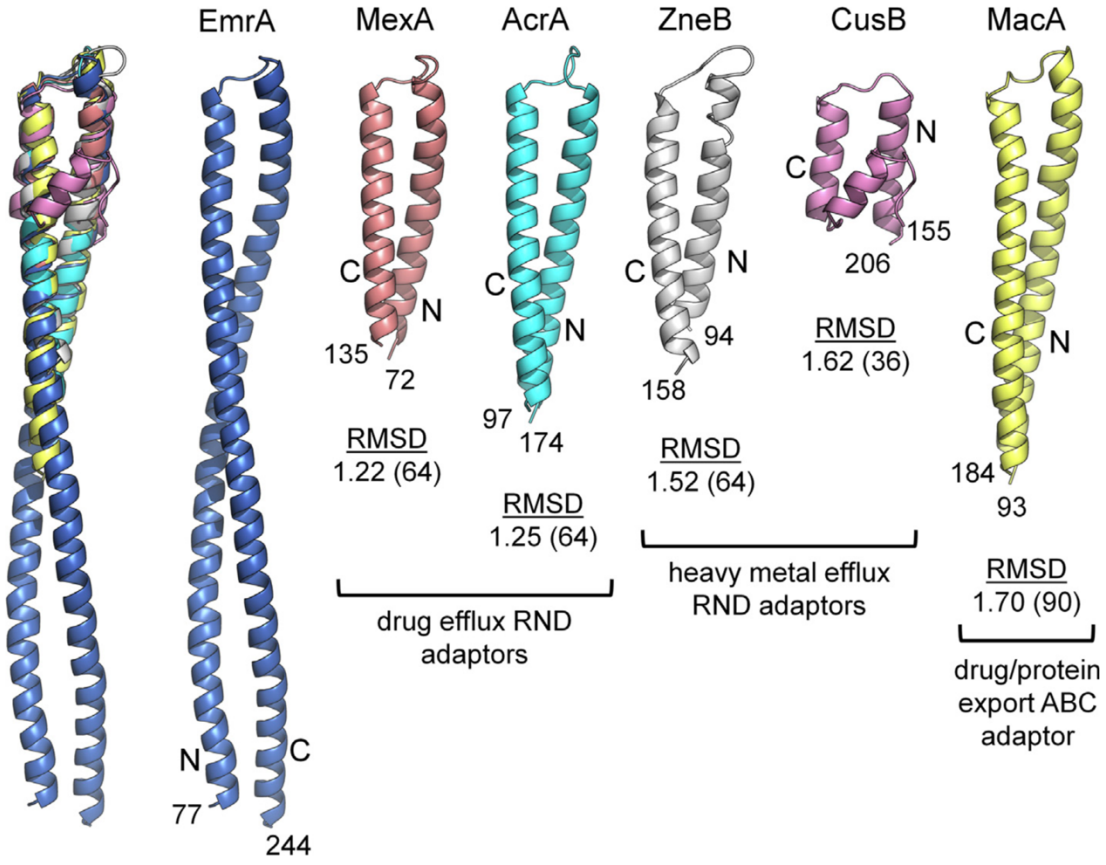

Figure 48. Comparison of the α-helical coiled-coil domains of different adaptors.

(Left) Structural overlay of the α-helical coiled-coil domains of EmrA from *Aquifex aeolicus* (blue), MexA from *Pseudomonas aeruginosa* (red, PDB ID: 2V4D), AcrA from *Escherichia coli* (cyan, PDB ID: 2F1M), ZneB from *Cupriavidus metallidurans* (grey, 3LNN), CusB from *Escherichia coli* (purple, 3OOC), and MacA from *Escherichia coli* (yellow, 3FPP). (Right) Adaptors compared to EmrA with the root mean square deviations (RMSDs) indicated in Å of the structural alignments. The number of residues aligned are indicated in brackets. The N- and C-terminal sides are labeled. Adapted from reference 215.

Furthermore, it is interesting to note the important size difference between the α-helical coiled-coil domains of EmrA and other adaptors. However, care should be taken before the consideration of the similarity of this domain with the one of EmrA from *Escherichia coli*. Indeed, a sequence alignment shown in Figure 49 suggests that this α-helical coiled-coil domain must be shorter in EmrA from *Escherichia coli*.

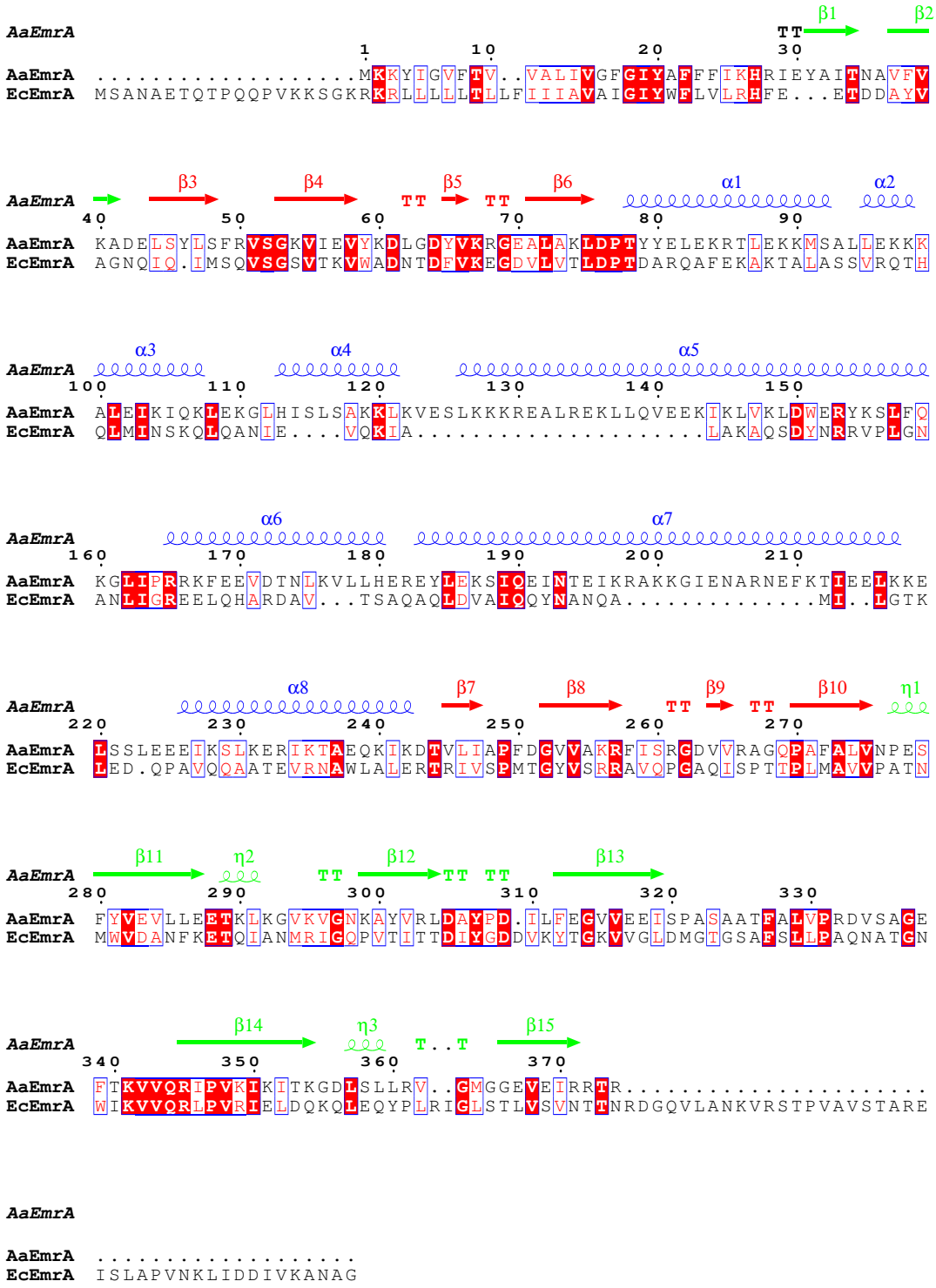

Figure 49. Sequence alignment of EmrA from *Aquifex aeolicus* (AaEmrA) and from *Escherichia coli* (EcEmrA).

Identical residues are marked in white on a red background. Highly conserved residues are marked in red on a white background. The secondary structure of EmrA from *Aquifex aeolicus* is indicated above and colored according to the different domains (β-barrel, green; lipoyl, red; and α-helical coiled-coil, blue).

The lipoyl and β-barrel domains of EmrA also align well structurally with the same domains of other adaptors. Interestingly, when the β-barrel domains of different adaptors are overlaid with the one from EmrA (Figure 50), the modeled 23 Å long loop region of the β-barrel domain of EmrA is approximately more than twice as long as the loop region from other adaptors. Furthermore, the high conservation of amino acid residues in this region could possibly indicate that it may have a functional importance in the MFS-type efflux systems[215].

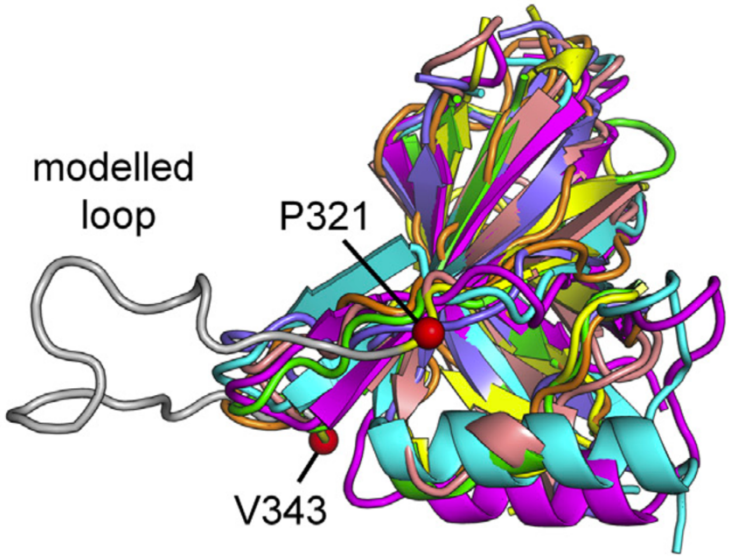

Figure 50. Structural overlay of β-barrel domains from different adaptors.

EmrA (yellow), MexA (cyan, PDB ID: 2V4D), AcrA (pink, PDB ID: 2F1M), ZneB (blue, PDB ID: 3LNN), CusB (red, PDB ID: 3OOC), BesA (green, PDB ID: 4KKS), and MacA (orange, PDB ID: 3FPP). Adapted from reference 215.

Studies in the literature showed that EmrA from *Escherichia coli* can form both dimers and trimers *in vitro* [184,225]. However, within the entire tripartite efflux system its oligomeric state together with the inner membrane transporter are still under debate as shown by the previous negative staining EM analyses as well as the proposed models.

Interestingly, it has been suggested that EmrA could possibly interact with EmrB via its N-terminal transmembrane helix[215]. Indeed, this would lead to the formation of a helical bundle in the inner membrane composed of the TMSs of EmrB and EmrA. Furthermore, using potassium iodide-induced quenching of four tryptophan residues present in the periplasmic region of EmrA in the absence and presence of drugs it has been suggested that EmrA could bind substrates transferring these from EmrB

to TolC[184]. Nevertheless, drug binding to EmrA could not be detected using isothermal titration calorimetry (ITC) or by crystallography in the study reporting the crystal structure of EmrA from *Aquifex aeolicus* [215].

3.2.3. Functional insights of tripartite MFS members

The first description of the *emr* locus already indicated that both EmrB and EmrA were required for resistance to CCCP[144]. Minimum inhibitory concentration (MIC) measurements in another study concerning the EmrAB from *Escherichia coli* also indicated that EmrB alone could not confer drug resistance[184]. Therefore, within the bacterium, EmrB must be associated to EmrA and TolC for the drug efflux function conferring resistance.

Contrary to the inner membrane, the outer membrane of gram-negative bacteria is a major barrier against hydrophobic compounds. Therefore, the association of EmrB to EmrA and TolC constitutes an efficient solution for the protection of the bacterium against hydrophobic noxious compounds which would directly be expelled from the cytoplasmic side to the extracellular medium[178]. It is likely that within the cell, such MFS-type tripartite efflux systems complement the resistance profile of a given organism. In fact, these systems could function together with RND and ABC type tripartite systems transporting substrates from the periplasmic space to the extracellular medium as well as single component MFS members but also (ABC, SMR, AbgT, PACE transporters) transferring drugs from the cytoplasmic side towards the periplasm.

Finally, even though the exact mechanistic details of the tripartite EmrAB-TolC system are not described, the transport mechanism at the level of EmrB must be similar to the previously described transport mechanism of single component MFS members[110].

4. Strategies employed for the isolation of the EmrAB-TolC system

As shown previously, only structural models of the tripartite system EmrAB-TolC and another homologous tripartite system FarAB-MtrE from *Neisseria gonorrhoeae* are available thus far. Therefore, the principle aim of the present Ph.D. project was to isolate the EmrAB-TolC system directly from bacterial cells and as stated earlier, study its structure using EM.

For the direct isolation purpose, two different cloning strategies were employed.

The first strategy consisted in the preparation of constructs for the co-expression and co-purification of the three protein partners forming the tripartite system. In this case, different affinity tags were added at the respective C-termini of all three protein partners. Furthermore, the inner membrane transporter EmrB and either the periplasmic adaptor EmrA or the outer membrane exit duct TolC were fluorescently labeled at their respective C-termini as well. The affinity tags and fluorescent labels were used as analytical indicators for different co-expression and co-purification trials.

The second strategy consisted in the preparation of constructs for the co-expression of EmrAB fusion chimeras together with TolC. Also here, different affinity tags were added at the respective C-termini of the EmrAB fusion chimeras as well as TolC. The affinity tags were used as analytical indicators during co-expression and co-purification trials. Compared to the first strategy, the aim here was to stabilize the tripartite assembly for the preparation of concentrated samples in a straightforward manner.

Finally, individual constructs using the three ORFs were also prepared to have an alternative solution in the case were none of the first two strategies would yield promising results.

METHODS AND RESULTS

*Chapter II: Identification of '*emrAB*' and '*emrAB-tolC*' operons and subsequent cloning*

In the present chapter, the identification of genes coding for EmrA and EmrB homologs from various organisms is described.

The FX-cloning methodology used overall in different types of strategies for the preparation of variable constructs will be described.

Firstly, a cloning strategy for the co-expression and co-purification trials of the EmrAB-TolC tripartite system with the help of fluorescent labels and affinity tags will be described.

A Second strategy was used in parallel to prepare various constructs for the co-expression of EmrAB fusion chimeras together with TolC.

Finally, an alternative strategy, namely the expression of each individual gene within *E. coli* for purification and subsequent reconstitution of the tripartite system will be presented.

1. Identification of emrAB-tolC *ORFs from* Escherichia coli *and other gram-negative bacteria*

Because EmrA and EmrB remain poorly studied in general, a so-called homologous screening strategy was first adopted in order to maximize the chances to obtain tripartite MFS-type systems which would be highly expressed and remain stable during purification procedures. For that purpose, based on information from the database (TransportDB)[226,227], numerous *emrAB*-type and *tolC*-type genes from different gram-negative bacteria putatively encoded via '*emrAB*' and '*emrAB-tolC*' operons were selected (Table 3).

Table 3. Various operons selected for subsequent PCR amplifications.

Number	Organism (strain)	Operon type* (efflux system)	Genes	NCBI gene ID	NCBI gene location	NCBI protein accession number
1	*Escherichia coli* (BL21DE3)	*emrAB* (EmrAB-TolC)	*emrB*	8116566	NC_012892.2 (2652576..2654114)	WP_001295176.1
			emrA	8112639	NC_012892.2 (2,651,387..2,652,559)	WP_001295175.1
			tolC	8113021	NC_012892.2 (3,045,546..3,047,027)	WP_000735278.1
2		*emrKY* (EmrKY-TolC)	*emrY*	8114560	NC_012892.2 (2,358,520..2,360,058)	WP_001018731.1
			emrK	8116571	NC_012892.2 (2,360,058..2,361,221)	WP_000435167.1
			tolC	8113021	NC_012892.2 (3,045,546..3,047,027)	WP_000735278.1

Number	Organism (strain)	Operon type* (efflux system)	Genes	NCBI gene ID	NCBI gene location	NCBI protein accession number
3	*Vibrio cholerae* (O395)	*vceABC* (VceABC)	*vceB*	5135707	NC_009457.1 (1,062,761..1,064,296)	WP_000019056.1
			vceA	5137704	NC_009457.1 (1,061,531..1,062,751)	WP_001087672.1
			vceC	5136171	NC_009457.1 (1,060,096..1,061,550)	WP_000798634.1
4	*Halomonas elongata* (DSM2581)	*emrAB* (EmrAB-TolC)	*emrB*	9747126	NC_014532.1 (4,028,578..4,030,125)	WP_013334195.1
			emrA	9747127	NC_014532.1 (4,030,122..4,031,189)	WP_013334196.1
			tolC	9747419	NC_014532.1 (2,506,524..2,507,972)	WP_013332884.1
5	*Shimwellia blattae* (DSM4481NBRC105725)	*emrAB* (EmrAB-TolC)	*emrB*	12956419	NC_017910.1 (915,415..916,950)	WP_002444409.1
			emrA	12954767	NC_017910.1 (916,966..918,141)	WP_002444407.1
			tolC	12953535	NC_017910.1 (569,561..570,967)	WP_002443304.1
6	*Desulfomonile tiedjei* (DSM6799)	*emrAB* (EmrAB-TolC)	*emrB*	13136809	NC_018025.1 (2,534,726..2,536,282)	WP_014809956.1
			emrA	13136808	NC_018025.1 (2,533,495..2,534,790)	WP_014809955.1
			tolC	13139409	NC_018025.1 (4,290,064..4,291,575)	WP_014811435.1

Number	Organism (strain)	Operon type* (efflux system)	Genes	NCBI gene ID	NCBI gene location	NCBI protein accession number
7	*Syntrophobacter fumaroxidans* (MPOB)	*emrAB* (EmrAB-TolC)	*emrB*	4458345	NC_008554.1 (4,059,195..4,060,775)	WP_011700115.1
			emrA	4458344	NC_008554.1 (4,057,937..4,059,208)	WP_011700114.1
			tolC	4458319	NC_008554.1 (4,145,189..4,146,685)	WP_011700176.1
8	*Thermovirga lienii* (DSM17291)	*emrAB* (EmrAB-TolC)	*emrB*	11377026	NC_016148.1 (1,927,381..1,928,886)	WP_014163964.1
			emrA	11377025	NC_016148.1 (1926267..1927321)	WP_052299903.1
			tolC	11377501	NC_016148.1 (461,411..462,781)	WP_014162590.1
9	*Thermocrinis albus* (DSM14484)	*emrAB-tolC* (EmrAB-TolC)	*emrB*	8813584	NC_013894.1 (93,874..95,400)	WP_012991144.1
			emrA	8813585	NC_013894.1 (95,397..96,521)	WP_012991145.1
			tolC	8813586	NC_013894.1 (96,518..97,783)	WP_012991146.1
10	*Klebsiella pneumoniae* (MGH78578)	*kpnGH-tolC* (KpnGH-TolC)	*kpnH*	5340206	NC_009648.1 (1,653,016..1,654,668)	WP_002902967.1
			kpnG	5340205	NC_009648.1 (1,651,902..1,652,987)	WP_015958345.1
			tolC	5340204	NC_009648.1 (1,650,526..1,651,905)	WP_004190382.1

Number	Organism (strain)	Operon type* (efflux system)	Genes	NCBI gene ID	NCBI gene location	NCBI protein accession number
11	*Salmonella enterica* (LT2)	*emrAB* (EmrAB-TolC)	*emrB*	1254338	NC_003197.2 (2,962,959..2,964,509)	NP_461741.1
			emrA	1254337	NC_003197.2 (2,961,770..2,962,954)	NP_461740.1
			tolC	1254709	NC_003197.2 (3,348,574..3,350,049)	NP_462101.3
12	*Cupriavidus metallidurans* (CH34)	*emrAB-tolC* (EmrAB-TolC n°1)	*emrB*	4038835	NC_007973.1 (2,193,553..2,195,166)	WP_011516743.1
			emrA	24152573	NC_007973.1 (2,192,292..2,193,533)	WP_011516742.1
			tolC	24152574	NC_007973.1 (2,190,731..2,192,245)	WP_035820801.1
13		*emrAB-tolC* (EmrAB-TolC n°2)	*emrB*	4041171	NC_007974.2 (906,714..908,261)	WP_011518802.1
			emrA	24154078	NC_007974.2 (908,306..909,577)	WP_011518803.1
			tolC	24154077	NC_007974.2 (909,607..911,073)	WP_035822472.1
14		*emrAB* (EmrAB-TolC n°3)	*emrB*	24153740	NC_007974.2 (1,282,836..1,284,473)	WP_080672771.1
			emrA	24153741	NC_007974.2 (1,281,697..1,282,839)	WP_011519088.1
			tolC	24153743	NC_007974.2 (1,279,530..1,281,017)	WP_011519086.1

Number	Organism (strain)	Operon type* (efflux system)	Genes	NCBI gene ID	NCBI gene location	NCBI protein accession number
15	*Cupriavidus metallidurans* (CH34)	*emrAB-tolC* (EmrAB-TolC n°4)	*emrB*	4042455	NC_007974.2 (2,344,887..2,346,461)	WP_011519999.1
			emrA	24152812	NC_007974.2 (2,346,454..2,347,599)	WP_011520000.1
			tolC	24152814	NC_007974.2 (2,343,427..2,344,884)	WP_011519998.1

* The operon type only describes the proximity of the different ORFs and does not represent the exact order (locations) in which the ORFs appear in the database.

As mentioned previously, the main rationale for the selection of the different ORFs was based on their organization into an operon supporting the hypothesis of a possible complex formation between the single protein components.

Based on amino acid sequence alignments of all the proteins using the multiple sequence alignment tool 'Clustal Omega' from the EMBL-EBI website[228,229] and the sequence manipulation suite 'SMS'[230,231], the identity and similarity values (compared to the EmrAB-TolC components from *E. coli*) for each of the three complex components from other gram-negative organisms were determined. These values are shown in Table 4.

Table 4. Identity and similarity values of the homologous proteins compared to the *E. coli* efflux system components.

EmrAB-TolC systems (Organism)	EmrB (*E. coli*)		EmrA (*E. coli*)		TolC (*E. coli*)	
	Identity (%)	Similarity (%)	Identity (%)	Similarity (%)	Identity (%)	Similarity (%)
EmrKY-TolC (*E. coli*)	EmrY 60.7	EmrY 73.8	EmrK 46.5	EmrK 62.9	TolC 100.0	TolC 100.0
VceABC (*V. cholerae*)	VceB 39.7	VceB 58.6	VceA 34.9	VceA 52.6	VceC 16.3	VceC 31.3
EmrAB-TolC (*H. elongata*)	EmrB 18.3	EmrB 36.3	EmrA 24.8	EmrA 46.8	TolC 25.8	TolC 44.2

EmrAB-TolC systems (Organism)	EmrB (*E. coli*)		EmrA (*E. coli*)		TolC (*E. coli*)	
	Identity (%)	Similarity (%)	Identity (%)	Similarity (%)	Identity (%)	Similarity (%)
EmrAB-TolC (*S. blattae*)	EmrB 90.6	EmrB 95.3	EmrA 74.7	EmrA 84.4	TolC 77.3	TolC 84.2
EmrAB-TolC (*D. tiedjei*)	EmrB 34.6	EmrB 56.2	EmrA 28.6	EmrA 42.0	TolC 12.6	TolC 25.9
EmrAB-TolC (*S. fumaroxidans*)	EmrB 33.5	EmrB 55.8	EmrA 31.9	EmrA 47.0	TolC 14.1	TolC 30.2
EmrAB-TolC (*T. lienii*)	EmrB 30.5	EmrB 51.8	EmrA 13.2	EmrA 25.1	TolC 16.7	TolC 32.5
EmrAB-TolC (*T. albus*)	EmrB 24.1	EmrB 45.2	EmrA 19.6	EmrA 37.0	TolC 16.7	TolC 29.4
KpnGH-TolC (*K. pneumoniae*)	KpnH 16.3	KpnH 34.5	KpnG 26.6	KpnG 44.3	TolC 17.0	TolC 33.8
EmrAB-TolC (*S. enterica*)	EmrB 95.7	EmrB 97.7	EmrA 90.0	EmrA 93.9	TolC 89.1	TolC 92.7
EmrAB-TolC n°1 (*C. metallidurans*)	EmrB 53.5	EmrB 66.7	EmrA 41.3	EmrA 56.5	TolC 16.4	TolC 32.6
EmrAB-TolC n°2 (*C. metallidurans*)	EmrB 47.7	EmrB 68.1	EmrA 42.6	EmrA 57.2	TolC 15.3	TolC 30.0
EmrAB-TolC n°3 (*C. metallidurans*)	EmrB 25.6	EmrB 46.6	EmrA 28.5	EmrA 43.6	TolC 14.9	TolC 28.3
EmrAB-TolC n°4 (*C. metallidurans*)	EmrB 27.4	EmrB 50.4	EmrA 29.9	EmrA 48.5	TolC 14.5	TolC 28.9

Given the different identity and similarity values, it is interesting to note that some systems seem to be closely comparable to EmrAB-TolC from *E. coli* whereas other systems are showing a more distant relationship. In general, a 30% identity threshold is used as a rule of thumb to confirm the homology between two proteins. However, this rule of thumb is actually too conservative and lower identity values do not exclude homology[232]. In addition, these greater differences (or greater evolutionary distances) would obviously be of great interest considering the subsequent screening analyses.

2. Fragment eXchange (FX) cloning

In order to generate numerous variable constructs in a straightforward manner, the so-called FX cloning methodology has been employed overall.

This efficient cloning tool was introduced in 2011 by Eric R. Geertsma and Raimund Dutzler[233] (Figure 51). The methodology uses the interesting properties of a type IIS restriction endonuclease namely SapI. This enzyme recognizes a nonpalindromic recognition site and cuts the DNA outside the recognition site (Figure 51A). Here, the main advantages are: (i) the production after cleavage of a variety of overhangs enabling directional cloning with one restriction endonuclease only and (ii) the diminution of the risks of gene truncations due to the occurrence in some cases of internal restriction sites compatible with the restriction sites of the plasmid.

The overall FX cloning strategy is divided into two steps (Figure 51B and C). During a so-called first 'initial cloning' step the ORF is amplified using primers introducing SapI restriction sites at the respective 5' ends. These sites are oriented in a manner to enable their removal after cleavage by the restriction endonuclease leaving only two overhangs of 3 nucleotides (Figure 51D). In order to ensure the directionality of the cloning procedure and to prevent self-ligation of the plasmid backbone the incompatible sequences AGT and GCA coding for serine and alanine respectively were selected. Therefore, the digested ORF can be correctly inserted in an intermediate sequencing plasmid named pINITIAL containing identical restriction sites in the same direction as the ORF flanking the counterselection gene *ccdB*. Thus, after the insertion of the ORF in the pINITIAL plasmid backbone the SapI restriction sites are conserved. During a so-called 'subcloning' step, the pINITIAL plasmid containing the ORF is simultaneously digested with a second plasmid used for protein expression named pEXPRESSION also having a *ccdB* gene with flanking SapI restriction sites. However, in contrast to the pINITIAL plasmid the restriction site directions are reversed in a manner to be removed after the digestion and ensure the addition of three nucleotides only to each end of the ORF. The second counterselection gene *sacB* is used to avoid the selection of clones having both pINITIAL and pEXPRESSION with the help of sucrose. Finally, the attractivity of the technique can be explained by the fact that both steps are realized in one cup without the need for an intermediate purification step of digested products. In addition, a single ORF can be inserted in multiple pEXPRESSIONs simultaneously for the screening of parameters such as promoter systems as well as affinity tag types and locations.

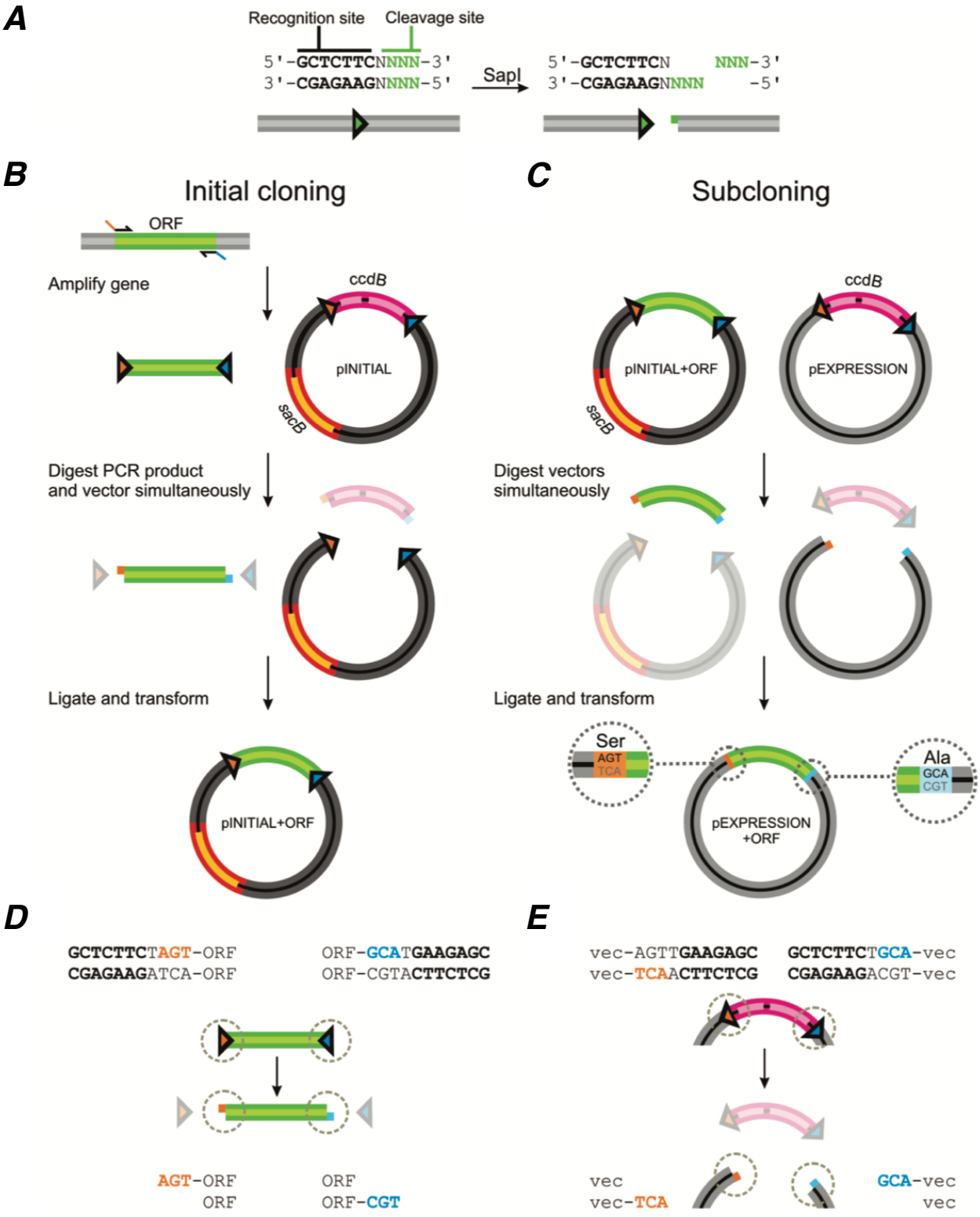

Figure 51. Schematic description of the FX cloning methodology.

(A) The SapI restriction site composed of the recognition site in bold letters and the cleavage site colored in green. N represents any of the four nucleotides. A schematic view of the cleavage is shown underneath. The arrow indicates the direction of the restriction site. (B) First cloning step of the amplified ORF colored in green into pINITIAL. The arrows indicating the directions of the SapI restriction sites are colored in orange and blue according to the three nucleotide overhangs added corresponding to the AGT and GCA sequences respectively. The counterselection marker genes *ccdB* and *sacB* are colored in magenta and orange respectively. (C) The second subcloning step from pINITIAL to pEXPRESSION. The color code is the same as in (B). The insets show the remaining additional nucleotides after the subcloning step. (D) Direction of the SapI restriction sites in the ORF and pINITIAL. (E) Direction of the SapI restriction sites in pEXPRESSION. vec, vector. Adapted from reference 233.

3. First cloning strategy for the subsequent use of the properties of fluorescent labels and affinity tags for screening purposes

In order to be able to choose the MFS-type tripartite efflux system amongst the different homologs selected previously that would be the most appropriate for structural studies (*i.e.* be both highly expressed and remain stable during purification procedures), a cloning strategy with the use of specific plasmid backbones for the three ORFs (which has been developed in the laboratory of Klaas Martinus Pos) was adopted. Thus, the following sections describe the overall methodology and the corresponding results that were obtained.

3.1. Cloning methodology for the production of proteins fused to fluorescent labels and affinity tags

During a first step, the various ORFs were amplified from genomic DNA via PCR. For that purpose, specific FX cloning primers were used for the subsequent insertion of the ORFs into pINITIAL (Table 5).

Table 5. Primers used for the insertion of the ORFs into pINITIAL.

Organism	Primer name*	Primer sequence (5' → 3')
E. coli	*emrB*-SapI-FW	ATATATGCTCTTCTAGTCAACAGCAAAAACCGCTGGA
	emrB-SapI-RV	TATATAGCTCTTCATGCGTGCGCACCGCCTCCGCCGC
	emrA-SapI-FW	ATATATGCTCTTCTAGTAGCGCAAATGCGGAGACTCA
	emrA-SapI-RV	TATATAGCTCTTCATGCGCCAGCGTTAGCTTTTACGA
	tolC-SapI-FW	ATATATGCTCTTCTAGTAAGAAATTGCTCCCCATTCT
	tolC-SapI-RV	TATATAGCTCTTCATGCGTTACGGAAAGGGTTATGAC
	emrY-SapI-FW	ATATATGCTCTTCTAGTGCAATCACTAAATCAACTC
	emrY-SapI-RV	TATATAGCTCTTCATGCCCCAACGCCTTTCGCTGTAAAC
	emrK-SapI-FW	ATATATGCTCTTCTAGTGAACAGATTAATTCAAATA
	emrK-SapI-RV	TATATAGCTCTTCATGCAAGTTGTCCATTATGCGAA

Organism	Primer name*	Primer sequence (5' → 3')
V. cholerae	*vceB*-SapI-FW	ATATATGCTCTTCTAGTAGTCATAACGCTGACAATGA
	vceB-SapI-RV	TATATAGCTCTTCATGCATGAACAGCAGAGGTATCCA
	vceA-SapI-FW	ATATATGCTCTTCTAGTAATTCAAATAATAGCAACAC
	vceA-SapI-RV	TATATAGCTCTTCATGCGCCTTGCTCTGATACTTTGG
	vceC-SapI-FW	ATATATGCTCTTCTAGTAAAAATAGCGTTCAAACGGT
	vceC-SapI-RV	TATATAGCTCTTCATGCAGATTCTGTTGTTTCAAAAC
H. elongata	*emrB*-SapI-FW	ATATATGCTCTTCTAGTTCCCTGCGACTTGTCCTCGGGC
	emrB-SapI-RV	TATATAGCTCTTCATGCTATGGGTCTGGGGCGGTTGAACC
	emrA-SapI-FW	ATATATGCTCTTCTAGTACCGATACCCAGACGCCCCAAGC
	emrA-SapI-RV	TATATAGCTCTTCATGCGGCGTCGGTATGGATGGTCGTC
	tolC-SapI-FW	ATATATGCTCTTCTAGTATGCCTTCTCGTCCCGTGACGCC
	tolC-SapI-RV	TATATAGCTCTTCATGCAGCATCGGGCCGGGGGGGTTCAC
S. blattae	*emrB*-SapI-FW	ATATATGCTCTTCTAGTGCACAGAAACCGCTGGAAGGCAC
	emrB-SapI-RV	TATATAGCTCTTCATGCGTGAGCGCCCCCGGCGCCGCCAC
	emrA-SapI-FW	ATATATGCTCTTCTAGTACCGGGTACTCGGATACTCAGGC
	emrA-SapI-RV	TATATAGCTCTTCATGCGCCAGCGTTTGTGCTGATTATC
	tolC-SapI-FW	ATATATGCTCTTCTAGTAAGAAACTGCTCCCACTACTTAT
	tolC-SapI-RV	TATATAGCTCTTCATGCCTGTAGCGGTACGTTCGCCATGC
D. tiedjei	*emrB*-SapI-FW	ATATATGCTCTTCTAGTGACAGGCCCCAGACAAACAAATG
	emrB-SapI-RV	TATATAGCTCTTCATGCGTGTGACGGCAAACCCGATTTAC
	emrA-SapI-FW	ATATATGCTCTTCTAGTGAATTTAAGCATGCTGACGCTAT
	emrA-SapI-RV	TATATAGCTCTTCATGCATGAGGGTTGGGAGCATCACTGC
	tolC-SapI-FW	ATATATGCTCTTCTAGTATAAGCGCTGTCAAATATAAAC
	tolC-SapI-RV	TATATAGCTCTTCATGCATCCTGGGGAACAGGTTCCTGG

Organism	Primer name*	Primer sequence (5' → 3')
S. fumaroxidans	*emrB*-SapI-FW	ATATATGCTCTTCTAGTCGGCGCCCTGAAACCATGAACG
	emrB-SapI-RV	TATATAGCTCTTCATGCATGCACGCCCTCCTTCCACCCG
	emrA-SapI-FW	ATATATGCTCTTCTAGTGCGGAAAACCAGAATCCGGAAAC
	emrA-SapI-RV	TATATAGCTCTTCATGCGGGCGCCGCATTCGAGGGAGG
	tolC-SapI-FW	ATATATGCTCTTCTAGTAAAAGTCAATCGGGGCGCGGTG
	tolC-SapI-RV	TATATAGCTCTTCATGCTTCATGCTTGGCCTCCAACC
T. lienii	*emrB*-SapI-FW	ATATATGCTCTTCTAGTATATTGGGAACGTTCATAG
	emrB-SapI-RV	TATATAGCTCTTCATGCTTCCATTAAACTCCTTATTTC
	emrA-SapI-FW	ATATATGCTCTTCTAGTGATAAGGAAACCGCCCAAC
	emrA-SapI-RV	TATATAGCTCTTCATGCGGGCGGCCTGGGCTTCCTTTG
	tolC-SapI-FW	ATATATGCTCTTCTAGTAACGTTCGTAGTAAATTTTTTG
	tolC-SapI-RV	TATATAGCTCTTCATGCGTTTCCGTTGCTTTGGAGC
T. albus	*emrB*-SapI-FW	ATATATGCTCTTCTAGTAGGGAAAATCTGCTTCTGAC
	emrB-SapI-RV	TATATAGCTCTTCATGCATCAGCTGGTGGGACGCGC
	emrA-SapI-FW	ATATATGCTCTTCTAGTAAAAAGCACGTGGCCATCAC
	emrA-SapI-RV	TATATAGCTCTTCATGCTGGAGATCGCTTTATCTCC
	tolC-SapI-FW	ATATATGCTCTTCTAGTATGTGGTGCGTCTTACTCAC
	tolC-SapI-RV	TATATAGCTCTTCATGCCGGAAGGACCCCCGCAGCC
K. pneumoniae	*kpnH*-SapI-FW	ATATATGCTCTTCTAGTCCCCGTCGCCAGGATAACC
	kpnH-SapI-RV	TATATAGCTCTTCATGCTTTGCTGCCGGTAGCGCGTTG
	kpnG-SapI-FW	ATATATGCTCTTCTAGTAGTCAGCAGGATGCGGCCAAAC
	kpnG-SapI-RV	TATATAGCTCTTCATGCTGGTTGTTTCTCCGCACGG
	tolC-SapI-FW	ATATATGCTCTTCTAGTATCCGCCCGGTCGCCCTTG
	tolC-SapI-RV	TATATAGCTCTTCATGCGCTCCCTGATGAGTCAGAC

Organism	Primer name*	Primer sequence (5' → 3')
S. enterica	*emrB*-SapI-FW	ATATATGCTCTTCTAGTCAACAGCAAAAACCGCTGG
	emrB-SapI-RV	TATATAGCTCTTCATGCGTGCGCGCCGCCACCGCCG
	emrA-SapI-FW	ATATATGCTCTTCTAGTAGCGCAAATGCGGAGATCC
	emrA-SapI-RV	TATATAGCTCTTCATGCACCCGCGTTAGCCTGTACG
	tolC-SapI-FW	ATATATGCTCTTCTAGTCAAATGAAGAAATTGCTCC
	tolC-SapI-RV	TATATAGCTCTTCATGCATGCCGGAATGGATTGCCG
C. metallidurans	*emrB1*-SapI-FW	ATATATGCTCTTCTAGTGCAGATTCCATCACCACAG
	emrB1-SapI-RV	TATATAGCTCTTCATGCGTGCGCTCCGGCTGCTTCC
	emrA1-SapI-FW	ATATATGCTCTTCTAGTAGCAACAACCAGCAATCGG
	emrA1-SapI-RV	TATATAGCTCTTCATGCGGTGTTGGCAGGCTTGGCG
	tolC1-SapI-FW	ATATATGCTCTTCTAGTAATCTCTCCCCTTCCTCCAC
	tolC1-SapI-RV	TATATAGCTCTTCATGCGCCCCGGGTGCCCTTTTCC
	emrB2-SapI-FW	ATATATGCTCTTCTAGTCCAAATCCGATGACAAGCG
	emrB2-SapI-RV	TATATAGCTCTTCATGCGTGGGCAGCGGATGCCGCAC
	emrA2-SapI-FW	ATATATGCTCTTCTAGTACCACGACCAACCCGAACC
	emrA2-SapI-RV	TATATAGCTCTTCATGCCGAGGCCCTGGCGAGTGCC
	tolC2-SapI-FW	ATATATGCTCTTCTAGTGTTTGGCAATCACCCCAGG
	tolC2-SapI-RV	TATATAGCTCTTCATGCCCGGGTTGTGTCCGCGCTTTG
	emrB3-SapI-FW	ATATATGCTCTTCTAGTAGCGCCGACGTCAGCACGC
	emrB3-SapI-RV	TATATAGCTCTTCATGCATGGGCATCCGCAGACGGC
	emrA3-SapI-FW	ATATATGCTCTTCTAGTACAAGCACCACCACCGCGAG
	emrA3-SapI-RV	TATATAGCTCTTCATGCTGACTTGCTCGCTGCGGTG
	tolC3-SapI-FW	ATATATGCTCTTCTAGTGCATCGAAAGCACTTCCCG
	tolC3-SapI-RV	TATATAGCTCTTCATGCGGACCATCCACCTCCCAGC
	emrB4-SapI-FW	ATATATGCTCTTCTAGTGCTGAAGCGGCGGCAATCG
	emrB4-SapI-RV	TATATAGCTCTTCATGCGTGACCGGTGGAAATAGGTG

Organism	Primer name*	Primer sequence (5' → 3')
C. metallidurans	*emrA4*-SapI-FW	ATATATGCTCTTCTAGTTCCGCCACCCACTCCGCCG
	emrA4-SapI-RV	TATATAGCTCTTCATGCGCCATGGGTGGCCTTGATG
	tolC4-SapI-FW	ATATATGCTCTTCTAGTTCTGTTTCCTTTCGTGCCAC
	tolC4-SapI-RV	TATATAGCTCTTCATGCGGAGGCTGGCGCCGCGGAAG

* FW, forward; RV, reverse.

The amplification reaction was performed with the Phusion polymerase (either Phusion Flash High-Fidelity PCR Master Mix or Phusion High-Fidelity DNA Polymerase, Thermo Fisher Scientific, USA). The compositions of the PCR reaction mixtures are shown in Tables 6, 7 and 8. The details of the PCR program used for all the target genes are mentioned in Table 9.

Table 6. PCR reaction mixture composition for the *E. coli* targets (*emrAB-tolC*).

Reagent	V (μL)	Final concentration
H_2O	19	-
2X Phusion Flash PCR Master Mix	25	1X
10 μM Primers	2.5	0.5 μM
Template DNA (genomic)	1	-
Total volume	50	-

Table 7. PCR reaction mixture composition for the *V. cholerae*, *H. elongata*, *D. tiedjei*, *T. lienii*, *T. albus* and *C. metallidurans* targets.

Reagent	V (μL)	Final concentration
H_2O	13.5-14.5	-
2X Phusion Flash PCR Master Mix	25	1X
10 μM Primers	2.5	0.5 μM

Template DNA (genomic)	1-2	-
DMSO	1.5	3%
Ethylene glycol	3	6%
Total volume	50	-

Table 8. PCR reaction mixture composition for the *E. coli* (*emrKY*), *S. blattae*, *S. fumaroxidans*, *K. pneumoniae*, and *S. enterica* targets.

Reagent	V (μL)	Final concentration
H_2O	28	-
5X GC Buffer	10	1X
10 mM dNTPs	1	200 μM
10 μM Primers	2.5	0.5 μM
Template DNA (genomic)	1	-
DMSO	1.5	3%
Ethylene glycol	3	6%
Phusion DNA polymerase	0.5	0.02 U/μL
Total volume	50	-

Table 9. PCR cycling program used for all the targets.

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	98	30 s	1
Denaturation	98	10 s	
Annealing	65* (Touch down, -0.5/cycle)	30 s	30
Extension	72	60 s	
Final extension	72	10 min	1
Hold	4	-	1

After the different amplification reactions, each target was analyzed by 1% (wt/vol) agarose gel electrophoresis at 120V for 30 min. The coloring agent used for the observation of the signals was the SERVA DNA stain G (SERVA Electrophoresis GmbH, Germany). For the determination of the band sizes, the GeneRuler 1kb Plus DNA Ladder (Thermo Fisher Scientific, USA) was used as a standard.

Each target band was subsequently purified from the agarose gel using the Zymoclean Gel DNA recovery kit (ZYMO RESEARCH, USA).

Initially, the goal was to use all the inserts for the initial FX cloning step. However, as will be explained in the following section, only some of the different inserts were cloned into the sequencing plasmid pINITIAL (pINIT_cat[233] (chloramphenicol resistance)). The single cup reaction mixture composition is shown in Table 10. The SapI enzyme used for the digestions was purchased from Thermo Fisher Scientific, USA.

Table 10. Reaction mixture composition of the initial FX cloning step.

Reagent	Amount/Volume
pINIT_cat	50 ng
PCR product	Adjusted to a final molar ratio of 1:5 (vector : insert)
10X Buffer Tango	1 μL
SapI (5 U/μL)	0.5 μL
H_2O	Adjusted to 10 μL
Total volume	10 μL

After the subsequent ligation of the vector with the insert molecule using the T4 DNA Ligase (Thermo Fisher Scientific, USA), the mixture was used for the transformation into *E. coli* MC1061[234] (ccdB sensitive) cells to select positive clones.

For the sequencing reactions, plasmid DNAs from different clones were isolated using the ZR Plasmid Miniprep-Classic kit (ZYMO RESEARCH, USA).

Similar to the initial cloning step, at the very beginning the aim was to insert the genes coding for the inner membrane, periplasmic and outer membrane components of all the different systems into FX compatible expression vectors p7XC3RH (derivative of p7XC3GH[233]), pRSFDMG and pRSFDM_G (derivatives of pRSFDuet[235]). However, as will be discussed in the following section, only the ORFs from *E. coli* and *V. cholerae* were inserted into these expression vectors.

Therefore, during a first step the genes coding for the inner membrane components from *E. coli* (*emrB*) as well as *V. cholerae* (*vceB*) were subcloned from pINIT_*emrB* and pINIT_*vceB* into the SapI site of p7XC3RH for the addition of a mRFP1[236] label and a His-tag to their C-termini. The general composition of the one cup FX subcloning reaction mixture in a given FX compatible expression vector is shown in Table 11. After the SapI (Thermo Fisher Scientific, USA) digestion and the ligation step using the T4 DNA Ligase (Thermo Fisher Scientific, USA), the mixture was transformed into *E. coli* DH5α[237] (ccdB sensitive) cells for the selection of positive clones. The ZR Plasmid Miniprep-Classic kit (ZYMO RESEARCH, USA) was used for the isolation of the final constructs which were further verified via restriction analyses and sequencing.

Table 11. Reaction mixture composition of the FX subcloning step.

Reagent	Amount/Volume
pExpression (p7XC3RH, pRSFDMG, and pRSFDM_G)	50 ng
pINIT_*gene*	Adjusted to a final molar ratio of 1:4 (pExpression : pINIT_*gene*)
10X Buffer Tango	1 µL
SapI (5 U/µL)	0.5 µL
H_2O	Adjusted to 10 µL
Total volume	10 µL

In a second step, the genes coding for the outer membrane components from *E. coli* (*tolC*) and *V. cholerae* (*vceC*) were traditionally cloned into the NdeI/PacI sites of the expression vectors pRSFDMG and pRSFDM_G leading after expression to the addition of either a Strep-tag only or a sfGFP[238] label and a Strep-tag respectively to the C-termini of both proteins.

The sequences of the primers used for the PCR amplification of *tolC* and *vceC* from pINIT_*tolC* and pINIT_*vceC* for the addition of NdeI/PacI sites at each end are shown in Table 12. The PCR reaction mixture composition for both cases is shown in Table 13. The PCR cycling program was identical to the previously described method (Table 9).

Table 12. Primers used for the traditional cloning of *E. coli tolC* and *V. cholerae vceC*.

Organism	Primer name*	Primer sequence (5' → 3')
E. coli	*tolC*-NdeI-FW	GGGAATTCCATATGAAGAAATTGCTCCCCATTCT
	tolC-PacI-RV	CCTTAATTAAGTTACGGAAAGGGTTATGAC
V. cholerae	*vceC*-NdeI-FW	GGGAATTCCATATGAAAAATAGCGTTCAAACGGT
	vceC-PacI-RV	CCTTAATTAAAGATTCTGTTGTTTCAAAAC

* FW, forward; RV, reverse.

Table 13. PCR reaction mixture composition for the amplification of *E. coli tolC* and *V. cholerae vceC* for traditional cloning.

Reagent	V (μL)	Final concentration
H_2O	32.5	-
5X GC Buffer	10	1X
10 mM dNTPs	1	200 μM
10 μM Primers	2.5	0.5 μM
Template DNA (pINIT_gene)	1	-
Phusion DNA polymerase	0.5	0.02 U/μL
Total volume	50	-

After the verification of the signals on a 1% (wt/vol) agarose gel (in an identical manner to the previous description), both inserts were purified using the Zymoclean Gel DNA recovery kit (ZYMO RESEARCH, USA). The empty plasmid backbones (pRSFDMG and pRSFDM_G) and both inserts were digested using the FastDigest NdeI/PacI restriction enzymes (Thermo Fisher Scientific, USA). In addition, the empty plasmid backbones were also dephosphorylated using the FastAP Thermosensitive Alkaline Phosphatase enzyme (Thermo Fisher Scientific, USA). After a second agarose gel purification step of the digested plasmids and inserts, the ligation reaction was performed using the T4 DNA Ligase (Thermo Fisher Scientific, USA). The ligation mixture was subsequently transformed into *E. coli* DB3.1[239] (ccdB resistant) cells for the selection of positive clones. Finally, the different plasmids were isolated using the ZR Plasmid Miniprep-Classic kit (ZYMO RESEARCH, USA) and verified via restriction analyses and sequencing.

In a third step, the genes coding for the periplasmic components from *E. coli* (*emrA*) as well as *V. cholerae* (*vceA*) were subcloned (in an identical manner as for the subcloning of *emrB* and *vceB*) into *tolC* and *vceC* gene containing pRSFDMG and pRSFDM_G for the addition after expression of either a sfGFP[238] label and Myc-tag or only a Myc-tag respectively to the C-termini of both proteins.

3.2. Results of the first cloning strategy employed for screening purposes

The aim of the homologous screening strategy was to choose orthologous MFS-type tripartite efflux systems besides of the EmrAB-TolC system from *E. coli* for subsequent structural analyses.

Overall, the amplification of numerous ORFs encoding different MFS-type tripartite efflux systems was a challenging task. Indeed by trial and error, I have learnt to avoid high GC content targets and to put strict criteria on amplified sequences and information from the DataBase (TransportDB)[226,227]. In total, all the ORFs of 15 homologous MFS-type tripartite efflux systems were successfully amplified from genomic DNA (Figure 52).

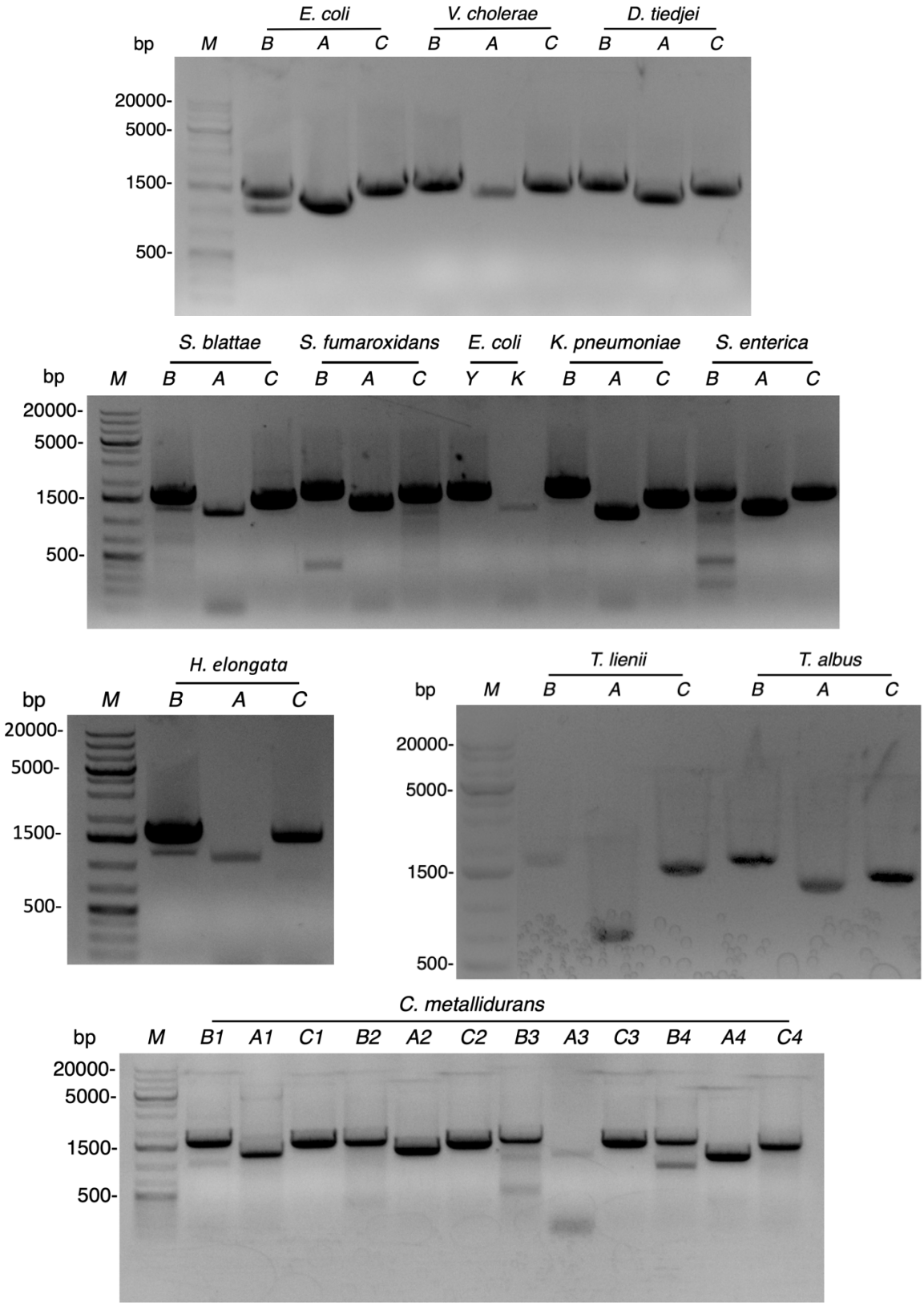

Figure 52. Agarose gel electrophoresis analysis of amplified *emrA*, *emrB* and *tolC* type genes.

B, *A* and *C* refer to the *emrB*- (~1500 bp), *emrA*- (~1200 bp) and *tolC*-type (~1500 bp) ORFs from various gram negative organisms.

The ORFs from *E. coli*, *V. cholerae*, *K. pneumoniae*, *S. enterica* and the ORFs coding for the first efflux system from *C. metallidurans* (*B1*, *A1* and *C1*) were successfully cloned into the sequencing plasmid pINITIAL (pINIT_cat). Due to the sequential cloning procedure, *emrAB-tolC*, *emrKY-tolC* from *E. coli* and *vceABC* from *V. cholerae* were subcloned into the FX compatible expression vectors p7XC3RH, pRSFDMG and pRSFDM_G (Figure 53).

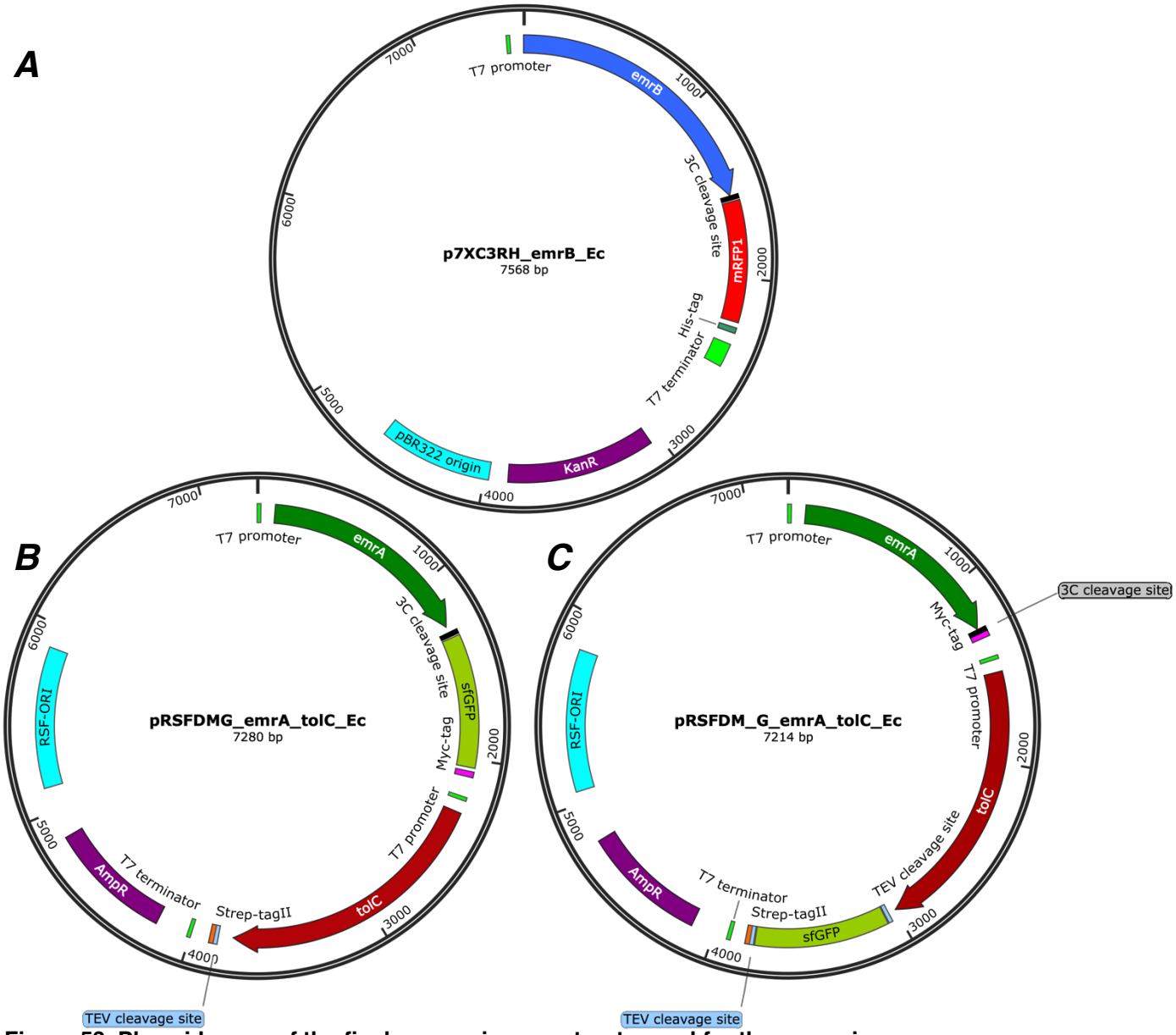

Figure 53. Plasmid maps of the final expression constructs used for the screening purposes.

(A) The gene *emrB* from *E. coli* is indicated in blue color inside the expression vector p7XC3RH. The FX-cloning procedure results in an *emrB*-mRFP1 fusion, with at the 3' end a coding region for 10x His-tag. The 3C cleavage site, T7 promoter, T7 terminator, pBR322 origin and Kanamycin resistance selection marker are indicated. (B) The *E. coli* genes *emrA* and *tolC* are indicated in dark green and brown colors, respectively inside the expression vector pRSFDMG. The FX-cloning procedure results in an *emrA*-sfGFP fusion, with at the 3' end a coding region for a Myc-tag. The traditional cloning procedure results in a *tolC* gene with at the 3' end a coding region for a Strep-tagII. The TEV cleavage site, T7 promoters, T7 terminator, RSF-origin and Ampicillin resistance selection marker are inidcated (C) The *E. coli* genes *emrA* and *tolC* and the previously described elements are colored as in (B). The position of the sfGFP encoding sequence is reversed in pRSFDM_G. The sequence positions coding for the Myc and Strep-tagII are as in pRSFDMG. Ec, *E. coli*.

The level of the cloning progress for the ORFs of 15 homologous EmrAB-TolC systems are summarized in Table 14.

Table 14. Cloning progress of the ORFs of 15 homologous EmrAB-TolC systems.

Organism	ORF	Amplification	pINIT_cat	pExpression
Escherichia coli (BL21DE3)	*emrB*	√	√	√
	emrA	√	√	√
	tolC	√	√	√
	emrY	√	√	√
	emrK	√	√	√
Vibrio cholerae (O395)	*vceB*	√	√	√
	vceA	√	√	√
	vceC	√	√	√
Halomonas elongata (DSM2581)	*emrB*	√	X	X
	emrA	√	X	X
	tolC	√	X	X
Shimwellia blattae (DSM4481NBRC105725)	*emrB*	√	X	X
	emrA	√	X	X
	tolC	√	X	X
Desulfomonile tiedjei (DSM6799)	*emrB*	√	X	X
	emrA	√	X	X
	tolC	√	X	X
Syntrophobacter fumaroxidans (MPOB**)**	*emrB*	√	X	X
	emrA	√	X	X
	tolC	√	X	X
Thermovirga lienii (DSM17291)	*emrB*	√	X	X
	emrA	√	X	X
	tolC	√	X	X
Thermocrinis albus (DSM14484)	*emrB*	√	X	X
	emrA	√	X	X
	tolC	√	X	X
Klebsiella pneumoniae (MGH78578)	*kpnH*	√	√	X
	kpnG	√	√	X
	tolC	√	√	X
Salmonella enterica (LT2)	*emrB*	√	√	X
	emrA	√	√	X
	tolC	√	√	X
Cupriavidus metallidurans (CH34)	*emrB1*	√	√	X
	emrA1	√	√	X
	tolC1	√	√	X
	emrB2	√	X	X
	emrA2	√	X	X

Organism	ORF	Amplification	pINIT_cat	pExpression
Cupriavidus metallidurans (CH34)	*tolC2*	√	X	X
	emrB3	√	X	X
	emrA3	√	X	X
	tolC3	√	X	X
	emrB4	√	X	X
	emrA4	√	X	X
	tolC4	√	X	X

The gene expressed from p7XC3RH_*emrB* encodes a EmrB-mRFP1-10x His fusion protein. Due to the topology of EmrB, the tag is located at the cytoplasmic side. In addition, using two alternative constructs (pRSFDMG and pRSFDM_G) the C-termini of EmrA was either tagged with sfGFP-Myc-tag or with Myc-tag only, and TolC was either tagged with Strep-tagII only or with sfGFP-Strep-tagII. The sfGFP is a version of GFP able to fold stably in the periplasm[240]. Both alternative labeling strategies were employed in case the sfGFP labeling of either the adaptor or outer membrane component would be deleterious for complex formation. As will be discussed in chapter III, these fluorescent labels, together with the three different affinity tags were used for small scale co-expression tests and preliminary analyses of the EmrAB-TolC and VceABC complexes from *E. coli* and *V. cholerae*, respectively.

4. Alternative cloning strategy for the production of genetically engineered EmrAB-TolC systems from E. coli

In parallel to the first strategy, an alternative cloning strategy was also employed to maximize the chances of the successful isolation of the entire EmrAB-TolC system from *E. coli*. This second strategy was adapted from the published work of Fitzpatrick and co-workers concerning the ABC-type tripartite MacAB-TolC efflux system from *E. coli* [241]. Therefore, the following sections describe the modifications of the 'MacAB-TolC strategy' for its adaptation for the EmrAB-TolC system from *E. coli* and the corresponding results that were obtained.

4.1. Cloning methodology for the co-production of fusion stabilized EmrAB-TolC systems from E. coli

The overall aim of the present strategy was to prepare affinity tagged and genetically fused EmrAB that could be co-expressed together with affinity tagged TolC using only one expression vector. Moreover, in order to maximize the chances of isolating a fusion stabilized efflux system without disturbing the expression and complex formation behaviors of the different protein components, multiple constructs had to be prepared (varying the affinity tag type, the promoter system and the poly-glycine-serine (GS)-linker size between EmrB and EmrA). Thus, a combination of 'modified FX cloning' and traditional cloning methodologies was employed (Figure 54).

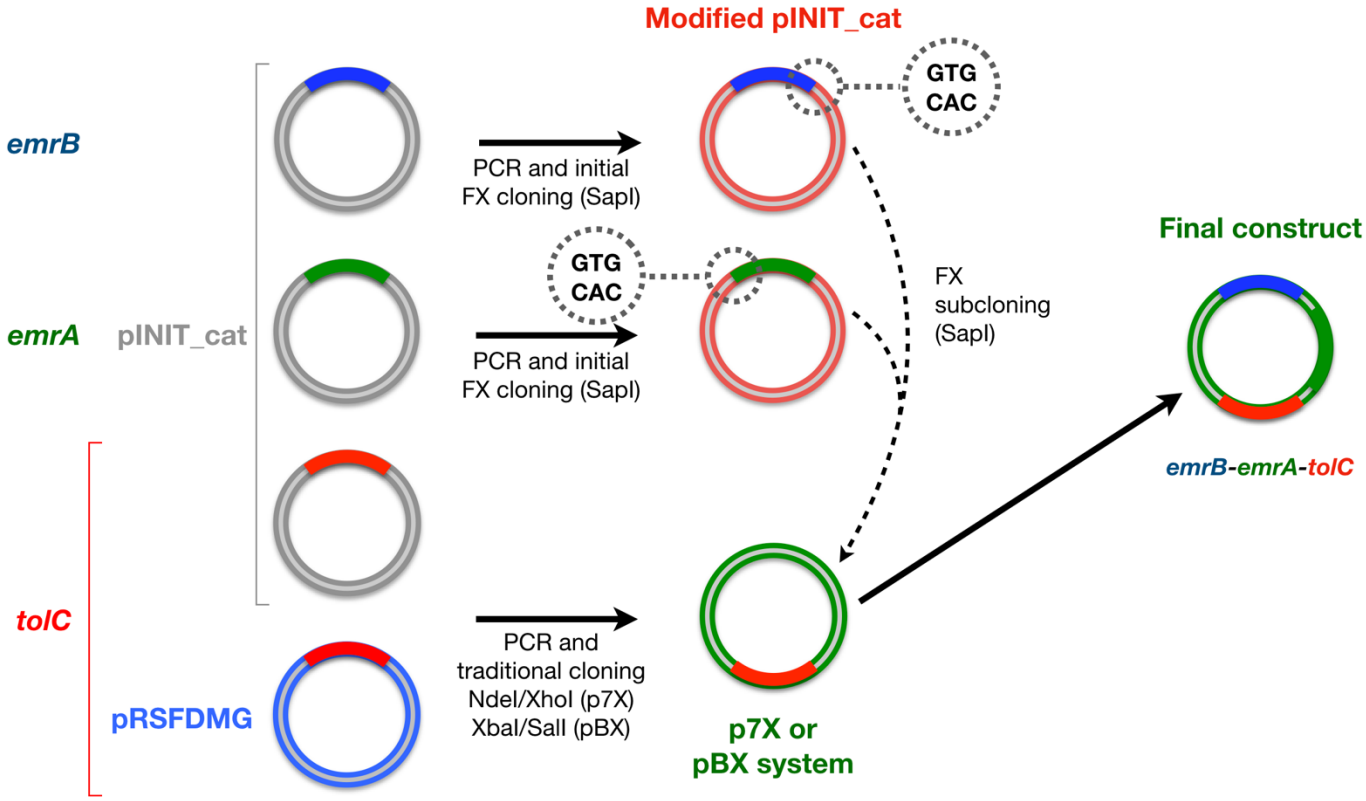

Figure 54. Cloning methodology used for the preparation of fusion stabilized EmrAB-TolC systems from *E. coli*.

The pINIT_cat vectors (colored in grey) containing the *emrAB-tolC* inserts from *E. coli* and the pRSFDMG vector (colored in blue) containing the *tolC* insert from *E. coli* were utilized for the PCR amplifications. The two modified pINIT_cat vectors having a GTG sequence instead of either the usual GCA or AGT sequence are colored in red. The expression vector (either p7XC3H_Duet (*T7* promoter) or pBXC3H (*araBAD* promoter)) used for traditional cloning and FX subcloning is colored in green. Each final expression construct contains all three inserts.

The first step consisted in a PCR amplification of the different inserts from pINIT_cat and pRSFDMG using various primers for the addition of different sequence types (restriction sites, GS-linker sequences, affinity tag sequences, ribosome binding sites, start and stop codons) (Table 15). For the amplification reactions, the mixture composition was similar to the description given in Table 6 only replacing the target by plasmid DNA. The PCR cycling program was identical to the previous descriptions (Table 9).

Table 15. Primers used for the 'modified FX cloning' and traditional clonings for the preparation of genetically engineered EmrAB-TolC systems from *E. coli*.

Primer name*	Features added	Primer sequence (5' → 3')
tolC-NdeI-p7X-FW	NdeI restriction site	GGGAATTCCATATGAAGAAATTGCTCCCCATTC
tolC-XhoI-p7X-RV	XhoI restriction site and 2 stop codons	CCGCTCGAGTTATTATTTCTCGAACTGCGGGTGG
tolC-XhoI-FLAG-p7X-RV	XhoI restriction site, 2 stop codons, and FLAG-tag encoding sequence	CCGCTCGAGTTATTACTTATCGTCGTCATCCTTGTAATCGTTA-CGGAAAGGGTTATGAC
emrB-SapI-pINIT-GTGend-FW	SapI restriction site	ATATATGCTCTTCTAGTCAACAGCAAAAACCGCTGG
emrB-SapI-pINIT-GTGend-RV	SapI restriction site with GTG overhang	ATATATGCTCTTCTCACGTGCGCACCGCCTCCGCCG
emrA-SapI-5GS-pINIT-GTGstart-FW	SapI restriction site with GTG overhang, 'GGGGS' encoding linker sequence	ATATATGCTCTTCTGTGGGATCCGGTGGGAGCAGCGCAAATG-CGGAGACTC
emrA-SapI-pINIT-GTGstart-RV	SapI restriction site	TATATAGCTCTTCATGCGCCAGCGTTAGCTTTTACG
emrA-SapI-10GS-pINIT-GTGstart-FW	SapI restriction site with GTG overhang, 'GGGGS × 2' encoding linker sequence	ATATATGCTCTTCTGTGGGATCCGGTGGGAGCGGCGGTGGTG-GCTCTAGCGCAAATGCGGAGACTC

Primer name*	Features added	Primer sequence (5' → 3')
emrA-SapI-15GS-pINIT-GTGstart-FW	SapI restriction site with GTG overhang, 'GGGGS × 3' encoding linker sequence	ATATATGCTCTTCTGTGGGATCCGGTGGGAGCGGCGGTGGTG-GCTCTGGTGGCGGCGGTAGTAGCGCAAATGCGGAGACTC
emrA-SapI-20GS-pINIT-GTGstart-FW	SapI restriction site with GTG overhang, 'GGGGS × 4' encoding linker sequence	ATATATGCTCTTCTGTGGGATCCGGTGGGAGCGGCGGTGGTG-GCTCTGGTGGCGGCGGTAGTGGCGGTGGTGGCTCTAGCGCA-AATGCGGAGACTC
tolC-XbaI-pBX-FW	XbaI restriction site and ribosome binding site (RBS)	GCTCTAGACAGGAGGAATTAACCATGAAGAAATTGCTCCCCAT-TC
tolC-SalI-pBX-RV	SalI restriction site and 1 additional stop codon	ACGCGTCGACTTATTATTTCTCGAACTGCGGG
tolC-XbaI-start-pBX-FW	XbaI restriction site, RBS and 1 start codon	GCTCTAGACAGGAGGAATTAACCATGAAGAAATTGCTCCCCAT-TC
tolC-SalI-FLAG-pBX-RV	SalI restriction site, 2 stop codons, and FLAG-tag encoding sequence	ACGCGTCGACTTATTACTTATCGTCGTCATCCTTGTAATCGTT-ACGGAAAGGGTTATGAC
tolC-XhoI-His-p7X-RV	XhoI restriction site, 2 stop codons, and His-tag encoding sequence	CCGCTCGAGTTATTAATGATGATGATGATGGTGATGATGATGG-TGGTTACGGAAAGGGTTATGAC

Primer name*	Features added	Primer sequence (5' → 3')
emrA-SapI-Strep-pINIT-GTGstart-RV	SapI restriction site, 2 stop codons, and Strep-tagII encoding sequence	TATATAGCTCTTCATGCTTATTATTTCTCGAACTGCGGGTGGC-TCCAGCCAGCGTTAGCTTTTACG
emrA-SapI-FLAG-pINIT-GTGstart-RV	SapI restriction site, 2 stop codons, and FLAG-tag encoding sequence	TATATAGCTCTTCATGCTTATTACTTATCGTCGTCATCCTTGTA-ATCGCCAGCGTTAGCTTTTACG
tolC-SalI-His-pBX-RV	SalI restriction site, 2 stop codons, and His-tag encoding sequence	ACGCGTCGACTTATTAATGATGATGATGATGGTGATGATGATG-GTGGTTACGGAAAGGGTTATGAC

* FW, forward; RV, reverse.

During a second step, the amplified *emrB* and *emrA* inserts were cloned into two modified pINIT_cat vectors. *emrB* was inserted into a pINIT_cat having a GTG sequence instead of the usual GCA sequence. In contrast, *emrA* was inserted into a pINIT_cat having a GTG sequence instead of the usual AGT sequence. The different FX initial cloning reactions were carried out in an identical manner as the previously described reactions (Table 10).

In a third step, the amplified *tolC* inserts were traditionally cloned into two different FX compatible expression vectors (either p7XC3H_Duet (p7XC3H derivative) or pBXC3H[233]) using the FastDigest NdeI/XhoI or XbaI/SalI restriction enzyme combinations respectively (Thermo Fisher Scientific, USA). The overall procedure was identical to the previously described traditional cloning methodology.

Finally, during a forth step the *emrB* and *emrA* inserts from the modified pINIT_cat vectors were simultaneously subcloned into the *tolC* containing expression vectors in an identical fashion as the previous subcloning reactions (Table 11) (with molar ratios for both inserts of 1:4 (pExpression :

pINIT_*gene*)). In fact, the specific and compatible GTG overhangs of the *emrB* and *emrA* inserts allowed their simultaneous subcloning into a single SapI site within a given expression vector.

4.2. Results of the alternative cloning strategy used for the stabilization of the EmrAB-TolC system from E. coli

The employed strategy as described in section 4.1 and Figure 54 resulted in 32 different expression constructs useful for the screening for stable EmrAB-TolC complexes (Figure 55).

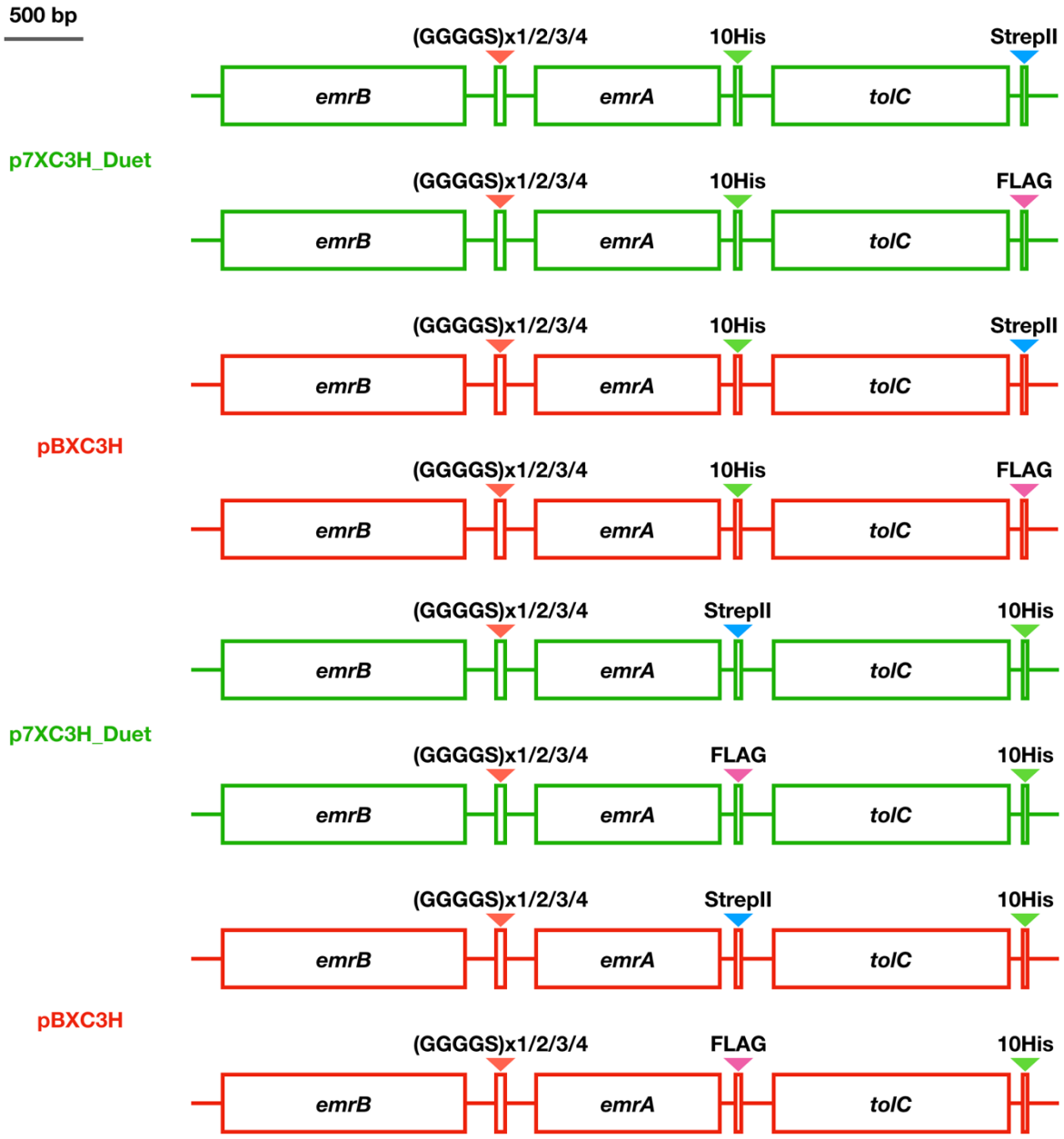

Figure 55. Schematic representations of the constructs encoding stabilized EmrAB-TolC efflux systems.

The differences between the constructs concern the promoter system type, the GS-linker length and the affinity tag positions.

As will be mentioned later, some of these constructs were further used for different co-expression and co-purification tests.

5. Cloning strategy for the individual expression of the emrA*,* emrB *and* tolC *genes*

An additional strategy was adopted in order to separately produce of all three components of the efflux system. In contrast to the two previous strategies, the subsequent goal would consist in the *in vitro* assembly of the entire tripartite system. The assembly technique with the purified components would be similar to either the 'Nanodisc methodology' reported by Daury and co-workers in 2016[242] or the 'Amphipol methodology' reported by Tsutsumi and co-workers in 2019[243] for the MexAB-OprM efflux system from *Pseudomonas aeruginosa*.

5.1. Cloning methodology for the preparation of individual expression constructs for the emrAB-tolC *inserts from* E. coli

As the *emrAB-tolC* inserts from *E. coli* where already cloned into the FX sequencing vector pINITIAL, 6 different expression vectors were prepared by subcloning the different inserts into the FX expression vectors (p7XC3H and pBXC3H[233]) in an identical manner to the first cloning strategy. Both p7XC3H and pBXC3H were chosen in order to be able to test the *T7* and *araBAD* promoters for the expression of all three proteins.

5.2. Results of the third cloning strategy representing an alternative method for the isolation of the EmrAB-TolC system from E. coli

The preparation of 6 expression constructs with the FX subcloning step was straightforward. The plasmid maps of both types of expression constructs are detailed in Figure 56.

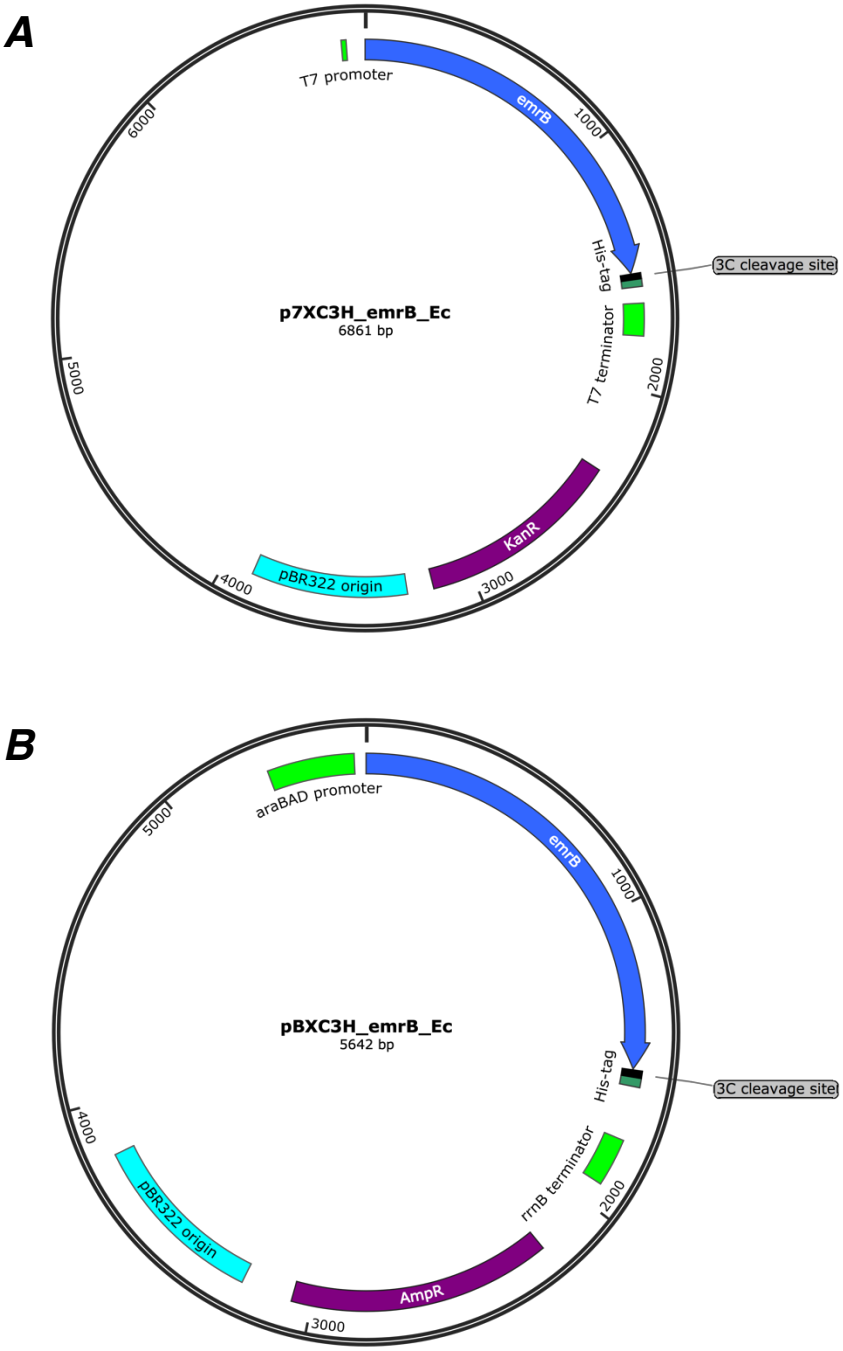

Figure 56. Plasmid maps of two types of individual expression constructs.

(A) The p7XC3H vector construct with the *emrB* insert shown as an example. In this case, the *T7* promoter system is used for the expression of the three inserts. (B) The pBXC3H vector construct with the *emrB* insert shown as an example. Here, in contrast to the p7XC3H construct the *araBAD* promoter system is used for the expression of the three inserts. Ec, *E. coli*.

Chapter III: Small-scale analyses of EmrAB-TolC from E. coli *and VceABC from* V. cholerae

In the present chapter, the main screening approach used (consisting of a high-throughput screening technique developed by Alina Ornik-Cha in the laboratory of Prof. Dr. Klaas Martinus Pos (first strategy)) will be described together with the corresponding results obtained.

In parallel, initial analyses of a second alternative strategy used for the isolation of the entire EmrAB-TolC system from *E. coli* (with EmrAB fusion chimeras) will also be mentioned.

1. A high-throughput screening pipeline for initial analyses of EmrAB-TolC systems

In the present section, the main screening approach employed with the use of specific expression constructs (first cloning strategy) will be presented. In addition, various experimental procedures used during the different steps of the approach will also be described together with the corresponding results obtained.

1.1. Presentation of the high-throughput screening pipeline

Based on previously reported information using a GFP label for screening purposes[244–247], the methodology was adapted for small scale co-expression and complex formation analyses. A schematic representation of the global strategy divided into two parts is shown in Figure 57.

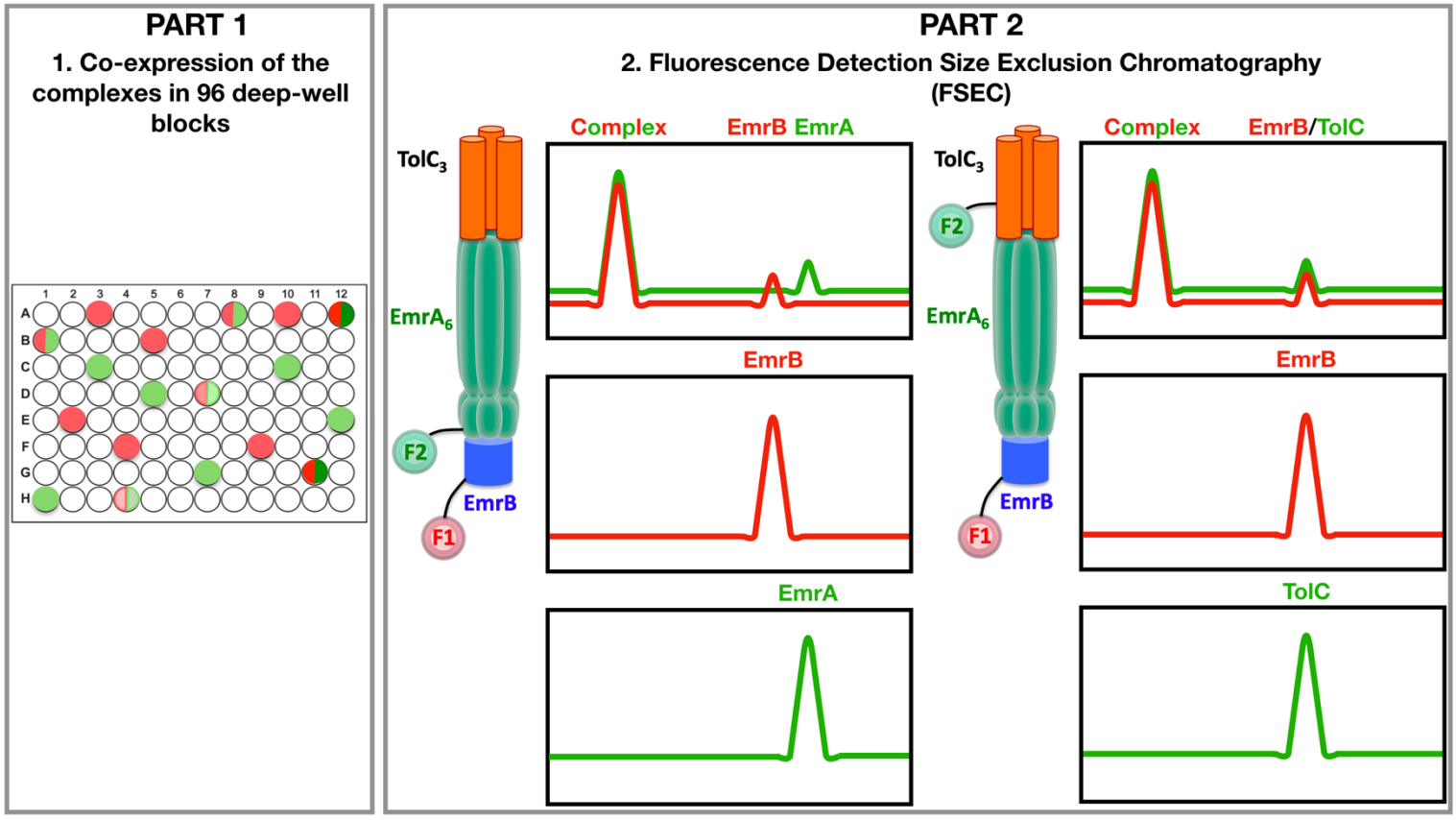

Figure 57. Schematic representation of the high-throughput screening pipeline.

(*Left panel*) During part 1, the expression levels of two of the partners from the tripartite system (either EmrB/EmrA or EmrB/TolC) can be analyzed via whole cell fluorescence measurements. At this step, different parameters can be easily tested for the co-expression optimizations. In some cases, only one of the two labeled components might be expressed whereas in other instances, both labeled components could be co-expressed. Dark colors illustrate high expression levels. (*Right panel*) During part 2, the complex formation and stability behaviors can be verified using FSEC. Theoretical examples of the red and green fluorescence chromatograms for the samples containing all three components from *E. coli* are shown in the first row. The second and third rows represent theoretical examples of the corresponding negative controls with samples containing either only EmrB (p7XC3RH) or only EmrA co-expressed with TolC (pRSFDMG or pRSFDM_G). F1, Fluorescent label 1 (mRFP1); F2, Fluorescent label 2 (sfGFP).

During a first step, the expression levels of the complex components can be easily tested using whole cell red and green fluorescence measurements. The mRFP1 and sfGFP fluorescence spectra[248,249] are shown in Figure 58. Thus, the co-expression of a given system can be optimized in a straightforward manner using the present methodology.

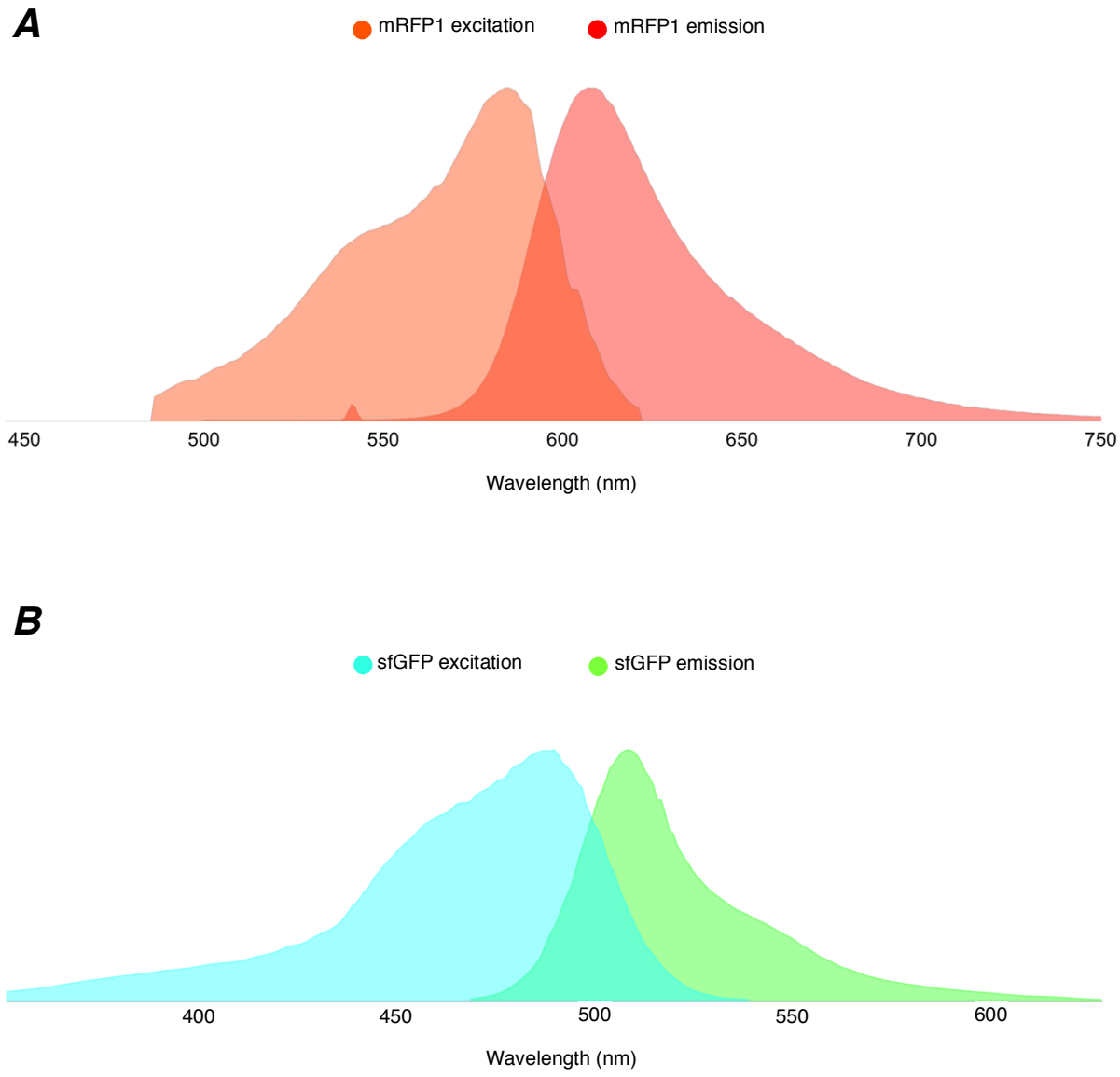

Figure 58. Fluorescence spectra of mRFP1 and sfGFP.

(A) Fluorescence spectra of mRFP1 (Excitation/Emission maxima are at 584/607 nm respectively). (B) Fluorescence spectra of sfGFP (Excitation/Emission maxima are at 485/510 nm respectively). Adapted from references 246 and 247.

After the co-expression optimization step, the complex formation and stability behaviors can be easily tested using FSEC[250–253]. Indeed, the aim is to decipher signal shifts between the elution volumes of the samples containing all three complex components versus the negative controls containing either only EmrB or only EmrA co-expressed with TolC.

Thus, the high-throughput screening pipeline represents a convenient tool for the identification of a suitable complex for subsequent structural analyses.

1.2. Experimental procedures for screening purposes

1.2.1. Co-expressions

As explained earlier, because of the hinderances encountered during the cloning step of the global homologous screening strategy, the different analyses will mainly mention the EmrAB-TolC system from *E. coli*. Nevertheless, the VceABC complex from *V. cholerae* was also included in some of the later analyses.

1.2.1.1. Co-expression level analyses of the EmrAB-TolC from **E. coli** *in 96 deep-well blocks*

Four different *E. coli* strains (BL21DE3[254], C41DE3*ΔacrAB*[255], C43DE3[255], and BW25113DE3*ΔacrAB*[256]) were used for cultivations in 2xYT medium supplemented with 50 μg/mL Kanamycin and 100 μg/mL Ampicillin at 37°C under vigorous aeration. The cultivations of the strains co-transformed with p7XC3RH_*emrB* and pRSFDMG_*emrA_tolC* were started with 1% (vol/vol) inoculums of overnight cultures prepared in LB medium supplemented with 50 μg/mL Kanamycin and 100 μg/mL Ampicillin. The growth was continued to reach an OD_{600}≈0.6. After a 20 min incubation step of the cultures on ice, ten different final concentrations (0.1-1mM) of Isopropylthiogalactopyranoside (IPTG) were used for the inductions. The growth of the cultures was subsequently continued overnight at 25°C.

After a harvesting step by centrifugation at 3000 *g* for 20 min at 4°C, the different cell pellets were resuspended in 200 μL PBS buffer (pH 7.4) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4). The final OD_{600} values as well as the red and green fluorescence intensities were measured in a TECAN reader infinite 200 (Tecan Trading AG, Switzerland).

1.2.1.2. Co-expressions of EmrAB-TolC from E. coli *and VceABC from* V. cholerae *for FSEC experiments*

1.2.1.2.1. Membrane preparation of cells co-expressing EmrAB-TolC from E. coli

In a first attempt to analyze the complex formation and stability behaviors of the EmrAB-TolC system from *E. coli*, a classical membrane protein production procedure was employed. Therefore, 2xYT medium supplemented with 50 μg/mL Kanamycin and 100 μg/mL Ampicillin was inoculated with an overnight culture in LB medium of *E. coli* C41DE3*ΔacrAB*[255] cells harboring the p7XC3RH_*emrB* and pRSFDMG_*emrA_tolC* constructs to a final OD_{600} of 0.05. The growth was continued at 37°C under vigorous aeration to reach an OD_{600}≈0.6. After a subsequent incubation step of the culture on ice for 20 min, 0.5 mM final concentration of IPTG was used for the induction and the growth was continued at 25°C overnight. Cells were harvested by centrifugation at 8,000 *g* for 25 min at 4°C and resuspended in a Tris buffer (50 mM Tris/HCl, pH 7.5, 400 mM NaCl, 30 U/mL of DNaseI, 2 mM EDTA, 10 mM $MgCl_2$, 2 mM $CaCl_2$, 0.2 mM DFP). The cell suspension mixture was stirred at 4°C for 20 min. Cells were subsequently disrupted by passing the mixture two times through a Stansted pressure cell homogenizer EP FPG12805 (Stansted Fluid Power LTD., United Kingdom) at 1.5 bar at 4°C. After a first centrifugation step at 25,000 *g* for 25 min at 4°C to remove the cell debris, the supernatant was ultracentrifuged at 125,000 *g* for 2 h at 4°C to pellet the cellular membrane. The cellular membrane was resuspended in a Tris buffer (50 mM Tris/HCl, pH 7.5, 400 mM NaCl) and divided into 1 mL aliquots which were conserved at -80°C until subsequent FSEC analyses.

In an identical manner, cellular membranes were also prepared from *E. coli* C41DE3*Δ*acrAB[255] cells harboring either only the p7XC3RH_*emrB* or only the pRSFDMG_*emrA_tolC* construct. As mentioned previously, these samples were used as negative controls for the FSEC experiments.

1.2.1.2.2. Small-scale co-expressions of EmrAB-TolC from E. coli *and VceABC from* V. cholerae

A second co-expression methodology for the complex formation and stability screenings was also employed. In that case, overnight cultures in LB media of *E. coli* C41DE3*Δ*acrAB[255] cells harboring either the p7XC3RH/pRSFDMG or p7XC3RH/pRSFDM_G vector combinations with the ORFs from *E.*

coli and *V. cholerae* were used for a 1% (vol/vol) inoculation of 2xYT media. The cultures were induced in an identical manner to the previous description. Cultivations were continued at 22°C overnight. Aliquots for the FSEC experiments (corresponding to a normalization of a total OD_{600}=10) were taken and centrifuged at 16,000 *g* for 2 min at 4°C to harvest the cells.

Also in this case, the FSEC aliquots of the corresponding negative controls were prepared in an identical manner.

Finally, as co-expression/expression level verifications, one aliquot from each sample was resuspended in 520 µL PBS buffer (pH 7.4) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) for red and green fluorescence measurements in a TECAN reader infinite 200 (Tecan Trading AG, Switzerland).

1.2.2. Small-scale preliminary verifications

1.2.2.1. In gel *mobility controls of EmrAB-TolC from* E. coli

1.2.2.1.1. In gel *fluorescence*

Whole cell samples from the 96 deep-well blocks corresponding to co-expressions of the *E. coli* EmrAB-TolC system within *E. coli* C41DE3⊿acrAB[255] cells were resuspended in 100 µL of 1X protein sample buffer (50 mM Tris/HCl, pH 6.8, 10 % (vol/vol) glycerol, 0.4 % (vol/vol) 2-Mercaptoethanol, 1 % (wt/vol) SDS, and 0.01 % (wt/vol) bromophenol blue). Samples were heated at 37°C for 10 min prior to use. 15 µL of the protein samples were analyzed by 12 % SDS-PAGE at 160 V for 60 min. *In gel* mRFP1 and sfGFP fluorescence measurements were immediately performed using a LAS-4000 imaging system and the provided software (GE Healthcare, USA).

1.2.2.1.2. Immunodetection

The gels used for the *in gel* fluorescence measurements, were submitted to semidry electroblotting and immunodetection. Protein bands were blotted on a Roti-NC 0.2 µm nitrocellulose membrane (Carl Roth GmbH + Co. KG) for 30 min at 25 V in transfer buffer (25 mM Tris/HCl, pH 8.3, 192 mM glycine, 20 % methanol). The membrane was subsequently blocked in 3 % BSA TBST (20 mM Tris/HCl, pH 8, 150 mM NaCl, 0.1 % Tween 20) at 4°C overnight. The membrane was then washed three times 5 min with TBST buffer. For the immunological detection of EmrB from *E. coli* an Alkaline

Phosphatase conjugated antibody raised against a His_6 tag (Merck KGaA, Germany) was used. EmrA from *E. coli* was detected with a primary antibody raised against a Myc tag (Merck KGaA, Germany) combined with a secondary Alkaline Phosphatase antibody raised against mouse IgG (Merck KGaA, Germany). For the detection of TolC from *E. coli*, a Strep-Tactin Alkaline Phosphatase conjugate (IBA Lifesciences, Germany) was used. Membranes were incubated with antibody and Strep-Tactin solutions for 1h each. Subsequently, membranes were washed three times 5 min with TBST. After an equilibration step in AP-buffer (100 mM Tris/HCl, pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$) for three times 5 min, the staining reaction was conducted in AP-buffer containing 90 mM nitro blue tetrazolium (NBT) and 135 mM 5-bromo-4-chloro-3-indoyl phosphate (BCIP). The reaction was stopped by washing the membranes with water.

1.2.2.2. Initial E. coli *EmrA integrity verification*

In contrast to EmrB, EmrA produced using the pRSFDMG construct during the 96 deep-well block expressions, was fused to sfGFP. As sfGFP does not represent a folding reporter, an additional procedure was used for the verification of the integrity of EmrA.

Whole cell samples from the *E. coli* hosts C41(DE3)*ΔacrAB*[255] and C43(DE3)[255] normalized to 2 mg of total protein were resuspended in 400 µL of Tris buffer (50 mM Tris/HCl, pH 7.5, 400 mM NaCl, 0.5 mM PMSF, 1 mM $MgSO_4$, 30 U/mL DNase I). Cells were subsequently disrupted with glass beads (300 mg, 0.1 mm diameter) by shaking the samples in a FastPrep-24 device (MP Biomedicals, LLC, USA). The procedure was realized for 20 s at force 6 and the samples were subsequently cooled on ice for 5 min. The same procedure was repeated once. After a prior centrifugation step at 16,000 *g* for 2 min at 4°C, 1 % (wt/vol) final concentration of DDM was added to the sample supernatants put under mild agitation for 1 h at 4°C for membrane protein solubilization. The samples were subsequently ultracentrifuged at 355,000 *g* for 10 min at 4°C. The red and green fluorescence signals of different aliquots taken at each step (Disrupted cells, samples before ultracentrifugation, and samples after ultracentrifugation) were measured in a microplate reader (TECAN reader infinite 200, Tecan Trading AG, Switzerland).

1.2.3. FSEC

1.2.3.1. FSEC starting from cellular membranes

Using the membrane preparations from the co-expressions of EmrAB-TolC from *E. coli* and the corresponding negative controls, several FSEC experiments were performed. Therefore, 180 µL aliquots of the three samples (EmrAB-TolC, EmrB and EmrA-sfGFP/TolC) were solubilized using either 1 % (wt/vol) DDM or 1.5 % (Triton X-100) with a gentle agitation for 1 h at 4°C. The samples were subsequently ultracentrifuged at 355,000 *g* for 10 min at 4°C. After the ultracentrifugation step, the FSEC experiments were performed using an Agilent 1200 series high-performance liquid chromatography (HPLC) system (Agilent Technologies, USA) equipped with an autosampler for 96-well microtiter plates. A Superose 6 increase 3.2/300 column (GE Healthcare, USA) was used for the different runs. All the FSEC runs were performed in a Tris buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, and containing either 0.03 % (wt/vol) DDM or 0.05 % (wt/vol) Triton X-100).

1.2.3.2. FSEC starting from whole cells

During a second trial, a modified FSEC methodology was employed. Therefore, whole cell samples (corresponding to a total OD_{600}=10) from the small-scale co-expressions (p7XC3RH and pRSFDMG constructs only) of the *E. coli* EmrAB-TolC, *V. cholerae* VceABC and from the preparations of their respective negative controls were resuspended in 520 µL of lysis buffer (50 mM Tris/HCl pH 7.5, 400 mM NaCl, 10 mM $MgCl_2$, 0.2 mM DFP, 1 mg/mL lysozyme, 210 U/mL DNase I). In each case a 100 µL aliquot was used for the global procedure. For cell lysis, the samples were first incubated for 30 min at 37°C under mild agitation. Subsequently, for membrane protein solubilization, 1 % (wt/vol) DDM or 1.3 % (wt/vol) ANAPOE-$C_{12}E_{10}$ were added to the samples incubated under mild agitation for 2 h at 4°C. After an ultracentrifugation step at 355,000 *g* for 10 min at 4°C, the FSEC analyses were performed in an identical manner as previously described using an Agilent 1200 series high-performance liquid chromatography (HPLC) system (Agilent Technologies, USA). The different FSEC runs were performed in a Tris buffer (50 mM Tris/HCl pH 7.5, 150 mM NaCl, and containing either 0.0174% (wt/vol) DDM or 0.026 % ANAPOE-$C_{12}E_{10}$).

1.3. Results of the screening procedures

1.3.1. Co-expressions

1.3.1.1. E. coli *C41 cells are best suited for the co-expression of EmrB and EmrA from* E. coli

The fluorescence levels of mRFP1 and sfGFP corresponding to the expression levels of EmrB and EmrA respectively within four different *E. coli* hosts co-expressing the entire EmrAB-TolC system (using the p7XC3RH and pRFSDMG vector combination) are illustrated in Figures 59 and 60.

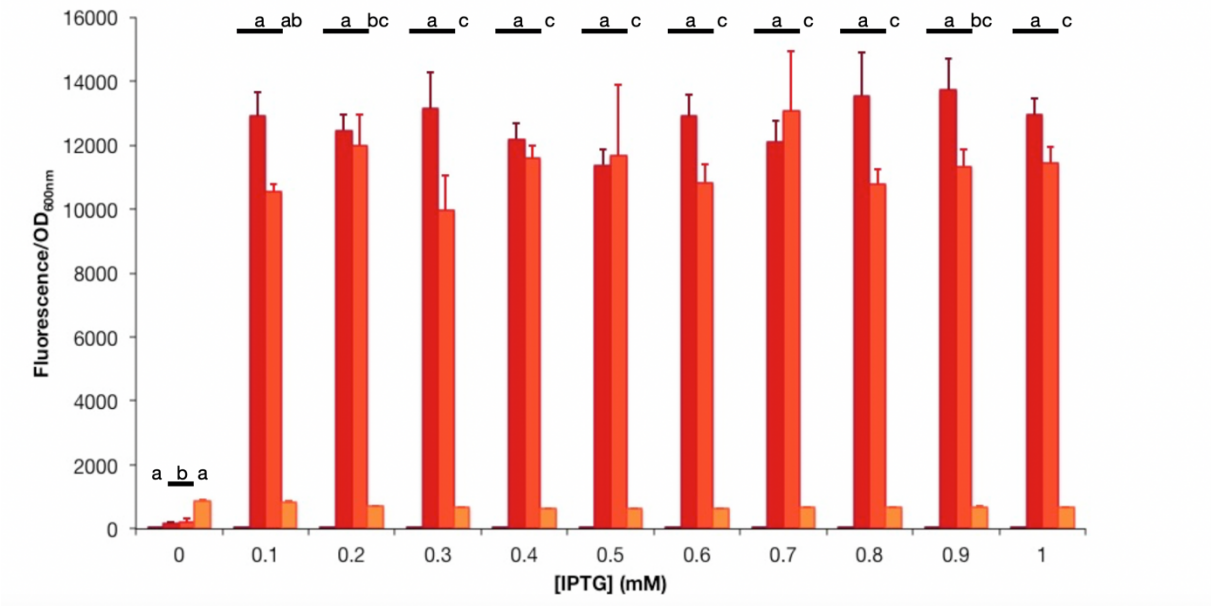

Figure 59. mRFP1 based expression screen of EmrB.

The overexpression of the inner membrane transporter EmrB as mRFP1 fusion was performed using the following *E. coli* strains (BL21(DE3) -■, C41(DE3)*ΔacrAB* -■, C43(DE3) -■, BW25113(DE3)*ΔacrAB* -■). 100 µL of the resuspended samples were used for OD_{600} and red fluorescence measurements (excitation wavelength at 576 nm). The expression tests were repeated four times. Values with the same letter for a given strain are not significantly different (ANOVA, Tukey HSD, $p<0.05$).

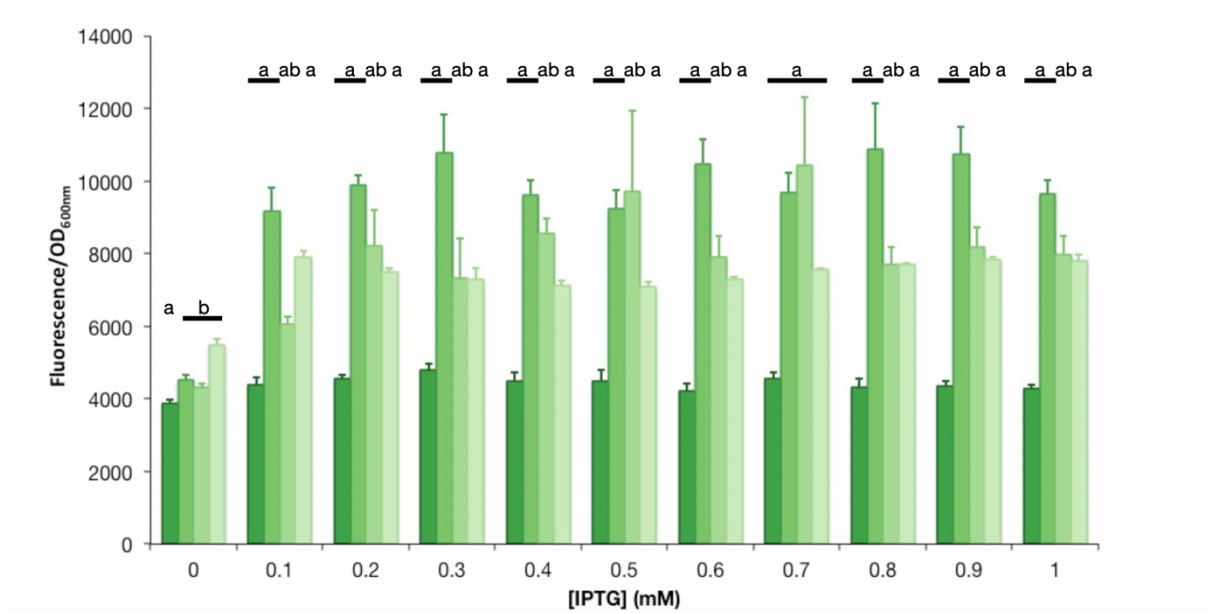

Figure 60. sfGFP based expression screen of EmrA.

The expression levels of the adaptor protein EmrA as sfGFP fusion were assayed in four different *E. coli* strains (BL21(DE3) -■, C41(DE3)*ΔacrAB* -■, C43(DE3) -■, BW25113(DE3)*ΔacrAB* -■). 100 µL of the resuspended samples were used for OD_{600} and green fluorescence measurements (excitation wavelength at 485 nm). The expression tests were repeated four times. Values with the same letter for a given strain are not significantly different (ANOVA, Tukey HSD, $p<0.05$).

As mentioned during the description of the high-throughput screening pipeline, the aim of this first co-expression screening experiment was to identify optimal conditions where both the inner membrane transporter EmrB as well as the periplasmic adaptor protein EmrA would be expressed in high amounts.

Considering both proteins and as already seen in previous reports, the concentration of the inducer (IPTG) did not seem to be a significant variable for the different strains tested[246] (Figures 59 and 60).

For EmrB, the expression levels were lowest in the BL21(DE3) and BW25113(DE3)*ΔacrAB* strains (Figure 59). Consistently, the widely used 'Walker strains' C41(DE3)*ΔacrAB* and C43(DE3) seemed to be best suited for its overexpression, with slightly higher expression levels for the C41(DE3)*ΔacrAB* strain.

Similar results were obtained for the overexpression of EmrA with the strains BL21(DE3), C41(DE3)*ΔacrAB*, and C43(DE3) (Figure 60). In this case the induction effect was visible for the

BW25113(DE3)*ΔacrAB* strain as well. The important background signal for the first expression condition (without inducer) could possibly be due to a less tight control level of the T7-promoter on the pRSFDMG_*emrA_tolC* construct.

Thus, as both EmrB and EmrA were highly expressed in the C41DE3*ΔacrAB* strain, this *E. coli* host was chosen for future co-expressions.

1.3.1.2. Comparison of the expression behaviors of the three partners forming the E. coli *efflux system*

During a second co-expression trial with the *E. coli* C41(DE3)*ΔacrAB* host, the expression levels of the *E. coli* tripartite system components were analyzed using all of the three DNA constructs prepared (EmrB-mRFP1/EmrA-sfGFP/TolC-sfGFP (*i.e.* both alternative sfGFP labeling strategies)). The results are shown in Figure 61.

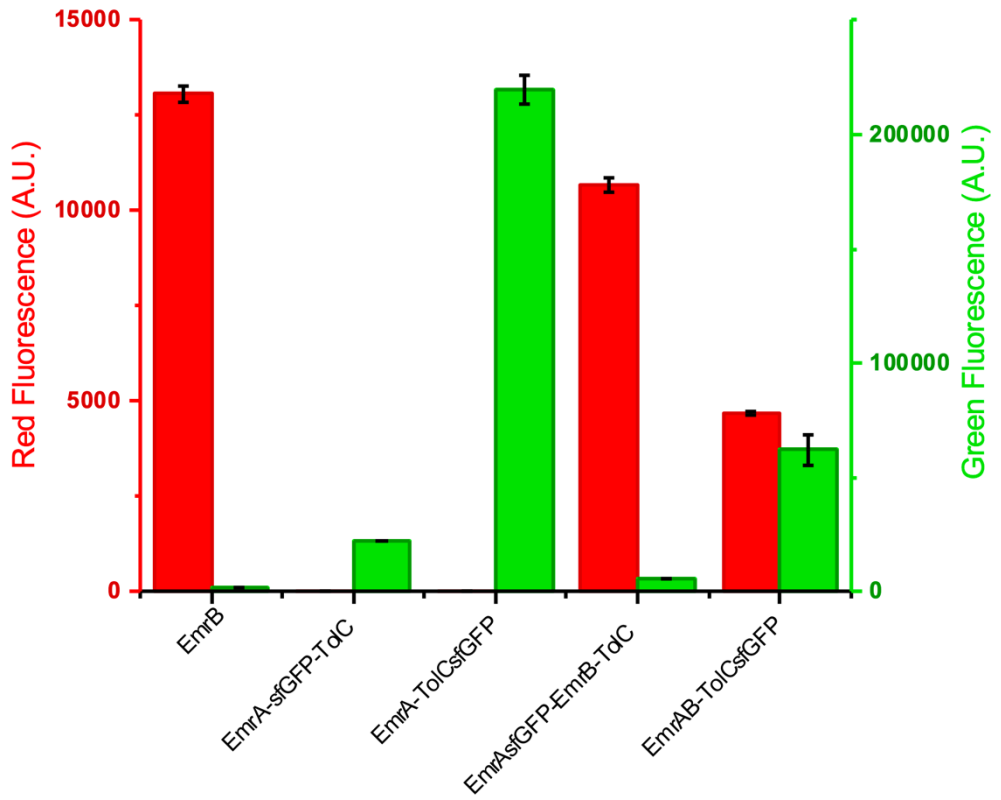

Figure 61. Expression level analyses of the *E. coli* efflux system components.

All the different constructs prepared for the expressions and co-expressions of the *E. coli* efflux system components were analyzed measuring the red and green fluorescence signals (excitation wavelengths at 576 and 485 nm respectively). The different values represent averages with the corresponding standard deviations of five simultaneous measurements of a 520 µL sample divided into 100 µL aliquots in each case.

First, all three complex components from the *E. coli* tripartite system were expressed. In addition, as expected all the negative controls showed that the red and green fluorescence signals did not overlap in each case.

Comparing the different expression levels, TolC seemed to be expressed at higher levels than EmrA in general both within the negative controls (EmrAsfGFP-TolC and EmrA-TolCsfGFP) as well as in the cells co-expressing all three proteins (EmrAsfGFP-EmrB-TolC and EmrAB-TolCsfGFP).

For the following FSEC analyses, the EmrB-mRFP1/EmrA-sfGFP labeling combination was chosen as other functional tests (not shown here) demonstrated that in the case of a RND system from *E. coli* (AcrAB-TolC) the labeling of TolC with sfGFP seemed to lower the antibiotic resistance capacity of cells harboring the p7XC3RH/pRSFDM_G construct combination.

1.3.2. Small-scale preliminary verifications

1.3.2.1. The two different fluorescent labels and three different affinity tags are correctly located on the three **E. coli** *protein partners*

After the 96 deep-well block expressions, the conditions with the highest expression levels for EmrB and EmrA were further analyzed for the *in gel* mobilities of all three proteins including TolC. In addition to *in gel* fluorescence measurements, a second elegant way of testing the presence of all three members forming the entire complex consisted in the performance of specific immunoblots towards the different affinity tags located at the C-terminus of each member (Figure 62).

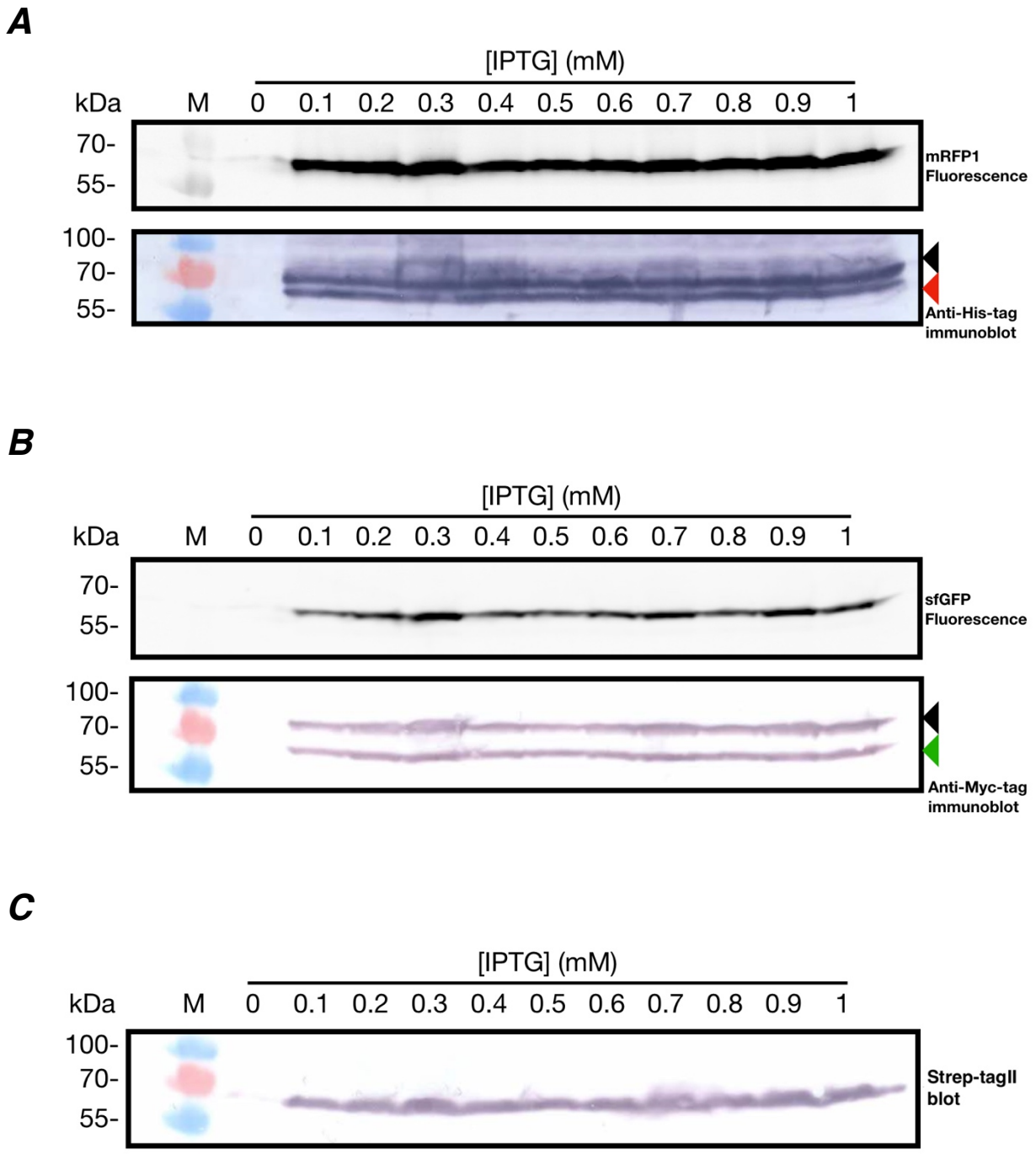

Figure 62. Electrophoretic mobility of mRFP1 and sfGFP fusion proteins expressed in *E. coli* C41(DE3)*ΔacrAB* cells.

(A) (Upper) *In gel* mRFP1 fluorescence used as marker for the expression of EmrB. (Lower) Immunoblot of the same gel decorated with anti-His-tag antibody to detect mRFP1-His-tag. (B) (Upper) *In gel* sfGFP fluorescence used as marker for the expression of EmrA. (Lower) Immunoblot of the same gel decorated with anti-Myc-tag antibody to detect sfGFP-Myc-tag. (C) Membrane decorated with Strep-Tactin Alkaline Phosphatase conjugate to detect TolC-Strep-tagII. Black and colored arrows indicate the positions of non-fluorescent and fluorescent species of mRFP1 and sfGFP fusion proteins respectively.

For EmrB-mRFP1-His-tag, the expression levels were visible as a single prominent fluorescent band. However, a second signal was also visible for the anti-His-tag immunoblot at about 10 kDa above the fluorescent signal (Figure 62A). Similar results were obtained for EmrA-sfGFP-Myc-tag where a second band was also visible at about 15 kDa above the fluorescent signal (Figure 62B). The second non-fluorescent signal in both cases might represent completely unfolded reporters. This signal could possibly be attributed to a misfolded fraction of each fusion protein expressed under these conditions, similar to previously described results for other membrane proteins[257]. Nevertheless, the unfolding of sfGFP remains intriguing. Finally, the expression of TolC-Strep-tagII was also checked with the anti-Strep-tagII Western blot signal (Figure 62C).

1.3.2.2. E. coli *EmrA fused to sfGFP is correctly folded*

In a second step, the aim was to check whether EmrA fused to sfGFP was correctly folded or not. Thus, whole cell samples (C41(DE3)*ΔacrAB* and C43(DE3)) from the 96 deep-well block expressions were selected for solubilization tests using DDM. The results obtained are shown in Figure 63.

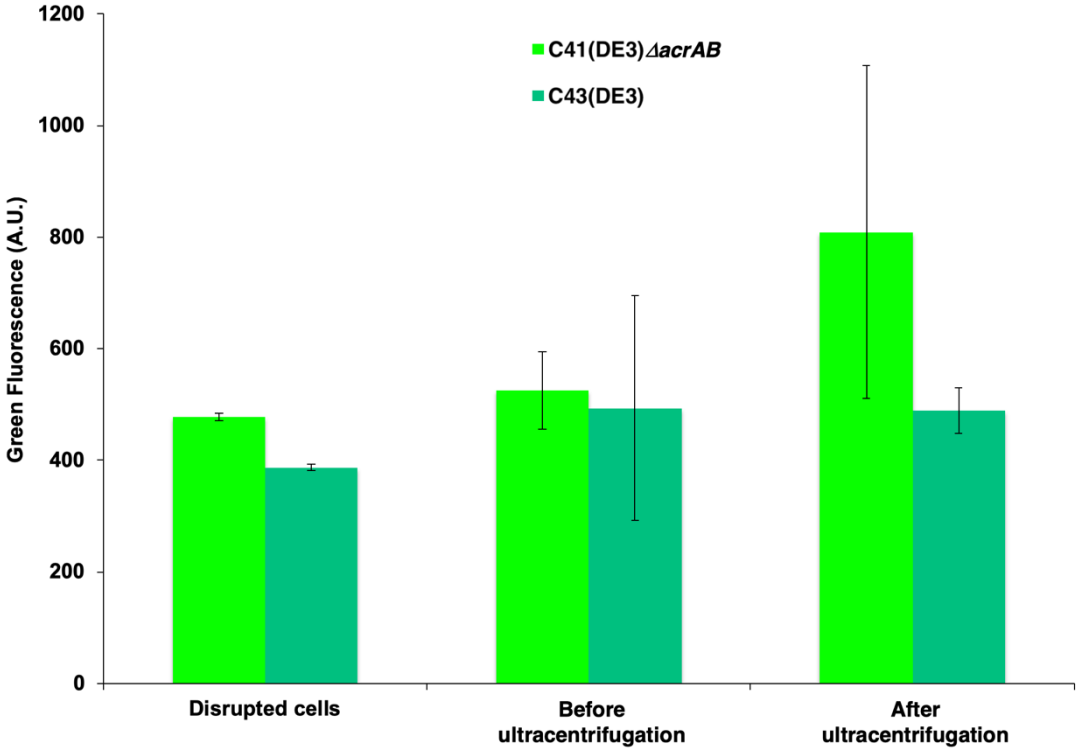

Figure 63. Comparison between green fluorescence measurements of whole cell samples at different solubilization test steps.

200 µL samples taken at each step were used for green fluorescence measurements (excitation wavelength at 485 nm). The different values represent averages with the corresponding standard deviations of two different solubilization tests for each expression strain. For the sample called 'After ultracentrifugation' the green fluorescence level of the supernatant was measured.

The comparison between the different green fluorescence values (whole cell samples, solubilized samples before and after ultracentrifugation) suggests that correctly folded EmrA can be produced using both *E. coli* expression hosts.

1.3.3. EmrAB-TolC from E. coli *is a tripartite complex*

Figure 64 shows the different FSEC results obtained for the EmrAB-TolC complex from *E. coli* using the detergent ANAPOE-$C_{12}E_{10}$ starting from whole cells and with a mild lysis.

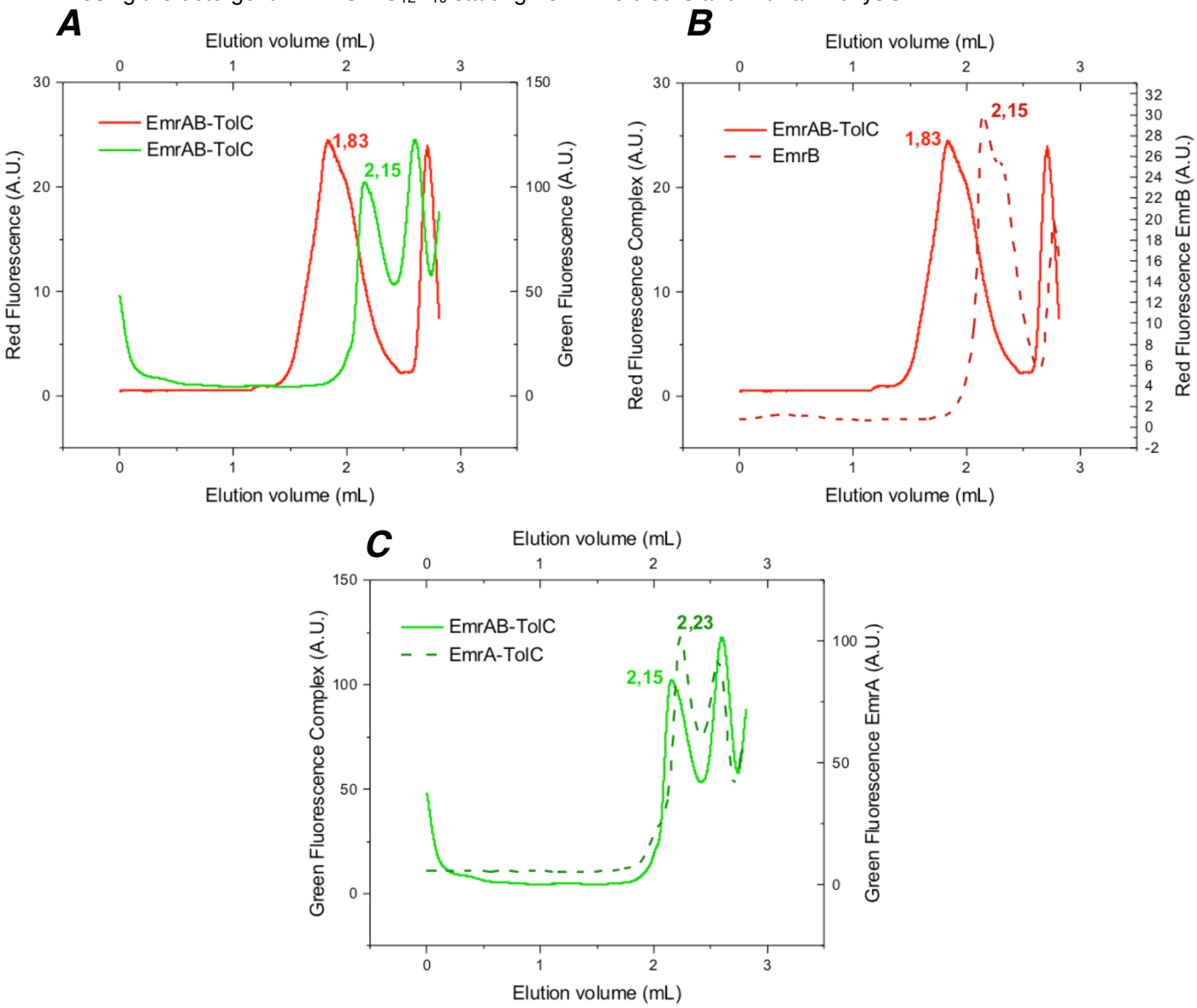

Figure 64. FSEC chromatograms of the EmrAB-TolC system from *E. coli* compared to each negative control.

(A) Red and green fluorescence chromatograms of the samples having co-expressed all three protein components (excitation wavelengths at 576 and 480 nm respectively). (B) Comparison of the red fluorescence elution profiles of EmrAB-TolC (red continuous line) with EmrB (dark red dashed line). (C) Comparison of the green fluorescence elution profiles of EmrAB-TolC (green continuous line) with the sample having co-expressed EmrA-sfGFP and TolC (dark green dashed line).

First, both the red and green fluorescence size-exclusion profiles were detected for the samples co-expressing all three proteins (Figure 64A). The slight shift in the elution volumes visible for the red fluorescence size-exclusion profiles might possibly indicate the presence of a tripartite complex (Figure 64B). However, no signal shift was visible for the green fluorescence size-exclusion profiles (Figure 64C). Furthermore, the non-superimposition of the red and green fluorescence size-exclusion profiles (Figure 64A) correlates with this second observation. No analyses could be made for identical experiments using the detergent DDM as the results obtained for the negative controls were not conclusive.

Interestingly, when the FSEC experiments were performed using cellular membranes as starting material, no signal shifts for the elution volumes were observed in general both with DDM and Triton X-100.

Moreover, noticeable variations of the elution profiles (with the detergent DDM) were seen in general between the FSEC experiments starting from cellular membranes (Figure 65A) and the FSEC analyses starting from whole cells with a mild lysis (Figure 65B).

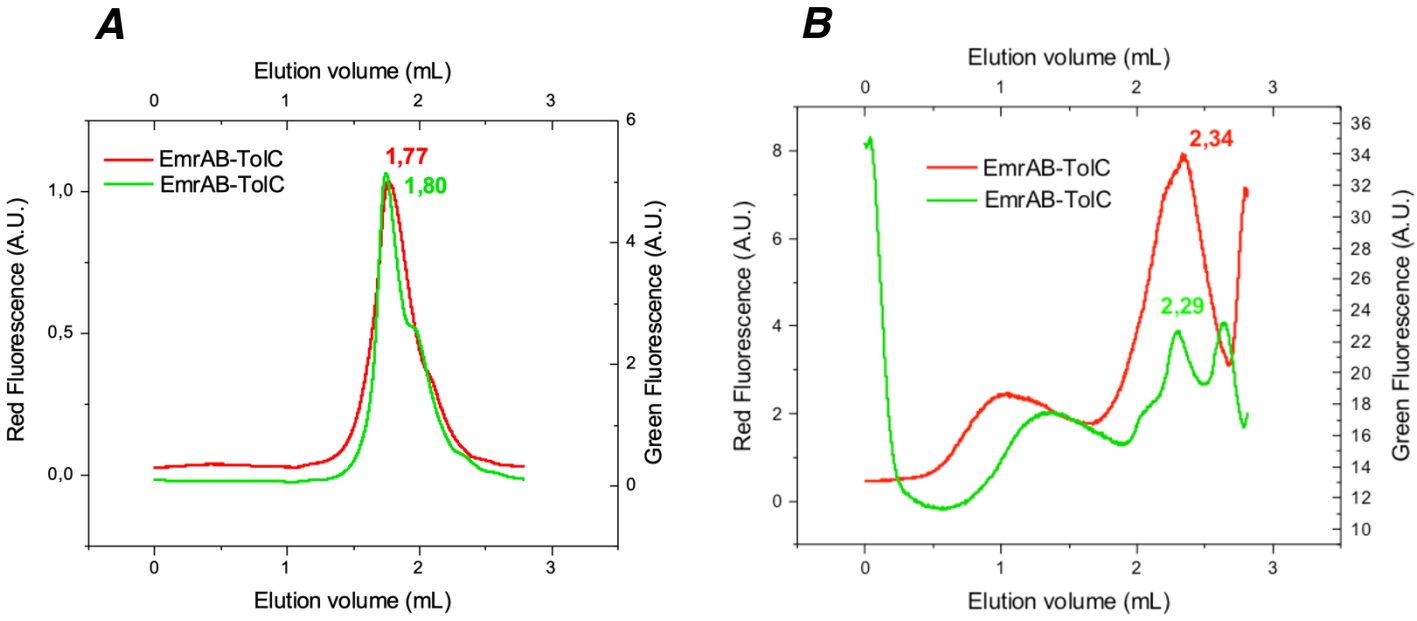

Figure 65. FSEC chromatograms of the EmrAB-TolC system from *E. coli* obtained via two different methods.

(A) Red and green fluorescence chromatograms of the cellular membrane samples containing all three protein components (excitation wavelengths at 576 and 480 nm respectively). (B) Red and green fluorescence chromatograms of the whole cell samples containing all three protein components.

Thus, considering the present observation and the previous FSEC results it was hypothesized that mechanical cell disruption could possibly have a deleterious effect for the isolation of the entire tripartite complex.

1.4. Comparison of the expression and complex formation behaviors of the E. coli *and* V. cholerae *efflux systems*

Figure 66 shows the different expression levels of the complex components from *E. coli* and *V. cholerae*.

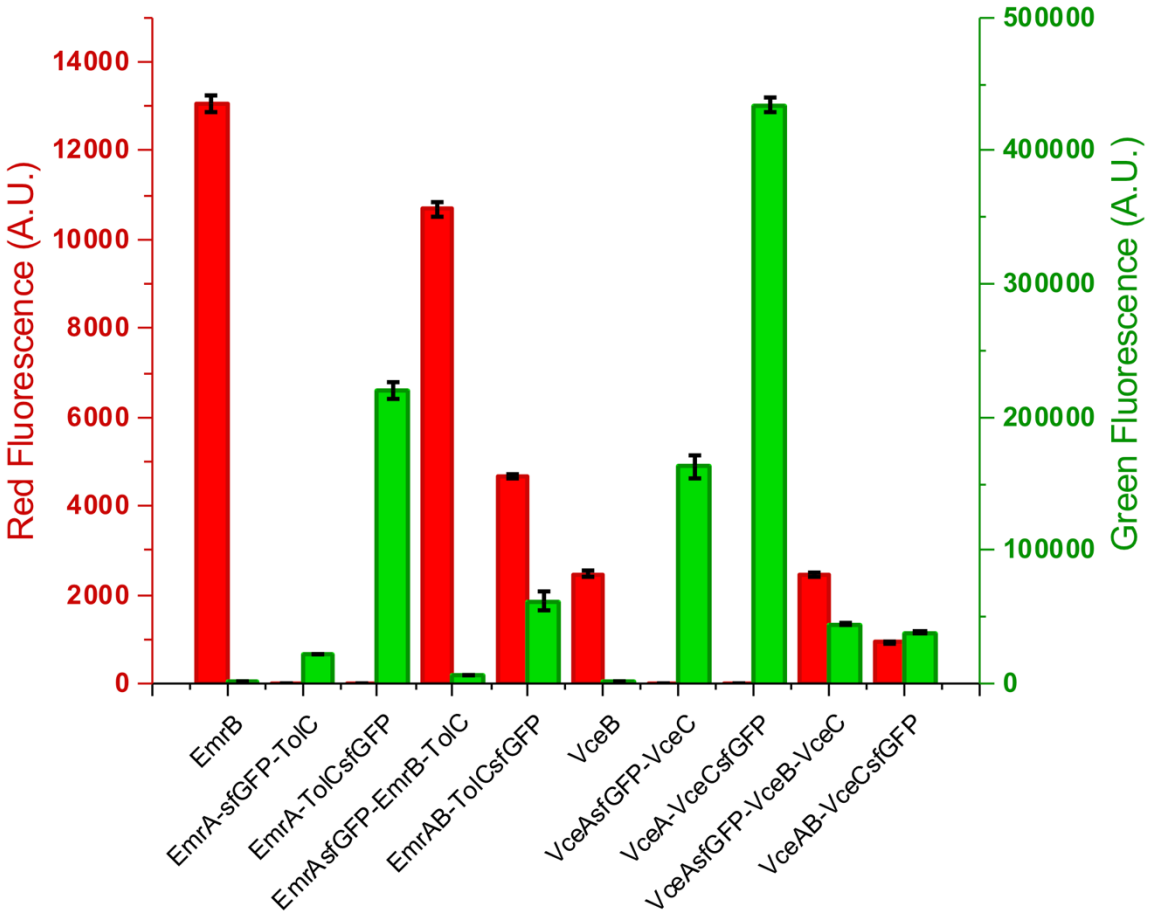

Figure 66. Comparison of the expression levels of the *E. coli* and *V. cholerae* systems.

The expression levels of the *V. cholerae* system components were analyzed measuring the red and green fluorescence signals for each sample in an identical manner as previously described for the *E. coli* system (Figure 61).

Overall, EmrB was expressed to higher extents than VceB. On the contrary, VceA was produced to higher amounts than EmrA both within the negative control VceAsfGFP-VceC as well as within cells co-expressing all three system components (VceAsfGFP-VceB-VceC). Finally, the comparison of the

negative controls VceA-VceCsfGFP and EmrA-TolCsfGFP showed that VceC was produced to higher amounts than TolC. However, the expression levels of VceC and TolC were comparable within cells co-expressing all three system components (VceAB-VceCsfGFP and EmrAB-TolCsfGFP respectively).

Similar to the *E. coli* system, VceC was expressed to higher amounts than VceA within the negative controls VceA-VceCsfGFP and VceAsfGFP-VceC respectively. However, contrary to the results obtained for the *E. coli* system, cells co-expressing all three *V. cholerae* proteins (VceAsfGFP-VceB-VceC and VceAB-VceCsfGFP) seemed to express the VceA and VceC components to similar amounts.

As previously explained for the *E. coli* system, also in this case the VceB-mRFP1/VceA-sfGFP labeling combination was chosen for different FSEC analyses. Figure 67 shows a comparison of the red fluorescence chromatograms of both systems starting from whole cells with a mild lysis and subsequent analyses using the detergent ANAPOE- $C_{12}E_{10}$.

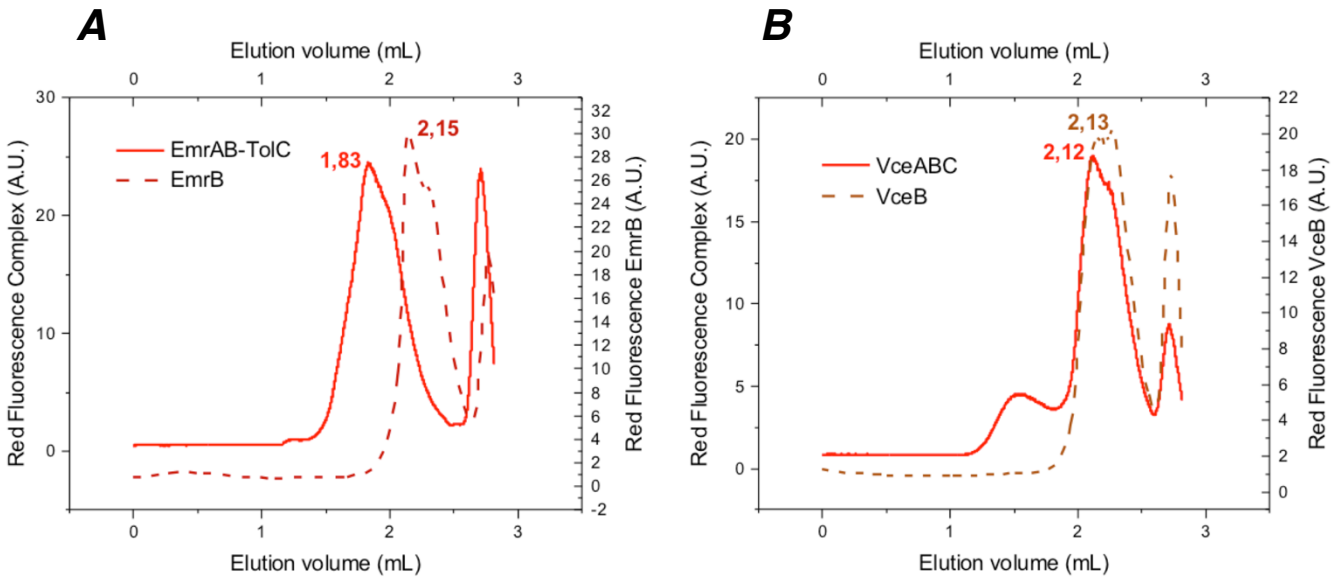

Figure 67. Comparison of the FSEC chromatograms of the *E. coli* and *V. cholerae* systems.

(A) Comparison of the red fluorescence elution profiles of EmrAB-TolC (red continuous line) with EmrB (dark red dashed line) (excitation wavelength at 576 nm). (B) Comparison of the red fluorescence elution profiles of VceABC (red continuous line) with VceB (dark red dashed line).

In contrast to the results obtained for the EmrAB-TolC system from *E. coli* (Figure 67A), no signal shift was observed for the VceABC system from *V. cholerae* when the red fluorescence signals were compared (Figure 67B). Similar results (not shown here) were also obtained with the detergent

DDM. Thus, no possible complex formation indications could be observed for the VceABC system from *V. cholerae* with the present methodology.

The EmrAB-TolC system from *E. coli* was therefore chosen for the following upscaling experiments for structural analyses.

2. Initial analyses of fusion stabilized EmrAB-TolC systems from E. coli

In a first attempt to biochemically analyze genetically stabilized EmrAB-TolC systems, four expression constructs (for the addition of variable GS-linkers) were chosen (Figure 68). Thus, in the present section the different methods employed for the small-scale analyses of these fusion stabilized complexes from *E. coli* will be mentioned with the corresponding results obtained in each case.

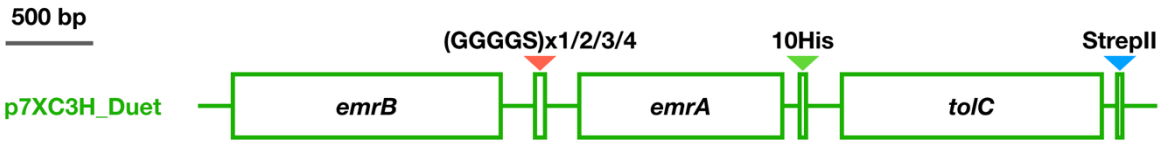

Figure 68. Schematic representation of the four constructs selected for biochemical analyses.

All the constructs present a T7 promoter system for the co-expression of the different EmrAB fusion chimeras together with TolC.

2.1. Experimental procedures

2.1.1. Small-scale co-expressions

2.1.1.1. Co-expressions for Western blot analyses

Similar to the previously described expression conditions, small-scale cultures of *E. coli* C41DE3*ΔacrAB* cells harboring the different constructs were prepared in 2xYT medium (1% (vol/vol) inoculation) supplemented with Kanamycin. The cultures were induced with 0.25 mM IPTG and further incubated overnight at 20°C. Cell samples were pelleted and normalized to a total OD_{600}=2 for subsequent Western blot analyses which were performed in a similar manner to the previously described method.

2.1.1.2. Co-expressions for a pull down assay

Small-scale cultures of *E. coli* C41DE3*ΔacrAB* cells harboring the EmrAB fusion construct with (GGGGS)x3 as well as cells transformed with the pRSFDMG_*emrA_tolC_Ec* construct used as a negative control were prepared in 2xYT medium. The cultures were induced with 1 mM IPTG and further incubated overnight at 25°C. Cell samples were pelleted and normalized to a total OD_{600}=10 for a subsequent pull down assay.

2.1.2. Pull down assay

In order to verify if the fusion stabilized systems formed complexes which could possibly be isolated, a His-tag based pull down assay was used. Therefore, cell samples were resuspended in 400 µL of lysis buffer (50 mM Tris/HCl pH 8, 400 mM NaCl, 10 mM $MgCl_2$, 2 mM $CaCl_2$, 0.2 mM DFP, 25 000 U/mL DNase I, and 5 mg/mL lysozyme). Samples were subsequently incubated at 4°C for 1h under constant agitation. For the solubilization, 2% (wt/vol) DDM was added to each sample with a subsequent incubation at 4°C for 2h under constant agitation. After a prior ultracentrifugation step at 356,000 *g* for 10 min at 4°C, HisPur Ni-NTA magnetic beads (Thermo Fisher Scientific, USA) were used for the pull down assay following the manufacturer's instructions for each supernatant. Finally, the eluted samples were analyzed via Western blotting in an identical manner as the previously described methodology.

2.2. Results

2.2.1. The EmrAB fusion chimeras are co-expressed with TolC

During a first step, the co-expression behavior of the EmrAB fusion chimeras and TolC was checked using four different expression constructs (Figure 68). The corresponding Western blot analyses are shown in Figure 69. Besides the EmrAB fusion with (GGGGS)x2, all the other fusion chimeras were co-expressed with TolC.

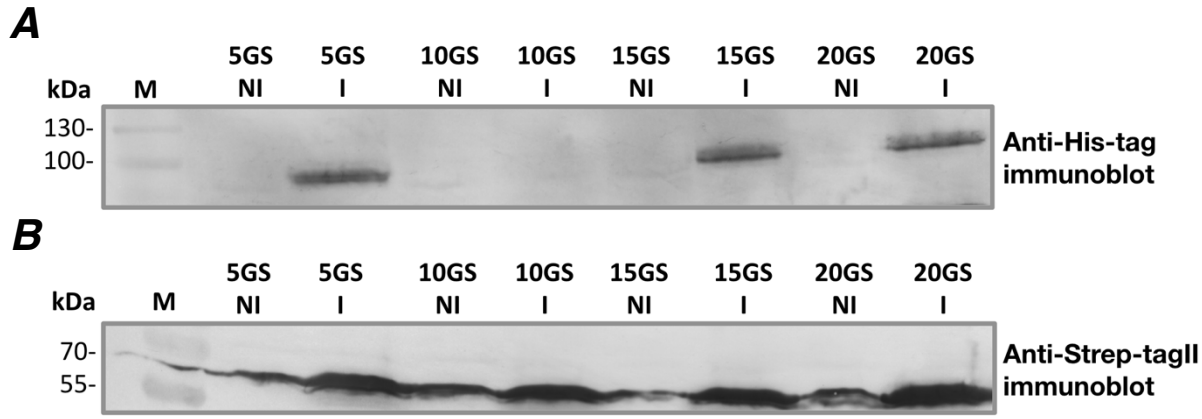

Figure 69. Western blot analyses of the EmrAB fusion chimeras co-expressed with TolC.

(A) Signals corresponding to the different EmrAB fusion chimeras with different linker sizes (either 5, 10, 15 or 20 Glycine-Serine (GS) repeats). (B) Signals corresponding to TolC of the four different constructs. NI, not induced. I, induced.

2.2.2. The 15GS EmrAB fusion can be isolated with TolC

In order to check whether the fusion stabilized EmrAB-TolC system could be isolated, a pull down assay was used (Figure 70). The aim was to isolate the entire complex via the His-tag located at the C-terminus of the EmrB-15GS-EmrA fusion chimera. The EmrA-Myc-tag and TolC-Strep-tagII (co-expressed using the contruct pRSFDMG_*emrA_tolC_Ec*) were used as a negative control during the assay. As seen in Figure 70, the EmrB-15GS-EmrA fusion chimera can possibly be isolated with TolC (Elution sample). Nevertheless, a signal was also visible for the negative control which could be linked to a non-specific binding of TolC-Strep-tagII with the Ni^{2+}-NTA magnetic beads.

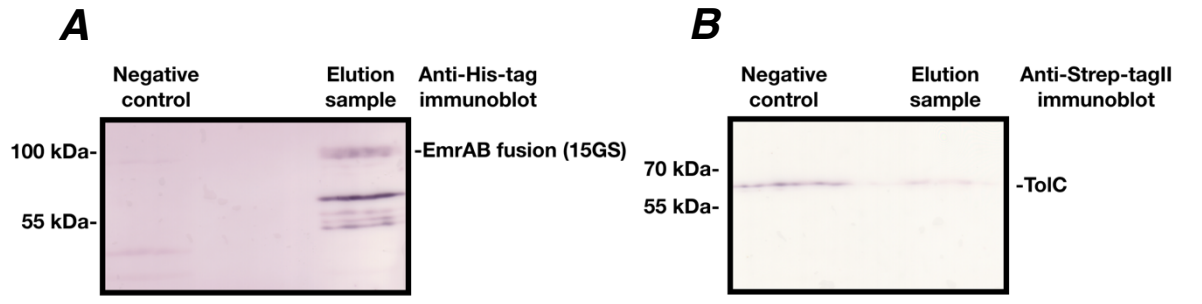

Figure 70. Western blot analyses of the pull down assay samples (fusion chimeras).

(A) Anti His-tag immunoblot of the eluted samples using Ni^{2+}-NTA magnetic beads. (B) Anti-Strep-tagII immunoblot of the eluted samples using Ni^{2+}-NTA magnetic beads. In both cases 28 µL of the elution samples were loaded.

Chapter IV: Large-scale co-expression, co-purification and EM analysis of the EmrAB-TolC complex from E. coli

In this chapter, the first strategy with the red and green fluorescent labels used for purification analyses will be mentioned together with the corresponding results obtained.

Similar to the previous chapter, the second alternative strategy making use of the 15GS EmrAB fusion chimera will also be described.

1. First strategy making use of the red and green fluorescent labels

Here, the different methodologies employed for the study of the EmrAB-TolC system from *E. coli* will be mentioned together with the corresponding results obtained.

1.1. Experimental procedures

1.1.1. Co-expression of EmrAB-TolC

The following description corresponds to the optimized procedure for the co-expression of the entire EmrAB-TolC system from *E. coli*. For the production, 1 L of 2xYT medium (in a 3L Fernbach flask) was inoculated with an overnight culture of *E. coli* C41DE3*ΔacrAB* [255] cells harboring the constructs p7XC3RH_emrB and pRSFDMG_emrA_tolC to a final OD_{600} of 0.05. The growth was continued at 30°C under vigorous aeration to reach an OD_{600} of 0.6. The culture was subsequently cooled at 4°C for 30 min. After the incubation period, 0.5 mM IPTG was used for the induction and the culture was continued at 25°C overnight. Cells were harvested by centrifugation at 5,525 *g* for 20 min at 4°C.

The other procedures tested for the optimization of the co-expression of the entire EmrAB-TolC system are listed in Table 16.

Table 16. Other co-expression conditions tested for the EmrAB-TolC complex from *E. coli*.

Number	Strain	Medium	IPTG (mM)	Temperature before induction (°C)	Temperature after induction (°C)	Time (hours)	Flask type
1	C41(DE3)*ΔacrAB*	2xYT (1L)	0.5	37	25	~18	5L Baffled Erlenmeyer flask
2	C41(DE3)*ΔacrAB*	2xYT (2L)	0.5	37	25	~18	5L Baffled Erlenmeyer flask
3	C41(DE3)*ΔacrAB*	2xYT (2L)	0.5	37	22	~18	5L Baffled Erlenmeyer flask
4	C41(DE3)*ΔacrAB*	2xYT (1L)	0.5	37	22	~18	5L Baffled Erlenmeyer flask
5	C41(DE3)*ΔacrAB*	TB (1L)	0.05	37	25	~18	2,5 L Tunair Baffled flask
6	C41(DE3)*ΔacrAB*	TB (1L)	0.3	37	25	~18	2,5 L Tunair Baffled flask

Number	Strain	Medium	IPTG (mM)	Temperature before induction (°C)	Temperature after induction (°C)	Time (hours)	Flask type
7	C41(DE3)*ΔacrAB*	TB (1L)	1	37	25	~18	2,5 L Tunair Baffled flask
8	C43(DE3)*ΔacrABD*	2xYT (2L)	0.25	37	20	~18	5L Baffled Erlenmeyer flask
9	C41(DE3)*ΔacrAB*	2xYT (0.5L)	1	30	25	~18	3L Baffled Erlenmeyer flask
10	C41(DE3)*ΔacrAB*	2xYT (1L)	1	30	25	~18	3L Fernbach flask

1.1.2. Co-purification of EmrAB-TolC

In a similar fashion to the previous section, the following description corresponds to the optimized co-purification procedure of EmrAB-TolC. The cell pellet of a 1L culture was resuspended in buffer A (50 mM Tris/HCl pH 7.5, 400 mM NaCl, 2 mM EDTA, 10 mM $MgCl_2$, 2 mM $CaCl_2$, 0.2 mM DFP and 952 U/mL DNaseI). Cells were gently disrupted with the addition of 1 mg/mL lysozyme and the further incubation at 37°C for 30 min. After a centrifugation step at 25,000 *g* the pellet was resuspended in buffer B (50 mM Tris/HCl pH 7.5, 400 mM NaCl and 714 U/mL DNaseI). The EmrAB-TolC system was solubilized using 2 % (wt/vol) DDM for 2 h at 4°C. The insoluble fraction was removed by ultracentrifugation at 108,800 *g* for 1 h at 4°C. Prior to the purification, the supernatant was passed through a 0.22 μm filter (Sartorius, Germany) and 20 mM Imidazole pH 7 was added to it. The supernatant was subsequently injected on a 5 mL Ni^{2+}-NTA column (GE Healthcare, USA) equilibrated with buffer I (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 20 mM Imidazole and 0.025 % (wt/vol) DDM). The elution of EmrAB-TolC was performed using a linear gradient of buffer II (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 500 mM Imidazole, and 0.025 % (wt/vol) DDM) ranging from 0 to 100 %. The eluted complex was subsequently concentrated using a 300 kDa cutoff concentrator (Sartorius, Germany) and subjected to size exclusion chromatography with a Superose 6 HR 10/30 column (Amersham Biosciences, UK) (Flow rate: 0.5 mL/min). The buffer used for the size exclusion chromatography was buffer III (50 mM Tris/HCl pH 7.5, 150 mM NaCl, and 0.025 % (wt/vol) DDM). Different fractions containing the EmrAB-TolC system were pooled and concentrated to about 1 mg/mL using a 100 kDa

cutoff concentrator (Merck KGaA, Germany). Amphipol A8-35 (10 mg/mL) (Anatrace, USA) was mixed with the protein solution with a mass ratio of amphipol A8-35 to protein of 4:1. The mixture was subsequently incubated at 4°C for 2h. Detergent was removed with the addition of SM2 Bio-beads (Bio-Rad, USA) into the mixture which was further incubated at 4°C for 3h with a gentle shaking. The mixture was afterwards subjected to size exclusion chromatogramphy (with a flow rate of 0.05 mL/min) on a Superose 6 3.2/300 column (GE Healthcare, USA) equilibrated with buffer IV (50 mM Tris/HCl pH 7.5, 150 mM NaCl, and 0.01 % (wt/vol) NaN_3). The different *in gel* fluorescence and immunodetection analyses during the purification procedure were performed in an identical manner to the previous descriptions. The silver staining procedure was performed using the PlusOne Silver staining Kit (GE Healthcare, USA) following the manufacturer's instructions. The other procedures tested for the co-purification optimization of EmrAB-TolC are summarized in Figure 71.

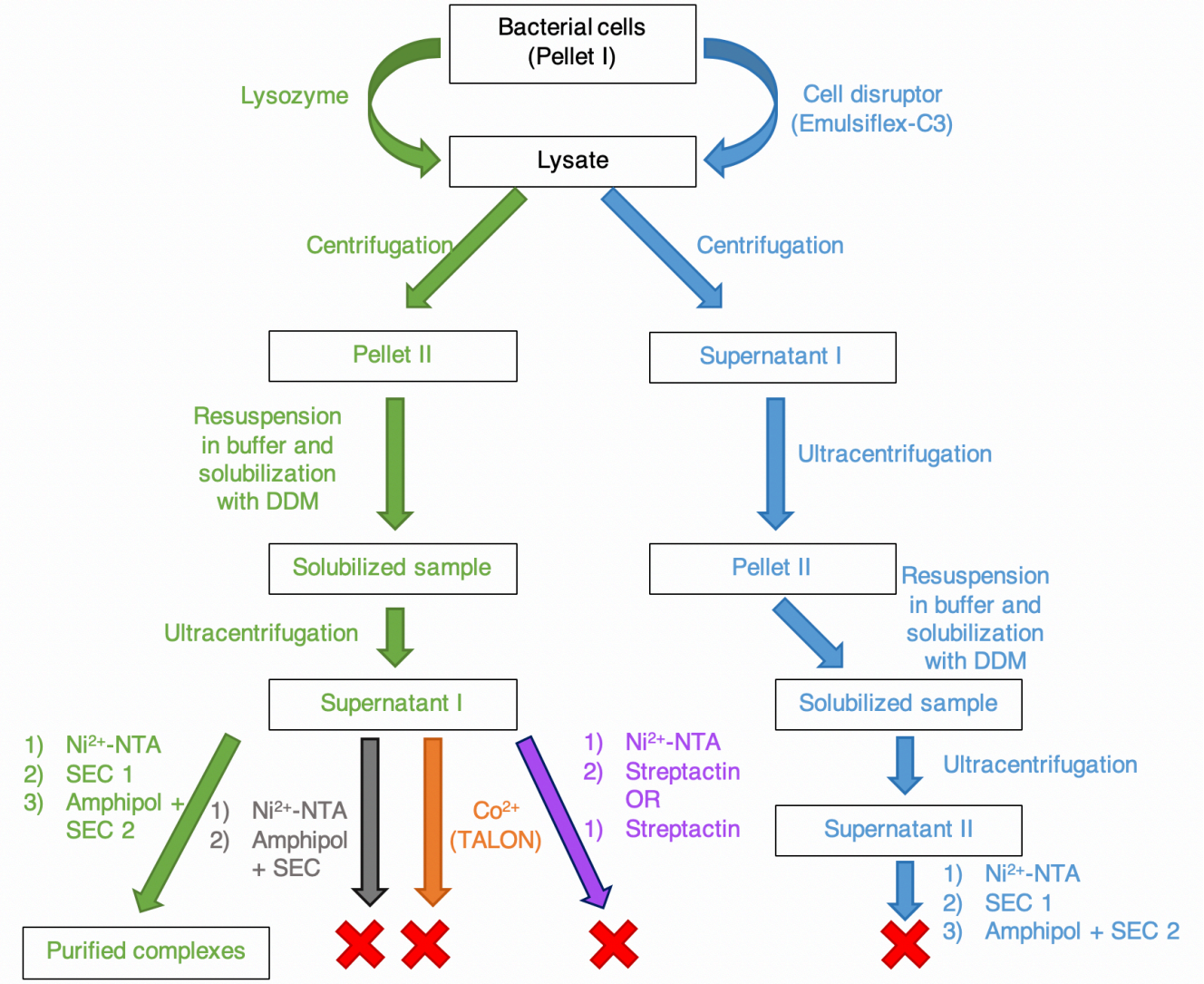

Figure 71. Scheme describing the purification procedures tested for the isolation of EmrAB-TolC.

The green pathway (green arrows) represents the methodology enabling the isolation of the EmrAB-TolC system.

1.1.3. Negative staining EM analysis of EmrAB-TolC

For EM grid preparations, the sample suspension diluted three times in 50 mM Tris/HCl pH 7.5, 150 mM NaCl, and 0.01 % (wt/vol) NaN_3 was applied to a glow-discharged carbon-coated copper 300 mesh grids and stained with 2 % (wt/vol) uranyl acetate solution. Images were acquired on a Tecnai F20 electron microscope (ThermoFisher FEI, USA) operated at 200 kV using a Eagle 4k_4k camera (ThermoFisher FEI, USA) with a nominal magnification of 80,000 and using serial EM software for automatic acquisition. Micrographs were collected with a nominal defocus range of -1.5 to -3 μm and a low dose of ~30 electrons/Å. All images were binned (2x2 pixels) to obtain a pixel size of 2.94 Å on the specimen level. Images were processed with EMAN2 software suite. A total of 518 particles were manually picked with a box size of 192x192 pixels. Following contrast transfer function (CTF) fitting particle sets were built and 2D reference-free alignement and classification of particle projections was performed. For AcrAB-TolC a set of 900 particles was manually picked.

1.1.4. Identification of pumps by labelling with nitrilotriacetic acid-nanogold

Samples containing the tripartite complex were diluted four times in 50 mM Tris/HCl pH 7.5, 150 mM NaCl and were further loaded on glow-discharged carbon coated Nickel 300 mesh EM grids. Excess liquid was removed using Whatman filter paper number 5 (GE Healthcare, USA). Grids were then washed two times quickly with 50 mM Tris/HCl pH 7.5, 150 mM NaCl containing 20 mM imidazole. Grids were placed upside-down on a droplet of 5 nm Ni-NTA-Nanogold (Nanoprobes, USA) diluted at 50 nM in 50 mM Tris/HCl pH 7.5, 150 mM NaCl containing 20 mM imidazole after a 5 min spinning of the solution at 2,000 *g*. After a 15 min incubation step at room temperature, grids were washed upside-down with two droplets of 50 mM Tris/HCl pH 7.5, 150 mM NaCl for 2 min at room temperature followed by one droplet of water. Finally, negative staining was performed placing grids upside-down on two droplets of Uranyl Formate for one min before removing the excess using Whatman filter paper number 5 (GE Healthcare, USA). After drying, images were acquired in an identical manner as previously described at a magnification of 50,000 x.

1.2. Results

1.2.1. The entire EmrAB-TolC complex can be isolated directly from bacterial cells

For the co-expression of EmrAB-TolC, different procedures tested showed that aeration seemed to be the most important variable. Indeed the best expression results were obtained using 3L Fernbach flasks. Furthermore, 2xYT medium seemed to be better suited than TB for the production. Finally, the best expression results were obtained using *E. coli* C41DE3*ΔacrAB* as host.

Isolation of the entire EmrAB-TolC complex for structural studies was challenging as the assembly presented a limited stability during the procedure. Therefore the strategy employed made use of the fluorescent reporters and the affinity tags to detect all three components during co-purification in a straightforeward manner. The presence of the fluorescent labels did not seem to impair the formation of the entire complex. Mechanical cell disruption seemed to alter the tripartite system and was therefore replaced by a mild chemical lysis of cells with lysozyme (Figure 71). Despite numerous efforts to perform a tandem-affinity co-purification of the entire system (by the His-tag and the Strep-tagII), the yields of the materials recovered at the end of the procedure were very low regardless of the performance order of the affinity chromatography techniques (*i.e.* Ni^{2+}-NTA and Streptactin). Similarly, replacing the Ni^{2+} ions by Co^{2+} ions for the His-tag based affinity purification did not yield enough material for the structural analysis (Figures 71 and 72).

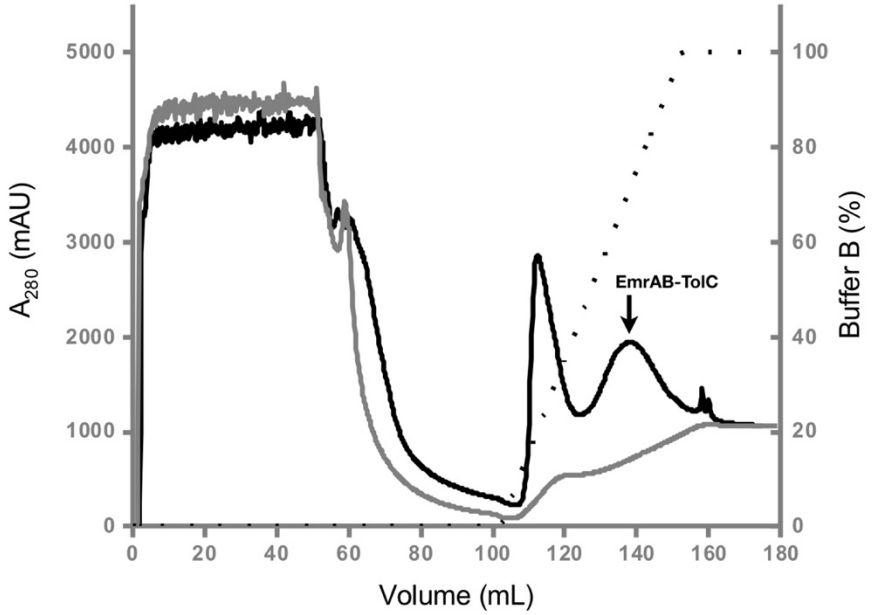

Figure 72. Comparison of the Ni^{2+} and Co^{2+} based affinity purifications.

The elution of the Ni^{2+} based affinity chromatography is shown as a black curve. The second peak corresponds to the EmrAB-TolC co-purification. The result of the Co^{2+} based affinity purification is evidenced as a grey curve. The second peak was not observed in this case. The dotted line indicates the linear grandient of imidazole for both cases.

Therefore only the Ni^{2+}-NTA affinity purification step was chosen for the isolation of the complex. The second purification step using size-exclusion chromatography with the Superose 6 10/300 or Superose 6 HR 10/30 column is necessary for the further isolation of the entire system (Figures 71 and 73). No enrichment of the entire system was visible if this step was removed from the purification procedure.

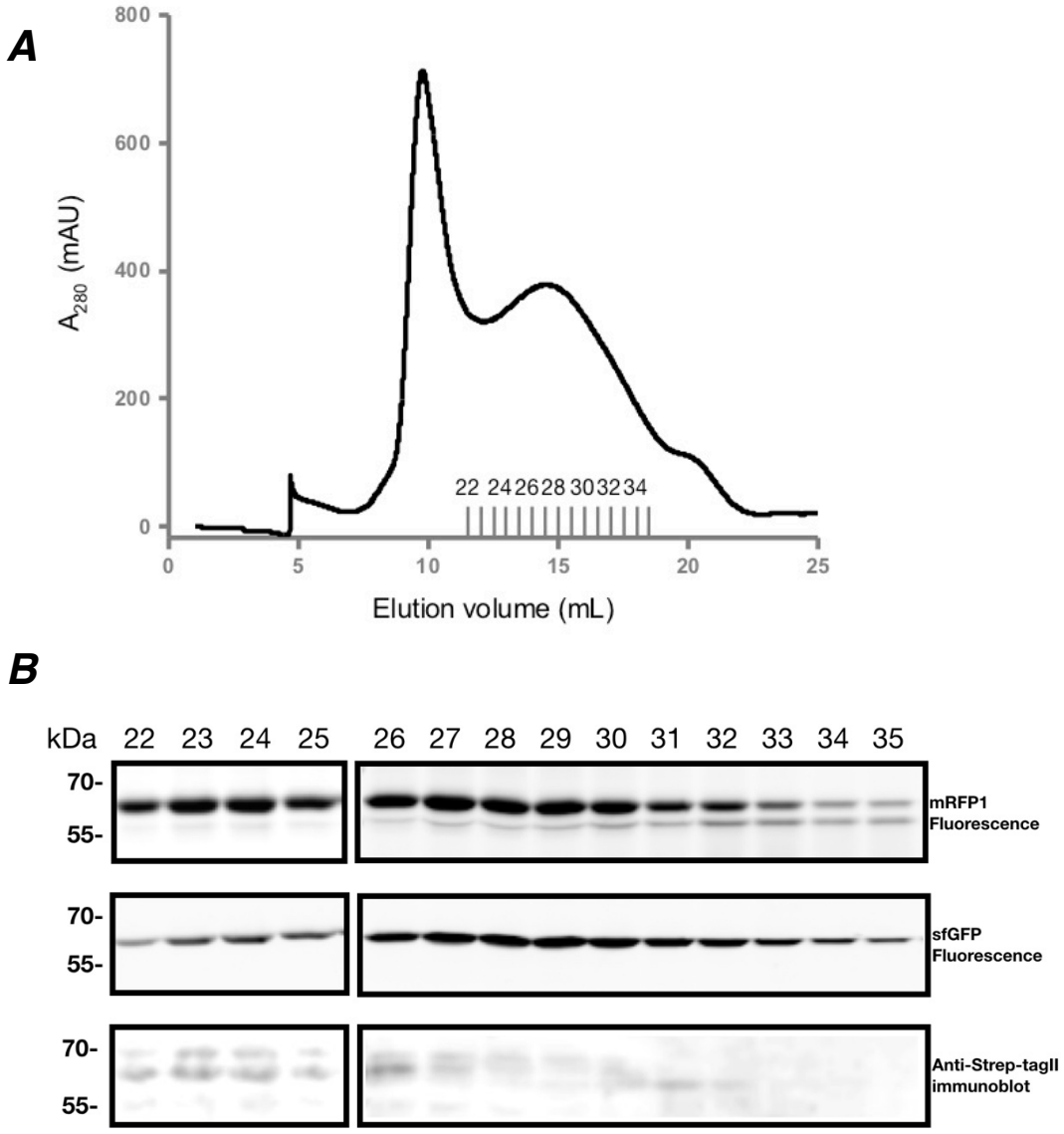

Figure 73. Large-scale co-purification analysis of EmrAB-TolC.

(A) Size-exclusion chromatography profile of the co-purification sample from the Ni-NTA purification obtained with a Superose 6 10/300 column. (B) Electrophoretic mobility of the indicated SEC fractions. The first peak corresponds to aggregates. Fractions 22 to 26 contained all three protein partners. 12 μL from 500 μL fractions were loaded in each case. mRFP1 fluorescence (excitation: 630 nm, emission: 670 nm). sfGFP fluorescence (excitation: 460 nm, emission: 515 nm).

Amphipol A8-35 was chosen for the stabilization of the tripartite system, as the procedure needed limited optimization compared to other detergent removal methods. After reconstitution into amphipol A8-35 and a second size exclusion chromatography step with a Superose 6 3.2/300 column to remove the excess amphipol, a stable tripartite system was obtained (Figure 74).

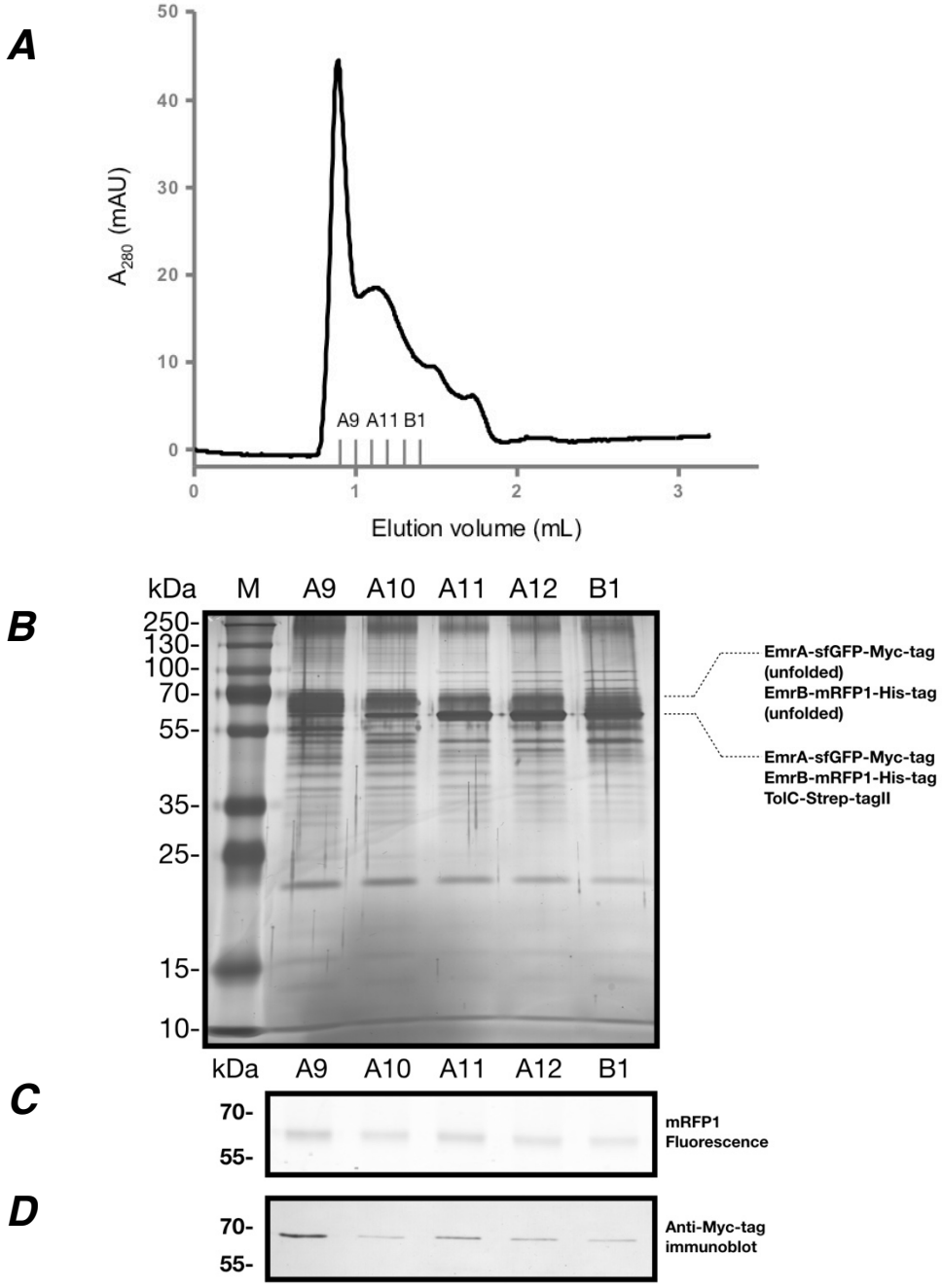

Figure 74. Size-exclusion chromatography profile and electrophoretic mobility analysis of EmrAB-TolC stabilized with Amphipol A8-35.

(A) Size-exclusion chromatography profile of EmrAB-TolC stabilized with Amphipol A8-35, fractions A9-A12 and B1 were collected and subjected to SDS-PAGE-, *in gel* fluorescence- and Western Blot-analysis. (B) Silver-stained SDS-polyacrylamide gel of fractions A9-A12 and B1 from the SEC profile in (A). 200 ng of each sample was loaded on the gel. (C) *In gel* mRFP1 fluorescence signal of fractions A9-A12 and B1(excitation: 633 nm, emission: 670 nm). 12 μL from 100 μL fractions were loaded in each case. (D) Anti-Myc-tag immunoblot of fractions A9-A12 and B1. 12 μL from 100 μL fractions were loaded in each case.

1.2.2. The entire EmrAB-TolC complex exhibits an elongated structure

Negative staining EM analysis of the A10, A11, A12 and B1 fractions (see Figure 74) of the size-exclusion chromatography after reconstitution in Amphipol A8-35 revealed elongated structures viewed from their sides (Figure 75A). An average image revealed the structure of the complex of about 33 nm in length with an Amphipol belt at both extremitites (Figure 75B). This finding is in consistency with the length of the AcrAB-TolC complex (Figure 75C) determined in similar EM conditions[242]. Both averages revealed similar features at the top corresponding to the densitites of TolC comprising a β-barrel channel and a periplasmic α-helical barrel. In contrast to the upper parts, the lower parts exhibited different features. The EmrAB part looked thinner than the AcrAB part. Unlike AcrB, EmrB does not contain any periplasmic domain, the densities in between the lower end of TolC and the second Amphipol belt were 15 nm in length and most likely correspond to the periplasmic part of EmrA including the α-helical coiled-coil, the lipoyl and the β-barrel domains. The transmembrane α-helices of EmrB and the N-terminal transmembrane α-helices of EmrA embedded in the Amphipol belt were not visible.

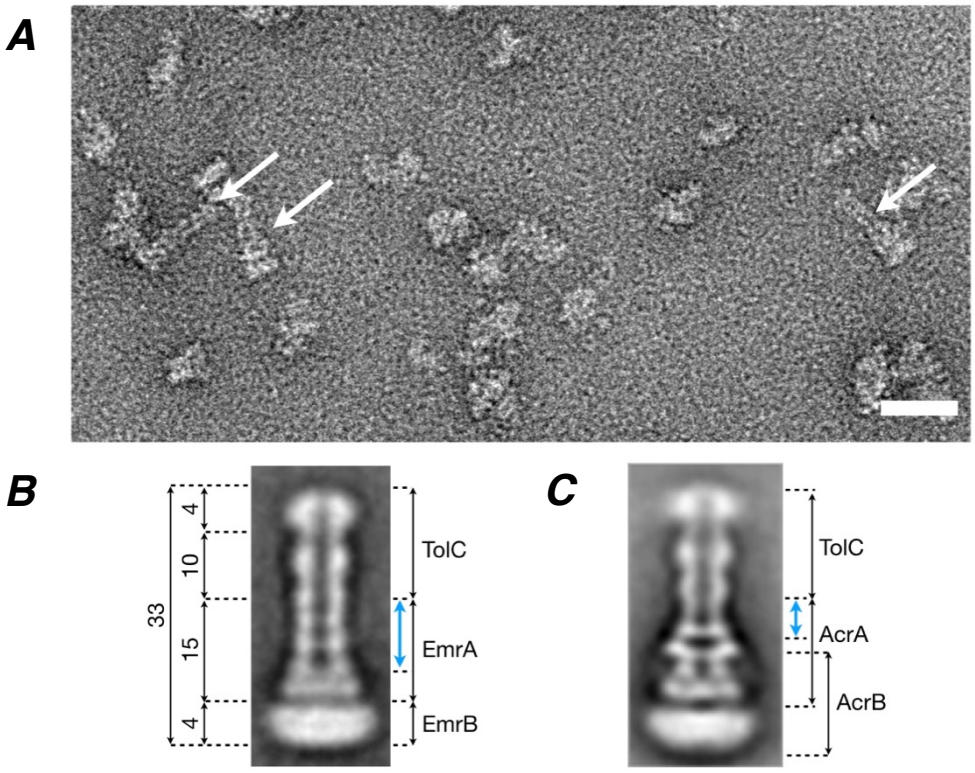

Figure 75. TEM analysis of the tripartite EmrAB-TolC efflux system.

(A) Field of view showing side views of EmrAB-TolC assemblies evidenced by the white arrows (scale bar, 30 nm). (B) Average image of the EmrAB-TolC complex showing densities corresponding to TolC, EmrA and the Amphipol belts at both ends (distances between the different components are indicated in nm). (C) For comparison, an average image obtained under similar EM conditions of the AcrAB-TolC complex reconstituted as published previously (reference 242) is shown. The densities corresponding to the three components are indicated. In the periplasmic part, note that the density overlap encountered for AcrA and AcrB is not predicted for EmrA and EmrB. The blue arrows indicate the α-helical coiled-coil domains.

1.2.3. The identity of the entire EmrAB-TolC efflux system could be confirmed by gold labelling

For the verification of the identity of the EmrAB-TolC efflux system, a gold labelling method was employed. As the elongated structures described previously were successfully labeled with nitilotriacetic acid-nanogold (Figure 76 and 77), these were attributed to the entire EmrAB-TolC complex. Note that care must be taken when interpreting the stoichiometry of the EmrB component as the sensitivity of the assay was not sufficient for such purposes. The remaining interaction of the gold labels can be due to either to a non specific electrostatic interaction with the carbon coating or a specific interaction with the excess free EmrB present in the sample.

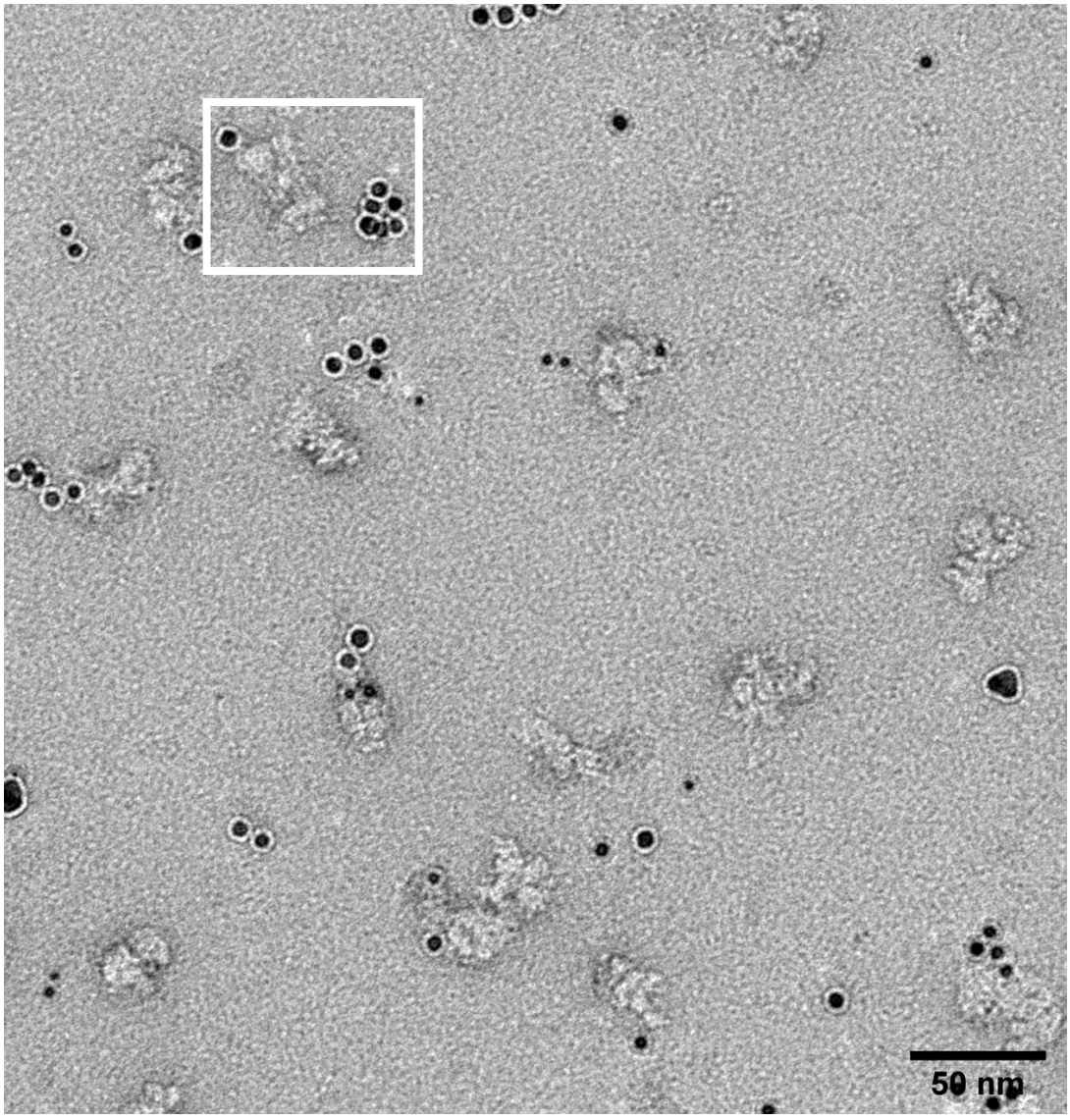

Figure 76. Wide field image of negatively stained and gold labeled sample containing EmrAB-TolC complexes.

Two EmrAB-TolC complexes labeled with nitrilotriacetic acid-nanogold are indicated with the white rectangle.

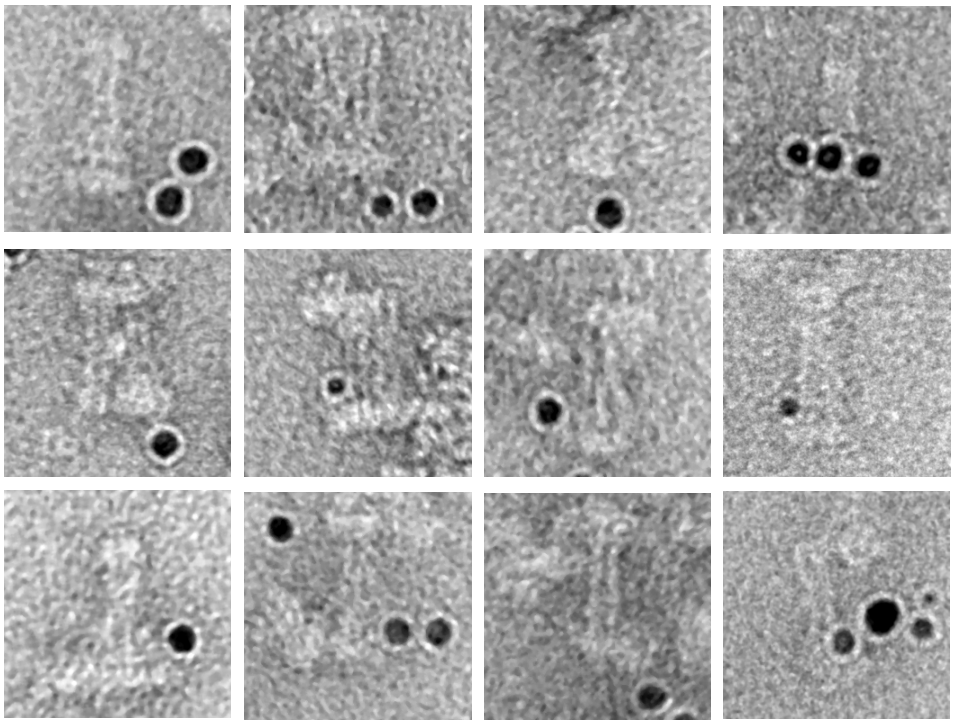

Figure 77. A gallery of electron micrographs of negatively stained and gold labeled EmrAB-TolC complexes.

The 5 nm-diameter gold particles (black) were attached to the C-terminal His-tag of EmrB. Each panel is cropped from the wide field images at 50 x 50 nm to show multiple objects of similar composition.

2. Second strategy making use of the EmrAB fusion chimera

Here, the different methodologies employed for the study of the 15 GS EmrAB fusion-TolC system from *E. coli* will be described together with the corresponding results obtained.

2.1. Experimental procedures

2.1.1. Co-expression of the 15 GS EmrAB fusion-TolC system

The co-expression of the 15GS EmrAB fusion-TolC system was performed in a similar manner to the previous description for the mRFP1/sfGFP labeled EmrAB-TolC system. Briefly, 1 L of 2xYT medium was inoculated with a pre-culture at a final OD_{600} of 0.05. The growth was continued at 37°C under vigorous aeration to reach an OD_{600} of 0.6. The culture was induced with 0.25 mM IPTG and further incubated overnight at 20°C.

2.1.2. Co-purification of the 15 GS EmrAB fusion-TolC system

For the co-purification, the protocol used was adapted from the previously published methodology used for the MacAB-TolC system from *E. coli* [241]. Briefly, the cell pellet of a 1 L culture was resuspended in 20 mM Tris/HCl pH 8, 400 mM NaCl, 0.2 mM DFP, 5 mg/mL lysozyme and 5 U/mL DNaseI. After an incubation step of 1 h at 4°C, cells were broken using a Constant systems OS cell disruptor (Constant systems Ltd., UK) (1 passage at 2,000 bar and at room temperature). After a pre-centrifugation step at 9,000 *g* for 30 min at 4°C, the supernatant was subjected to an ultracentrifugation step at 108,800 *g* for 4 h at 4°C to pellet the cellular membrane. Subsequently, the cellular membrane was resuspended in 20 mM Tris/HCl pH 8, 400 mM NaCl, 0.2 mM DFP, and 1.5 % (wt/vol) DDM and mixed gently at 4°C for 3 h for complex solubilization. The insoluble fraction was removed by centrifugation at 108,800 g for 30 min at 4°C. Subsequently, 10 mM imidazole pH 7 were added to the supernatant. A 1 mL Ni^{2+}-NTA column (GE Healthcare, USA) was equilibrated with buffer I (20 mM Tris/HCl pH 8, 400 mM NaCl, 20 mM imidazole and 0.05 % (wt/vol) DDM). The supernatant was passed through the column one time. Subsequently, the column was washed with 25 mL of buffer II (20 mM Tris/HCl pH 8, 400 mM NaCl, 50 mM imidazole, and 0.03 % (wt/vol) DMNG). For the elution, 10 mL of buffer III (20 mM Tris/HCl pH 8, 400 mM NaCl, 500 mM imidazole and 0.03 % (wt/vol) DMNG) was used. Different samples taken at each step were analyses via Western blotting in an identical manner to the previous methodologies.

2.2. Results

2.2.1. The EmrAB-fusion and TolC protein partners did not seem to form stable complexes

As a prior verification step, anti His-tag and anti Strep-tagII Western blots were performed (Figure 78). As shown by the figure, the His-tagged 15 GS EmrAB fusion partner was detected within the membrane and elution fractions. On the contrary the TolC protein partner having a Strep-tagII was detected within the membrane and flow through fraction of the Ni^{2+}-NTA purification. Thus, no stable complexes were present for the 15 GS EmrAB fusion-TolC construction. The covalent linker between EmrA and EmrB was not sufficient to maintain all the protein partners together.

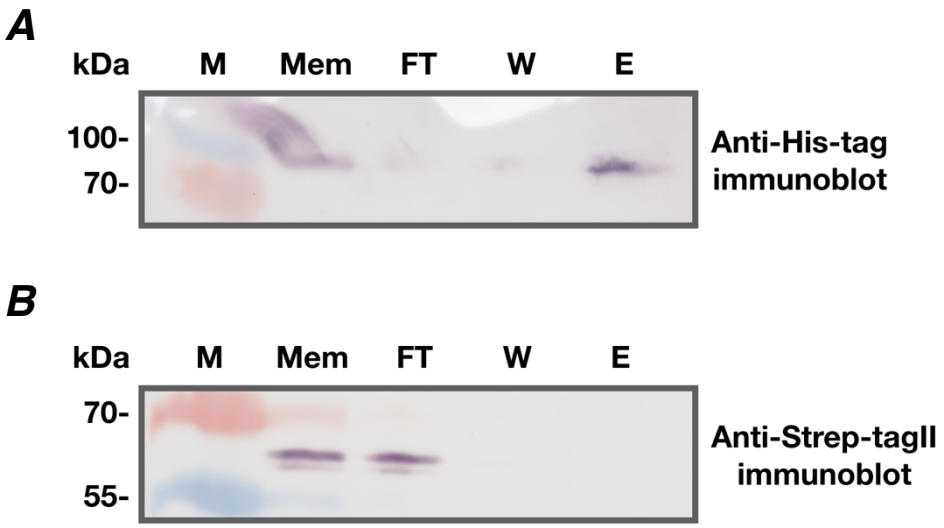

Figure 78. Western blot analyses of the Ni^{2+}-NTA purification of the 15 GS EmrAB fusion-TolC system.

(A) Signals corresponding to the EmrAB fusion chimera within the different samples taken at various stages during the purification. (B) Signals corresponding to TolC within the different samples taken at various stages during the purification. Mem, membrane (10 μg). FT, flow-through (10 μg). W, wash (10 μg). E, elution (2.1 μg).

Chapter V: Large-scale expression, purification and EM analysis of EmrB from E. coli

Within the present chapter, different experimental procedures employed and the corresponding results obtained for the EmrB construct having the mRFP1 label will be mentioned.

1. Experimental procedures

Here, the different expression, purification, proteomic analysis and EM study methodologies employed for EmrB will be described.

1.1. Expression of EmrB

E. coli strain C41(DE3)*ΔacrAB* harboring the p7XC3RH_*emrB* construct was cultivated as described for EmrAB-TolC.

1.2. Purification of EmrB

Cells harvested by centrifugation were resuspended in Tris buffer (50 mM Tris/HCl pH 7.5, 400 mM NaCl, 2 mM EDTA, 10 mM $MgCl_2$, 2 mM $CaCl_2$, 0.2 mM DFP, and 952 U/mL DNase I) and further incubated at 4°C for 30 min under constant mixing. Cells were disrupted by three passages through a French pressure cell (1,379 bar). After a 10,000 *g* centrifugation step to eliminate cell debris and unbroken cells (4°C, 30 min), the supernatant was ultracentrifuged at 100,000 *g* for 2 h at 4°C. The resulting pellet containing the *E. coli* membranes was resuspended in Tris buffer (50 mM Tris pH 8, 200 mM NaCl containing 2 % (wt/vol) DDM and cocktail inhibitor tablets (Roche, Switzerland) for 1 h at 4°C under constant mixing. The membrane lysate was diluted 5 times in Tris buffer without DDM and ultracentrifuged at 100,000 *g* for 1 h at 4°C. The resulting supernatant was used to purify EmrB using Ni^{2+}-NTA affinity purification. After a washing step with 80 mM imidazole buffer containing 5 mM ATP, EmrB was eluted from the Ni^{2+}-NTA column with 400 mM imidazole. The eluate was subsequently dialyzed overnight in Tris buffer containing 0.025 % (wt/vol) DDM and concentrated to 1 mg/mL using a 10 kDa cutoff concentrator (Merck KGaA, Germany).

To remove mRFP1, 1.5 mg of EmrB-mRFP1 was incubated overnight at 4°C with 15 μg of HIS10-HRV 3C (rhinovirus 3C) protease (Pierce, USA). The mixture was subsequently loaded on a Ni^{2+}-NTA column. The cleaved protein was found in the resulting flow-through, while the HRV 3C protease and mRFP1 were retained on the column. The concentrated EmrB (0.5 mg/mL) in 0.025% (wt/vol) DDM was reconstituted in Amphipol A8-35 (Anatrace, USA) as previously described. Fractions containing EmrB were further analyzed by SDS-PAGE and Coomassie staining.

1.3. Proteomic analysis

Protein digestion by chymotrypsin was performed as previously described[258]. NanoLC-MS/MS analysis were performed using an Ultimate 3000 RSLC Nano-UPHLC system (Thermo Scientific, USA) coupled to a nanospray Orbitrap Fusion Lumos Tribrid Mass Spectrometer (ThermoFisher Scientific, USA). Mascot, Sequest and Amanda algorithms through Proteome Discoverer 2.3 Software (ThermoFisher Scientific, USA) were used for protein identification in batch mode by searching against a merge of protein databases: Uniprot *Escherichia coli BL21_DE3* database (UP000002032, release 18/11/09, 31587 entries) and the sequence of the recombinant protein.

1.4. Negative staining EM analysis of EmrB

For the analysis of EmrB the data acquisition and analyses were performed as previously described for EmrAB-TolC. Briefly, the EmrB sample (fractions B3 and B4) was diluted four times in 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.01 % (wt/vol) NaN_3 and applied to a glow-discharged grid. The staining reaction was performed with uranyl fromate pH 7. Automatic acquisition was performed using serialEM software with a pixel size of 2.161 Å (box size of 165x165 pixels). In total, a set of 9732 particles was picked and submitted to 2D reference-free alignment and classification. As a comparison, data were also acquired and analyzed for a sample containing Amphipol A8-35 only (at 100 µg/mL). In this case, a set of 4822 particles was automatically picked with a box size of 118x118 pixels.

2. Results

Here the different results obtained with the use of the previous methodologies will be described.

2.1. Biochemical characterizations of EmrB

The SDS-PAGE analyses (Figure 79A and 79B) indicated that EmrB-mRFP1 had an apparent molecular weight of 60 kDa. After digestion by 3C protease, EmrB was found at about 40 kDa which is inferior to its molecular weight (56 kDa), but in line with the aberrant behaviour of membrane proteins subjected to SDS-PAGE analysis. This finding is in agreement with a previously published work[214].

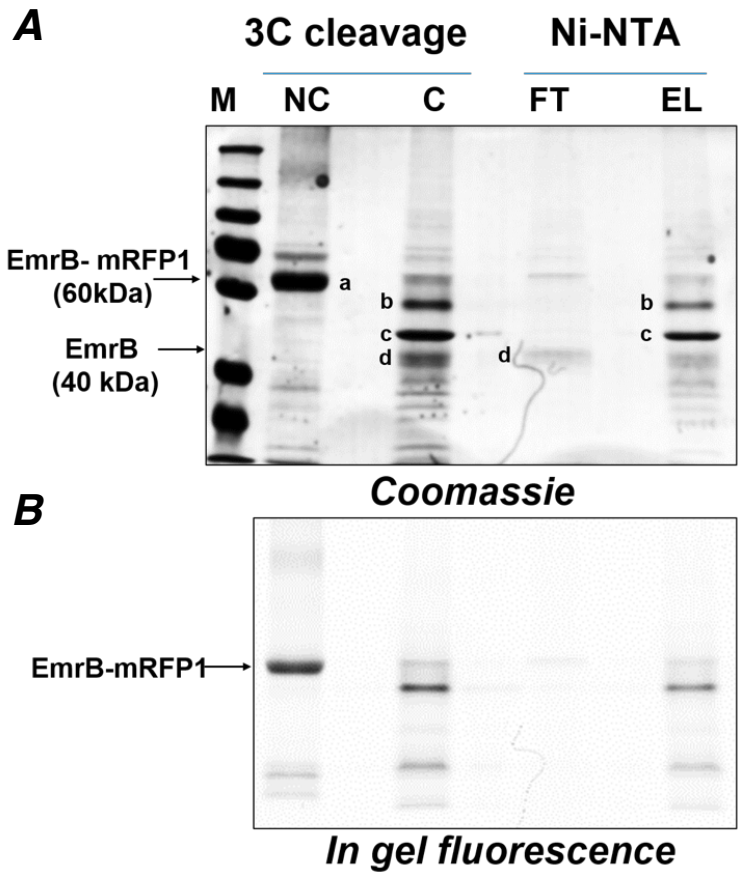

Figure 79. SDS-PAGE analysis of EmrB after mRFP1 cleavage.

(A) Coomassie blue-stained SDS-PAGE analysis of affinity purified EmrB-mRFP1 before 'NC' (40 μg) and after protease 3C cleavage 'C' (40 μg) followed by an additional reverse Ni^{2+}-NTA affinity purification. 'FT', flow-through (9 μg). 'EL', elution (30 μL from a 3 mL fraction were loaded on the gel). Bands are annotated as follows: a. EmrB-mRFP1-His fusion protein, b. mRFP1-His, c. 3C protease, d. EmrB. (B) In gel mRFP1 fluorescence analysis of the same SDS-PAGE gel shown in (A). The migration profile of EmrB-mRFP1-His corresponds to 60 kDa and that of EmrB to 40 kDa before and after 3C protease cleavage respectively. After cleavage, no fluorescence signal is recovered in FT.

After Amphipol A8-35 reconstitution and size exclusion chromatography, EmrB was recovered in fractions B3 and B4 as shown by SDS-PAGE (Figure 80A and 80B). The identity of EmrB was confirmed by mass spectrometry (Figure 81). The elution profile of EmrB suggested that it would be present as a monomer.

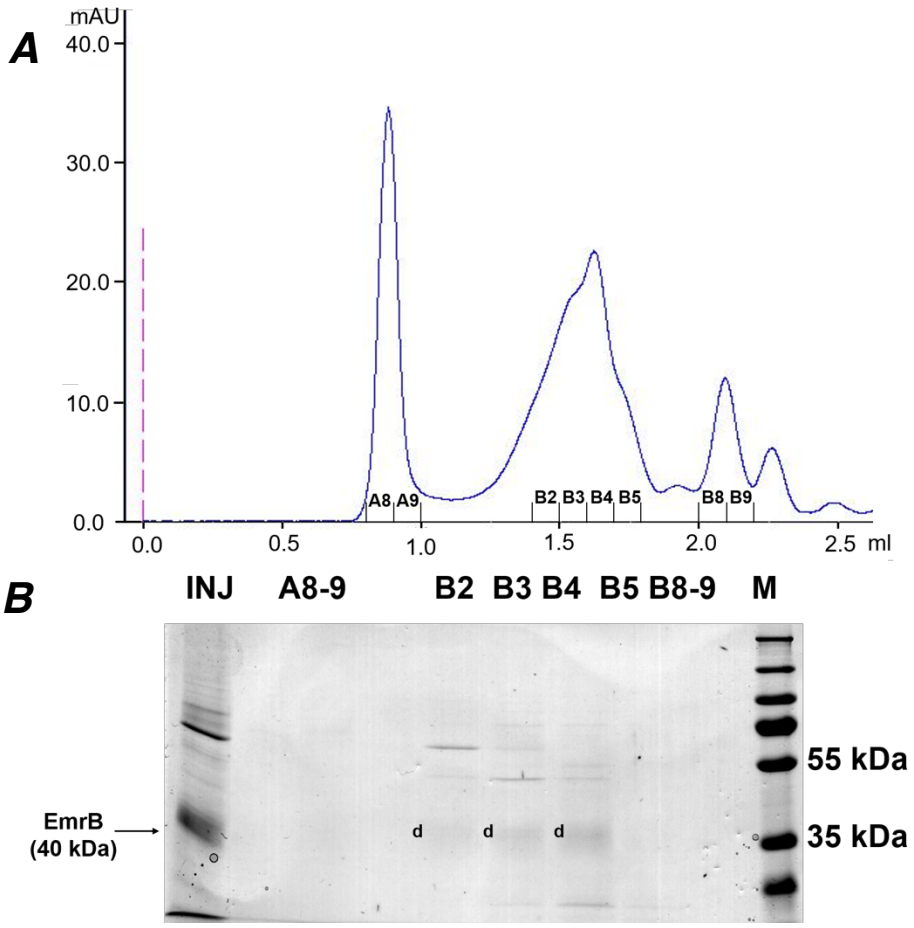

Figure 80. Reconstitution of EmrB in Amphipol A8-35.

(A) Size exclusion chromatography profile of EmrB after mRFP1 cleavage using a Superose 6 column. 35 μg of EmrB was injected (dashed line). (B) Coomassie blue stained gel after SDS-PAGE analysis of the indicated SEC fractions. Fractions B2 to B4 contained EmrB (arrow and annotated as d). INJ corresponds to cleaved EmrB sample stabilized with amphipol A8-35. 20 μL from the 100 μL fractions were loaded in each case.

```
MSQQQKPLEG AQLVIMTIAL SLATFMQVLD STIANVAIPT IAGNLGSSLS QGTWVITSFG    60
VANAISIPLT GWLAKRVGEV KLFLWSTIAF AIASWACGVS SSLNMLIFFR VIQGIVAGPL   120
IPLSQSLLLN NYPPAKRSIA LALWSMTVIV APICGPILGG YISDNYHWGW IFFINVPIGV   180
AVVLMTLQTL RGRETRTERR RIDAVGLALL VIGIGSLQIM LDRGKELDWF SSQEIIILTV   240
VAVVAICFLI VWELTDDNPI VDLSLFKSRN FTIGCLCISL AYMLYFGAIV LLPQLLQEVY   300
GYTATWAGLA SAPVGIIPVI LSPIIGRFAH KLDMRRLVTF SFIMYAVCFY WRAYTFEPGM   360
DFGASAWPQF IQGFAVACFF MPLTTITLSG LPPERLAAAS SLSNFTRTLA GSIGTSITTT   420
MWTNRESMHH AQLTESVNPF NPNAQAMYSQ LEGLGMTQQQ ASGWIAQQIT NQGLIISANE   480
IFWMSAGIFL VLLGLVWFAK PPFGAGGGGG GAHALEVLFQ                         520
```

Figure 81. Mass spectrometry sequence coverage for EmrB.

Peptides identified are highlighted in green. EmrB was identified with 30 specific and unique peptides with a coverage rate of 47.69%.

2.2. Structural characterization of EmrB

Figure 82 compares the negative stain EM analyses of EmrB to a sample containing amphipol A8-35 only. The comparison reveals some structural differences with objects having globally more homogenous morphologies in the case of EmrB. Class averages from single-particle image analysis of the EmrB sample revealed two basic forms: a 12 x 12 nm square-shaped structure with densities delineating an apparent central hole and a 12 x 17 nm rectangular-shaped structure with densities delimiting an elongated hole (Figure 82A). The densities observed correspond most likely to an amphipol belt in each case surrounding EmrB as EmrB only presents a small hydrophilic loop protruding from the lipid membrane. The size of these amphipol belts surrounding EmrB suggests that the square shaped structure likely contained one EmrB molecule while the rectangular structure may comprise two EmrB molecules.

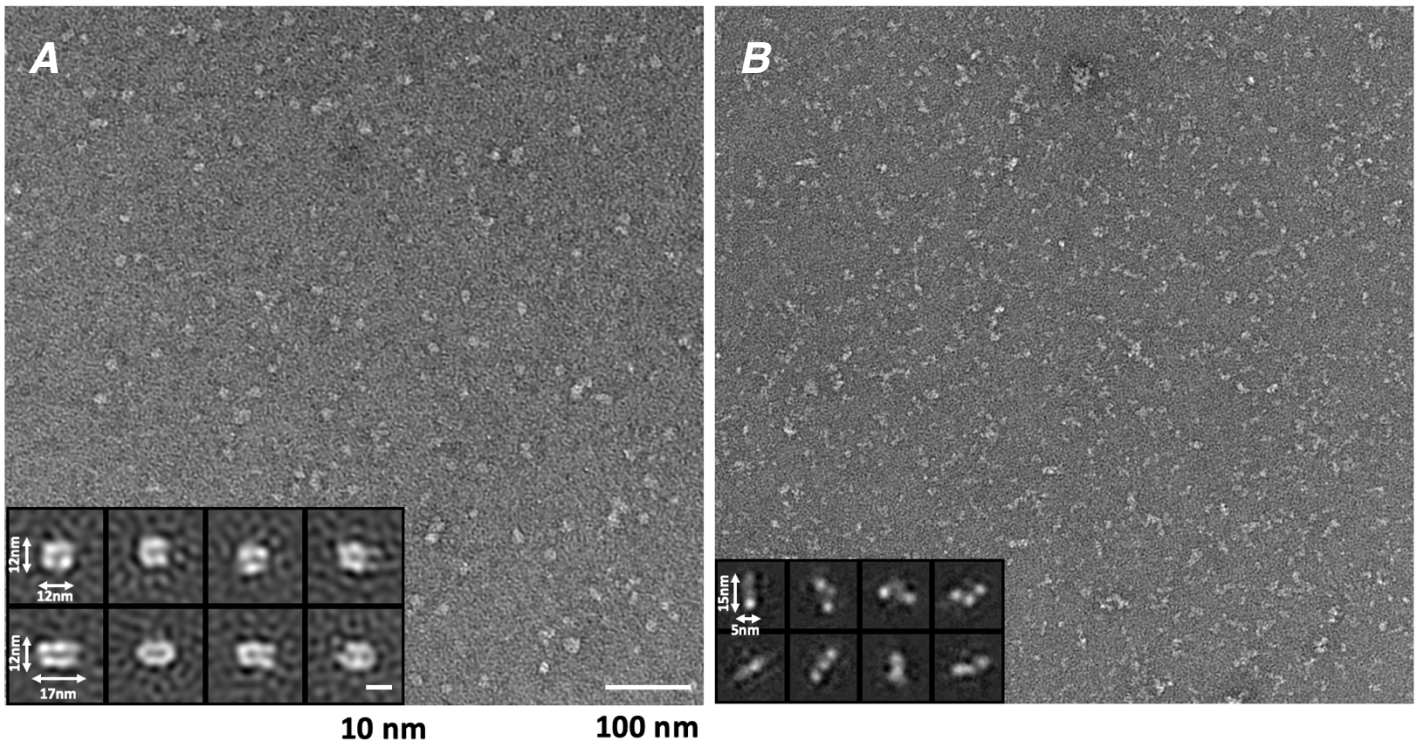

Figure 82. Negative stain EM analysis of EmrB and Amphipol A8-35.

(A) Representative negative stain image and corresponding 2D class averages of purified EmrB reconstituted in Amphipol A8-35. Two different particle sizes are visible indicated by the distance measurements with the white arrows (B) Representative negative stain image and corresponding 2D class averages of Amphipol A8-35.

DISCUSSION AND CONCLUSION

Chapter VI: Global discussion and conclusion

Whithin the present chapter, the different results concerning the preparation of various DNA constructs, the use of a high-throughput screening pipeline, the co-expression/co-purification/EM analysis of EmrAB-TolC and the expression/purification/EM analysis of EmrB will be discussed together with their respective perspectives.

1. Global discussion and conclusion

Membrane proteins are of great importance because of their involvement in the transport of molecules, energy and information across the membrane barrier. Thus, they are the subjects of many biophysical and biochemical studies.

Even if numerous research works concerning integral membrane protein overexpression optimization are present in the literature[259–264], the production of sufficient amounts of integral membrane proteins still represents a hurdle as no general guidelines are available thus far. In *E. coli* the difficulties may be linked to the correct targeting towards the inner membrane, membrane integration and folding. Thus, the study of each of these steps represents an important goal in order to improve our knowledge about membrane protein production.

Whithin the present Ph.D. project, the overall aim was to use a high-throughput screening pipeline in order to identify suitable targets which are highly expressed and stable during purification for further structural characterizations. Furthermore the goal was to establish a suitable methodology for the direct purification of the EmrAB-TolC complex from bacterial cells.

Thus within the present section, all the different steps leading to the structural characterizations of EmrAB-TolC as well as EmrB will be discussed in detail.

1.1. Generation of different DNA constructs

Overall, the identification of different operons within the genomes of various gram negative bacteria was a challenging task as *emrB* genes are usually present in multiple copies. Furthermore, most of the *emrB* genes do not form an operon with a *emrA* and a *tolC* gene. Thus, each of the targets found within the database were verified one by one in order to identify *emrAB* and *emrAB-tolC* operons.

For the cloning, a high-throughput approach had to be adopted because of the important number of genes selected and because of the low amount of expression and purification data available. Thus, the FX-cloning method seemed to be best suited for the overall project. However, important limitations were encountered at the level of the PCR amplifications as well as sequencing results. Thus, other targets were chosen to successfully amplify the ORFs of 15 homologous EmrAB-TolC systems. Because of the first difficulties encountered at the level of the PCR amplifications only the *emrAB* and *tolC* ORFs from *E. coli* and *V. cholerae* were further inserted in the corresponding expression vectors

p7XC3RH, pRSFDMG and pRSFDM_G. In parallel, for the ORFs from *E. coli* other cloning strategies were also employed: (i) generation of constructs for the expression of EmrAB fusion chimeras, and (ii) generation of single expression constructs for the separate expression of all three genes. Overall these two additional preparations were straightforward thanks to the FX-compatible expression vectors.

As a future perspective, all the other ORFs from the homologous EmrAB-TolC systems could be inserted into pINIT_cat and subsequently subcloned and traditionnaly cloned into the expression vectors.

1.2. Utilization of a high-throughput screening pipeline

After the subcloning of the *E. coli* and *V. cholerae* ORFs in the expression vectors p7XC3RH, pRSFDMG and pRSFDM_G the different proteins were expressed as fusion proteins with C-terminal fluorescent labels (mRFP1 and sfGFP) and affinity tags (His-tag, Myc-tag, and Strep-tagII). Because of the low amount of expression and purification data available, the high-throughput screening pipeline represented an elegant methodology in order to test the expression and purification behaviours of both complexes. Overall, the testing of the expression behaviour of both complexes was straightforward with the measurement of the red and green fluorescent signals linked to the expression level of each protein partner. However, the testing of the complex formation behaviour using the FSEC[250–253] methodology seemed to be challenging. Indeed, even if in one case a signal shift was visible for the red fluorescence signal of the EmrAB-TolC complex from *E. coli*, no signal shift was visible for the corresponding green fluorescence signal. Thus even if the methodology was straightforward, it did seem to present some limitations and could only give some indications about the possible complex formation behaviour of each complex. Nevertheless it remains unclear why the observation of a clear signal shift between a ~ 700 kDa complex and a 76 kDa EmrA monomer could not be seen. Indeed a ten fold size difference should be sufficient to see a clear signal shift. Such results were readily obtained for smaller complexes in the literature[250,251,265,253]. Nevertheless, the FSEC method enabled the identification of an effect linked to the preparation of *E. coli* membranes which could possibly be deleterious for the complex stability during purification. As a future perspective, it would be interesting to test the complex formation behaviour using multiple buffers, and detergents. Furthermore the effect of the formation of styrene maleic acid copolymer lipid particles (SMALPs)[266] for the extraction and purification of the complexes could also be

tested. This would indeed yield important information about the correct purification methodology to be employed.

1.3. Co-expression, co-purification and EM analysis of EmrAB-TolC

Overall, the first strategy employed with the fluorescent labeling of EmrB and EmrA with mRFP1 and sfGFP respectively led to the correct isolation of the entire complex for further sturctural characterization. The presence of the fluorescent labels and the affinity tags (His-tag, Myc-tag and Strep-tagII) enabled the specific detection of the different protein partners at any step during co-purification. For the co-purification the best suited procedure consisted of a mild chemical lysis of cells with Lysozyme followed by membrane protein solubilization with DDM and a Ni^{2+}-NTA affinity co-purification. As a second co-purification step a SEC step was also included. Finally, purified complexes were stabilized using Amphipol A8-35[267,268] for structural analysis. Despite numerous trials an orthogonal co-purification methodology making use of the His and Strep-tagII could not be established even if such a methodology was successfully employed by Parcej and co-workers[251]. This observation could possibly be linked to the low amount of TolC present within the sample after the solubilization or the first Ni^{2+}-NTA co-purification step. The EM average image of the entire efflux system showed an elongated structure of about 33 nm, which is similar to the length of AcrAB-TolC analyzed in similar conditions[242] and by cryoEM[269]. In addition, this result was also similar to the length of MacAB-TolC complexes[241]. The MFS transporter EmrB has 14 predicted TMSs embedded into the membrane. It only presents a predicted loop containing 53 amino acid residues located between TMS 13 and TMS 14 towards the periplasm, much smaller compared to the periplasmic loops present in the RND transporters (approx. 600 amino acid residues per protomer, *e.g.* trimeric AcrB[270], MexB[271]) or dimeric ABC transporter MacB (approx. 220 amino acid residues per protomer)[241] (Figure 83)[272].

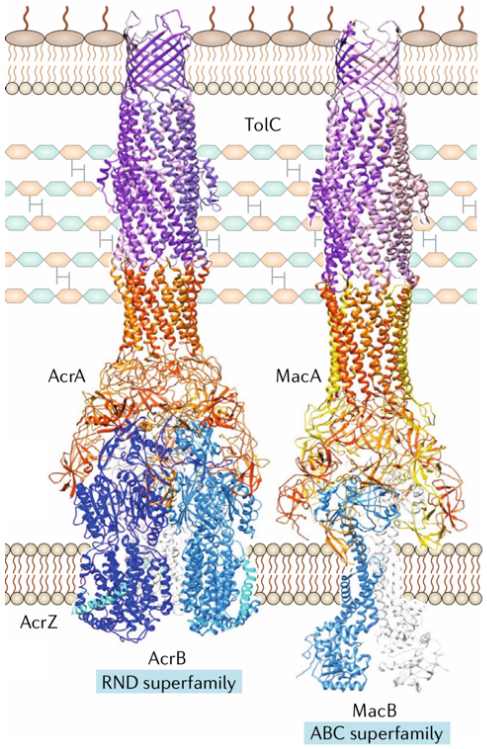

Figure 83. Structures of RND and ABC superfamily tripartite efflux systems.

AcrAB-TolC (EMD-8640); MacA-TolC (EMD-3652); MacB (EMD-3653). Adapted from reference 272.

Therefore, the densities observed between the periplasmic end of TolC and the second amphipol belt (corresponding to the inner membrane) can be assigned to EmrA alone (Figure 75B).

The length of the periplasmic region of EmrA (about 15 nm) is shorter than the one described for the EmrA homolog from *Aquifex aeolicus* (aaEmrA, 18.5 nm)[215]. Sequence alignement and structure modeling (using the I-Tasser server[273–275]) predicted that EmrA from *E. coli* (ecEmrA) presents an α-helical coiled-coil domain (120 amino acid residues) shorter than that of aaEmrA (165 amino acid residues) (Figure 49 and 84A)[215].

The structures of two EmrAB-TolC models prepared using the I-Tasser[273–275] and Phyre2[276] servers are shown in Figure 84B. Given the length of the ecEmrA molecules, the interaction with the periplasmic α-helical barrel of TolC could possibly correspond to a 'tip-to-tip' interaction as observed for other tripartite efflux systems[269,241,243] (Figure 84B). To further evaluate the residue pairs from EmrA and TolC involved in the interaction interface, a sequence covariation analysis was performed using the online tool GREMLIN[277]. A set of four covarying and interacting residue pairs were identified (Table 17).

Table 17. Top 4 covarying residue pairs of EmrA and TolC predicted by GREMLIN.

The GREMLIN scores are expressed as probabilities that the predicted residue pairs are covarying and interacting. As a high-confidence prediction the threshold value of 0.70 was used. The default E-value of 1e-20 and the default HHblits method were used for the analysis.

EmrA residue	TolC residue	GREMLIN score (rank)
A152	A382	0.99 (1)
A152	Q164	0.98 (2)
P148	V168	0.93 (3)
G156	V171	0.79 (4)

As a control the same analysis was also performed for EmrB and TolC which do not directly interact with each other. Thus all the residue pairs identified had GREMLIN scores of 0. As an example the top four predictions are shown in Table 18.

Table 18. Predictions of covarying residue pairs between EmrB and TolC by GREMLIN.

The GREMLIN scores are given as probabilities that predicted residues are covarying and interacting with a threshold value of 0.70 for high confidence predictions. All the GREMLIN scores for the different residue pairs of EmrB and TolC corresponded to 0. As an example four of the residue pairs are mentioned. The default E-value of 1e-20 and the default HHblits method were used for the analysis.

EmrB residue	TolC residue	GREMLIN score
I91	A132	0.0
L432	A32	0.0
G486	Q261	0.0
M483	Q27	0.0

The stoichiometry of the EmrAB complex was previously analysed corresponding to a so called 'dimer-of-dimers' in a physiological state[214]. However, it is unclear how such organization could interact with a trimeric TolC. The side view of the EmrAB-TolC complex did not resemble the 'dimer-of-dimers'

of the EmrAB complex but is in favour of a hexameric arrangement of EmrA similar to other types of adaptor proteins (*e.g.* AcrA, MacA) in a 'tip-to-tip' contact with TolC (Figure 84).

As EmrB was completely embedded in the inner amphipol belt, no direct interaction between EmrB and TolC is expected. The oligomerization state of EmrB remains speculative, however, from the dimensions of the β-barrel domain of EmrA at the inner membrane proximal side, one EmrB protomer was tentatively modelled (in the first model) similar to a previously proposed model[215]. A second model including a dimeric EmrB is also represented in Figure 84B. Even if tripartite systems thus far include multimeric inner membrane transporters, various single component MFS members (*e.g.* MdfA and EmrD) were shown to be monomeric.

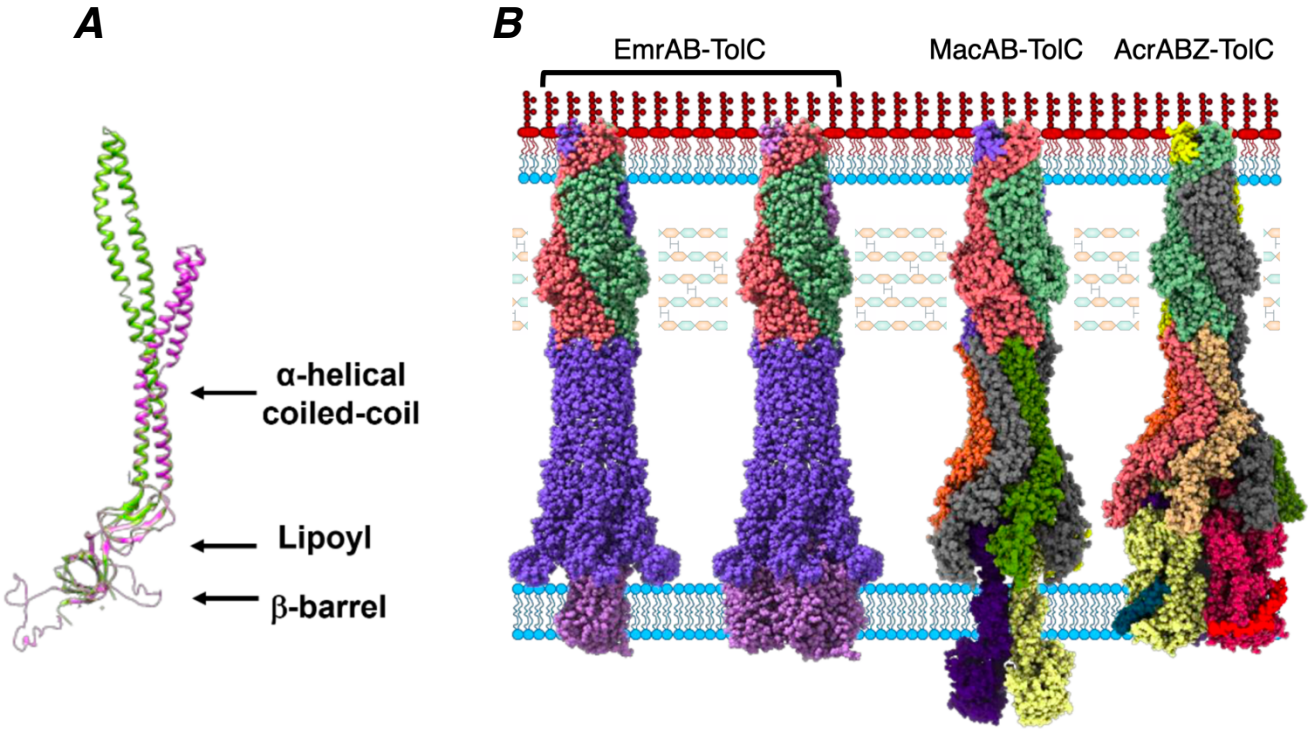

Figure 84. Structural model of ecEmrA and overview of tripartite efflux systems from Gram-negative bacteria.

(A) Predicted structure of ecEmrA (purple) using I-Tasser server (references 273-275) and the structure of *Aquifex aeolicus* (aaEmrA) (green) (pdb: 4TKO) superimposed using Chimera. The α-helical coiled-coil of ecEmrA is shorter by 45 amino acids. The β-barrel, lipoyl and α-helical coiled-coil domains are indicated. (B) Putative assembly of EmrAB-TolC (left). EmrA is shown as a hexameric ring structure (blue) forming a channel through the periplasm. The six α hairpins of EmrA were modelled in a 'tip to tip' interaction with TolC. The β-barrel and lipoyl domains are close to the inner membrane component EmrB (purple), which was modelled from its primary sequence using the Phyre2 server (reference 276). Since the oligomeric state of EmrB is unknown, the views of the two structures (from the left) represent monomeric or dimeric EmrB. *E. coli* MacAB-TolC (pdb: 5NIK) and AcrABZ-TolC (pdb: 5NG5) complexes are shown for comparison.

Even if complex formation of the 15 GS EmrAB fusion-TolC system was possibly observed with the small scale pull-down analysis, no complexes could be isolated at large scale. This result indicates that contrary to the MacAB-TolC[241] system from *E. coli* such chimera does not stabilize the entire complex in the case of EmrAB-TolC.

As a future perspective, the co-purification protocol of the fluorescently labeled complex could be further optimized with additional purification steps (*e.g.* ion exchange chromatography). Furthermore, a cleavage of the fluorescent labeles and an inverse Ni^{2+}-NTA step could also improve the purity of the sample. It would also be interesting to test the formation of SMALPs[266] for the extraction and isolation of the entire complex. Finally, concerning the fusion stabilized chimera, other linker sizes, affinity tags as well as the arabinose promoter system could also be tested.

1.4. Expression, purification and EM analysis of EmrB

For the purification of EmrB only, *E. coli* membranes could be prepared as the goal was not to isolate any complexes. Once the mRFP1 label was cleaved a signal was visible at about 40 kDa which is in consistency with a previously published work[214].

The EM analysis of EmrB did not yield clear results possibly because of the relatively small size of EmrB. Indeed the dye could not be deposited correctly around the different objects and therefore the signal obtained was not sufficient for the structural determination. Nevertheless, comparing the class averages of EmrB and Amphipol A8-35 more homogenous objects were visible in the case of EmrB.

As a future perspective, it would be interesting to study the structure of EmrB by X-ray crystallography because of its small size.

1.5. Conclusion

Overall, the co-purification strategy employed provides new insights on the isolation and structure of EmrAB-TolC system without the need of artificial linking between the subunit components of the pump as is the case for the recent AcrAB-TolC and MacAB-TolC structures[278,241]. The overall length of EmrAB-TolC complex is similar to that of AcrAB-TolC with a likely 'tip-to-tip' interaction between EmrA and TolC forming an extended periplasmic canal with at least similar length as the periplasmic tunnels shown on other known tripartite complexes. Future research will be to determine the interaction

sites of EmrA with EmrB. Since EmrA contains a monotopic helical anchor in the inner membrane, possible interaction between EmrA and EmrB might be between the transmembrane regions. In addition, a small periplasmic loop of EmrB predicted between TMS 13 and TMS 14 might be another contact region between the components of this tripartite setup. To analyse these interactions, higher resolution Cryo-EM structures are needed.

REFERENCES

References

1. No time to wait: securing the future from drug-resistant infections. *World Health Organization* https://www.who.int/docs/default-source/documents/no-time-to-wait-securing-the-future-from-drug-resistant-infections-en.pdf?sfvrsn=5b424d7_6 (2019).
2. Tackling drug-resistant infections globally: final report and recommendations. *AMR Review* https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf (2016).
3. The evolving threat of antimicrobial resistance. Options for action. *World Health Organization* https://apps.who.int/iris/bitstream/handle/10665/44812/9789241503181_eng.pdf?sequence=1 (2012).
4. The bacterial challenge: time to react. *European Centre for Disease Prevention and Control* https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/0909_TER_The_Bacterial_Challenge_Time_to_React.pdf (2009).
5. Li, B. & Webster, T. J. Bacteria antibiotic resistance: New challenges and opportunities for implant-associated orthopedic infections. *J. Orthop. Res. Off. Publ. Orthop. Res. Soc.* **36**, 22–32 (2018).
6. Fleming, A. Penicillin. https://www.nobelprize.org/uploads/2018/06/fleming-lecture.pdf (1945).
7. Ventola, C. L. The Antibiotic Resistance Crisis. *Pharm. Ther.* **40**, 277–283 (2015).
8. Hede, K. Antibiotic resistance: An infectious arms race. *Nature* **509**, S2–S3 (2014).
9. Antibiotic resistance threats in the United States. *Centres for Disease Control and Prevention.* https://www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf (2013).
10. Dantas, G. & Sommer, M. O. A. How to fight back against antibiotic resistance. *Am. Sci.* **102**, 42–51 (2014).
11. Sales of veterinary antimicrobial agents in 25 EU/EEA countries in 2011. *European Medicines Agency* https://www.ema.europa.eu/en/documents/report/sales-veterinary-antimicrobial-agents-25-european-union/european-economic-area-countries-2011-third-european-surveillance-veterinary-antimicrobial_en.pdf (2013).
12. Sales of veterinary antimicrobial agents in 30 European countries in 2016. *European Medicines Agency* https://www.ema.europa.eu/en/documents/report/sales-veterinary-antimicrobial-agents-30-european-countries-2016-trends-2010-2016-eighth-esvac_en.pdf (2018).

13. Marshall, B. M. & Levy, S. B. Food Animals and Antimicrobials: Impacts on Human Health. *Clin. Microbiol. Rev.* **24**, 718–733 (2011).
14. Martin, M. J., Thottathil, S. E. & Newman, T. B. Antibiotics Overuse in Animal Agriculture: A Call to Action for Health Care Providers. *Am. J. Public Health* **105**, 2409–2410 (2015).
15. Cogliani, C., Goossens, H. & Greko, C. Restricting Antimicrobial Use in Food Animals: Lessons from Europe: Banning nonessential antibiotic uses in food animals is intended to reduce pools of resistance genes. *Microbe Mag.* **6**, 274–279 (2011).
16. Holmes, A. H. *et al.* Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet Lond. Engl.* **387**, 176–187 (2016).
17. Fact Sheet: Veterinary Feed Directive Final Rule and Next Steps. *US Food and Drug Administration* https://www.fda.gov/animal-veterinary/development-approval-process/fact-sheet-veterinary-feed-directive-final-rule-and-next-steps (2019).
18. Coyne, L. *et al.* Characterizing Antimicrobial Use in the Livestock Sector in Three South East Asian Countries (Indonesia, Thailand, and Vietnam). *Antibiot. Basel Switz.* **8**, (2019).
19. May, M. Drug development: Time for teamwork. *Nature* **509**, S4-5 (2014).
20. Santos, L. & Ramos, F. Antimicrobial resistance in aquaculture: Current knowledge and alternatives to tackle the problem. *Int. J. Antimicrob. Agents* **52**, 135–143 (2018).
21. Baquero, F., Martínez, J.-L. & Cantón, R. Antibiotics and antibiotic resistance in water environments. *Curr. Opin. Biotechnol.* **19**, 260–265 (2008).
22. Perry, J., Waglechner, N. & Wright, G. The Prehistory of Antibiotic Resistance. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
23. Benveniste, R. & Davies, J. Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 2276–2280 (1973).
24. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
25. Santiago-Rodriguez, T. M. *et al.* Gut Microbiome of an 11th Century A.D. Pre-Columbian Andean Mummy. *PloS One* **10**, e0138135 (2015).
26. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344 (2014).

27. Bartoloni, A. *et al.* Antibiotic resistance in a very remote Amazonas community. *Int. J. Antimicrob. Agents* **33**, 125–129 (2009).

28. Boucher, H. W. *et al.* Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **48**, 1–12 (2009).

29. Rice, L. B. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. *J. Infect. Dis.* **197**, 1079–1081 (2008).

30. Santajit, S. & Indrawattana, N. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *BioMed Res. Int.* **2016**, 2475067 (2016).

31. Esposito, S. & De Simone, G. Update on the main MDR pathogens: prevalence and treatment options. *Infez. Med. Riv. Period. Eziologia Epidemiol. Diagn. Clin. E Ter. Delle Patol. Infett.* **25**, 301–310 (2017).

32. El-Mahallawy, H. A., Hassan, S. S., El-Wakil, M. & Moneer, M. M. Bacteremia due to ESKAPE pathogens: An emerging problem in cancer patients. *J. Egypt. Natl. Cancer Inst.* **28**, 157–162 (2016).

33. Arias, C. A. & Murray, B. E. The rise of the Enterococcus: beyond vancomycin resistance. *Nat. Rev. Microbiol.* **10**, 266–278 (2012).

34. Ranotkar, S. *et al.* Vancomycin-resistant enterococci: Troublemaker of the 21st century. *J. Glob. Antimicrob. Resist.* **2**, 205–212 (2014).

35. Galloway-Peña, J. R., Nallapareddy, S. R., Arias, C. A., Eliopoulos, G. M. & Murray, B. E. Analysis of Clonality and Antibiotic Resistance among Early Clinical Isolates of Enterococcus faecium in the United States. *J. Infect. Dis.* **200**, 1566–1573 (2009).

36. Vigani, A. G., Oliveira, A. M. de, Bratfich, O. J., Stucchi, R. S. B. & Moretti, M. L. Clinical, epidemiological, and microbiological characteristics of bacteremia caused by high-level gentamicin-resistant Enterococcus faecalis. *Braz. J. Med. Biol. Res. Rev. Bras. Pesqui. Medicas E Biol.* **41**, 890–895 (2008).

37. Haaber, J., Penadés, J. R. & Ingmer, H. Transfer of Antibiotic Resistance in Staphylococcus aureus. *Trends Microbiol.* **25**, 893–905 (2017).

38. Lakhundi, S. & Zhang, K. Methicillin-Resistant Staphylococcus aureus: Molecular Characterization, Evolution, and Epidemiology. *Clin. Microbiol. Rev.* **31**, (2018).

39. Chandrakanth, R. K., Raju, S. & Patil, S. A. Aminoglycoside-resistance mechanisms in multidrug-resistant Staphylococcus aureus clinical isolates. *Curr. Microbiol.* **56**, 558–562 (2008).
40. Hyo, Y., Yamada, S., Fukutsuji, K. & Harada, T. Thickening of the cell wall in macrolide-resistant Staphylococcus aureus. *Med. Mol. Morphol.* **46**, 217–224 (2013).
41. Huys, G., D'Haene, K., Van Eldere, J., von Holy, A. & Swings, J. Molecular Diversity and Characterization of Tetracycline-Resistant Staphylococcus aureus Isolates from a Poultry Processing Plant. *Appl. Environ. Microbiol.* **71**, 574–579 (2005).
42. Gu, B., Kelesidis, T., Tsiodras, S., Hindler, J. & Humphries, R. M. The emerging problem of linezolid-resistant Staphylococcus. *J. Antimicrob. Chemother.* **68**, 4–11 (2013).
43. Dortet, L., Anguel, N., Fortineau, N., Richard, C. & Nordmann, P. In vivo acquired daptomycin resistance during treatment of methicillin-resistant Staphylococcus aureus endocarditis. *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.* **17**, e1076-1077 (2013).
44. McGuinness, W. A., Malachowa, N. & DeLeo, F. R. Vancomycin Resistance in Staphylococcus aureus. *Yale J. Biol. Med.* **90**, 269–281 (2017).
45. Chambers, H. F. & Deleo, F. R. Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat. Rev. Microbiol.* **7**, 629–641 (2009).
46. Navon-Venezia, S., Kondratyeva, K. & Carattoli, A. Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiol. Rev.* **41**, 252–275 (2017).
47. Veeraraghavan, B. *et al.* Carbapenem resistant Klebsiella pneumoniae isolated from bloodstream infection: Indian experience. *Pathog. Glob. Health* **111**, 240–246 (2017).
48. Galimand, M., Courvalin, P. & Lambert, T. Plasmid-mediated high-level resistance to aminoglycosides in Enterobacteriaceae due to 16S rRNA methylation. *Antimicrob. Agents Chemother.* **47**, 2565–2571 (2003).
49. Higashino, M. *et al.* Fluoroquinolone resistance in extended-spectrum β-lactamase-producing Klebsiella pneumoniae in a Japanese tertiary hospital: silent shifting to CTX-M-15-producing K. pneumoniae. *J. Med. Microbiol.* **66**, 1476–1482 (2017).
50. Pintado, V. *et al.* Intravenous colistin sulphomethate sodium for therapy of infections due to multidrug-resistant gram-negative bacteria. *J. Infect.* **56**, 185–190 (2008).

51. Fritsche, T. R., Strabala, P. A., Sader, H. S., Dowzicky, M. J. & Jones, R. N. Activity of tigecycline tested against a global collection of Enterobacteriaceae, including tetracycline-resistant isolates. *Diagn. Microbiol. Infect. Dis.* **52**, 209–213 (2005).

52. Rojas, L. J. *et al.* Colistin Resistance in Carbapenem-Resistant Klebsiella pneumoniae: Laboratory Detection and Impact on Mortality. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **64**, 711–718 (2017).

53. Hladicz, A., Kittinger, C. & Zarfel, G. Tigecycline Resistant Klebsiella pneumoniae Isolated from Austrian River Water. *Int. J. Environ. Res. Public. Health* **14**, (2017).

54. Lee, C.-R. *et al.* Biology of Acinetobacter baumannii: Pathogenesis, Antibiotic Resistance Mechanisms, and Prospective Treatment Options. *Front. Cell. Infect. Microbiol.* **7**, 55 (2017).

55. Rezaee, M. A. *et al.* Prevalence of Ambler class A β-lactamases and ampC expression in cephalosporin-resistant isolates of Acinetobacter baumannii. *Diagn. Microbiol. Infect. Dis.* **76**, 330–334 (2013).

56. Evans, B. A., Hamouda, A. & Amyes, S. G. B. The rise of carbapenem-resistant Acinetobacter baumannii. *Curr. Pharm. Des.* **19**, 223–238 (2013).

57. Lari, A. R., Ardebili, A. & Hashemi, A. AdeR-AdeS mutations & overexpression of the AdeABC efflux system in ciprofloxacin-resistant Acinetobacter baumannii clinical isolates. *Indian J. Med. Res.* **147**, 413–421 (2018).

58. Lin, T. *et al.* Identification of aac(2')-I type b aminoglycoside-modifying enzyme genes in resistant Acinetobacter baumannii. *Genet. Mol. Res. GMR* **14**, 1828–1835 (2015).

59. Fang, Y.-Q., Zhan, R.-C., Jia, W., Zhang, B.-Q. & Wang, J.-J. A case report of intraventricular tigecycline therapy for intracranial infection with extremely drug resistant Acinetobacter baumannii. *Medicine (Baltimore)* **96**, e7703 (2017).

60. Dong, X. *et al.* In vitro activities of rifampin, colistin, sulbactam and tigecycline tested alone and in combination against extensively drug-resistant Acinetobacter baumannii. *J. Antibiot. (Tokyo)* **67**, 677–680 (2014).

61. Qureshi, Z. A. *et al.* Colistin-resistant Acinetobacter baumannii: beyond carbapenem resistance. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **60**, 1295–1303 (2015).

62. Trebosc, V. *et al.* A Novel Genome-Editing Platform for Drug-Resistant Acinetobacter baumannii Reveals an AdeR-Unrelated Tigecycline Resistance Mechanism. *Antimicrob. Agents Chemother.* **60**, 7263–7271 (2016).

63. Thapa, B. *et al.* Rifampin resistance in carbapenem-resistant Acinetobacter baumannii in Siriraj Hospital, Thailand. *Nepal Med. Coll. J. NMCJ* **11**, 232–237 (2009).

64. Chatterjee, M. *et al.* Antibiotic resistance in Pseudomonas aeruginosa and alternative therapeutic options. *Int. J. Med. Microbiol. IJMM* **306**, 48–58 (2016).

65. Pan, Y.-P., Xu, Y.-H., Wang, Z.-X., Fang, Y.-P. & Shen, J.-L. Overexpression of MexAB-OprM efflux pump in carbapenem-resistant Pseudomonas aeruginosa. *Arch. Microbiol.* **198**, 565–571 (2016).

66. Barnes, M. D. *et al.* Deciphering the Evolution of Cephalosporin Resistance to Ceftolozane-Tazobactam in Pseudomonas aeruginosa. *mBio* **9**, (2018).

67. Peng, J., Cao, J., Ng, F. M. & Hill, J. Pseudomonas aeruginosa develops Ciprofloxacin resistance from low to high level with distinctive proteome changes. *J. Proteomics* **152**, 75–87 (2017).

68. Francisco, G. R., Nora, S. T. R., Bueno, M. F. C., da Silva Filho, L. V. R. F. & de Oliveira Garcia, D. Identification of aminoglycoside-resistant Pseudomonas aeruginosa producing RmtG 16S rRNA methyltransferase in a cystic fibrosis patient. *Antimicrob. Agents Chemother.* **59**, 2967–2968 (2015).

69. Johansen, H. K., Moskowitz, S. M., Ciofu, O., Pressler, T. & Høiby, N. Spread of colistin resistant non-mucoid Pseudomonas aeruginosa among chronically infected Danish cystic fibrosis patients. *J. Cyst. Fibros. Off. J. Eur. Cyst. Fibros. Soc.* **7**, 391–397 (2008).

70. Potron, A., Poirel, L. & Nordmann, P. Emerging broad-spectrum resistance in Pseudomonas aeruginosa and Acinetobacter baumannii: Mechanisms and epidemiology. *Int. J. Antimicrob. Agents* **45**, 568–585 (2015).

71. Cabral, A. B. *et al.* Clonal spread and accumulation of β-lactam resistance determinants in Enterobacter aerogenes and Enterobacter cloacae complex isolates from infection and colonization in patients at a public hospital in Recife, Pernambuco, Brazil. *J. Med. Microbiol.* **66**, 70–77 (2017).

72. Liu, C. *et al.* New Delhi Metallo-β-Lactamase 1(NDM-1), the Dominant Carbapenemase Detected in Carbapenem-Resistant Enterobacter cloacae from Henan Province, China. *PLoS ONE* **10**, (2015).

73. Babouee Flury, B. *et al.* The differential importance of mutations within AmpD in cephalosporin resistance of Enterobacter aerogenes and Enterobacter cloacae. *Int. J. Antimicrob. Agents* **48**, 555–558 (2016).
74. Fernández, J. *et al.* Dissemination of multiresistant Enterobacter cloacae isolates producing OXA-48 and CTX-M-15 in a Spanish hospital. *Int. J. Antimicrob. Agents* **46**, 469–474 (2015).
75. Davin-Regli, A. & Pagès, J.-M. Enterobacter aerogenes and Enterobacter cloacae; versatile bacterial pathogens confronting antibiotic treatment. *Front. Microbiol.* **6**, 392 (2015).
76. Norgan, A. P. *et al.* Carbapenem- and Colistin-Resistant Enterobacter cloacae from Delta, Colorado, in 2015. *Antimicrob. Agents Chemother.* **60**, 3141–3144 (2016).
77. Cha, M. K. *et al.* Genetic characterisation of tigecycline-resistant Enterobacter spp. in blood isolates causing bacteraemia. *J. Glob. Antimicrob. Resist.* **13**, 115–118 (2018).
78. Bentley, R. & Meganathan, R. Biosynthesis of vitamin K (menaquinone) in bacteria. *Microbiol. Rev.* **46**, 241–280 (1982).
79. Hudault, S., Guignot, J. & Servin, A. Escherichia coli strains colonising the gastrointestinal tract protect germfree mice against Salmonella typhimurium infection. *Gut* **49**, 47–55 (2001).
80. Antimicrobial resistance: global report on surveillance. *World Health Organization* https://apps.who.int/iris/bitstream/handle/10665/112642/9789241564748_eng.pdf?sequence=1 (2014).
81. Surveillance of antimicrobial resistance in Europe 2017. *European Centre for Disease Prevention and Control* https://www.ecdc.europa.eu/sites/portal/files/documents/EARS-Net-report-2017-update-jan-2019.pdf (2018).
82. Roer, L. *et al.* WGS-based surveillance of third-generation cephalosporin-resistant Escherichia coli from bloodstream infections in Denmark. *J. Antimicrob. Chemother.* **72**, 1922–1929 (2017).
83. Cantón, R. *et al.* Rapid evolution and spread of carbapenemases among Enterobacteriaceae in Europe. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **18**, 413–431 (2012).
84. Korona-Glowniak, I., Skrzypek, K., Siwiec, R., Wrobel, A. & Malm, A. Fluoroquinolone-resistance mechanisms and phylogenetic background of clinical Escherichia coli strains isolated in south-east Poland. *New Microbiol.* **39**, 210–215 (2016).

85. Haldorsen, B. C., Simonsen, G. S., Sundsfjord, A., Samuelsen, O. & Norwegian Study Group on Aminoglycoside Resistance. Increased prevalence of aminoglycoside resistance in clinical isolates of Escherichia coli and Klebsiella spp. in Norway is associated with the acquisition of AAC(3)-II and AAC(6')-Ib. *Diagn. Microbiol. Infect. Dis.* **78**, 66–69 (2014).
86. Mediavilla, J. R. *et al.* Colistin- and Carbapenem-Resistant Escherichia coli Harboring mcr-1 and blaNDM-5, Causing a Complicated Urinary Tract Infection in a Patient from the United States. *mBio* **7**, (2016).
87. Wang, Q. *et al.* Emergence of tigecycline resistance in Escherichia coli co-producing MCR-1 and NDM-5 during tigecycline salvage treatment. *Infect. Drug Resist.* **11**, 2241–2248 (2018).
88. Seiple, I. B. *et al.* A platform for the discovery of new macrolide antibiotics. *Nature* **533**, 338–345 (2016).
89. Rohde, C., Wittmann, J. & Kutter, E. Bacteriophages: A Therapy Concept against Multi-Drug-Resistant Bacteria. *Surg. Infect.* **19**, 737–744 (2018).
90. Mahlapuu, M., Håkansson, J., Ringstad, L. & Björn, C. Antimicrobial Peptides: An Emerging Category of Therapeutic Agents. *Front. Cell. Infect. Microbiol.* **6**, (2016).
91. Calvert, M. B., Jumde, V. R. & Titz, A. Pathoblockers or antivirulence drugs as a new option for the treatment of bacterial infections. *Beilstein J. Org. Chem.* **14**, 2607–2617 (2018).
92. Campfield, B., Chen, K. & Kolls, J. K. Vaccine approaches for multidrug resistant Gram negative infections. *Curr. Opin. Immunol.* **28**, 84–89 (2014).
93. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
94. Chellat, M. F., Raguž, L. & Riedl, R. Targeting Antibiotic Resistance. *Angew. Chem. Int. Ed Engl.* **55**, 6600–6626 (2016).
95. Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12**, 371–387 (2013).
96. Coates, A., Hu, Y., Bax, R. & Page, C. The future challenges facing the development of new antimicrobial drugs. *Nat. Rev. Drug Discov.* **1**, 895–910 (2002).
97. Munita, J. M. & Arias, C. A. Mechanisms of Antibiotic Resistance. *Microbiol. Spectr.* **4**, (2016).
98. Antão, E.-M. *et al.* Antibiotic resistance, the 3As and the road ahead. *Gut Pathog.* **10**, 52 (2018).

99. Peterson, E. & Kaur, P. Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. *Front. Microbiol.* **9**, 2928 (2018).

100. Iredell, J., Brown, J. & Tagg, K. Antibiotic resistance in Enterobacteriaceae: mechanisms and clinical implications. *BMJ* **352**, h6420 (2016).

101. Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**, 67–113 (1992).

102. Pao, S. S., Paulsen, I. T. & Saier, M. H. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev. MMBR* **62**, 1–34 (1998).

103. Jack, D. L., Yang, N. M. & Saier, M. H. The drug/metabolite transporter superfamily. *Eur. J. Biochem.* **268**, 3620–3639 (2001).

104. Hvorup, R. N. *et al.* The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *Eur. J. Biochem.* **270**, 799–813 (2003).

105. Tseng, T. T. *et al.* The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.* **1**, 107–125 (1999).

106. Delmar, J. A. & Yu, E. W. The AbgT family: A novel class of antimetabolite transporters. *Protein Sci. Publ. Protein Soc.* **25**, 322–337 (2016).

107. Hassan, K. A., Liu, Q., Henderson, P. J. F. & Paulsen, I. T. Homologs of the Acinetobacter baumannii AceI transporter represent a new family of bacterial multidrug efflux systems. *mBio* **6**, (2015).

108. Zgurskaya, H. I. Multicomponent drug efflux complexes: architecture and mechanism of assembly. *Future Microbiol.* **4**, 919–932 (2009).

109. Yan, N. Structural Biology of the Major Facilitator Superfamily Transporters. *Annu. Rev. Biophys.* **44**, 257–283 (2015).

110. Quistgaard, E. M., Löw, C., Guettou, F. & Nordlund, P. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat. Rev. Mol. Cell Biol.* **17**, 123–132 (2016).

111. TC-SUPERFAMILIES. *Transporter Classification Database* http://www.tcdb.org/superfamily.php (2019).

112. Saier, M. H. & Paulsen, I. T. Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* **12**, 205–213 (2001).

113. Chen, D. E., Podell, S., Sauer, J.-D., Swanson, M. S. & Saier, M. H. The phagosomal nutrient transporter (Pht) family. *Microbiol. Read. Engl.* **154**, 42–53 (2008).

114. Lorca, G. L. *et al.* Transport capabilities of eleven gram-positive bacteria: comparative genomic analyses. *Biochim. Biophys. Acta* **1768**, 1342–1366 (2007).

115. Yen, M. R., Chen, J. S., Marquez, J. L., Sun, E. I. & Saier, M. H. Multidrug resistance: phylogenetic characterization of superfamilies of secondary carriers that include drug exporters. *Methods Mol. Biol. Clifton NJ* **637**, 47–64 (2010).

116. Saier, M. H. *et al.* The major facilitator superfamily. *J. Mol. Microbiol. Biotechnol.* **1**, 257–279 (1999).

117. Saier, M. H. Tracing pathways of transport protein evolution. *Mol. Microbiol.* **48**, 1145–1156 (2003).

118. Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I. & Saier, M. H. The major facilitator superfamily (MFS) revisited. *FEBS J.* **279**, 2022–2035 (2012).

119. Saier, M. H. Transport protein evolution deduced from analysis of sequence, topology and structure. *Curr. Opin. Struct. Biol.* **38**, 9–17 (2016).

120. Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* **38**, 151–159 (2013).

121. Kumar, A. & Schweizer, H. P. Bacterial resistance to antibiotics: active efflux and reduced uptake. *Adv. Drug Deliv. Rev.* **57**, 1486–1513 (2005).

122. Li, X.-Z., Elkins, C. A. & Zgurskaya, H. I. *Efflux-Mediated Antimicrobial Resistance in Bacteria Mechanisms, Regulation and Clinical Implications*. (Springer International Publishing, 2016).

123. Hayashi, M., Tabata, K., Yagasaki, M. & Yonetani, Y. Effect of multidrug-efflux transporter genes on dipeptide resistance and overproduction in Escherichia coli. *FEMS Microbiol. Lett.* **304**, 12–19 (2010).

124. Yamada, S. *et al.* Effect of drug transporter genes on cysteine export and overproduction in Escherichia coli. *Appl. Environ. Microbiol.* **72**, 4735–4742 (2006).

125. Phadtare, S., Yamanaka, K., Kato, I. & Inouye, M. Antibacterial activity of 4,5-dihydroxy-2-cyclopentan-1-one (DHCP) and cloning of a gene conferring DHCP resistance in Escherichia coli. *J. Mol. Microbiol. Biotechnol.* **3**, 461–465 (2001).

126. Naroditskaya, V., Schlosser, M. J., Fang, N. Y. & Lewis, K. An E. coli gene emrD is involved in adaptation to low energy shock. *Biochem. Biophys. Res. Commun.* **196**, 803–809 (1993).

127. Matsumura, K., Furukawa, S., Ogihara, H. & Morinaga, Y. Roles of multidrug efflux pumps on the biofilm formation of Escherichia coli K-12. *Biocontrol Sci.* **16**, 69–72 (2011).

128. Koita, K. & Rao, C. V. Identification and analysis of the putative pentose sugar efflux transporters in Escherichia coli. *PloS One* **7**, e43700 (2012).

129. Blickwede, M. & Schwarz, S. Molecular analysis of florfenicol-resistant Escherichia coli isolates from pigs. *J. Antimicrob. Chemother.* **53**, 58–64 (2004).

130. Cloeckaert, A. *et al.* Plasmid-mediated florfenicol resistance encoded by the floR gene in Escherichia coli isolated from cattle. *Antimicrob. Agents Chemother.* **44**, 2858–2860 (2000).

131. Edgar, R. & Bibi, E. MdfA, an Escherichia coli multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. *J. Bacteriol.* **179**, 2274–2280 (1997).

132. Bohn, C. & Bouloc, P. The Escherichia coli cmlA gene encodes the multidrug efflux pump Cmr/MdfA and is responsible for isopropyl-beta-D-thiogalactopyranoside exclusion and spectinomycin sensitivity. *J. Bacteriol.* **180**, 6072–6075 (1998).

133. Mine, T., Morita, Y., Kataoka, A., Mizushima, T. & Tsuchiya, T. Evidence for chloramphenicol/H+ antiport in Cmr (MdfA) system of Escherichia coli and properties of the antiporter. *J. Biochem. (Tokyo)* **124**, 187–193 (1998).

134. Krulwich, T. A., Lewinson, O., Padan, E. & Bibi, E. Do physiological roles foster persistence of drug/multidrug-efflux transporters? A case study. *Nat. Rev. Microbiol.* **3**, 566–572 (2005).

135. Nishino, K. & Yamaguchi, A. Analysis of a complete library of putative drug transporter genes in Escherichia coli. *J. Bacteriol.* **183**, 5803–5812 (2001).

136. Fàbrega, A., Martin, R. G., Rosner, J. L., Tavio, M. M. & Vila, J. Constitutive SoxS expression in a fluoroquinolone-resistant strain with a truncated SoxR protein and identification of a new member of the marA-soxS-rob regulon, mdtG. *Antimicrob. Agents Chemother.* **54**, 1218–1225 (2010).

137. Holdsworth, S. R. & Law, C. J. The major facilitator superfamily transporter MdtM contributes to the intrinsic resistance of Escherichia coli to quaternary ammonium compounds. *J. Antimicrob. Chemother.* **68**, 831–839 (2013).

138. Holdsworth, S. R. & Law, C. J. Functional and biochemical characterisation of the Escherichia coli major facilitator superfamily multidrug transporter MdtM. *Biochimie* **94**, 1334–1346 (2012).

139. Holdsworth, S. R. & Law, C. J. Multidrug resistance protein MdtM adds to the repertoire of antiporters involved in alkaline pH homeostasis in Escherichia coli. *BMC Microbiol.* **13**, 113 (2013).

140. Paul, S. *et al.* A single-component multidrug transporter of the major facilitator superfamily is part of a network that protects Escherichia coli from bile salt stress. *Mol. Microbiol.* **92**, 872–884 (2014).

141. Liu, J., Keelan, P., Bennett, P. M. & Enne, V. I. Characterization of a novel macrolide efflux gene, mef(B), found linked to sul3 in porcine Escherichia coli. *J. Antimicrob. Chemother.* **63**, 423–426 (2009).

142. Cattoir, V., Poirel, L. & Nordmann, P. Plasmid-mediated quinolone resistance pump QepA2 in an Escherichia coli isolate from France. *Antimicrob. Agents Chemother.* **52**, 3801–3804 (2008).

143. Levy, S. B. Active efflux mechanisms for antimicrobial resistance. *Antimicrob. Agents Chemother.* **36**, 695–703 (1992).

144. Lomovskaya, O. & Lewis, K. Emr, an Escherichia coli locus for multidrug resistance. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8938–8942 (1992).

145. Elkins, C. A. & Mullis, L. B. Mammalian steroid hormones are substrates for the major RND- and MFS-type tripartite multidrug efflux pumps of Escherichia coli. *J. Bacteriol.* **188**, 1191–1195 (2006).

146. Deininger, K. N. W. *et al.* A requirement of TolC and MDR efflux pumps for acid adaptation and GadAB induction in Escherichia coli. *PloS One* **6**, e18960 (2011).

147. Furukawa, H., Tsay, J. T., Jackowski, S., Takamura, Y. & Rock, C. O. Thiolactomycin resistance in Escherichia coli is associated with the multidrug resistance efflux pump encoded by emrAB. *J. Bacteriol.* **175**, 3723–3729 (1993).

148. Elkins, C. A. & Mullis, L. B. Substrate competition studies using whole-cell accumulation assays with the major tripartite multidrug efflux pumps of Escherichia coli. *Antimicrob. Agents Chemother.* **51**, 923–929 (2007).

149. Nishino, K. & Yamaguchi, A. Overexpression of the response regulator evgA of the two-component signal transduction system modulates multidrug resistance conferred by multidrug resistance transporters. *J. Bacteriol.* **183**, 1455–1458 (2001).

150. Han, X. *et al.* Escherichia coli genes that reduce the lethal effects of stress. *BMC Microbiol.* **10**, 35 (2010).

151. Nishino, K. & Yamaguchi, A. Role of histone-like protein H-NS in multidrug resistance of Escherichia coli. *J. Bacteriol.* **186**, 1423–1429 (2004).

152. Eguchi, Y. *et al.* Transcriptional regulation of drug efflux genes by EvgAS, a two-component system in Escherichia coli. *Microbiol. Read. Engl.* **149**, 2819–2828 (2003).

153. Yoshida, H., Bogaki, M., Nakamura, S., Ubukata, K. & Konno, M. Nucleotide sequence and characterization of the Staphylococcus aureus norA gene, which confers resistance to quinolones. *J. Bacteriol.* **172**, 6942–6949 (1990).

154. Neyfakh, A. A., Borsch, C. M. & Kaatz, G. W. Fluoroquinolone resistance protein NorA of Staphylococcus aureus is a multidrug efflux transporter. *Antimicrob. Agents Chemother.* **37**, 128–129 (1993).

155. Kaatz, G. W., Seo, S. M. & Ruble, C. A. Efflux-mediated fluoroquinolone resistance in Staphylococcus aureus. *Antimicrob. Agents Chemother.* **37**, 1086–1094 (1993).

156. Kaatz, G. W., Thyagarajan, R. V. & Seo, S. M. Effect of promoter region mutations and mgrA overexpression on transcription of norA, which encodes a Staphylococcus aureus multidrug efflux transporter. *Antimicrob. Agents Chemother.* **49**, 161–169 (2005).

157. Brown, M. H. & Skurray, R. A. Staphylococcal multidrug efflux protein QacA. *J. Mol. Microbiol. Biotechnol.* **3**, 163–170 (2001).

158. Mitchell, B. A., Brown, M. H. & Skurray, R. A. QacA multidrug efflux pump from Staphylococcus aureus: comparative analysis of resistance to diamidines, biguanidines, and guanylhydrazones. *Antimicrob. Agents Chemother.* **42**, 475–477 (1998).

159. Mitchell, B. A., Paulsen, I. T., Brown, M. H. & Skurray, R. A. Bioenergetics of the staphylococcal multidrug export protein QacA. Identification of distinct binding sites for monovalent and divalent cations. *J. Biol. Chem.* **274**, 3541–3548 (1999).

160. Littlejohn, T. G. *et al.* Substrate specificity and energetics of antiseptic and disinfectant resistance in Staphylococcus aureus. *FEMS Microbiol. Lett.* **74**, 259–265 (1992).

161. Floyd, J. L., Smith, K. P., Kumar, S. H., Floyd, J. T. & Varela, M. F. LmrS is a multidrug efflux pump of the major facilitator superfamily from Staphylococcus aureus. *Antimicrob. Agents Chemother.* **54**, 5406–5412 (2010).

162. Srinivasan, V. B., Singh, B. B., Priyadarshi, N., Chauhan, N. K. & Rajamohan, G. Role of novel multidrug efflux pump involved in drug resistance in Klebsiella pneumoniae. *PloS One* **9**, e96288 (2014).

163. Roca, I. *et al.* CraA, a major facilitator superfamily efflux pump associated with chloramphenicol resistance in Acinetobacter baumannii. *Antimicrob. Agents Chemother.* **53**, 4013–4014 (2009).

164. Zhu, L. *et al.* Complete genome analysis of three Acinetobacter baumannii clinical isolates in China for insight into the diversification of drug resistance elements. *PloS One* **8**, e66584 (2013).

165. Rajamohan, G., Srinivasan, V. B. & Gebreyes, W. A. Molecular and functional characterization of a novel efflux pump, AmvA, mediating antimicrobial and disinfectant resistance in Acinetobacter baumannii. *J. Antimicrob. Chemother.* **65**, 1919–1925 (2010).

166. Colinon, C. *et al.* Genetic analyses of Pseudomonas aeruginosa isolated from healthy captive snakes: evidence of high inter- and intrasite dissemination and occurrence of antibiotic resistance genes. *Environ. Microbiol.* **12**, 716–729 (2010).

167. Jonas, B. M., Murray, B. E. & Weinstock, G. M. Characterization of emeA, a norA Homolog and Multidrug Resistance Efflux Pump, in Enterococcus faecalis. *Antimicrob. Agents Chemother.* **45**, 3574–3579 (2001).

168. Van Bambeke, F., Pagès, J.-M. & Lee, V. J. Inhibitors of bacterial efflux pumps as adjuvants in antibiotic treatments and diagnostic tools for detection of resistance by efflux. *Recent Patents Anti-Infect. Drug Disc.* **1**, 157–175 (2006).

169. Ahmed, M. *et al.* Two highly similar multidrug transporters of Bacillus subtilis whose expression is differentially regulated. *J. Bacteriol.* **177**, 3904–3910 (1995).

170. Baranova, N. N., Danchin, A. & Neyfakh, A. A. Mta, a global MerR-type regulator of the Bacillus subtilis multidrug-efflux transporters. *Mol. Microbiol.* **31**, 1549–1559 (1999).

171. Bolhuis, H. *et al.* The Lactococcal lmrP gene encodes a proton motive force-dependent drug transporter. *J. Biol. Chem.* **270**, 26092–26098 (1995).

172. Nishino, K., Latifi, T. & Groisman, E. A. Virulence and drug resistance roles of multidrug efflux systems of Salmonella enterica serovar Typhimurium. *Mol. Microbiol.* **59**, 126–141 (2006).

173. Rensch, U., Nishino, K., Klein, G. & Kehrenberg, C. Salmonella enterica serovar Typhimurium multidrug efflux pumps EmrAB and AcrEF support the major efflux system AcrAB in decreased susceptibility to triclosan. *Int. J. Antimicrob. Agents* **44**, 179–180 (2014).

174. Smith, K. P., Kumar, S. & Varela, M. F. Identification, cloning, and functional characterization of EmrD-3, a putative multidrug efflux pump of the major facilitator superfamily from Vibrio cholerae O395. *Arch. Microbiol.* **191**, 903–911 (2009).

175. Colmer, J. A., Fralick, J. A. & Hamood, A. N. Isolation and characterization of a putative multidrug resistance pump from Vibrio cholerae. *Mol. Microbiol.* **27**, 63–72 (1998).

176. Woolley, R. C. *et al.* Characterization of the Vibrio cholerae vceCAB multiple-drug resistance efflux operon in Escherichia coli. *J. Bacteriol.* **187**, 5500–5503 (2005).

177. Chang, Y.-C., Tsai, M.-J., Huang, Y.-W., Chung, T.-C. & Yang, T.-C. SmQnrR, a DeoR-type transcriptional regulator, negatively regulates the expression of Smqnr and SmtcrA in Stenotrophomonas maltophilia. *J. Antimicrob. Chemother.* **66**, 1024–1028 (2011).

178. Huang, Y.-W., Hu, R.-M., Chu, F.-Y., Lin, H.-R. & Yang, T.-C. Characterization of a major facilitator superfamily (MFS) tripartite efflux pump EmrCABsm from Stenotrophomonas maltophilia. *J. Antimicrob. Chemother.* **68**, 2498–2505 (2013).

179. Rossbach, S., Kunze, K., Albert, S., Zehner, S. & Göttfert, M. The Sinorhizobium meliloti EmrAB efflux system is regulated by flavonoids through a TetR-like regulator (EmrR). *Mol. Plant-Microbe Interact. MPMI* **27**, 379–387 (2014).

180. Maggiorani Valecillos, A., Rodríguez Palenzuela, P. & López-Solanilla, E. The role of several multidrug resistance systems in Erwinia chrysanthemi pathogenesis. *Mol. Plant-Microbe Interact. MPMI* **19**, 607–613 (2006).

181. Ravirala, R. S. *et al.* Efflux pump gene expression in Erwinia chrysanthemi is induced by exposure to phenolic acids. *Mol. Plant-Microbe Interact. MPMI* **20**, 313–320 (2007).

182. Lee, E. H. & Shafer, W. M. The farAB-encoded efflux pump mediates resistance of gonococci to long-chained antibacterial fatty acids. *Mol. Microbiol.* **33**, 839–845 (1999).

183. Tait-Kamradt, A. *et al.* mefE is necessary for the erythromycin-resistant M phenotype in Streptococcus pneumoniae. *Antimicrob. Agents Chemother.* **41**, 2251–2255 (1997).

184. Borges-Walmsley, M. I. *et al.* Identification of oligomerization and drug-binding domains of the membrane fusion protein EmrA. *J. Biol. Chem.* **278**, 12903–12912 (2003).

185. Guan, L. & Kaback, H. R. Lessons from Lactose Permease. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 67–91 (2006).

186. Abramson, J. *et al.* Structure and mechanism of the lactose permease of Escherichia coli. *Science* **301**, 610–615 (2003).

187. Mirza, O., Guan, L., Verner, G., Iwata, S. & Kaback, H. R. Structural evidence for induced fit and a mechanism for sugar/H+ symport in LacY. *EMBO J.* **25**, 1177–1183 (2006).

188. Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H. R. Structural determination of wild-type lactose permease. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15294–15298 (2007).

189. Chaptal, V. *et al.* Crystal structure of lactose permease in complex with an affinity inactivator yields unique insight into sugar recognition. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9361–9366 (2011).

190. Kumar, H. *et al.* Structure of sugar-bound LacY. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1784–1788 (2014).

191. Kumar, H., Finer-Moore, J. S., Kaback, H. R. & Stroud, R. M. Structure of LacY with an α-substituted galactoside: Connecting the binding site to the protonation site. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9004–9009 (2015).

192. Jiang, X. *et al.* Crystal structure of a LacY-nanobody complex in a periplasmic-open conformation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12420–12425 (2016).

193. Kumar, H. *et al.* Crystal Structure of a ligand-bound LacY-Nanobody Complex. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 8769–8774 (2018).

194. Heng, J. *et al.* Substrate-bound structure of the E. coli multidrug resistance transporter MdfA. *Cell Res.* **25**, 1060–1073 (2015).

195. Paulsen, I. T., Brown, M. H. & Skurray, R. A. Proton-dependent multidrug efflux systems. *Microbiol. Rev.* **60**, 575–608 (1996).

196. Zomot, E. *et al.* A New Critical Conformational Determinant of Multidrug Efflux by an MFS Transporter. *J. Mol. Biol.* **430**, 1368–1385 (2018).

197. Nagarathinam, K. *et al.* Outward open conformation of a Major Facilitator Superfamily multidrug/H+ antiporter provides insights into switching mechanism. *Nat. Commun.* **9**, 4005 (2018).

198. Wu, H.-H., Symersky, J. & Lu, M. Structure of an engineered multidrug transporter MdfA reveals the molecular basis for substrate recognition. *Commun. Biol.* **2**, 210 (2019).

199. Yin, Y., He, X., Szewczyk, P., Nguyen, T. & Chang, G. Structure of the multidrug transporter EmrD from Escherichia coli. *Science* **312**, 741–744 (2006).

200. Newstead, S. *et al.* Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* **30**, 417–426 (2011).

201. Fowler, P. W. *et al.* Gating topology of the proton-coupled oligopeptide symporters. *Struct. Lond. Engl. 1993* **23**, 290–301 (2015).

202. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).

203. Law, C. J., Maloney, P. C. & Wang, D.-N. Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **62**, 289–305 (2008).

204. del Castillo, I., Gómez, J. M. & Moreno, F. mprA, an Escherichia coli gene that reduces growth-phase-dependent synthesis of microcins B17 and C7 and blocks osmoinduction of proU when cloned on a high-copy-number plasmid. *J. Bacteriol.* **172**, 437–445 (1990).

205. Lomovskaya, O., Lewis, K. & Matin, A. EmrR is a negative regulator of the Escherichia coli multidrug resistance pump EmrAB. *J. Bacteriol.* **177**, 2328–2334 (1995).

206. Brooun, A., Tomashek, J. J. & Lewis, K. Purification and ligand binding of EmrR, a regulator of a multidrug transporter. *J. Bacteriol.* **181**, 5131–5133 (1999).

207. Xiong, A. *et al.* The EmrR protein represses the Escherichia coli emrRAB multidrug resistance operon by directly binding to its promoter region. *Antimicrob. Agents Chemother.* **44**, 2905–2907 (2000).

208. Lomovskaya, O., Kawai, F. & Matin, A. Differential regulation of the mcb and emr operons of Escherichia coli: role of mcb in multidrug resistance. *Antimicrob. Agents Chemother.* **40**, 1050–1052 (1996).

209. Glaser, P., Sakamoto, H., Bellalou, J., Ullmann, A. & Danchin, A. Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of Bordetella pertussis. *EMBO J.* **7**, 3997–4004 (1988).

210. Felmlee, T., Pellett, S., Lee, E. Y. & Welch, R. A. Escherichia coli hemolysin is released extracellularly without cleavage of a signal peptide. *J. Bacteriol.* **163**, 88–93 (1985).

211. Gilson, L., Mahanty, H. K. & Kolter, R. Genetic analysis of an MDR-like export system: the secretion of colicin V. *EMBO J.* **9**, 3875–3884 (1990).

212. Lewis, K. Multidrug resistance pumps in bacteria: variations on a theme. *Trends Biochem. Sci.* **19**, 119–123 (1994).

213. Lewis, K. Translocases: a bacterial tunnel for drugs and proteins. *Curr. Biol. CB* **10**, R678-681 (2000).

214. Tanabe, M. *et al.* The multidrug resistance efflux complex, EmrAB from Escherichia coli forms a dimer in vitro. *Biochem. Biophys. Res. Commun.* **380**, 338–342 (2009).

215. Hinchliffe, P. *et al.* Structure of the periplasmic adaptor protein from a major facilitator superfamily (MFS) multidrug efflux pump. *Febs Lett.* **588**, 3147–3153 (2014).

216. Pei, X.-Y. *et al.* Structures of sequential open states in a symmetrical opening transition of the TolC exit duct. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2112–2117 (2011).

217. Symmons, M. F., Marshall, R. L. & Bavro, V. N. Architecture and roles of periplasmic adaptor proteins in tripartite efflux assemblies. *Front. Microbiol.* **6**, (2015).

218. Shafer, W. M. *et al.* Efflux Pumps in Neisseria gonorrhoeae: Contributions to Antimicrobial Resistance and Virulence. in *Efflux-Mediated Antimicrobial Resistance in Bacteria: Mechanisms, Regulation and Clinical Implications* (eds. Li, X.-Z., Elkins, C. A. & Zgurskaya, H. I.) 439–469 (Springer International Publishing, 2016). doi:10.1007/978-3-319-39658-3_17.

219. Lei, H.-T. *et al.* Crystal structure of the open state of the Neisseria gonorrhoeae MtrE outer membrane channel. *PloS One* **9**, e97475 (2014).

220. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914–919 (2000).

221. Higgins, M. K. *et al.* Structure of the ligand-blocked periplasmic entrance of the bacterial multidrug efflux protein TolC. *J. Mol. Biol.* **342**, 697–702 (2004).

222. Bavro, V. N. *et al.* Assembly and Channel Opening in a Bacterial Drug Efflux Machine. *Mol. Cell* **30**, 114–121 (2008).

223. Koronakis, V., Andersen, C. & Hughes, C. Channel-tunnels. *Curr. Opin. Struct. Biol.* **11**, 403–407 (2001).

224. Andersen, C., Koronakis, E., Hughes, C. & Koronakis, V. An aspartate ring at the TolC tunnel entrance determines ion selectivity and presents a target for blocking by large cations. *Mol. Microbiol.* **44**, 1131–1139 (2002).

225. Tikhonova, E. B., Dastidar, V., Rybenkov, V. V. & Zgurskaya, H. I. Kinetic control of TolC recruitment by multidrug efflux complexes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 16416–16421 (2009).

226. Ren, Q., Chen, K. & Paulsen, I. T. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* **35**, D274–D279 (2007).

227. TransportDB Genomic Comparisons of Membrane Transport Systems. http://www.membranetransport.org/transportDB2/index.html.

228. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

229. Clustal Omega. *EMBL-EBI* https://www.ebi.ac.uk/Tools/msa/clustalo/.

230. Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **28**, 1102, 1104 (2000).

231. Sequence Manipulation Suite: Ident and Sim. https://www.bioinformatics.org/sms2/ident_sim.html.

232. Pearson, W. R. An Introduction to Sequence Similarity ("Homology") Searching. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **0 3**, (2013).

233. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50**, 3272–3278 (2011).

234. Wertman, K. F., Wyman, A. R. & Botstein, D. Host/vector interactions which affect the viability of recombinant phage lambda clones. *Gene* **49**, 253–262 (1986).

235. Tolia, N. H. & Joshua-Tor, L. Strategies for protein coexpression in Escherichia coli. *Nat. Methods* **3**, 55–64 (2006).

236. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7877–7882 (2002).

237. Taylor, R. G., Walker, D. C. & McInnes, R. R. E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res.* **21**, 1677–1678 (1993).

238. 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**, 79–88 (2006).

239. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795 (2000).

240. Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H. & Phillips, G. J. Green fluorescent protein functions as a reporter for protein localization in Escherichia coli. *J. Bacteriol.* **182**, 4068–4076 (2000).

241. Fitzpatrick, A. W. P. *et al.* Structure of the MacAB-TolC ABC-type tripartite multidrug efflux pump. *Nat. Microbiol.* **2**, 17070 (2017).

242. Daury, L. *et al.* Tripartite assembly of RND multidrug efflux pumps. *Nat. Commun.* **7**, 10731 (2016).

243. Tsutsumi, K. *et al.* Structures of the wild-type MexAB–OprM tripartite pump reveal its complex formation and drug efflux mechanism. *Nat. Commun.* **10**, (2019).

244. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**, 691–695 (1999).

245. Drew, D. E., von Heijne, G., Nordlund, P. & de Gier, J. W. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Lett.* **507**, 220–224 (2001).

246. Hsieh, J. M. *et al.* Bridging the gap: a GFP-based strategy for overexpression and purification of membrane proteins with intra and extracellular C-termini. *Protein Sci. Publ. Protein Soc.* **19**, 868–880 (2010).

247. Drew, D. *et al.* A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci. Publ. Protein Soc.* **14**, 2011–2017 (2005).

248. mRFP1. *FPbase* https://www.fpbase.org/protein/mrfp1/.

249. Superfolder GFP. *FPbase* https://www.fpbase.org/protein/superfolder-gfp/.

250. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Struct. Lond. Engl. 1993* **14**, 673–681 (2006).

251. Parcej, D., Guntrum, R., Schmidt, S., Hinz, A. & Tampé, R. Multicolour fluorescence-detection size-exclusion chromatography for structural genomics of membrane multiprotein complexes. *PloS One* **8**, e67112 (2013).

252. Sjöstrand, D., Diamanti, R., Lundgren, C. A. K., Wiseman, B. & Högbom, M. A rapid expression and purification condition screening protocol for membrane protein structural biology. *Protein Sci. Publ. Protein Soc.* **26**, 1653–1666 (2017).

253. Hattori, M., Hibbs, R. E. & Gouaux, E. A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Struct. Lond. Engl. 1993* **20**, 1293–1299 (2012).

254. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).

255. Miroux, B. & Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).

256. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).

257. Geertsma, E. R., Groeneveld, M., Slotboom, D.-J. & Poolman, B. Quality control of overexpressed membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5722–5727 (2008).

258. Rascle, C. *et al.* The pH regulator PacC: a host-dependent virulence factor in Botrytis cinerea. *Environ. Microbiol. Rep.* **10**, 555–568 (2018).

259. Schlegel, S. *et al.* Revolutionizing membrane protein overexpression in bacteria. *Microb. Biotechnol.* **3**, 403–411 (2010).

260. Wagner, S., Bader, M. L., Drew, D. & de Gier, J.-W. Rationalizing membrane protein overexpression. *Trends Biotechnol.* **24**, 364–371 (2006).

261. Scott, D. J., Kummer, L., Tremmel, D. & Plückthun, A. Stabilizing membrane proteins through protein engineering. *Curr. Opin. Chem. Biol.* **17**, 427–435 (2013).

262. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).

263. Asial, I., Nordlund, P. & Dahlroth, S.-L. Hot CoFi Blot: A High-Throughput Colony-Based Screen for Identifying More Thermally Stable Protein Variants. *Methods Mol. Biol. Clifton NJ* **2025**, 299–320 (2019).

264. Lee, C. *et al.* MemStar: a one-shot Escherichia coli-based approach for high-level bacterial membrane protein production. *FEBS Lett.* **588**, 3761–3769 (2014).

265. Kawate, T., Michel, J. C., Birdsong, W. T. & Gouaux, E. Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. *Nature* **460**, 592–598 (2009).

266. Postis, V. *et al.* The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. *Biochim. Biophys. Acta* **1848**, 496–501 (2015).

267. Picard, M., Duval-Terrié, C., Dé, E. & Champeil, P. Stabilization of membranes upon interaction of amphipathic polymers with membrane proteins. *Protein Sci. Publ. Protein Soc.* **13**, 3056–3058 (2004).

268. Zoonens, M. & Popot, J.-L. Amphipols for each season. *J. Membr. Biol.* **247**, 759–796 (2014).

269. Wang, Z. *et al.* An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *eLife* **6**, (2017).

270. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587–593 (2002).

271. Sennhauser, G., Bukowska, M. A., Briand, C. & Grütter, M. G. Crystal structure of the multidrug exporter MexB from Pseudomonas aeruginosa. *J. Mol. Biol.* **389**, 134–145 (2009).

272. Du, D. *et al.* Multidrug efflux pumps: structure, function and regulation. *Nat. Rev. Microbiol.* **16**, 523–539 (2018).

273. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).

274. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).

275. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).

276. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

277. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030 (2014).

278. Du, D. *et al.* Structure of the AcrAB-TolC multidrug efflux pump. *Nature* **509**, 512–515 (2014).

ANNEXES

Curriculum Vitae

Narek Yousefian

12, Rue André Messager (Apt. 15)
64000 Pau, France
Mobile: (+33) 7 69 44 46 34
Email: narek.yousefian@gmail.com

PERSONAL INFORMATION

Nationality: French
Born: 10/09/1992 (in Tehran)
Age: 27

EDUCATION

2016-2020 **Ph.D. in Biochemistry in cotutelle** (Thesis title: "The three-component multidrug MFS-type efflux pump EmrAB-TolC from *Escherichia coli*: from cloning to structural analysis") directed by Dr. Olivier Lambert (Laboratory CBMN CNRS UMR-5248) from the University of Bordeaux, France and Prof. Dr. Klaas Martinus Pos (Institute of Biochemistry) from Goethe University, Germany

2014-2016 **Master's degree** in Biochemistry, University of Bordeaux, France. (Theoretical knowledge of Structural and Chemical Biology).

2011-2014 **Bachelor's degree** in Microbiology, University of Pau, France. (Theoretical knowledge of Microbiology, Cell Biology, and Molecular Biology).

2010-2011 **Baccalaureate S**, A-Level equivalent specialized in Mathematics, Physics and Biology, Lycée Louis Barthou, Pau, France.

TECHNICAL SKILLS

Biotechnology Techniques

Molecular Biology
Primer Design, traditional cloning, fragment exchange cloning, site directed mutagenesis, PCR optimisation

Microbiology
Soluble and membrane protein expression optimisation in *E. coli*, high-throughput screening techniques for the study of the expression levels, protein production on large scale

Biochemical Techniques

Protein purification
Liquid chromatography techniques (affinity, size exclusion, hydrophobic interaction), high-throughput screening techniques using Fluorescent detection size-exclusion chromatography, recombinant protein detection with fluorescence reporter systems, qualitative and quantitative assays, spectroscopy, determination of the kinetic parameters of an enzyme, determination of pre-steady-state parameters of an enzyme by stopped-flow methods, experience with both soluble and membrane proteins

Membrane protein reconstitution
Removal of detergent with the use of amphipols, nanodiscs, SMA polymers

Protein structure study
Negative staining TEM

WORK EXPERIENCE

04/07/16-29/07/16

Training period
Institute of Chemistry & Biology of Membranes & Nano-objects (CBMN UMR-5248)
Allée Geoffroy Saint Hilaire, Bât B14 – 33600 Pessac, France.
Research team: Architecture of membrane complexes and cellular processes
Research topic: Study of multi-component efflux pumps in molecular detail.

04/01/16-30/06/16

Training period
Paul Pascal Research Center (CRPP UPR-8641)
115 Avenue Albert Schweitzer – 33600 Pessac, France.
Research team: Biofuel cells and Biosensors
Research topic: Biochemical and enzymatic characterization of Bilirubin oxidases for the elaboration of enzymatic biofuel cells.

01/04/15-01/06/15

Training period
Paul Pascal Research Center (CRPP UPR-8641)
115 Avenue Albert Schweitzer – 33600 Pessac, France.
Research team: Biofuel cells and Biosensors
Research topic: Determination of the rate-limiting step of the reaction catalyzed by the Bilirubin oxidases from *Bacillus pumilus* and *Magnaporthe oryzae*.

LANGUAGES

English	Fluent
German	Fluent
French	Native speaker
Armenian	Fluent
Persian	Fluent

COMPUTER SKILLS

General software: Word, Excel and PowerPoint.
Chemistry software: HyperChem, ChemSketch
Biology software: RStudio, Origin, Bio-Kine 32, PyMOL, BioEdit, Chimera, ImageJ

List of publications

Structural characterization of the EmrAB-TolC efflux complex from *E. coli*, Yousefian, N., Ornik-Cha, A., Poussard, S., Decossas, M., Berbon, M., Daury, L., Taveau, J.-C., Dupuy, J.-W., , Lambert, O., Pos, K.M., (BBA)-Biomembranes, **submitted**