Functional and Structural Characterization of

Aquifex aeolicus sulfide:quinone oxidoreductase

vom Fachbereich Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe Universität als Dissertation angenommen.

Dekan: Prof. Dr. Dieter Steinhilber

1. Gutachter: Prof. Dr. Bernd Ludwig
2. Gutachter: Prof. Dr. Hartmut Michel

Datum der Disputation:

Diese Doktorarbeit wurde vom 16. Januar 2006 bis zum 15. Januar 2010 unter Leitung von Prof. Dr. Hartmut Michel in der Abteilung für Molekulare Membranbiologie am Max-Planck-Institut für Biophysik in Frankfurt am Main durchgeführt.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig angefertigt habe und keine weiteren Hilfsmittel und Quellen als die hier aufgeführten verwendet habe.

Marco MARCIA

Frankfurt am Main

Publications

Marcia M., Ermler U., Peng G., Michel H. (2009) The structure of *Aquifex aeolicus* sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration. *Proc Natl Acad Sci USA* 106:9625-9630.

Marcia M., Ermler U., Peng G., Michel H. (2010) A new structure-based classification of sulfide:quinone oxidoreductases. *Proteins* 78:1073-1083.

Marcia M., Langer J.D., Parcej D., Vogel V., Peng G., Michel H. (2010) Characterizing a monotopic membrane enzyme. Biochemical, enzymatic and crystallization studies on *Aquifex aeolicus* sulfide:quinone oxidoreductase. *Biochim Biophys Acta* doi:10.1016/j.bbamem.2010.07.033.

Table of content

List of Figures

List of Tables

List of Equations

List of Appendix tables

Ausführliche deutschsprachige Zusammenfassung

Die experimentelle Bestimmung der Struktur und Funktion von Makromolekülen ist von hohem biologischen, mikrobiologischen, pharmakologischen und medizinischen Interesse. Die erfolgreichste Methode, um biologische Moleküle auf atomarer Ebene zu untersuchen, ist die Röntgenbeugung an 3-D Kristallen, durch die Strukturen von Makromolekülen mit hoher Auflösung ermittelt werden können. Diese Methode wird durch andere biochemische, genetische oder strukturelle Methoden wie Massenspektrometrie, analytische Ultrazentrifugation, Mutagenese, Elektronenmikroskopie und verschiedenen spektroskopischen Techniken ideal ergänzt.

Unter allen biologisch relevanten Makromolekülen stellen Membranproteine eine besondere Herausforderung für strukturelle Untersuchungen dar. Es ist schwierig, sie in ausreichenden Mengen zu gewinnen, sie sind instabil, nur durch milde amphipatische Detergenzien in Lösungen zu halten und neigen wenig dazu, gut geordnete Kristalle zu bilden. Trotzdem sind sie eine überaus interessante Proteingruppe, da sie rund 20 – 25 % aller Proteome ausmachen und 60 % aller Pharmaka an Membranproteinen angreifen.

Diese Arbeit beschreibt die strukturellen und funktionellen Eigenschaften eines membrangebundenen Enzyms, der Sulfid:Chinon Oxidoreduktase (SQR), die aus den Zellen des hyperthermophilen Bakteriums *Aquifex aeolicus* isoliert wurde.

SQRs gehören zusammen mit der Glutathion-Reduktase, der Lipoamid-Dehydrogenase, der Thioredoxin-Reduktase und der Trypanothion-Reduktase zur Superfamilie der Flavoprotein-Disulfid-Reduktasen (FDR). Für die ersten Lebensformen war die sulfidabhängige Chinon-Reduktion durch die SQRs essentiell, um in ihrem sauerstoffarmen, marinen Lebensraum Sulfid zur Energiegewinnung zu nutzen. Die Gene dieser Proteine sind durchweg in der Evolution erhalten geblieben und heute in allen Domänen des Lebens, auch im Menschen, jedoch nicht in Pflanzen, zu finden. Sie spielen eine physiologische Rolle bei der Sulfidentgiftung (in allen Organismen), bei der sulfidabhängigen Atmung und Photosynthese (in Bakterien) und bei der Schwermetalltoleranz (in Hefen). In Eukaryoten könnten sie an sulfidabhängigen Signalprozessen beteiligt sein, da Sulfid in den Zellen produziert wird und als „Gasotransmitter" wirkt.

Zu Beginn dieser Arbeit gab es keine der strukturellen Informationen über eine SQR, und nur einige wenige Homologe waren funktionell charakterisiert. Darüber hinaus zeigen SQRs eine geringe Sequenzähnlichkeit, so dass die Klassifizierung der SQR-Familie nur teilweise möglich und auch widersprüchlich war, da die Bestimmung der funktionell relevanten Sequenzbereiche ungenau war. Bisher wurde der katalytische Mechanismus auf Basis der Struktur der *Allochromatium vinosum* Flavocytochrom *c*:Sulfid Dehydrogenase (FCSD) vorhergesagt, ein Protein der FDR Familie mit nur 24 % Sequenzidentität zur *A. aeolicus* SQR.

Die SQR wurde in den Membranen von *A. aeolicus* durch denaturierende Polyacrylamid Gelelektrophorese (SDS-PAGE), Peptidmassen-Fingerabdruck-Massenspektrometrie (PMF-MS) und spektrophotometrische Aktivitätstests identifiziert, mit dem Detergenz Dodecyl-β-D-maltosid (DDM) solubilisiert und durch fünf aufeinander folgende chromatographische Schritte mit einer abschließenden Ausbeute von 32 % gereinigt. Das Protein hat einen berechneten isoelektrischen Punkt von 4,2 - 4,5, und durch Massenspektrometrie wurde für das Proteinmonomer eine Masse von 47.388 Da berechnet. Das Absorptionsspektrum im sichtbaren Wellenlängenbereich deutet auf ein fest gebundenes Flavin-Adenin-Dinukleotid (FAD) Molekül hin. Überraschenderweise wurde der Oligomerzustand der gereinigten SQR mit analytischer Ultrazentrifugation (AUC), Crosslinking und Einzelpartikelanalyse als Trimer bestimmt. Die Totalreflektionsröntgenfluoreszenz (TXRF) zeigte interessanterweise, dass die SQR eine höhere Anzahl an Schwefelatomen besitzt, als man von der Aminosäuresequenz her erwarten würde. Experimente mit hoch auflösender Massenspektrometrie deuten darauf hin, dass diese Schwefelatome in Ketten von mindestens vier Atomen angeordnet und kovalent an das Protein gebunden sind. Der gereinigte Proteinkomplex ist aktiv, zeigt eine mikromolare Affinität zu seinen Substraten Sulfid und Ubichinon und kann durch Chinon-Analoge kompetitiv inhibiert werden. Die größte Aktivität wurde bei der höchsten getesteten Temperatur von 80 °C gemessen. Die Aktivierungsenergie des Enzyms beträgt rund 14 kJ/(K·mol). Besonders bemerkenswert ist die Stabilität der isolierten SQR bei hoher Temperatur, denn selbst nach 24 Stunden bei 80 °C hat das Protein noch immer eine 50%ige Aktivität.

Die SQR wurde durch 3D Kristallisation und Röntgenbeugung strukturell untersucht. Kristalle konnten in hängenden und sitzenden Tropfen in Dampfdiffusion, im Mikroansatz unter Paraffinöl und in der so genannten Schwammphase (engl.: sponge phase) erhalten werden. Die erfolgreichste Technik war die Dampfdiffusion, durch die zwei verschiedene Kristallformen erhalten wurden, die als hexagonale und nadelförmige Kristalle beschrieben werden können und Röntgenstrahlen mit einer Auflösung bis zu mehr als 2 Å beugen. In beiden Kristallformen zeigt das Enzym die gleichen elektrophoretischen, spektroskopischen, massenspektrometrischen und enzymatischen Eigenschaften wie in Lösung. Obwohl beide mit den Präzipitationsmitteln Ammoniumsulfat und Polyethylenglykol (PEG) gezüchtet wurden, unterscheiden sie sich jedoch signifikant in den Beugungseigenschaften. Das Beugungsbild der hexagonalen Kristalle ist durch eine hohe Mosaizität gekennzeichnet und daher für die Datenauswertung unbrauchbar. Dahingegen konnten die Nadelkristalle zur Strukturlösung verwendet werden. Sie gehören mit einer Einheitszelle von 112 x 154 x 178 Å^3 der Raumgruppe $P2_12_12_1$ an, enthalten 6 Moleküle in der asymmetrischen Einheit und weisen einen Solvensgehalt von 53 % auf. Das Kristallgitter ist aus Molekülschichten und dünnen Detergenzschichten aufgebaut und imitiert in gewisser Weise die Anordnung der Proteinkomplexe über der zellulären Membran.

Die Struktur der SQR wurde mit einer Auflösung von 2,3 Å durch *de novo* Phasenbestimmung mit der Methode des Multiplen-Isomorphen-Ersatzes mit anomaler Streuung von Osmium- und Gold-Derivaten ermittelt (die MIRAS Methode). Das Modell wurde weitestgehend per Hand gebaut und bis zu einem R_{free} von 23,5 % (R_{work} 19,2 %) verfeinert. Kristalle der SQR wurden mit dem Substrat Decylubichinon und mit dem Inhibitor Aurachin C inkubiert und die Strukturen der Substrat- und Inhibitorkomplexe mit einer Auflösung von 2,0 Å bzw. 2,9 Å bestimmt. Sie zeigen im Vergleich zur Chinon-gebundenen Struktur keine Unterschiede in der Konformation der Polypeptidkette, aber sie erlauben die eindeutige Identifizierung der Chinon-Bindestelle, die so zum ersten Mal für SQRs bestimmt werden konnte.

Die sechs SQR Monomere in der asymmetrischen Einheit sind nahezu identisch. Alle enthalten die für FDRs typischen zwei FAD-bindenden Rossmann Domänen und eine C-terminale Domäne, die für jede FDR Unterfamilie spezifisch ist. Die C-terminale Domäne der SQR besteht aus zwei amphipatischen Helices gefolgt von einer Schleife, die aus 18 Aminosäuren besteht, und durch zwei Disulfidbrücken stabilisiert wird. Die Existenz der strukturell wichtigen Disulfidbrücken deutet auf eine Lokalisation des Proteins im Periplasma hin. Zudem tragen die Disulfide zusammen mit dem ausgeprägten Netzwerk an Wasserstoffbrückenbindungen zwischen den Untereinheiten zu der groβen „thermischen" Stabilität des Proteins bei.

Die sechs Monomere in der asymmetrischen Einheit lagern sich zu zwei homotrimeren Komplexen zusammen, jeder mit einer ausgeprägten elliptischen Form und einer starken Polarität. Die Seite, welche durch die beiden Rossmann Domänen gebildet wird, ist vorwiegend negativ geladen, während die andere Seite durch einen Überschuss an positiver Ladung und hydrophobe Stellen charakterisiert ist und deshalb die Interaktionsfläche mit der Membran bildet. Aus der Kristallstruktur kann gefolgert werden, dass sich die SQR symmetrisch mit Hilfe der zwei C-terminalen amphipatischen Helices bis zu einer Tiefe von 12 Å in die Lipiddoppelschicht einlagert. Folglich ist die SQR ein integrales monotopisches Membranprotein. Trp 377 und Trp 391 binden zwei Sulfatgruppen, welche chemisch analog zu den Phosphatgruppen von Lipiden sind, und könnten daher so genannte „Lipidklammern" (engl.: lipid clamps) darstellen, die eine hohe Affinität zu zwitterionischen Phospholipiden besitzen. Die Bindung von zwei weiteren Sulfaten zwischen benachbarten Monomeren deutet darauf hin, dass die Membraneinlagerung die Oligomerisierung stabilisiert. Diese Verknüpfung von Membraneinlagerung, Trimerisierung und enzymatischer Aktivität könnte sicherstellen, dass die SQR ausschließlich hydrophobe Chinone reduziert. Vielleicht wird dadurch auch die Abgabe des Polyschwefelproduktes direkt in die unpolare Umgebung der Lipiddoppelschicht erleichtert.

Auch eine ausführliche Beschreibung des aktiven Zentrums, das um die FAD Bindetasche lokalisiert ist, wird durch die Struktur ermöglicht. Wie die FCSD bindet auch die SQR FAD kovalent. Unerwarteterweise besteht die Verbindung nicht aus einer klassischen Thiol 8-α-S_γ-

Cysteinylbindung sondern aus einer indirekten 8-α-X-S_γ-Cys124-Bindung, wobei die Gruppe X aufgrund einer experimentellen anomalen Differenz-Elektronendichtekarte als Schwefelatom interpretiert wurde. Diese ungewöhnliche Art der Bindung könnte das Reduktionspotential des FAD beeinflussen, wichtig für die richtige Positionierung des Cofaktors im aktiven Zentrum sein oder sogar eine aktive Rolle in der Katalyse spielen.

Der Isoalloxazinring des FAD trennt die Bindetasche in zwei Hohlräume. Die Chinon-Bindestelle liegt auf der *si*-Seite des FAD. Diese Seite ist von der Membran über einen Kanal zugänglich, der von den oben beschriebenen amphipatischen Helices geformt wird. Die kürzeste Entfernung zwischen dem Chinon und dem FAD ist etwa 3.5 Å, was einen schnellen Elektronentransfer erlaubt. Der aromatische Ring des Substrats ist zwischen der Benzylgruppe von Phe 385 und der Seitenkette von Ile 346 gebunden, während Glu 318, Lys 382 und andere Wassermoleküle Wasserstoffbrückenbindungen mit dem Chinon bilden und in den Protonentransfer involviert sein könnten. Die Struktur der SQR Chinon-Bindestelle ist neu und zeigt keine Ähnlichkeit zu der von anderen FDRs. Zum Beispiel hat die chinon-bindende Proteindomäne der FCSD, dem nächsten strukturellen Homolog, eine völlig andere Architektur, welche an die Interaktion mit Cytochrom *c* angepasst ist, das anstelle von Chinon der Elektronenakzeptor dieses Proteins ist. Auch die Sulfid-Oxidationsstelle zeigt neue Eigenschaften und unterscheidet sich deutlich von der der FCSD, obwohl beide Proteine die gleiche reduktive Semireaktion katalysieren. Während sich in der FCSD die Sulfid-Oxidationsstelle auf der Oberfläche des Proteins befindet, ist diese in der SQR im Innern des Proteins auf der *re*-Seite des FAD verborgen, wo sie mit der umgebenden Flüssigkeit durch einen Kanal verbunden ist. Dieser besitzt ein positives elektrostatisches Oberflächenpotential, wird von der konservierten essentiellen Aminosäure Val 294 flankiert, und ist desweiteren mit Wassermolekülen, welche den Pfad des Substrates H_2S imitieren könnten, besetzt. Außerdem liegen an der Sulfidoxidationsstelle der SQR drei – und nicht zwei, wie in FCSD – konservierte Cysteine. Cys 124 bindet FAD wie oben beschrieben. Cys 156 und Cys 347 entsprechen dem aktiven Redoxpaar der FCSD, sind aber, anders als bei der FCSD, nicht durch eine Disulfidbrücke verbunden und liegen zu weit entfernt vom FAD, um ein für den Reaktionszyklus der FDR typisches kovalentes Intermediat zu bilden. Die Sulfid-Oxidationstasche in der Struktur ist unerwarteterweise nicht leer, sondern enthält eine kovalent gebundene Polyschwefelkette, die durch TXRF und Massenspektrometrie nachgewiesen wurde und welche vermutlich das Produkt der Reaktion darstellt. In den meisten Monomeren, ist die Kette an das S_γ Atom von Cys 156 gebunden, in Richtung Cys 347 orientiert, umgeben von hydrophoben Aminosäuren und in der Nähe des Proton Donor/Akzeptor Paares Tyr 161 – Glu 162 lokalisiert. In manchen Monomeren wird eine geschlossene ringförmige Elektronendichte gefunden, die sich mit einem Ring aus acht Schwefelatomen (S_8) erklären läßt. Die Kette in anderen Monomeren verlängert nicht die Seitenkette von Cys 156, sondern von Cys 347.

Die Struktur ermöglicht es, eine Hypothese für den katalytischen Mechanismus zu formulieren. Entsprechend dieser Hypothese folgt die SQR einer ungewöhnlichen Schwefelpolymerisationsreaktion, welche von allen bekannten FDRs abweicht. Im Allgemeinen wird das aktive Cystein/Cystin Redoxpaar oxidiert/reduziert und gibt die Elektronen über ein Cys-S_γ-C4A-FAD Ladungstransfer-Intermediat an FAD ab. Im Gegensatz dazu werden in dieser Arbeit zwei unterschiedliche, alternative Mechanismen für die SQR vorgeschlagen, welche im Folgenden als „C4A-S"- und „C8M"-Mechanismen bezeichnet werden. Der C4A-S Mechanismus postuliert die Bildung eines kovalenten Cys156-S_γ-S-C4A-FAD Zwischenprodukts, ähnlich zu dem allgemein angenommenen FDR Zwischenprodukt, aber mit einem zusätzlichen Schwefelatom, welches durch das Substrat Sulfid beigesteuert wird und die Protein-FAD Verbindung vermittelt. Der C4A-S Mechanismus würde geringe Konformationsänderungen während des Reaktionszyklus erfordern, und nur Cys 156 und Cys 347 wären beteiligt. Im Gegensatz dazu würde der C8M Mechanismus auch Cys 124 eine aktive Rolle zuordnen und würde die ungewöhnliche Bindung des FAD erklären. Dieser Mechanismus schlägt vor, dass Cys 124, durch Sulfid zu der nukleophilen Form Cys124-S_γ-S^- aktiviert, das Chinon-Methid-Tautomer von FAD in Position C8M angreift. Cys 156 und Cys 347 wären indessen nötig, um die länger werdenden Polyschwefelprodukte zu übernehmen und zu beherbergen. Obwohl diese Hypothese mit den strukturellen Beobachtungen vereinbar ist, hat der C8M Mechanismus auch Schwachstellen. Erstens spiegeln sich die erwarteten weitaus größeren Konformationsänderungen im Vergleich zum C4A-S Mechanismus in der Struktur nicht wider, da die niedrigen Temperaturfaktoren eine geringe Flexibilität der Polypeptidkette zeigen. Zweitens ist Cys 124 zwar tatsächlich essentiell für einige Arten von SQRs, jedoch nicht durchgehend konserviert.

Unabhängig vom Mechanismus wird der Abbruch der Polymerisierung vermutlich durch den limitierenden Platz im aktiven Zentrum ausgelöst. Das tatsächliche Produkt der SQR Reaktion ist unbekannt. Trotzdem zeigt die Struktur der *A. aeolicus* SQR, dass das katalytische Zentrum nur ein katalytisch produktives Milieu für Ketten aus bis zu 9-S-Atomen bietet, welche vermutlich abschließend zu einem S_8 Ring geschnitten werden. Diese Form ist sehr apolar und instabil in Wasser und könnte durch einen hydrophoben Kanal abgegeben werden, welcher sich in Richtung der in die Membran eingebetteten Region zieht. Auf diesem Weg würde sich der S_8 Ring direkt zur Membran bewegen und könnte abschließend in so genannten „Schwefelkügelchen" (engl.: „sulfur globules") eingelagert werden, welche der Zelle, sollte der Nährstoffgehalt in der Umgebung abfallen, als Energie Reserve dienen. In *A. aeolicus* befinden sich diese Kügelchen im Cytoplasma. Deshalb könnte die SQR, obwohl sie ein monotopisches und kein polytopisches Membranprotein ist, an der zellulären Aufnahme von Schwefel effizient teilnehmen.

Es sollte erwähnt werden, dass die strukturellen Eigenschaften und die hypothetischen Mechanismen, die hier für die SQR von *A. aeolicus* vorgestellt werden, sich nicht nur von anderen

FDR unterscheiden, sondern auch nicht unbedingt universell für alle SQRs gelten. Zum Beispiel zeigt die Struktur der SQR des Archaeons *Acidianus ambivalens*, welche seit kurzem auch zur Verfügung steht, signifikante Unterschiede sowohl bei den Schleifen, welche die Sulfid-Zugangskanäle flankieren, als in der FAD Bindung als auch in der Architektur der Sulfid-Oxidationsstelle. Diese Daten zeigen eine überraschende und bisher unerwartete Variabilität zwischen SQRs, die bisher als homolog angesehen wurden. Deswegen ist es möglich, dass die *A. ambivalens* SQR zumindest in Bezug auf die Sulfid-Oxidation, einem anderen katalytischen Mechanismus folgt als die *A. aeolicus* SQR. Basierend auf dem strukturellen Vergleich, der hohen Anzahl an SQR Sequenzen, die nun verfügbar sind, und anhand neuester funktioneller Charakterisierung von anderen Mitgliedern der SQR Familie, konnten neue Struktur-basierte SQR Sequenzmotive identifiziert werden. Daher wurde die Klassifizierung der SQRs wieder aufgegriffen und verbessert.

Diese Arbeit liefert nach mehr als 20 Jahren weltweiter Forschung an SQRs die erste vollständige Struktur einer SQR und eine neue strukturbasierte Klassifizierung für diese Proteinfamilie. Die strukturellen Daten liefern zum ersten Mal einen näheren Einblick in die faszinierende, für Energieerhaltungs- und Entgiftungsprozesse äußerst wichtige, aber komplexe SQR-Reaktion und eröffnen neue Möglichkeiten für die zukünftige Forschung. Einerseits sollte der Mechanismus der SQR von *A. aeolicus* weiter untersucht werden, um zwischen den beiden aktuell vorgeschlagenen Hypothesen zu unterscheiden. Andererseits sollten auch die Unterschiede innerhalb der SQR Familie näher untersucht werden. Dabei sollte ein besonderer Augenmerk auf die Organismen gelegt werden, welche gleichzeitig SQRs unterschiedlicher Art besitzen. Schließlich sollten mit Nachdruck eukaryotische SQRs untersucht werden. Diese spielen vermutlich auch im Menschen eine Rolle bei der Homöostase von Sulfid, welches ein Vermittler bei der Signalübertragung im sympathischen Nervensystem und ein Schlüsselmetabolit bei neurodegenerativen Krankheiten ist.

Zusammenfassung

Diese Arbeit stellt die erste vollständige Röntgenkristallstruktur des Membranproteins Sulfid:Chinon Oxidoreduktase (SQR) vor. Die Beschreibung der Struktur, wird durch die biochemische und funktionelle Charakterisierung des Enzyms ergänzt.

SQRs sind ubiquitäre Flavoprotein-Disulfid-Reduktasen (FDRs), die in allen Domänen des Lebens, darunter auch im Menschen, vertreten sind. Ihre physiologische Funktion reicht von der Sulfidentgiftung bis hin zur sulfidabhängigen Atmung und Photosynthese (in Archaea und Bakterien), zur Schwermetalltoleranz (in Hefen) und vermutlich zur sulfidabhängigen Signalübertragung (in höheren Eukaryoten). Bis heute war es schwierig, die Funktion der SQRs zu verstehen, da diese Familie eine geringe Sequenzidentität aufweist, nur wenig funktionell charakterisiert war und keine strukturellen Daten zur Verfügung standen.

Die SQR wurde in den Plasmamembranen des hyperthermophilen Bakteriums *Aquifex aeolicus* durch einen Peptidmassen Fingerabdruck (PMF) und einen spektrophotometrischen Aktivitätstest identifiziert. Das Protein wurde mit dem Detergenz Dodecyl-β-D-maltosid (DDM) solubilisiert und in aktiver Form bis zur Homogenität gereinigt. Jedes Monomer bindet ein FAD Molekül, welches der einzige Cofaktor ist. Die Struktur der SQR wurde in der „wie-gereinigten", der Substrat-gebundenen und der Inhibitor-gebundenen Form mit einer Auflösung von 2,3, 2,0 bzw. 2,9 Å bestimmt. Sie besteht aus zwei Rossmann-Domänen und einer Membrankontaktregion. Obwohl die Architektur des Monomers derjenigen von FDRs ähnelt, zeigt die Struktur Merkmale, die bisher noch nicht beobachtet wurden, jedoch große Auswirkungen auf den katalytischen Mechanismus der SQRs haben.

Die SQR von *A. aeolicus* liegt überraschenderweise im Kristall als Trimer vor, was durch analytische Ultrazentrifugation (AUC), Crosslinking und Einzelpartikelanalyse auch als native Form in Lösung bestätigt wurde. Das Trimer erzeugt eine passende Oberfläche für die Bindung an die Membran und stellt dadurch sicher, dass die SQR ausschließlich membranassoziierte Chinone reduziert. Die SQR sitzt als integrales monotopisches Membranprotein bis zu einer Tiefe von etwa 12 Å in der Membran. Die Interaktion wird durch zwei amphipatische Helices und zwei spezifische Lipid-Bindestellen vermittelt. Ein Kanal, der von der Membranbindedomäne zur *si*-Seite des FAD reicht, repräsentiert die Chinon-Bindestelle. Der Chinon-Ring ist zwischen den zwei konservierten Aminosäuren Phe 385 und Ile 346 gebunden und wird nach der Reduktion vermutlich von Glu 318, Lys 382 und/oder den in der Nähe lokalisierten Wassermolekülen protoniert. Die Sulfid-Polymerisierung hingegen geschieht auf der *re*-Seite des FAD. Hier sind die hoch konservierten Aminosäuren Cys 156 und Cys 347 kovalent an eine Polyschwefelkette gebunden, die in manchen Monomeren die Form eines S_8 Rings annimmt und vermutlich das Produkt der Reaktion ist.

Schließlich zeigt die Struktur eine ungewöhnliche kovalente Bindung der 8-Methylgruppe von FAD zu Cys 124, welche vermutlich eine Disulfidbrücke darstellt.

Der detaillierte Einblick in das Protein und alle unerwarteten strukturellen Beobachtungen, die hier in dieser Arbeit gezeigt werden, deuten auf einen katalytischen Mechanismus der SQRs hin, der sich signifikant zu dem der anderen FDRs unterscheidet. In Übereinstimmung mit den strukturellen und funktionellen Daten werden zwei Reaktionsmechanismen für SQRs vorgeschlagen. Beide erklären wie Sulfid und Chinon in das aktive Zentrum gelangen und binden, wie Elektronen vom Sulfid über FAD zum Chinon übertragen werden und wie das Polyschwefelprodukt, das an die Polypeptidkette gebunden ist, verlängert und schließlich abgegeben wird. Der Unterschied beider Hypothesen liegt in dem kovalenten Protein-FAD Intermediat, das sich während des Reaktionszyklus bildet und dessen Struktur experimentell noch nicht bestimmt wurde.

Interessanterweise zeigt die Struktur der SQR von *Acidianus ambivalens*, welche seit kurzem auch zur Verfügung steht, Unterschiede im aktiven Zentrum und in der FAD-Bindestelle zur SQR von *A. aeolicus*. Diese Variabilität der SQRs deutet darauf hin, dass nicht alle diese Enzyme dem gleichen katalytischen Mechanismus folgen, obwohl sie bisher als homolog betrachtet wurden. Folglich, wurden die verfügbaren aber widersprüchlichen sequenzbasierten Klassifizierungen der SQR Familie wieder aufgegriffen. Da die entscheidenden Sequenzabschnitte nun bekannt waren, erfolgte die Einteilung der Gruppen durch einen strukturbasierten Abgleich der steigenden Anzahl an verfügbaren Sequenzen und wurde mit den vorhandenen strukturellen und funktionellen Daten in Einklang gebracht.

Diese Arbeit liefert erstmals einen tieferen Einblick in den faszinierenden aber komplexen Reaktionen, die SQRs katalysieren, und bietet eine Grundlage für weitere genetische, biochemische und strukturelle Untersuchungen.

Detailed English Summary

Experimental determination of the structures and functions of macromolecules is of high biochemical, microbiological, pharmacological and medical interest. The most successful approach to study biological targets at a molecular level is X-ray diffraction of 3-D crystals, which provides high resolution structural information. This method is efficiently complemented by other biochemical, genetic or structural tools, such as mass spectrometry, analytical ultracentrifugation, mutagenesis, electron microscopy, and a wide variety of spectroscopic techniques.

Among all biologically relevant macromolecules, membrane proteins are particularly challenging targets for structural investigation. It is difficult to express them in high amount and they are unstable, soluble only in solutions containing mild amphipathic detergents and not prone to form well-ordered crystals. Nonetheless, they are an exceptionally interesting group of proteins, as they comprise 20 – 25 % of total proteomes and 60 % of the current drug targets.

This work describes the structural and functional properties of a membrane enzyme, the sulfide:quinone oxidoreductase (SQR) isolated from the native cells of the hyperthermophilic bacterium *Aquifex aeolicus*.

SQRs belong to the flavoprotein disulfide reductase (FDR) superfamily, together with well-characterized soluble proteins like glutathione reductase, lipoamide dehydrogenase, thioredoxin reductase and trypanothione reductase. At the origin of life, SQRs were essential for catalyzing sulfide-dependent quinone reduction, an important reaction that allowed organisms to thrive in the sulfidic anoxic conditions of early oceanic environments. They were conserved throughout evolution and are currently found in all domains of life, including humans, although being absent in plants. They play a physiological role in sulfide detoxification, in sulfide-dependent respiration and photosynthesis (in bacteria) and in heavy metal tolerance (in yeast). In eukaryotes, they may be involved in sulfide-dependent signalling processes, since sulfide is endogenously produced and acts as a "gasotransmitter".

At the beginning of this work, no structure was available for any SQR and only few homologues had been functionally characterized. Furthermore, sequence conservation among SQRs is very poor, so that their classification was partial and contradictory because the determination of the functionally relevant sequence fingerprints was uncertain. Until now, their catalytic mechanism had been hypothesised on the basis of the structure of *Allochromatium vinosum* flavocytochrome *c*:sulfide dehydrogenase (FCSD), a protein of the FDR family with only 24 % sequence identity to *A. aeolicus* SQR.

A. aeolicus SQR was identified in total membrane preparations by denaturing polyacrylamide electrophoresis (SDS-PAGE), peptide mass fingerprint mass spectrometry (PMF-MS) and a spectrophotometric activity assay. It was solubilized in the detergent dodecyl-β-D-maltoside (DDM)

and purified by conventional chromatographic methods in 5 steps, with a final yield of 32 %. The protein has a calculated isoelectric point of 4.2 – 4.5, a monomeric molecular mass of 47,388 Da, as judged by full-length matrix-assisted laser-desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS) and harbours one tightly-bound flavin adenine dinucleotide molecule (FAD), as determined from the absorption spectrum in the visible wavelength range. Surprisingly, purified SQR was shown to be trimeric in solution by density-matching analytical ultracentrifugation (AUC), crosslinking and single particle electron microscopy. Interestingly, total X-ray fluorescence (TXRF) showed that SQR binds a higher number of sulfur atoms than expected from the protein sequence. High-resolution mass spectrometric experiments suggest that such sulfurs are organised in chains of at least 4 atoms and are covalently attached to the protein. Finally, the protein complex is functionally active, possesses micromolar affinity to its substrates, sulfide and ubiquinone, and can be competitively inhibited by quinone analogues. The maximal activity was detected at 80 °C, the highest temperature tested. The activation energy of the enzyme is of about 14 kJ/(K·mol). Particularly remarkable is the stability of the purified complex over time at high temperature. Incubated at 80 °C, it retains 50 % activity after 24 h.

The structure of SQR was studied by 3-D crystallization and X-ray diffraction. Crystals were obtained by hanging- and sitting-drop vapour diffusion, by microbatch under paraffin oil and in lipidic sponge phase. The hanging-drop vapour diffusion technique was the most successful, yielding two crystal forms, hereafter referred to as the hexagonal and the needle crystals, which both diffract X-rays at a resolution of better than 2 Å. The protein in both the hexagonal and the needle crystals show the same electrophoretic, spectroscopic, mass spectrometric and enzymatic properties as the protein in solution and both crystal forms were obtained using ammonium sulfate and PEG as precipitating agents. However, their X-ray diffraction properties differ significantly. The diffraction of the hexagonal crystals is characterized by high mosaicity and is consequently not suitable for data processing, while the needle crystals could be used for structure determination. They belong to the space group $P2_12_12_1$, with unit cell dimensions of approximately 112 x 154 x 178 $Å^3$. They possess 6 molecules in the asymmetric unit and a solvent content of 53 %. The molecules are arranged over layers of detergent-filled regions in the crystal lattice, mimicking the arrangement of the protein complexes over the cellular membranes.

The SQR structure was solved to 2.3 Å resolution by *de novo* phasing using multiple isomorphous replacement and anomalous scattering of osmium and gold derivatives (the MIRAS method). Model building was done mostly by hand and the initial model was then refined to a final R_{free} of 23.5 % (R_{work} is 19.2 %). Soaking with the substrate decylubiquinone and with the inhibitor aurachin C was also successful and the corresponding structures were solved by rigid body refinement at 2.0 Å and 2.9 Å resolution, respectively. They do not show significant movements of

the protein chain, but they do allow unequivocal identification of the SQR quinone binding site, which is described for the first time in this work.

The six SQR monomers in the asymmetric unit are nearly identical. They are each composed of two FAD-binding Rossmann-fold domains typical of FDRs and a C-terminal domain distinctive of each FDR subfamily. The latter protein region consists of an amphipathic helix-turn-helix motif followed by an 18-amino-acid long loop stabilized by two disulfide bridges. The presence of these structurally important disulfide bonds suggests that the protein is localized in a periplasmic oxidising environment. These disulfides, together with the tight hydrogen-bonding network that mediates oligomeric contacts and the large solvent-filled cavities which confer resilience to SQR, explain the protein's high thermal stability.

The six monomers in the asymmetric unit are associated in two homotrimeric complexes, that have a pronounced elliptical shape and a strong polarity. One side, formed by the two Rossmann-fold domains, is predominantly negative, while the other side is characterized by clusters of positive charges and hydrophobic patches and mediates the interaction with the membrane. From the crystal structure, it can be concluded that SQR inserts symmetrically to a depth of about 12 Å into one leaflet of the bilayer through the C-terminal amphipathic helices and is therefore classified as an integral monotopic membrane protein. Two sulfate ions – chemically analogous to lipidic phosphate head groups – are bound to Trp 377 and Trp 391, which might thus represent two clamps with affinity for zwitterionic phospholipids. Furthermore, additional sulfates bind between neighbouring monomers suggesting that membrane insertion stabilizes oligomerization and, *vice versa*, that the trimer creates an appropriate surface for binding lipids. The coupling between membrane insertion, trimerization and enzymatic activity might ensure that SQR exclusively reduces hydrophobic quinones and, perhaps, might also facilitate the release of the water-labile polysulfur product directly into the apolar lipid bilayer environment.

The structure also allows a detailed description of the active sites, located around the FAD binding pocket. SQR binds FAD covalently, like FCSD. Unexpectedly though, the connection does not consist of a classic thiol 8-α-S_γ-cysteinyl bond, but of an indirect 8-α-X-S_γ-Cys124 bond, where group X was interpreted as a sulfur atom on the basis of an experimental anomalous difference electron density map. This unusual binding mode may affect the reduction potential of FAD, may be important for a correct localization of the cofactor in the active site or may even play an active role in catalysis.

The isoalloxazine ring of FAD divides the binding pocket into two cavities. The quinone binding site is on the *si*-face of FAD and is accessible from the membrane through a channel formed between the amphipathic helices described above. The shortest distance between quinone and FAD is about 3.5 Å, which permits fast electron transfer. The aromatic ring is bound between the benzene moiety of Phe 385 and the side chain of Ile 346, while Glu 318, Lys 382 and other water

molecules form hydrogen bonds with the quinone and might be involved in proton transfer. The fold of the SQR quinone binding site is new and has no similarity with any domain of other FDRs. In fact, the corresponding region in FCSD, the closest SQR structural homologue, has a completely different architecture, suited for interaction with cytochrome *c*, the electron acceptor replacing quinone in that protein. Moreover, also the SQR sulfide oxidation site shows surprising novel features and differs significantly from that of FCSD, despite both proteins catalysing the same reductive semireaction. While in FCSD the sulfide oxidation site is exposed to the bulk solvent, in SQR it is buried in the protein core on the *re*-face of FAD. It is accessible through a channel characterized by a positive electrostatic surface potential, flanked by the conserved essential residue Val 294 and occupied in the structure by four solvent molecules that may mimic the trace of the substrate, H_2S, approaching the active pocket. Furthermore, three – and not two, like in FCSD – conserved cysteines face the sulfide oxidation site. Cys 124 forms the unprecedented indirect bond with FAD, as described above. Instead, Cys 156 and Cys 347 correspond to the redox active couple of FCSD, but in contrast to that protein, they are not connected by a disulfide bridge and they are too far from FAD to form the covalent intermediate typical of the FDR reaction cycle. In fact, the sulfide oxidation pocket is unexpectedly not empty in the structure but contains the covalently attached polysulfur chain identified by TXRF and mass spectrometry and presumably corresponding to the product of the reaction. In most of the monomers, the chain extends from the S_γ atom of Cys 156, is oriented towards Cys 347, is surrounded by hydrophobic amino acids and is near the proton donor/acceptor couple Tyr 161 – Glu 162. In some other monomers, it takes the form of a ring of 8 sulfur atoms (S_8). Finally, in few monomers, it does not extend from Cys 156 but from Cys 347.

The structure allows the formulation of detailed hypotheses on the SQR catalytic mechanism and suggests that the enzyme performs an unusual sulfur polymerization reaction rather different from that of all known FDRs. In general in FDRs, a redox active cystine/cysteine couple is reduced/oxidised and transfers electrons to FAD via a Cys-S_γ-C4A-FAD charge-transfer intermediate. Instead, two alternative mechanisms are proposed here for SQR and are referred to as the "C4A-S" and the "C8M" mechanism, respectively. The C4A-S mechanism postulates the formation of a Cys156-S_γ-S-C4A-FAD covalent intermediate, similar to the common FDR intermediate, but with an additional sulfur atom contributed by the substrate sulfide to mediate the protein-FAD linkage. The C4A-S mechanism would require moderate conformational changes during the reaction cycle and would directly involve only Cys 156 and Cys 347. On the contrary, the C8M mechanism would attribute an active role also to Cys 124 and would explain its unusual linkage to FAD. This mechanism postulates that Cys 124, activated by sulfide to the nucleophilic form Cys124-S_γ-S^-, would then attack the quinone methide tautomer of FAD in position C8M. Cys 156 and Cys 347 would then be necessary to take over and harbour the elongating polysulfur

product. Although compatible with all structural observations, the C8M mechanism presents two inconsistencies. First, it would require more pronounced conformational changes than the C4A-S mechanism, though the structure (low B factors) does not suggest flexibility of any polypeptide segment. Second, it is limited by the fact that Cys 124, while essential for some SQR types, is not strictly conserved.

Independent of the mechanism, the termination of the elongation is presumably triggered by the limited space of the active site. The true product of the SQR reaction is unknown. Nonetheless, the *A. aeolicus* SQR structure shows that the active site can only provide a catalytically productive environment for up to a 9-S-atom chain which is finally most likely cleaved to an S_8 ring. The latter species is very apolar and water-labile and may be released through a hydrophobic channel conducting towards the membrane embedded region. Using this pathway, S_8 would move directly towards the membrane interior and could finally be incorporated into so called "sulfur globules", particles that the cells may use as reserve when energetic supplies become short. In *A. aeolicus* these particles are accumulated in the cytoplasm so that – while not spanning the bilayer – SQR would participate in cellular uptake of sulfur in an efficient manner.

It must be said that the structural features and the mechanistic hypotheses presented for *A. aeolicus* SQR do not only differ from the other FDRs, but may also not be universal for all SQRs. In particular, the structure of the SQR from the archaeon *Acidianus ambivalens*, which also became available recently, shows pronounced differences in the loops that flank the sulfide access channel, in the binding mode of FAD and in the architecture of the sulfide oxidation site. These data show surprising and previously unexpected variability among SQRs until now considered homologous. Therefore, it is possible that *A. ambivalens* SQR follows a catalytic mechanism different from that of *A. aeolicus* SQR, at least with respect to sulfide oxidation. Based on this structural comparison, on the large number of SQR sequences now available and on recent functional characterization of other members of the SQR family, new structure-based SQR sequence fingerprints have been identified. Consequently, the SQR classification has been revised.

In conclusion, after more than 20 years of worldwide research, this work contributes the first complete structural description of an SQR and a new structure-based classification for this protein family. The structural data offer for the first time a detailed look into the intriguing but complicated SQR reaction, crucial for energy conservation and detoxification processes and open up new possibilities for future research. On one hand, the mechanism of *A. aeolicus* SQR must be studied further in order to discriminate among the currently proposed catalytic hypotheses. On the other hand, the variety in the SQR family must be further investigated, with particular emphasis on those organisms which possess SQRs of different types. Finally, a high priority is the study of eukaryotic SQRs, potentially important in humans for the homeostasis of sulfide, a mediator of sympathetic neurotransmission and a key metabolite in neurodegenerative diseases.

Abstract

This work presents the first complete structure of the membrane protein sulfide:quinone oxidoreductase (SQR), obtained by X-ray crystallography. Its description is complemented by the results of biochemical and functional experiments.

SQRs are ubiquitous flavoprotein disulfide reductases (FDRs), present in all domains of life, including in humans. Their physiological role extends from sulfide detoxification to sulfide-dependent respiration and photosynthesis (in archaea and bacteria), to heavy metal tolerance (in yeast) and possibly to sulfide signalling (in higher eukaryotes). Until now understanding the function of SQRs was difficult because of the poor level of sequence conservation in this enzyme family, the limited functional characterization available and the absence of any structural data.

SQR was identified in the native membranes of the hyperthermophilic bacterium *Aquifex aeolicus* by peptide mass fingerprinting (PMF) and by a spectrophotometric activity assay. The protein was solubilized in the detergent dodecyl-β-D-maltoside (DDM) and purified to homogeneity in a functionally active state. It binds one FAD molecule per protein monomer and FAD is its only cofactor. Its structure was determined in the "as-purified", substrate-bound and inhibitor-bound forms at resolutions of 2.3, 2.0 and 2.9 Å, respectively. It is composed of two Rossmann-fold domains and of one membrane-attachment region. Despite the overall monomeric architecture being similar to that of FDRs, the structure reveals properties that had not been observed in FDRs until now and that have strong implications for the SQR catalytic mechanism.

Surprisingly, *A. aeolicus* SQR is trimeric in the crystal structure and in solution, as determined by density-matched analytical ultracentrifugation, cross-linking and single particle electron microscopy. The trimer creates an appropriate surface for binding lipids and thus ensures that SQR exclusively reduces hydrophobic quinones. SQR inserts to a depth of about 12 Å into the membrane as an integral monotopic membrane protein. The interaction is mediated by an amphipathic helix-turn-helix tripodal motif and two lipid clamps. A channel in the membrane-binding domain extends towards the *si*-side of FAD and represents the quinone-binding site. The quinone ring is sandwiched between the conserved amino acids Phe 385 and Ile 346 and is possibly protonated upon reduction via Glu 318, Lys 382 and/or neighboring solvent molecules. Sulfide polymerization occurs on the *re*-side of FAD, where the highly conserved Cys 156 and Cys 347 appear to be covalently bound to the putative product of the reaction, a polysulfur chain which takes the form of an S_8 ring in some monomers. Finally, the structure shows that FAD is covalently connected to the protein in an unprecedented way, via a putative disulfide bridge between the 8-methyl group of the isoalloxazine moiety and Cys 124.

The high resolution insight into the protein and all unexpected structural observations presented in this work suggest that the catalytic mechanism of SQRs is significantly different from that of FDRs.

In agreement with the structural and functional data, two reaction schemes are proposed for *A. aeolicus* SQR. They both provide a detailed description of how sulfide and quinones reach and bind the active site, how electrons are transferred from sulfide to quinone via FAD and how the elongating polysulfur product is attached to the polypeptide and is finally released. The two hypotheses differ in defining the structure of the covalent protein-FAD intermediate that forms during the reaction cycle and whose identity still remains experimentally undetermined.

Remarkably, the structure of the active site and the FAD-binding mode of *A. aeolicus* SQR are not conserved in another SQR structure which also became available recently, that of the archaeon *Acidianus ambivalens*. The variability in SQRs suggests that not all of these enzymes follow the same catalytic mechanism, despite having been considered homologous. Consequently, the currently available but contradictory sequence-based classifications of the SQR family were revised. A structure-based alignment calculated on the increasing number of available sequences allowed to define new SQR groups and their characteristic sequence fingerprints in agreement with the reported structural and functional data.

In conclusion, the results obtained in this work offer for the first time a detailed look into the intriguing but complicated reactions catalysed by SQRs and provide a stimulus for further genetic, biochemical and structural investigation.

Abbreviations

The abbreviations used in this work are listed in the following table.

Table of abbreviations.

1.	MEASURES AND UNITS DERIVED FROM OR NOT BELONGING TO THE INTERNATIONAL SYSTEM.
Å	angstrom
°C	degrees celsius
cal	calories
Da	dalton
ε	extinction coefficient
eV	electron volt
h	hour
K_M	Michaelis-Menten constant
L	liter
λ	absorption wavelength
M	molar
min	minute
Pa	pascal
ppm	parts per million
rad	radian
rpm	rotations per minute
S	svedberg
U	(enzymatic) unit
V (as unit)	volt
V (as measure)	volume
Vpp	peak power point voltage (also peak-to-peak voltage)
v/v	volume/volume
w/v	weight/volume
w/w	weight/weight

2.	BIOMOLECULES AND NON-CONVENTIONAL CHEMICALS.
ACN	acetonitrile
AHC	ammonium hydrogen carbonate
BCA	bicinchonin acid
BisTris	1,3-bis(tris(hydroxymethyl)methylamino)propane
CHCA	α-cyano-4-hydroxy-cinnamic acid
ddH_2O	bidistilled water (Millipore)
DDM (also LM)	n-dodecyl-β-D-maltoside
EDTA	ethylendiaminetetracetic acid
FAD	flavin adenine dinucleotide (oxidised)
$FADH_2$	flavin adenine dinucleotide (reduced)

FCSD	flavocytochrome c:sulfide dehydrogenase
FDR	flavoprotein disulfide reductase
H_2S (or HS^- or S^{2-})	sulfide
LM	see DDM
βME	2-mercapto-ethanol
MES	2-(N-morpholino)-ethanesulfonic acid
MPD	2-methyl-2,4-pentanediol
PEG	polyethylene glycol
sDHB	super-DHB, mixture of 2,5-dihydroxybenzoic acid and 5-methoxysalicylic acid
SDS	sodium dodecyl sulfate
SQR	sulfide:quinone oxidoreductase
SQRDL	sulfide:quinone oxidoreductase dehydrogenase-like
TEMED	N,N,N',N'-tetramethylethylenediamine
TFA	trifluoroacetic acid
Tris	Tris-hydroxymethyl-aminomethane
UQ	ubiquinone
UQH_2	ubiquinol
Zwittergent	N-alkyl-N,N-dimethyl-3-amino-1-alkane sulfonate

3.   TECHNIQUES AND INSTRUMENTATION.

AUC	analytical ultracentrifugation
BN	blue-native
CN	clear-native
HPLC	high performance (pressure) liquid chromatography
EM	electron microscopy
ESI	electrospray ionisation
IEF	isoelectric focusing
LC	liquid chromatography
LILBID	laser-induced liquid-beam ionization / desorption
MALDI	matrix-assisted laser desorption ionisation
MIRAS	multiple isomorphous replacement with anomalous scattering
MS	mass spectroscopy
NCS	non-crystallographic symmetry
PAGE	polyacrylamide gel electrophoresis
PMF	peptide mass fingerprint
SEC	size-exclusion chromatography
SMART	simple modular architecture research tool
TLS	translation, libration, screw rotation
TOF	time of flight
TXRF	total reflection X-ray fluorescence
XRD	X-ray diffraction

4.   DATABASES AND SOFTWARE.

3D-PSSM	Three-Dimensional Position-Specific Scoring Matrix
ARP/wARP	Automated Refinement Procedure
BLAST	Basic Local Alignment Search Algorithm
CCP4	Collaborative Computational Project number 4
CNS	Crystallography and NMR System
DM	Density Modification
EBI	European Bioinformatics Institute
FASTA	FAST-All
GRASP	Graphical Representation and Analysis of Surface Properties
NCBI	National Centre for Biotechnology Information
PDB	Protein Data Bank
SCOP	Structural Classification of Proteins
SHARP	Statistical Heavy-Atom Refinement and Phasing
T-COFFEE	Tree-based Consistency Objective Function For AlignmEnt Evaluation
TMHMM	Transmembrane Hidden Markov Model

5.   GENERAL ABBREVIATIONS.

2-D	bi-dimensional
3-D	tri-dimensional
a_W	water activity
CCD	charged-coupled device
CMC	critical micelle concentration
conc	concentration
ED	electron density
e.g.	exempli gratia (lat., engl.: for example)
et al.	et alii (lat., engl.: and others)
eq.	equation
ESRF	European Synchrotron Radiation Facility
i.e.	in exemplum (lat., engl.: for example)
LD	lethal dose
(f)M	(fluorescent) protein molecular weight marker
MMP	monotopic membrane protein
MW	molecular weight
p.a.	per analysis
pI	isoelectric point
RMSD	root mean square deviation
ROI	region of interest
SLS	Swiss Light Source
t_R	retention time
UV/Vis or Vis/UV	ultraviolet/visible

Glossary

amphipathic sequence	polypeptide sequence folding as a helix with one polar and one apolar face
assimilatory sulfur metabolism	metabolic pathway involving sulfur compounds and used to produce organic S-containing biomolecules (i.e. the amino acid cysteine and methionine, coenzyme A, biotin, lipoic acid, molybdopterin, thionucleotides and thiamine)
chemolithoautotrophic organism	organism using only inorganic carbon sources (i.e. CO_2) for biosynthesis and inorganic chemical energy sources (i.e. H_2, NH_3, NO_2^-, Fe^{2+}, S^{2-}, $S_2O_3^{2-}$ …)
cyclooctasulfur	the most stable form of zero valent sulfur, taking the structure of an orthorhombic S_8 ring; it is a very hydrophobic molecule, soluble in organic solvents but not in water and it has aromatic features because one electron pair of each sulfur atom is delocalised across the ring
dissimilatory sulfur metabolism	metabolic pathway involving sulfur compounds as electron donors or acceptors and used to produce cellular energy
flavoprotein	protein containing an isoalloxazine chromophore
gasotransmitter	small molecule present in a gaseous state at standard temperature and pressure and acting as a tissue hormone
hyperthermophilic organism	organism thriving at temperatures higher than 80 °C
isoalloxazine	tricyclic heteronuclear organic molecule derived from pteridine and composed of a pyrimidine-2,4-dione, a pyrazine and a dimethylbenzyl ring
integral monotopic membrane protein	membrane protein constitutively bound to one leaflet of biological membranes, but not crossing the lipidic bilayer
protein domain	an independently folding part of a protein with defined secondary structure
sulfide	the most reduced form of sulfur, present as a gas at standard temperature and pressure in the non-ionised form H_2S, or as monovalent (HS^-) or divalent anions (S^{2-}) in solution, with pK_A values of 6.89 and > 14, respectively

1. Introduction

1.1. Structure and function of proteins

Proteins were recognised to be essential macromolecules almost two centuries ago, when G. J. Mulders and J. J. Berzelius described in their correspondence the existence of "animal substances" formed by a large number of carbon, hydrogen, oxygen and nitrogen atoms and containing traces of sulfur and phosphorus (Whitford, 2005). In 1934, it became clear that proteins have a regular structure when J.D. Bernal and D. Crowfoot showed that crystals of pepsin can diffract X-rays (Bernal and Crowfoot, 1934; Perutz, 1992). It was not earlier than 1958, though, that the first model of the structural architecture of a protein was obtained, thanks to the work of J. C. Kendrew on myoglobin (Kendrew *et al.*, 1958). As of February 23rd, 2009, the latest SCOP database release 1.75 (Murzin *et al.*, 1995; Andreeva *et al.*, 2008) showed that 38,221 unique entries of protein coordinates were available on the Protein Data Bank (PDB)[1]. They can be classified in 1,195 folds, 1,962 superfamilies, 3,902 families and 110,800 domains.

As John Kendrew and coworkers wrote about the structure of myoglobin, "complexity and lack of symmetry" are certainly remarkable in protein structures (Kendrew *et al.*, 1958). These features reflect though the requirements for proteins to be endowed with diverse functions (Branden and Tooze, 1998). There is a strict interdependence between the amino acid sequence of a protein, the structures it can adopt and the functions it performs. While it has to be considered that evolution acts on genotypes, in other words on the protein sequences, natural selection actually occurs on phenotypes, that is to say on the protein function. Because of this discrepancy, the paradigm "similar sequence – similar structure – similar function" does not universally hold (Martin *et al.*, 1998) (Table 1.1).

In general, homologous proteins[2] tend to develop similar folds and functions but exceptions are numerous. For instance, human α-lactalbumin and chicken egg-white lysozyme are homologous, sharing 40 % sequence identity and the same structural fold, but lactalbumin does not possess the hydrolase activity of lysozyme (Kumagai *et al.*, 1992). Even 100 % identical protein sequences can possess different functions. This extreme is the case of "moonlighting proteins" (Jeffery, 1999) and of "promiscuous enzymes" (Khersonsky *et al.*, 2006). In addition, polypeptides with identical amino acid sequences may adopt different tertiary folds and even different secondary structures, as is the case for the so called "chameleon sequences" (Minor

---

[1] The total number of PDB entries was 55,842 on Feb 23rd, 2009 and it has increased to 62,634 until Jan 15th, 2010.

[2] Two proteins are defined as homologous if they have a common ancestor (Whitford, 2005).

Table 1.1: Sequence-structure-function relationships[a].

evolution	sequence	structure	function	examples
divergent	+	+	+	general paradigm
	+	+	−	lactalbumin *vs* lysozyme, moonlighting proteins, promiscuous enzymes
	+	−	+/−	chameleon sequences
convergent	−	+	+	*V. stercoraria vs P. marinus* haemoglobin
	−	+	−	TIM barrels
	−	−	+	serine proteases
	−	−	−	general paradigm

[a] + indicates similarity, − indicates difference.

and Kim, 1996) (Figure 1.1).

The examples mentioned above represent cases of divergent evolution of similar protein sequences. The opposite case of non-homologous sequences evolving in a convergent manner has also been widely reported. For example, TIM-barrel proteins – named after the triosephosphateisomerase (Banner *et al.*, 1975) – are enzymes whose sequences are strongly divergent, such that it is unlikely that they descend all from the same ancestor. They also catalyse different reactions with representatives in almost 30 different enzyme classes. Nonetheless, they all present the same fold, an eight-stranded α/β-barrel, which was probably selected during evolution because it constitutes a favourable structural solution in forming a closed stable β-barrel (Lesk *et al.*, 1989). A different case of convergent evolution is represented by serine proteases (i.e. trypsin and subtilisin) that possess significantly different sequences and structures but catalyze their enzymatic reaction following a very similar mechanism (Robertus *et al.*, 1972). Last but not least, proteins may also substantially differ in their primary structure but have both a similar fold and a similar function, like *Vitreoscilla stercoraria* and *Perkinsus marinus* haemoglobin that share only 8% sequence identity.

In the absence of a general paradigm that relates sequences to structures and to functions and given the complication of macromolecular folding processes, accurate *de novo* structure prediction is difficult and

Figure 1.1: Chameleon sequences. Chameleon sequences represent one of the many exceptions to the paradigm "similar sequence – similar structure – similar function". The Figure shows a ribbon diagram in blue of the structure of the IG-binding domain of protein G (GB1). The same 11-amino-acid sequence AWTVEKAFKTF (light blue) was alternatively substituted in two different segments of the native GB1 sequence. Depending on the position, it adopts two different secondary structure folds. The Figure is reproduced from Minor and Kim, 1996.

unreliable. Therefore, experimental determination of the structures and functions of biological macromolecules remains necessary. Ultimately, the description of structural-functional relationships is relevant to many scientific fields. Combined structural and functional data contribute to explaining the dynamics of biological and cellular processes. Additionally, they reveal a physico-chemical description of natural reactions providing ideas for reproducing them artificially. Finally, they are highly valuable from a medical and pharmacological perspective, because they provide the basis for the study of pathological states at the molecular level, thereby opening up possibilities for rational drug development.

1.1.1. Structure and function of enzymes

Enzymes (from Greek, *en* + *zyme*, "in yeast") are particularly attractive targets in the study of structural-functional relationships because specific structural features determine their high specificity in substrate binding and their high efficiency in catalysing chemical reactions otherwise impossible at standard temperature and pressure (Fersht, 1999). The general shape of the active site, the physico-chemical properties of the catalytic residues, their spatial localization and in some cases the presence of cofactors offer an optimal environment for reactions to occur with lowered activation energy and therefore with higher probability and speed. Nonetheless, the number of catalytically important amino acids in one enzyme is very low with respect to the total size of the protein. Therefore, enzymes, more than other proteins, are likely to represent exceptions to the paradigm "similar sequence – similar structure – similar function" (Martin *et al.*, 1998). More often, enzymes respect a different paradigm "similar cofactor/ligand – similar structure – similar function" (Martin *et al.*, 1998). For instance, the three-layer αβα Rossmann fold (Rossmann *et al.*, 1974) is a widely preferred structural motif in proteins binding nucleotides (NADH, NADPH, flavins) (Martin *et al.*, 1998). However, this paradigm is by far also not universal (see following Chapter 1.1.2).

Therefore, structural and functional investigation on enzymes is particularly necessary for a detailed understanding of their catalytic cycles. Ideally, a comprehensive view would be obtained from the structures of the enzyme and of its complexes with substrates, intermediates, products and inhibitors.

Systematic scientific approaches for characterizing enzymes can be said to have started in 1897 when H. and E. Buchner discovered that cell-free yeast extracts could transform sugar into alcohol (Fersht, 1999). Structural investigation was opened in the late 1920s by the successful crystallization of urease by J.B. Sumner and of pepsin by J.H. Northrop. The first three-dimensional structure of an enzyme, lysozyme, was solved in 1965 by D. C. Philips and coworkers (Blake *et al.*, 1965). Starting from the years 1999 – 2000, crystal

structures beyond 3 Å resolution became available for ribosomes, the most complex protein-RNA catalytic assemblies, thereby achieving a milestone in the field of macromolecular crystallography (Ramakrishnan and Moore, 2001). Nonetheless, despite the tremendous developments of X-ray crystallography and the increasing number of available structural as well as enzymological data, further and deeper investigation is required because exciting and often unpredictable properties of enzyme functions are still continuously being revealed (De Colibus and Mattevi, 2006).

1.1.2. Structure and function of flavoenzymes

Flavoenzymes are enzymes containing a heterocyclic isoalloxazine chromophore (Figure 1.2), a remarkably versatile and still poorly understood molecule in terms of chemical reactivity and mechanistic properties.

Flavoenzymes can catalyse a wide variety of reactions, including dehydrogenation, one- and two-electron transfer reactions, light emission, oxygen activation and hydroxylation (Fraaije and Mattevi, 2000).
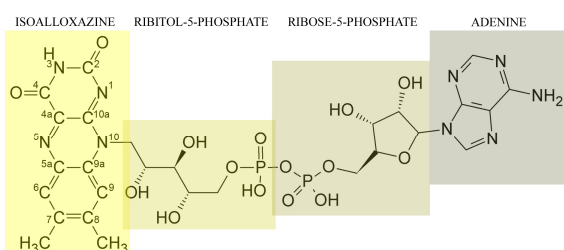
Figure 1.2: The structure of flavomolecules. The representation shows the oxidised form of flavin adenine dinucleotide (FAD). It is composed of the isoalloxazine group, a triciclic aromatic ring (light yellow), the ribitol-5-phosphate group (dark yellow), the ribose-5-phosphate group (light brown) and the adenine group (grey). Other flavomolecules of biological relevance are riboflavin, composed of the isoalloxazine and the ribitol group, and flavin adenine mononucleotide (FMN), consisting of the isoalloxazine and the ribitol-5-phosphate group.

They display a variety of folds, including the $(\alpha/\beta)_8$ barrel fold, the p-hydroxybenzoate hydroxylase (PHBH) fold, the flavodoxin-like fold, two- or three-$(\alpha+\beta)$ domains or other dinucleotide-binding multidomains (Fraaije and Mattevi, 2000) (Figure 1.3).

Conserved features in the cofactor binding pocket of different flavoenzymes are a positive charge or dipole near the N^1-C^2=O^2 flavin region and a hydrogen-bond donor near the N^5 isoalloxazine atom (Fraaije and Mattevi, 2000). However, some flavoenzymes bind their cofactor in a covalent manner, others not. In addition, covalent binding may occur in different positions of the isoalloxazine group (Heuts *et al.*, 2009).

Concerning substrate binding and substrate-flavin interactions some features are conserved among flavoenzymes. The location of the atom attacking the flavin is typically located 3.5 Å from the N^5 atom, forming an angle of 96 – 117° with atoms N^5 and N^{10}. In addition, substrate binding to flavoenzymes quite commonly occurs through a channel and the active site is often shielded from the bulk solvent (Fraaije and Mattevi, 2000). However, there are exceptions to this rule and the substrates of flavoenzymes are also very different in chemistry and size, requiring completely different architectures of the binding pockets.
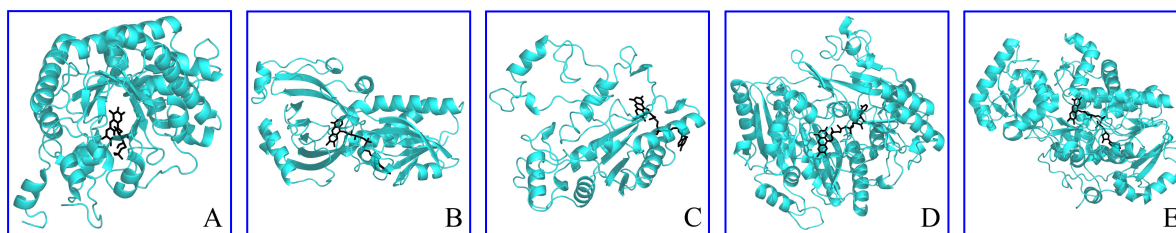
Figure 1.3: Structural folds of flavoenzymes. The proteins are shown as light blue cartoon representations. The flavin ligand (FMN or FAD), represented in black sticks, is approximately oriented in the same direction in all panels. Old yellow enzyme (PDB id.: 1OYB) possesses the $(\alpha/\beta)_8$ barrel fold (A), D amino-acid oxidase (1DDO) the PHBH fold (B), NAD(P)H quinone reductase (1QRD) the flavodoxin-like fold (C), vanillyl-alcohol oxidase (2VAO) the two-$(\alpha+\beta)$ domain (D), and flavocytochrome c_3 (1QJD) the dinucleotide-binding multidomain (E).

In conclusion, flavoenzymes are outstanding exceptions to the paradigms "similar sequence – similar structure – similar function" and "similar cofactor / ligand – similar structure – similar function" (Fraaije and Mattevi, 2000). Consequently, this protein group has been recently indicated as an exciting research area for the investigation of structural-functional relationships (De Colibus and Mattevi, 2006).

1.2.   Three-dimensional crystallography and X-ray diffraction

Currently, three techniques are most suitable for obtaining high-resolution structures of macromolecules[3], namely X-ray diffraction (XRD) (54,018)[4], nuclear magnetic resonance spectroscopy (NMR) in solution (8,175) or in solid state (35) and electron microscopy (EM) (290). Other experimental models available on the PDB were additionally obtained by fiber diffraction (35), neutron diffraction (37), solution scattering (35), powder diffraction (18), infrared spectroscopy (4) and fluorescence transfer (1).

The statistics show that XRD has been by far the most successful method used to solve macromolecular structures. This is due to a number of reasons. XRD studies can be performed (i) with the majority of biomolecules, (ii) with macromolecules of a broad molecular mass range[5], (iii) to high resolution limits[6] and (iv) with advanced high-throughput technologies and

---

[3] The methodologies considered hereafter are all those used to create models of macromolecular structures accepted as entries in the PDB.

[4] The numbers in parentheses refer to the number of all PDB depositions obtained with the respective technique as of Jan, 15th 2010. On the same day, the total number of PDB entries was 62,634. The sum of the values in parentheses exceeds this number because some entries were obtained by a combination of methodologies.

[5] One of the largest available structures is the *Thermus thermophilus* 70 S ribosome bound to EF-Tu and tRNA, whose coordinates are split among PDB entries 2WRN, 2WRO, 2WRQ and 2WRR. The biological unit of this complex has a molecular weight of 2.3 MDa.

[6] As of Jan, 15th 2010 the highest resolution X-ray structure deposited in the PDB was 1EJG (Jelsch *et al.*, 2000), corresponding to the 46-residue-long amphipathic crambin, a membrane-toxic-like protein from the plant *Crambe abyssinica*. It was solved at a resolution of 0.54 Å.

dedicated facilities.

However, XRD studies present disadvantages, too. They require (i) target crystallizability, (ii) sophisticated experimental stations, (iii) acquisition of complete and high quality diffraction data, (iv) experimental phase determination and most often (v) complementary information for interpreting the electron density map and for establishing structural-functional relationships.

1.2.1. Three-dimensional crystallography

The prerequisite for successful 3-D crystallization experiments is a sample of high purity and homogeneity in milligram amounts[7]. Such a sample is then subjected to sparse matrix or customised screening to search for the optimal crystallization condition. Most commonly, macromolecular crystals are grown by vapour diffusion, by evaporation, by free interface diffusion, by dialysis or in batch under oil (McPherson, 1999; 2009).

Vapour diffusion is the most widely used technique. A typical vapour diffusion setup consists of a sealed chamber. In one region of the chamber, there is a reservoir solution generally composed of a precipitating agent, salts, a pH-buffering system and optional additives in water. In a separate region of the chamber, a drop of reservoir solution is mixed with the purified sample. According to whether the drop sits on or is hanging from a support, these are referred to as sitting- or hanging-drop vapour diffusion experiments. Figure 1.4 shows the time course of the phase diagram in such a crystallization setup. At initial time t_0, the reservoir solution has a lower water activity (a_w)[8] than the sample-reservoir drop. Water evaporates from the drop until an equilibrium is established (generally within

Figure 1.4: Phase diagram for vapour diffusion crystallization. At t_0 the protein solution is mixed with buffer containing a precipitating agent (0) and equilibrated against the same buffer. During equilibration the concentration increases. If the protein solubility barrier is not crossed (1), the protein remains in solution. If the solution becomes saturated (2), the protein precipitates. In favourable cases, the solubility barrier is crossed slowly and a supersaturated solution is formed (3, 4 and 5) In the supersaturated labile region nuclei form spontaneously (3 and 4). In the metastable region, nucleation can be induced by seeding, mechanical stimulation or other techniques (5). Ideally, few nucleation points form and few but large crystals grow in the drop (4). The representation of the structure and of the crystals are taken from this work (see Results section).

---

[7] 2 mg of sample (that is e.g. 200 µL at 10 mg/mL) was recently suggested as a starting point for initial sparse-matrix screening (Newby *et al.*, 2009).

[8] Water activity is the vapour pressure of water in a given medium over the vapour pressure of pure water at the same temperature.

hours or days). The a_w at equilibrium tends to correspond to the a_w of the reservoir solution, because the volume of the reservoir solution is much greater than that of the drop (typically by a factor of 100 or 1000). Over time (weeks – months), water escapes the whole chamber despite it being tightly closed, so that the equilibrium shifts towards even lower a_w values. During the process, the protein and precipitant concentrations increase and may cross the solubility barrier, so that the protein precipitates to a solid state. If the process occurs in an appropriate buffer environment and with appropriate kinetics, a metastable supersaturation condition forms temporarily. In such state, nuclei can form and induce precipitation of the sample in a regular array, a three-dimensional crystal.

The formation of crystal contacts is a delicate and unpredictable process. Sample instability and flexible regions can hamper crystallization. A number of physical and chemical parameters also affect the crystallization process and need to be controlled, i.e. temperature, pressure, mechanical stimulation, the geometry of the setup, the chemical structure and concentration of the components in the crystallization drop and the pH and ionic strength of the solution.

Crystals should be optimized to a size of approximately 0.1 – 1 mm in the three dimensions to be suitable for X-ray exposure, although it is nowadays already possible to obtain useful diffraction data with microcrystals of 5 – 20 μm in thickness. For X-ray crystallographic analysis, the crystals need to be extracted from the mother liquor with appropriate nylon loops or with glass capillaries. Exposure to the X-ray beams can be performed at room temperature or in a frozen state. Measurements at low temperatures reduce the damaging effect that radiation has on the crystal. However, freezing may distort crystal packing and lower diffraction quality, so appropriate cryo-protection is necessary. For initial tests, new robotic setups also offer the possibility to test diffraction by direct exposure of the crystallization plate to the X-ray beam.

1.2.2. *Crystallization of membrane proteins*

Crystallization of membrane proteins is particularly challenging because of the intrinsic properties of these proteins (Michel, 1991; McPherson, 1999; Iwata, 2003). In the first place, the production of correctly folded, stable and abundant membrane proteins is difficult (Ostermeier and Michel, 1997; Michel, 2006). Additionally, membrane proteins which are composed of two hydrophilic surfaces on the two sides of the membrane, separated by a hydrophobic core, need to be solubilized using mild detergents. The choice of the optimal detergent is crucial for the stability of the protein and is a determining factor of its "crystallizability". The presence of the detergent might hamper the formation of crystal contacts, it might distort the crystal lattices or it might induce phase separation in

the crystallization drops. Moreover, the detergent itself may crystallize in place of the protein (Michel, 2006).

In order to facilitate and improve the crystallization of membrane proteins, a number of dedicated strategies were developed. (i) Small amphiphiles, compounds like heptane-1,2,3-triol or benzamidine which form mixed micelles with the detergents, can be used to modulate the shape and the size of the micelles in a favourable way (Michel, 1983). (ii) The hydrophilic surfaces of membrane proteins, which form the crystal contacts, may be extended by co-crystallization with antibody fragments (Ostermeier *et al.*, 1995; Hunte *et al.*, 2000) or by insertion of large soluble domains into intra- or extra-cellular loops (Cherezov *et al.*, 2007). (iii) Addition of lipids may facilitate the crystallization process, because lipids bind membrane proteins, contribute to their functional and structural integrity and occasionally provide lattice contacts, too (Palsdottir and Hunte, 2004; Hunte and Richers, 2008). (iv) Two crystallization techniques were specifically established to obtain crystals of membrane proteins, namely the bicontinous lipidic cubic phase (Landau and Rosenbusch, 1996) and the sponge phase (Wadsten *et al.*, 2006) methods.

1.2.3. *Experimental setup for X-ray diffraction*

A typical experimental setup for XRD is composed of an X-ray source, a crystal holder and an X-ray detector.

The spectrum of X-ray radiation used to measure macromolecular crystals is the wavelength range 0.5 – 1.9 Å, corresponding to the size and length of atoms and molecular bonds. Such X-ray beams are sufficiently strong to penetrate the samples, but soft enough to interact with matter and to be scattered (Blow, 2002)[9]. Two sources of this X-ray waverange are widely used. In a rotating anode generator, an electron beam is emitted from a heated wire (cathode) at -30 / -40 kV towards a copper anode[10]. Excitation of copper leads to emission of an X-ray spectrum characterized by two particularly intensive bands. Through mirrors, the K_α band is selectively filtered and used for measurement. The X-ray beam produced by rotating anode generators[11] has a diameter of approximately 200 μm at the sample, a fixed wavelength (1.54178 Å for copper) and a brightness of $3 \cdot 10^9 - 8 \cdot 10^9$

---

[9] Instead, X-rays used for medical radiography are more penetrating and more weakly interacting with matter ($\lambda < 0.2$ Å).

[10] Anodes made of chromium, molybdenum, copper/chromium or copper/cobalt are also available but less common.

[11] The following specifications refer to the rotating anode generators used in this work, namely the MicroMax-007 HF source equipped with the RaxisIV++ image plate detector and the FR-E+ SuperBright source equipped with the Saturn 944+ CCD detector, all produced by Rigaku. Full specifications can be found at http://www.rigaku.com.

photon/s/(0.1 % BW)/mm^2/mrad^2. In a synchrotron[12], the electrons are produced in bunches by a tungsten cathode at 90 – 100 keV, accelerated in a linear accelerator (LINAC), further accelerated in a circular booster and finally stored in a ring of approximately 30 – 2000 m in diameter at an energy of 2 – 6 GeV. The storage ring is composed of linear segments where the electrons are accelerated and of turns where the electron beam is bent by magnets. When their trajectory deviates, electrons lose energy by emitting X-rays tangentially. Sophisticated optical elements modulate the emitted X-ray beam to optimize its focus, monochromaticity and intensity. In third generation synchrotrons, particularly bright beams are not produced with bending magnets, but with undulators, so called insertion devices consisting of permanent magnets arranged in a periodic array and inserted into the straight sections of the storage ring. The X-ray beam produced in such a synchrotron beamline has a diameter of 5 – 50 μm, an often tunable wavelength in the range of 0.6 – 2 Å (20 – 6 keV) and a brightness of 10^{18} – 10^{21} photon/s/(0.1 % BW)/mm^2/mrad^2.

Another essential component of an X-ray generator setup is the goniometer head, which holds and rotates the crystal during exposure. With the help of a digital camera, the crystal can be centred and aligned in the X-ray beam. A few millimetres before contacting the crystal, the X-ray beam is additionally focused by a collimator. Behind the crystal, a beam stop absorbs the unscattered beam, 99 % of the total incident beam intensity, thereby improving the signal-to-noise ratio of the diffraction image.

The third essential component of an X-ray diffractometer is the detector that collects the scattered beam. Three kinds of detectors are currently used in macromolecular crystallography and have replaced the photographic films used in earlier times (Blow, 2002). They differ in sensitivity, in recording speed and in dynamic range. Image plates (IP) contain phosphorus that absorbs X-ray photons and relaxes by fluorescence. Image plates are relatively slow (read-out time of 100 s) and have a limited dynamic range, up to 1:50,000. Charge-coupled devices (CCDs) are more rapid (read-out time of 1 s) and sensitive, with a dynamic range up to 1:150,000. They capture X-rays through a photoactive region, a layer of silicon acting as a series of capacitors, that accumulate an electric charge proportional to the X-ray intensity. The capacitors are all integrated in a controlling circuit, which digitizes the electrical signal and "reads" the image. Finally, pixel detectors, like the Pilatus 6M detector developed at the Swiss Light Source, are also silicon detectors. However, differently than in CCDs, each pixel (170 x 170 μm^2) is

---

[12] The following specifications refer to the synchrotron sources used in this work, namely the third generation synchrotrons Swiss Light Source (SLS), Villigen, Switzwerland (http://sls.web.psi.ch) and European Synchrotron Radiation Facility (ESRF), Grenoble, France (http://www.esrf.eu).
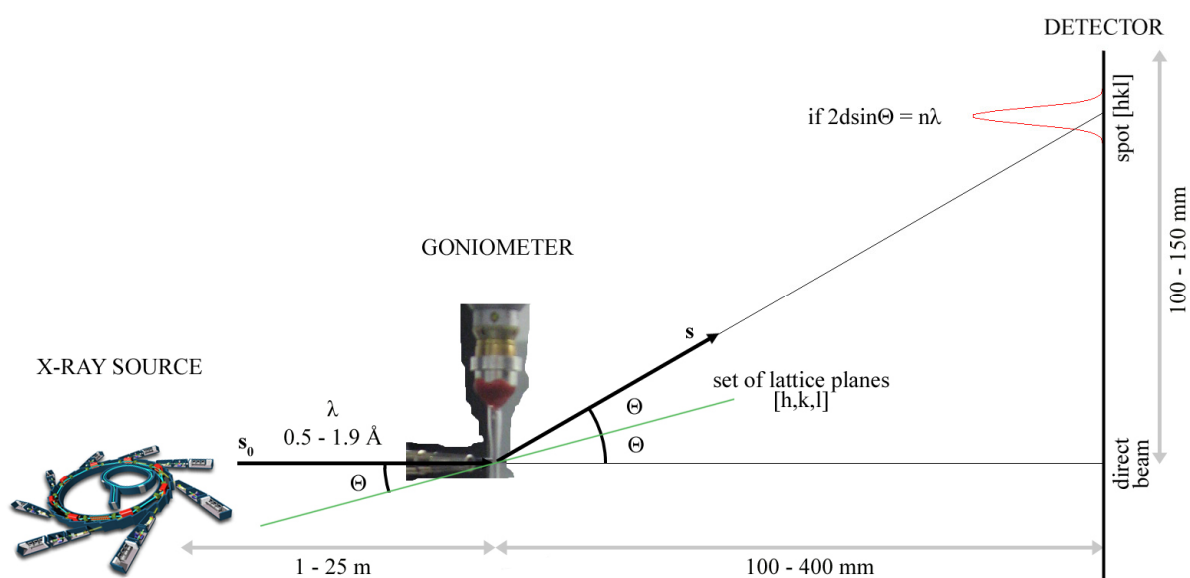
Figure 1.5: X-ray diffractometer set-up. X-rays produced by a source (in this representation, a synchrotron) pass through a crystal hold by a loop or a capillary and hanging at a goniometer. The scattered beams interfere constructively if the Bragg's law is satisfied and produce diffraction spots on a detector. The spatial scale of the experiment is not respected, therefore approximate distances and sizes of the components are indicated.

associated to its own amplifier, discriminator and counter circuit, thus conferring on the hardware much higher sensitivity and a dynamic range of 1:1,000,000. Pixel detectors are also faster in data acquisition (read-out time of 5 ms), because they are associated to a number of read-out channels 10 – 100 times higher than CCDs.

Finally, additional equipment of an X-ray diffraction setup are an optional flux of nitrogen vapour at approximately 100 K for measurements in the frozen state and an X-ray fluorescence detector to measure the fluorescence emission properties of the crystal. Figure 1.5 summarizes the components of an X-ray diffractometer.

1.2.4. X-ray diffraction theory

Atoms are surrounded by electronic orbitals that are distributions of the probability of finding an electron in a given region around the nucleus and that therefore reflect the shape of the atomic electron density (ED). When more atoms are bound to each other to form a molecule, the ED distribution takes a complicated three-dimensional (3D) shape, but its intensity can still be described as a function of the molecular volume. In a regular 3D array of molecules, the ED function is periodic.

A macromolecular crystal is a regular 3D array. Its minimal unit is called the unit cell. By simple translation of the unit cell along the crystal axis, the whole crystal volume can be filled. A unit cell may contain multiple asymmetric units related to each other by crystallographic symmetry elements (rotation or screw rotation axis) that define the space group of the crystal. The asymmetric units may themselves contain multiple copies of the

target molecule, related to each other by non-crystallographic symmetry (NCS), which refers to symmetry operations not compatible with the periodicity of the crystal.

A crystal can be represented as a lattice, with each lattice point corresponding to a unit cell and separated from the neighbouring ones by distances corresponding to the unit cell dimensions (indicated by a, b and c). A plane drawn through the lattice points is defined as lattice plane. Moreover, a lattice plane set is a series of lattice planes parallel and equidistant from each other, such that all lattice points are intersected by one plane in the set and that each unit cell axis is divided by the set into equal parts. The dividers are called Miller indices and are indicated with h, k and l for the dimensions a, b and c, respectively. In a lattice plane set hkl, the distance between the planes is indicated with the symbol d_{hkl} and it is measured in angstroms (Å). Given a crystal (its unit cell dimensions) and a minimal spacing $d_{hkl\ min}$, it is possible to draw a fixed number of lattice plane sets, independent of the structure of the crystallized macromolecule and given by:

$$N_{sets} = 4\pi V_{uc}/3(d_{hkl\ min})^3 \qquad \text{(eq. 1)}$$

where V_{uc} indicates the volume of the unit cell.

When an X-ray beam interacts with an atom, a small percentage of photons excites its electrons and it is re-emitted in all directions as a beam of the same energy and frequency of the incident beam (elastic or Thomson scattering). The intensity of the scattered beam is proportional to the atomic ED. The scattered beam in one direction is composed of all photons scattered by the different atom positions. While all incident photons are in phase, the scattered photons are out of phase, since they have run a different light path, so they interfere with each other. The phase difference determines whether the interference event is constructive or destructive. Therefore, it also contributes to defining the intensity of the scattered beam. The function describing the beam scattered by an atom is called the atomic structure factor, f_A:

$$f_A = {}_r\!\int\rho(\mathbf{r})\exp(2\pi i\mathbf{r}\cdot\mathbf{S})d\mathbf{r} \qquad \text{(eq. 2)}$$

where $\rho(\mathbf{r})$ is the ED function and $2\pi i\mathbf{r}\cdot\mathbf{S}$ is an expression for the phase of the beams scattered by each position ($\mathbf{r}$) around the atom.

The beam scattered by a unit cell is the sum of the structure factors of each atom composing it, corrected by their phase difference relative to the origin of the unit cell. Its function is called the structure factor, $\mathbf{F(S)}$:

$$\mathbf{F(S)} = {}_R\!\int f_A\exp(2\pi i\mathbf{R}\cdot\mathbf{S})d\mathbf{R} \qquad \text{(eq. 3)}$$

where f_A is the atomic structure factor defined above and $\mathbf{R}$ indicates the distance from the origin of the unit cell.

Analogously, the beam scattered by a crystal is the sum of all beams scattered by each unit cell and corrected by their phase difference. As a reference, it can be considered that macromolecular crystals have unit cell dimensions in the range of 100 Å and that the diameter of the incident beams produced i.e. by synchrotron sources is approximately 10 µm. Therefore, approximately 1000 unit cells in one direction are contemporarily exposed. Because each position of a crystal scatters the X-ray beam in all directions and because all scattered beams have different phases, their interference is in most orientations a destructive event and the signal is cancelled out. Interference of the scattered beams is constructive only in precise directions, defined by Bragg's law:

$$n\lambda = 2d_{hkl}\sin\theta \qquad\qquad (eq.\ 4)$$

where n is an integer, λ the wavelength of the incident beam, d_{hkl} the spacing of a lattice plane set and θ the angle between the lattice planes and the incident beam. During exposure, rotation of the crystal brings successively each set of lattice planes to form an angle θ with the incident beam satisfying Bragg's law.

The scattered beam in each of the directions that satisfy Bragg's law is recorded on the X-ray detector as a spot, called a reflection. Ideally, in each direction, the scattered beam should have the same diameter of the incident beam (5 – 200 µm) and therefore its signal should be recorded as a sharp peak on one single pixel on the detector (100 x 100 µm^2). However, due to imperfections in the crystal, broadening occurs and the intensity of each reflection spreads as a Gaussian distribution over few detector pixels.

Each reflection is indicated with the same indices h, k and l as the lattice plane set from which it is produced and it corresponds to a precise d_{hkl}. Reflections originating from sets of lattice planes very close to each other (small d_{hkl}) carry high resolution information, while reflections originating from sets of lattice planes far from each other (large d_{hkl}) carry low resolution information. Geometrically, sets of lattice planes close to each other form large θ angles with the incident beam and, *vice versa*, sets of lattice planes far apart from each other form smaller θ angles. Consequently, high resolution reflections are formed by the beams scattered further from the direction of the incident radiation whereas the low resolution reflections are found in the centre of the detector (Figure 1.6). The intensity of each reflection hkl is proportional to the second power of the structure factor and consequently to the ED across the set of lattice planes hkl:

$$I_{hkl} = k|F_{hkl}|^2 \qquad\qquad (eq.\ 5)$$
$$k = [\lambda^3/(\omega V_{uc}^2)][e^2/(m_e c^2)]V_c I_0 LPA$$

where k is a parameter that depends on the wavelength of the incident beam (λ), on the

angular speed of rotation (ω), on the volume of the unit cell (V_{uc}), on the charge of the electron (e), on the mass of the electron (m_e), on the speed of light (c), on the volume of the crystal (V_c), on the intensity of the incident beam (I_0), on the Lorentz factor (L), on a polarization factor (P) and on an absorption factor (A).

As can be seen from Bragg's law (eq. 4), each reflection originates from one lattice plane set. Therefore, each contains partial information about the whole unit cell. Combining all pieces of information

Figure 1.6: X-ray diffraction from a crystal lattice. For clarity, the representation has been drawn in two dimensions only. The set of lattice planes (1,1), separated by $d_{1,1} < d_{1,0}$ carries information at higher resolution than the set (1,0). It corresponds to an incident angle $\theta_{1,1} > \theta_{1,0}$ and therefore it generates a diffraction spot further from the beam centre. The insets show the geometrical construction of the Bragg's law for the two lattice plane sets, respectively.

from all (at least 90 %) possible reflections that can be generated by a crystal (see eq. 1) is necessary to reconstruct the full structure of the target. This process consists of indexing the diffraction images to determine the space group which the crystal belong to, integrating the intensities of each reflection and finally scaling them to compensate for small beam fluctuations during data collection. The result is a collection of intensities and hkl indices of all reflections produced by the target crystal. Such a data set is characterized by its completeness and its redundancy, defined by the following equations:

$$completeness = N_{ur}/N_{up} \qquad \text{(eq. 6)}$$

$$redundancy = N_{tr}/N_{ur} \qquad \text{(eq. 7)}$$

where N_{ur} is the number of unique recorded reflections, N_{up} the number of unique possible reflections and N_{tr} the number of total recorded reflections.

1.2.5. *The phase problem*

In a classical light microscope, the beams scattered by the sample are converged by lenses and the original image is recomposed. Instead, it is not possible to produce appropriate lenses for X-rays because (i) it is impossible to shape materials, that are formed by atoms, so accurately to recompose sharp images at atomic size, and (ii) the X-ray refractive index of all material is very close to that of a vacuum (Blow, 2002). Consequently, an X-ray detector cannot record a picture of the electron density of the exposed crystal, but only the

scattered beams. In other words, in a diffraction experiment the periodic ED function, characterized by amplitudes and phases, is "transformed" into a list of reflections, indicating the intensities and positions of the scattered beams. This operation corresponds in mathematics to the Fourier Transform (FT) operation. Therefore, it can be said that the diffraction pattern of a crystal is the FT of its electron density distribution. The FT is a reversible operation, so that applying a reverse FT to the diffraction data set enables the recalculation of the ED function. However, the phase information of the interfering scattered beams is lost during the diffraction experiment and has to be independently determined. This is called the phase problem in X-ray crystallography.

Three major strategies have been developed to solve the phase problem. These are molecular replacement, isomorphous replacement and anomalous dispersion.

Molecular replacement is an indirect method that can be used if a structure of a template similar[13] to the target is available. The method consists of finding the best superposition of the template over the target molecule. This is a six-dimensional problem. Three rotational and three translational parameters (or, in other words, a rotation and a translation function) have to be defined to fit the template optimally into the target unit cell. The calculation is possible by comparing the Patterson functions[14] of target and template, respectively, because the Patterson function can be calculated without additional phase information. Once the best template orientation is found, the phases of the template are used to calculate an initial ED of the target.

Isomorphous replacement and anomalous dispersion are two *ab initio* methods used to solve the phase problem. They exploit two different properties of heavy atoms[15] naturally present or artificially introduced into the target crystal.

Isomorphous[16] replacement exploits the high scattering intensity of heavy atoms. It combines information from the data set of the untreated target and one or more data sets of the target derivatized with heavy atoms. The presence of additional atoms in the crystal changes its structure factor, and so the recorded intensities of the reflections. The

---

[13] "Similar" means here similar in 3D structure. A cut-off of sequence similarity between target and template cannot be confidently defined because the paradigm "similar sequence – similar structure" does not thoroughly hold (see Chapter 1.1). However, it can be said that molecular replacement has good probability of success if template and target share above 25 % sequence identity and if the RMSD between their C_α traces is lower than 2.0 Å (Taylor, 2003).

[14] The Patterson function is the Fourier Transform of the diffraction intensities and not of the structure factors.

[15] Atoms are considered significantly "heavy" in macromolecular crystallography if their atomic number Z is 26 (Fe) or higher (Blow, 2002).

[16] Structurally isomorphous means that there should not be significant conformational changes between the different forms of the target, that the space group should be conserved and that the crystallographic parameters, such as the unit cell dimensions, should not vary significantly (indicatively less than 1 %).

difference between the structure factor measured for the protein only ($\mathbf{F}_P$) and the structure factor of the protein plus derivative ($\mathbf{F}_{PH}$) corresponds to the contribution of the derivative alone ($\mathbf{F}_H$). Therefore, the difference Patterson function $\mathbf{F}_{PH} - \mathbf{F}_P$ corresponds to the Patterson function of the derivative, $\mathbf{F}_H$. In the presence of a limited number of atoms (in the range of 10 per unit cell), the Patterson function of $\mathbf{F}_H$ can be interpreted, the position of the derivative atoms localized in the unit cell and their ED assigned an amplitude and a phase. Once $\mathbf{F}_H$ is defined, it can be used to calculate $\mathbf{F}_P$ from:

$$\mathbf{F}_H = \mathbf{F}_{PH} - \mathbf{F}_P \qquad\qquad (eq.\ 8)$$

However, the same vector $\mathbf{F}_H$ can be derived both by the difference of $+|\mathbf{F}_{PH}|$ and $+|\mathbf{F}_P|$ or of $-|\mathbf{F}_{PH}|$ and $-|\mathbf{F}_P|$. In other words, defining $\mathbf{F}_H$ for one single derivative leaves an uncertainty on the phases of $\mathbf{F}_P$. The Harker construction (Harker, 1956; Turkenburg van Diepen, 1996) (Figure 1.7) helps to describe the relationship between $\mathbf{F}_{PH}$, $\mathbf{F}_P$ and $\mathbf{F}_H$ better. In order to solve this phase bias, either the calculation of the phase probability or the calculation of $\mathbf{F}_H$ from at least a second derivative are necessary. Then, $\mathbf{F}_P$ is defined in both its amplitude and phase and an initial ED can be calculated. The methods that make use of isomorphous
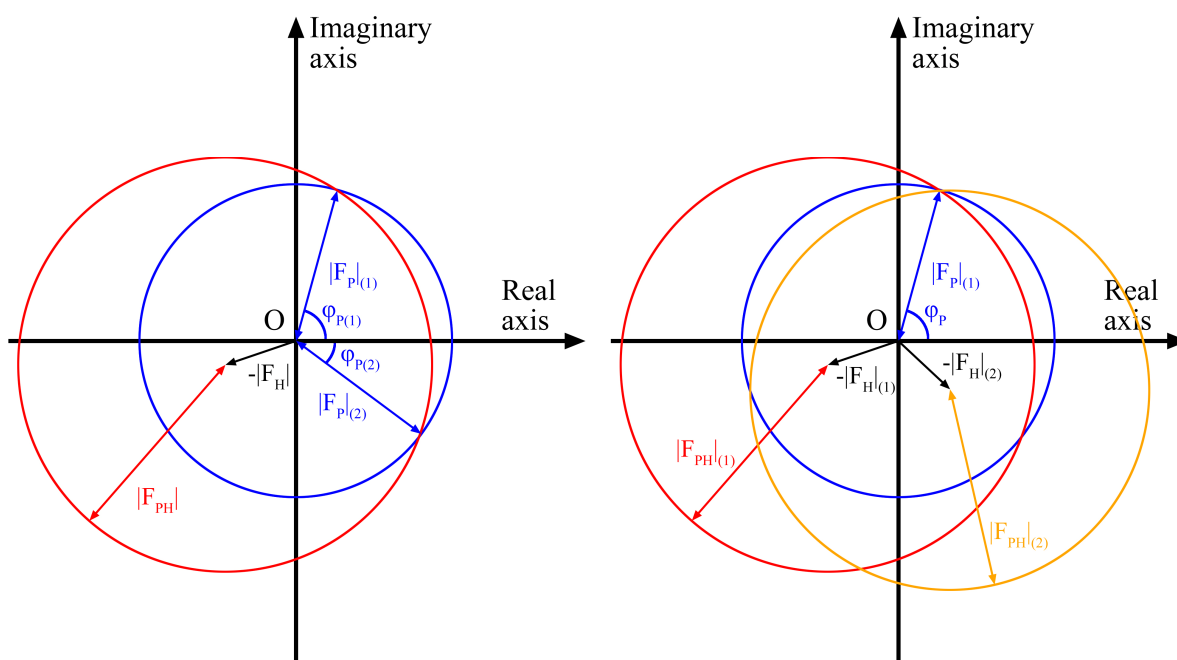
Figure 1.7: Harker's construction. The structure factor ($\mathbf{F}$) functions, which are imaginary sinusoidal functions (see eq. 3), can be represented as circles in a real *vs* imaginary coordinate plane. The amplitudes of $\mathbf{F}$ ($|\mathbf{F}|$) corresponds to the radius of the circle, while the phase (φ) is given by the angle that $|\mathbf{F}|$ form with the real axis. The blue circle represents the scattering of the target protein ($|\mathbf{F}_P|$, φ_P). The red and orange circle represent the scattering of isomorphous crystals of the same protein derivatized with different heavy metals ($|\mathbf{F}_{PH}|$). The difference $\mathbf{F}_H = (\mathbf{F}_{PH} - \mathbf{F}_P)$ is represented as a black vector, with amplitude $|\mathbf{F}_H|$. The scheme on the left shows the phase bias (φ_{P1}, φ_{P2}) when using a single derivative (SIR). The scheme on the right shows that adding a second derivative (MIR) aids in selecting between the two ambiguous phases. The figure is adapted from Turkenburg van Diepen, 1996 (available on line at: http://www.ysbl.york.ac.uk/~mgwt/thesis-tth/index.html).

/p

replacement for phase determination are the single and the multiple isomorphous replacement methods (SIR and MIR, respectively).

Anomalous dispersion requires a beam of tunable wavelength to exploit the anomalous scattering of (heavy) atoms around their absorption edges, which are specific for each element. In fact, X-rays at an appropriate energy can induce a transition of an electron of certain atoms to an excited state. An electron of an outer shell falls back and emits a photon observed as fluorescence, while the incident beam is scattered with lower intensity and a different phase. To include the anomalous scattering component, the expression of the atomic scattering factor, f_A (eq. 2), has to be corrected as a function of the wavelength λ to:

$$f_A(\lambda) = f_A - \delta f'(\lambda) + if''(\lambda) = f'(\lambda) + if''(\lambda) \qquad \text{(eq. 9)}$$

where $\delta f'(\lambda)$ and $f''(\lambda)$ define the intensity change and the phase change of the scattered beam at λ, respectively. A fluorescence detector near the goniometer can record the X-ray absorption spectrum of the derivatized crystal, from which f' and f'' can be calculated. Four characteristic wavelengths can be defined. First, the peak point is the wavelength at which f'' is maximal. Second, the inflection point is the wavelength at which f' is minimal. Finally, the high and low remote points are two wavelengths far from the absorption edge and at higher and lower energies than the peak point, respectively. Anomalous absorption by heavy atoms in the crystal has an influence on its diffraction pattern. When only normal scattering occurs, the X-ray diffraction pattern is centrosymmetric. Consequently, the structure factor from the plane h,k,l ($F_{h,k,l}$), and that from the plane -h,-k,-l ($F_{-h,-k,-l}$) are identical and have the same intensities ($I_{h,k,l} = I_{-h,-k,-l}$), despite having opposite phases ($\varphi_{h,k,l}$ = -$\varphi_{-h,-k,-l}$). The reflections h,k,l and -h,-k,-l are called Friedel's pairs. Instead, when anomalous scattering occurs, the intensities of the two reflections of a Friedel's pair are different. In particular, at the peak point the intensities are different because the scattered beams in the two directions have maximal phase difference, while at the inflection point the maximal difference is in the amplitudes of the two beams. Data can be collected on the same crystal at one or more of the four different wavelengths described above. From each resulting data set, the difference between the structure factors $F_{h,k,l}$ and $F_{-h,-k,-l}$ can be computed. The resulting value is the anomalous component of the structure factor, F_{HA}, contributed only by the heavy atoms. The heavy atom position in the unit cell can then be calculated from F_{HA} and thus initial phases can be obtained to compute a first ED map for the crystal. The methods that most typically make use of anomalous dispersion are the single-wavelength or the multiple-wavelength anomalous dispersion methods (SAD and MAD, respectively).

Anomalous dispersion and isomorphous replacement can also be combined to increase the

phasing power. The single and the multiple isomorphous replacement using anomalous scattering methods (SIRAS and MIRAS, respectively) combine the information of data sets from crystals derivatized with different metals and collected at their respective absorption edges.

1.2.6. Model building, refinement and validation

The initial phases obtained with one of the described methods can be used to trace the peptide chain and to build an initial model of the target. This procedure can be done manually or automatically and in general it requires that the amino acid sequence of the target is known.

Successively, the phases need to be iteratively improved by a refinement process. Refinement consists of (i) modifications of the model to improve the fitting to the observed structure factors[17], (ii) relaxation of the model by simulated annealing, (iii) inclusion of solvent molecules, ligands and cofactors, and also (iv) application of crystallographic restraints, such as translation-libration-screw rotation (TLS) and non-crystallographic symmetry (NCS) restraints. For TLS refinement, portions of the sample are defined as single rigid bodies and anisotropic refinement parameters are applied to each group. This procedure is advantageous for structures having domains with significantly different anisotropic motion. NCS restraints, instead, can be used if a molecule has identical copies within the asymmetric unit. NCS refinement takes into account these identities in the asymmetric unit and results in an improved signal-to-noise ratio.

The iterative process of refinement can be followed up by statistical parameters, called R values, that quantify the level of agreement between the recorded data and the model being created. The R_{factor} compares the observed amplitudes with the theoretical amplitudes of the model. It is defined by:

$$R_{factor} = \Sigma_{hkl} ||\mathbf{F}_{obs}| - |\mathbf{F}_{calc}|| / \Sigma_{hkl} |\mathbf{F}_{obs}| \qquad \text{(eq. 10)}$$

The R_{free} value is an R_{factor} calculated on a small percentage (5 %) of reflections that are excluded from refinement calculations. The R_{free} value is a more independent evaluation of the quality of the model and not prone to model bias. When the refinement iterations cannot improve the R values further (convergence) and their values are in the range of 20 % (R_{factor}) and 25 % (R_{free}) at a resolution of 2.0 – 2.5 Å, the model is ready and needs to be validated. To this purpose, a number of geometrical parameters are considered, i.e. the Ramachandran plot, the root mean square deviations (RMSD) of bonds and angles, the

---

[17] i.e.: changing the orientation of amino acid side chains or – for structures solved by molecular replacement – changing the sequence of the template to the correct sequence of the target.

conformation of the rotamers and the polypeptide bond stereochemistry (*trans* or *cis*), the orientation of the side chains and the correctness of the non-covalent interactions between the functional groups.

1.2.7. Techniques complementary to X-ray diffraction studies

For a successful and correct interpretation of the ED map, complementary characterization of the target is often useful or even necessary.

First of all, as already mentioned, the knowledge of the protein sequence is important for model building, especially at medium-low resolutions (above 1.5 Å) and with initial ED maps of poor quality. Many genomic projects in the last decades have exponentially increased the number of available protein sequences and consequently they have also contributed to a faster development of macromolecular crystallography.

Additionally, chemical knowledge on the structure of ligands or cofactors is necessary for modelling small molecules in the ED. If unknown ligands are present in the structure, techniques like mass spectrometry can be used to define their composition.

Besides information that facilitates model building and structure refinement, a detailed biochemical characterization of the protein is also helpful for understanding which structural segments determine the functionality of the target. These results are obtained for instance through mutagenesis studies and / or functional assays.

Finally, complementary techniques may be useful to compensate for two limitations of XRD studies. On one hand, XRD provides a model of a single conformation of the target. Obtaining structural data on multiple conformations or on intermediate states of reactions may be very difficult because of the instability, transience or flexibility of such species. This limitation can be overcome by spectrophotometric techniques, like Fourier transform infrared spectroscopy (FTIR), fluorescence resonance energy transfer (FRET), electron paramagnetic resonance (EPR), nuclear magnetic resonance (NMR) and Raman spectroscopy, and by computer calculations, like molecular dynamics simulations. On the other hand, certain structural features are difficult to obtain by X-ray crystallography. As an example, hydrogen bonds often play a crucial role in catalytic processes, but resolving hydrogen atoms by X-ray crystallography requires very high resolutions (< 1 Å). Neutron diffraction studies can instead reveal the positions of hydrogen atoms by making use of the neutron scattering properties of the hydrogen isotope deuterium (Blakeley *et al.*, 2008).

1.3.   Biological membranes

Biological membranes are amphipathic barriers defining cellular compartments in living organisms. They mediate communication between the interior and exterior phases, they ensure

selective permeability, they provide shape to cells and organells through interaction with the cytoskeleton, they maintain the cell electrical potential, they are the location of cellular energy production and they metabolize hydrophobic substrates.

Biological membranes are films of approximately 45 Å in thickness, with a central 20 – 30 Å broad hydrophobic core and two 10 – 15 Å broad interface regions on each side. Their actual size and morphology depend on the organism, on the subcellular compartment, on their composition and on the structure of the macromolecules embedded in them.

Biological membranes are composed on average of a 1:1 weight ratio of lipids and proteins, with carbohydrates being present in minor amounts (1 – 5 %). This ratio corresponds to a molar ratio of about 50 lipid molecules per membrane protein (Guidotti, 1972). However, the precise membrane composition varies strongly with the cell type. For instance, the myelin sheath (Schwann cells) contains about 80 % w/w lipids, while the mitochondrial inner membrane and the membrane of Gram positive bacteria contain only 25 % w/w lipids (Guidotti, 1972).

Both the lipidic and the protein component determine the membrane properties.

Lipids confer their amphipathic nature and their fluidity on the membranes. In addition, they interact directly with membrane proteins and contribute in an essential way to the stabilization of their structures and/or modulation of their functions (Palsdottir and Hunte, 2004; Hunte and Richers, 2008). Lipids are a class of biomolecules with heterogeneous structures, and include fatty acids, steroids, acylglycerols, phospholipids and sphingolipids as well as hydrophobic vitamins and quinones (Whitford, 2005). Phospholipids are the major components of biological membranes. They are amphipathic molecules with a glycerol-3-phosphate polar head group in general esterified with two $C_{16} - C_{18}$ saturated or unsaturated fatty acid chains.

Membrane proteins comprise 15 – 20 % of bacterial proteomes (Bendtsen *et al.*, 2005) and up to 27 % of the human proteome (Almen *et al.*, 2009). In absolute terms, their number in humans was recently predicted to be 6,718 (Almen *et al.*, 2009). Besides representing a large percentage of the protein pool of all organisms, membrane proteins are attractive for biochemical and pharmaceutical research being targets of approximately 60 % of the drugs currently on the market (Arinaminpathy *et al.*, 2009). Furthermore, among proteins, membrane proteins are outstanding because they occupy a complex biphasic cellular compartment. The peculiar physical and chemical nature of the lipidic bilayer limits the freedom of evolution of the membrane-embedded polypeptide portions. Consequently, it was said that the "folding space" of membrane proteins is "quasi-two-dimensional" and not three-dimensional like that of soluble proteins (Theobald and Miller, 2010). At least three different structural requirements specific for membrane proteins were identified. (i) Membrane proteins need to partition across five phases, namely two aqueous and one hydrophobic zones separated by two transition layers.

(ii) Transmembrane regions must be enriched in hydrophobic residues, while at the water-membrane interface the polar heads of the phospholipids preferentially interact with tryptophans, lysines and arginines (von Heijne and Gavel, 1988). (iii) Transmembrane segments all need to have a comparable length, because the thickness of the membrane is approximately constant in all compartments and organisms. Additionally, also the functions possessed by membrane proteins may impose constraints on their transmembrane topology. This idea is suggested for example by the observation that a number of transporters (Yamashita *et al.*, 2005; Faham *et al.*, 2008; Weyand *et al.*, 2008; Fang *et al.*, 2009; Ressl *et al.*, 2009) and one channel (Waight *et al.*, 2010) all adopt a "5+5" inverted structural symmetry motif (Krishnamurthy *et al.*, 2009) despite having no obvious sequence similarity and conducting different molecules across the bilayer. In summary, membrane proteins are subject to more structural constraints than soluble proteins and in particular they seem to be more prone to structural convergence. Consequently, exceptions in the paradigm "similar sequence – similar structure – similar function" are likely to be numerous for membrane proteins (Theobald and Miller, 2010), making structural-functional relationship studies on them particularly fascinating. Despite their exceptional interest, expression, extraction, solubilisation, purification and – for structure determination – crystallization remain difficult and expensive procedures for membrane proteins in comparison with soluble proteins. Because of these difficulties, studies on membrane proteins began later than those on soluble proteins and are still challenging. Initially, structural studies on membrane proteins were considered impossible, until the pioneering work by H. Michel and J. Deisenhofer in the early 1980s (Michel, 1982a; 1982b) led to solving the first three-dimensional membrane protein structure, *Rhodopseudomonas viridis* photosynthetic reaction centre, in 1985 (Deisenhofer *et al.*, 1985). Now in 2010, the unique membrane protein structures available still represent only less than 1 % of the number of unique protein structures deposited in the Protein Data Bank (PDB)[18].

1.3.1. Peripheral membrane proteins

Membrane proteins can be classified into two groups according to their topology (Blobel, 1980; Lomize *et al.*, 2007).

Polytopic membrane proteins (or transmembrane proteins, TMPs) are characterized by at least one transmembrane spanning segment. Therefore, they expose segments to each side

---

[18] A statistics of membrane protein structures is available from the Membrane Protein Data Bank [www.mpdb.ul.ie, (Raman *et al.*, 2006)], which covers integral, anchored and peripheral membrane proteins and peptides. According to this data bank, as of Jan, 15th 2010, the total number of membrane protein and peptide structures deposited on the PDB was 1,037 (corresponding to 267 unique structures), of which 843 were determined by XRD, 167 by NMR, and 27 by EM. The membrane proteins with known structure belong to 146 unique protein families.

of the membrane. TMPs are either α-helix-bundle or β-barrel proteins (von Heijne, 1994). These two assemblies guarantee an energetically favourable distribution of polar and apolar residues in the hydrophobic membrane environment, while a combination of α-helices and β-sheets has been observed more rarely[19].

On the other hand, peripheral membrane proteins (PMPs) interact with one membrane leaflet only. Among PMPs, amphitropic proteins (APs) bind the membrane transiently. They are soluble proteins that undergo a conformational change upon activation, acquire affinity to the lipid bilayers and consequently bind the membrane to perform their membrane-related functions. Activation can be stimulated by external factors, such as fluctuations of pH or ligand concentration, but also by changes in the membrane composition itself. The recent structure of *Escherichia coli* pyruvate oxidase well exemplifies the membrane-binding/release cycle of APs. Upon flavin reduction, the C-terminal segment of that enzyme rearranges from a half-barrel to a helix, thereby acquiring affinity to the lipidic bilayer and opening the active site to the second substrate, the lipophilic ubiquinone 8 (Neumann *et al.*, 2008). Another group of PMPs are monotopic membrane proteins (MMPs). They bind the membrane constitutively and cannot be isolated in a stable water-soluble form. Examples of MMPs with known structure are prostaglandin H_2 synthases 1 and 2 [20] (PDB id.: 1Q4G and 1CX2), microsomal prostaglandin E synthase (1Z9H), 11-β-hydrosteroid dehydrogenase (1XSE), apocarotenoid cleavage oxygenase (2BIX), electron transfer flavoprotein (ETF):ubiquinone oxidoreductase (2GMH), lanosterol synthase [21] (1W6K), squalene-hopene cyclase (2SQC), carnitine acyltransferase (1T7Q), carnitine O-palmitoyltransferase 2 (2H4T), cytochrome P450 (2BDM), monoamine oxidase B (1OJA), and fatty acid amide hydrolase (1MT5), [see http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html; http://opm.phar.umich.edu; (Lomize *et al.*, 2007)].

PMPs perform various physiological functions (Johnson and Cornell, 1999). Some catalyse reactions that may require membrane attachment, either because the catalysis itself is influenced by membrane components (i.e. protein kinase C, PKC; cytidylyltransferase, CT) or because the substrates of the protein are hydrophobic (i.e. phospholipases, lipid kinases). Others interact with membrane components (i.e. with TMPs) thus ensuring an effective

---

[19] For instance, it has been observed in some outer membrane proteins, such as porin OmpF (PDB id.: 2OMF), the ferric citrate transporter FecA (1PO0) and the colicin I receptor (2HDF). However, in those structures the membrane-interacting domain consists of an all-β barrel, while a distinct all-α or mixed α/β domain folds inside the central pore generated by the barrel itself.

[20] Also known as cyclooxygenases (COX) 1 and 2.

[21] Also known as oxidosqualene cyclase (OSC).

assembly of e.g. signalling complexes (i.e. G proteins, lipid kinases). Finally, PMPs may constitute a connection between membrane and cytoskeleton (i.e. vinculin, hisactophilin) or mediate endocytosis and other processes of membrane remodelling (Johnson and Cornell, 1999; Bhatia *et al.*, 2009).

The strength of a PMP binding to the membrane depends on different forces (Johnson and Cornell, 1999; Lomize *et al.*, 2007). (i) The electrostatic contribution is mainly due to the interactions of basic residues (arginine, lysine) with the polar head groups of anionic lipids. These interactions can be competed at high ionic strength and can vary depending on the protonation and phosphorylation state of surface amino acids. (ii) Hydrophobic interactions involve apolar residues and in particular tryptophans and tyrosines, which are often present at the water-lipid interface of all membrane proteins (Granseth *et al.*, 2005; Lomize *et al.*, 2007). Through hydrophobic interactions some proteins specifically target to membranes having a particular composition. (iii) The distortion of the lipidic bilayers by PMPs increases or decreases the freedom of the lipidic acyl side chains and changes the layer curvature. Therefore, it also contributes positively or negatively to the free energy of protein-membrane binding. (iv) For APs, membrane binding is additionally influenced by the free energy required by the protein to undergo the conformational change upon activation and by an entropy factor, due to the reduced degree of freedom of the protein bound to the membrane with respect to the protein in solution.

Depending on the interactions occurring between PMPs and lipid bilayers, the degree of penetration of PMPs into the membrane varies. Those proteins that only form electrostatic interactions with the lipids are buried approximately 3 Å under the membrane surface. If hydrophobic interactions are also formed, the protein can insert as much as 15 – 25 Å into the membrane. In some cases, the membrane-inserted form of some APs traverses both leaflets of the bilayer (i.e. SecA, colicins, diphtheria toxin).

On the basis of the structures known to date, it can be observed that the structural motifs for peripheral membrane binding are rather heterogeneous. However, some common features can be described (Figure 1.8).

(i) Lipid clamps are structural motifs which direct protein binding to a specific lipid in the membrane. Examples are the cysteine-rich domain (named CRD or C1) specific for diacylglycerol, the calcium-and-lipid-binding domain (CaLB or C2) interacting
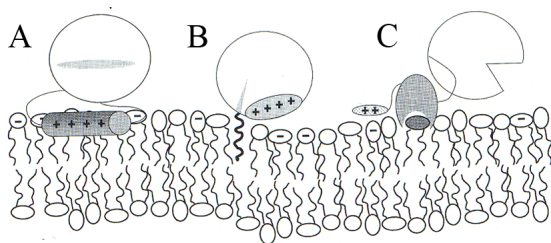
Figure 1.8: Peripheral membrane proteins. Three different strategies can be adopted by PMPs to bind the membrane: amphipathic in-plane helices (A), covalent binding to aliphatic chains (B), and lipid clamp domains (C). The Figure is reproduced from Johnson and Cornell, 1999.

with anionic phospholipids and the Pleckstrin Homology (PH) domain specific for phosphatidylinositol. (ii) Covalent binding to membrane-embedded aliphatic chains is another strategy used by soluble proteins to attach tightly to bilayers. Soluble proteins with covalently-bound lipids are i.e. Raf or the G_α subunit of G-proteins. (iii) The Bin, Amphiphysin, Rvs (BAR) and the Epsin NH_2-Terminal Homology (ENTH) domains are α-helical membrane binding modules that sense and regulate the curvature of the bilayer and play therefore a major role in endocytosis, actin regulation and signalling (Itoh and De Camilli, 2006). Their structure was first revealed by crystallographic work on *Drosophila* amphiphysin [PDB id.: 1URU, (Peter *et al.*, 2004)]. (iv) Amphipathic helices (AHs) also sense changes in the physical properties of the bilayer (electrostatic potential, lipid packing density, curvature strain). They occur in APs, like CT, DnaA, ADP-ribosylation factor (ARF), blood clotting factor VIII and vinculin, and in MMPs, like prostaglandin synthase or squalene hopene synthase. Each of these proteins displays a different helical length and a different tertiary structure arrangement in the amphipathic domain, but they also share some conserved features. In particular, the distinctive parameters of amphipathic helices are an overall positive net charge, a high hydrophobic moment and a relative hydrophobicity of $\Delta G_H/N_{residues} = 1 - 2$ Kcal/mol[22] (Eisenberg *et al.*, 1982; Engelman *et al.*, 1986; Johnson and Cornell, 1999). AHs typically intercalate in the polar head group region of the membrane within 15 Å of the water exposed surface. Finally, they are not specific for particular lipid classes, with the exception of the amphipathic domain of blood clotting factor VIII.

1.3.2. Enzymes in the membrane

A number of reactions are catalyzed by membrane proteins. Enzymes in the membranes take part in the lipid and quinone biosynthesis pathways, in the metabolism of neurotransmitters and of mediators of inflammatory processes, in signalling cascades, in synaptic activities, in the synthesis of the bacterial cell wall, in proteolytic processes, in photosynthesis and cellular respiration, in other oxidoreductive processes and in reactions associated to ATP hydrolysis (Lenaz and Degli Esposti, 1994).

A peculiarity of membrane-localized enzymes is that they typically catalyse the reaction between a soluble and a hydrophobic molecule. Consequently, their catalytic site must be accessible from the membrane as well as from the bulk solvent. Three strategies have been identified to satisfy this requirement (Forneris and Mattevi, 2008) (Figure 1.9). (i) In enzymes such as metalloproteases, the lipophilic molecule moves by lateral diffusion while

---

[22] ΔG_H is the energy of displacement from a membrane mimetic environment to water.
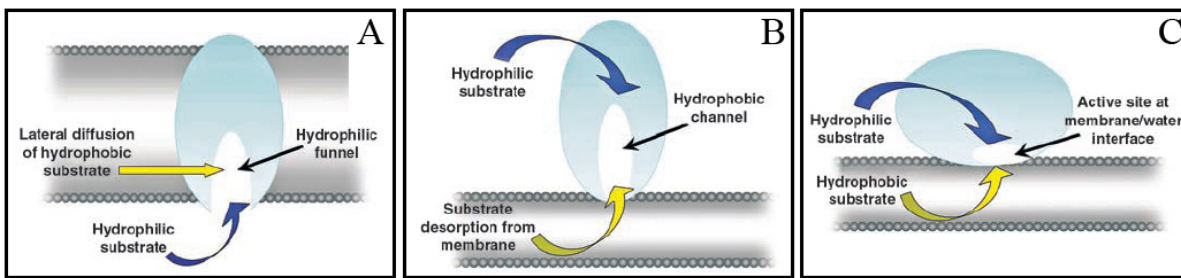
Figure 1.9: Membrane-bound enzymes. Three different strategies can be adopted by membrane enzymes for catalyzing reactions between lipophilic and water-soluble substrates: lateral diffusion (A), desorption (B) and catalysis at the water-membrane interface (C). The Figure is reproduced from Forneris and Mattevi, 2008.

the hydrophilic substrate reaches the membrane-buried active site through a specific funnel. (ii) In hydrosteroid dehydrogenases (HSDs), bacterial fatty acyl-CoA synthetase (FACS) and fatty acid amide hydrolase (FAAH), the lipophilic reactant can be "desorbed" from the membrane through a hydrophobic cavity and reaches the active site localized in a water-soluble domain of the protein. (iii) Enzymes like peptidoglycan transglycosylase and phospholipases perform their reactions at the water-membrane interface. Finally, an atypical role is played by i.e. saposins, non-enzymatic PMPs acting as "helpers" for water-soluble enzymes that process lipophilic substrates. Saposins extract glycosphingolipids from the membrane via large flexible domains, undergo a significant conformational change and present their ligand to the enzymes in the surrounding solution.

1.4. Extremophiles

It has been known for several decades that life is possible in very demanding environments such as geothermal sources, polar regions, acid and alkaline springs and the cold pressurized depths of the oceans. The discovery of life forms in harsh conditions has raised scientific curiosity as to the biochemical mechanisms of adaptation and the possibility of exploiting them for new chemical and biotechnological processes (Horikoshi and Grant, 1998).

"Extreme" habitats are populated by species representing all three domains of life, but prokaryotes and in particular archaea are predominant (Rothschild and Mancinelli, 2001). These organisms that tolerate or require and, so to say, "love" ("*philos*" in ancient Greek), "extreme" conditions were named "extremophiles" first by MacElroy (Macelroy, 1974). However, the definition of "extreme" is controversial. For example, in an anthropocentric view extreme conditions are those incompatible with the life of human beings. Instead, Rottschild and Mancinelli (Rothschild and Mancinelli, 2001) more objectively defined as extreme those environments in which biomolecules are unstable. The physical and chemical factors that the authors considered are temperature, pressure, pH, ionic strength, hydration, abundance of nutrients and exposure to toxic agents, including gases (O_2, CO_2, H_2S), heavy metals and

synthetic compounds. Consequently, a number of classes of extremophiles can be enumerated (Rothschild and Mancinelli, 2001; Gerday and Glansdorff, 2007). Thermophiles are resistant to temperatures above 60 °C, the highest temperature compatible with life being 113 °C at present knowledge. Psycrophiles survive cold temperatures between -2 °C and 20 °C. Alkalyphiles thrive at pH values above 9, while acidophiles at pH values below 4. Piezophiles live at high pressure, up to 70 – 80 MPa in the deepest oceanic trenches. Halophiles are organisms adapted to high ionic strengths, up to 5 M NaCl. Xerophiles tolerate desiccation and can live in environments with water activity well below 0.8 [23]. Among xerophiles, endolithic and hypolithic organisms can grow into or under the surface of stony substances. Oligotrophic life forms can grow in nutritionally limited environments. Radioresistant organisms survive exposures of up to i.e. 20 kGy of γ radiation or 1000 J/m^2 of UV radiation. Toxicotolerant species are able to withstand high levels of damaging agents and among them metallotolerants specifically tolerate high concentrations of heavy metals. Finally, aerobic organisms may also be considered extremophiles since they tolerate exposure to oxygen, a particularly reactive species especially in its reduced peroxide and hydroxyl radical forms and consequently an unquestionable source of macromolecular damage (Rothschild and Mancinelli, 2001).

1.4.1. *Thermophiles and proteins from thermophilic organisms*

Most thermophiles are found in early diverging branches of the 16 – 18 S rRNA universal phylogenetic tree (Hough and Danson, 1999) so that it is debated whether they constitute the ancestors of life forms on Earth (Gerday and Glansdorff, 2007). They occupy a variety of natural or artificial environmental niches, including hot natural or industrial waters, sun-heated soils, geothermal areas and heated waste dumps. Depending on the optimal growth temperature, they can be subdivided into moderate thermophiles, that live between 60 °C and 80 °C, and hyperthermophiles, that live above 80 °C (Horikoshi and Grant, 1998).

At high temperatures, the membrane fluidity increases to lethal levels and proteins and nucleic acids normally denature. Moreover, the solubility of gasses is altered, affecting O2- or CO2-dependent metabolic pathways. To our present knowledge, there is not a universal mechanism of thermal stabilization, but a number of different strategies used to cope with high temperature stresses can be described (Hough and Danson, 1999; Rothschild and Mancinelli, 2001; Jaenicke and Sterner, 2006; Gerday and Glansdorff, 2007).

(i)   The membrane composition is adjusted to maintain optimal fluidity and decrease permeability. The polar head group of the phospholipids is not always glycerol, as for mesophilic organisms, but it may also be a longer chain polyol, i.e. nonitol

---

[23] As reference, the a_w of a saturated solution of NaCl is 0.75.

(Jaenicke and Sterner, 2006). Polyols form a denser hydrogen bonding network, thereby increasing membrane compactness. Furthermore, ether bonds between the polar head groups and the fatty acid chains widely substitute the ester connection typical of mesophilic membranes. Besides conferring higher resistance to hydrolysis (Gerday and Glansdorff, 2007), the replacement of ester with ether bonds provides a barrier to transmembrane proton flux (Jaenicke and Sterner, 2006). Additionally, the acyl chains of thermophilic lipids are saturated so that stronger hydrophobic interactions can be formed. The mobility of the chains is decreased by the presence of methyl side groups at every fourth carbon atom in the backbone or by cyclization. Finally, the length of the acyl chains also differs with respect to those of lipids in mesophilic organisms. In the latter, the fatty acids have an even number of carbon atoms, C16 and C18 being the most common species. In thermophiles, instead, longer chains (i.e. C20-24) and chains with odd numbers of carbon atoms (i.e. C17 in Thermodesulfobacterium) are also observed (Jaenicke and Sterner, 2006). In some cases, the two membrane leaflets are covalently connected through dibiphytanyl-diglycerol-tetraether lipids, whose C30-40 acyl chains span the entire membranes. Such monolayers have a constant thickness of 25 – 30 Å and are particularly rigid (Jaenicke and Sterner, 2006; Gerday and Glansdorff, 2007) (Figure 1.10).

(ii)   DNA stability is enhanced by mono- or divalent salts that shield the negative charges of the phosphate groups and protect the DNA from depurination and hydrolysis. Formation of G-C pairs stabilizes ribosomal and transfer RNA, while an increase in G-C content is not observed in chromosomal DNA of thermophiles. The DNA is rather stabilized by a DNA reverse gyrase, a unique type-I DNA-topoisomerase that causes supertwists, and by very basic histones (Horikoshi and Grant, 1998).

(iii)   Proteins are stabilized by increasing ion-pair content and compactness. They also tend to form higher-order oligomers (Kumar *et al.*, 2000; Yano and Poulos, 2003; Tanaka *et al.*, 2004). The presence of additional inter-subunit interactions involving amino acid residues near the carboxyl terminus with respect to their mesophilic homologues has also been correlated to thermostability (Russell et al., 1997). Additionally, the secondary structural elements are stabilized by shorter surface loop connections, by optimized electrostatic and hydrophobic interactions, by disulfide bridges, by a more pronounced hydrophobicity in the protein core and by a higher helical propensity of residues in α-helices (Rothschild and Mancinelli, 2001; Jaenicke and Sterner, 2006; Gerday and Glansdorff, 2007).

GLYCEROL

ALIPHATIC CHAIN

glycerol ester 18:0

glycerol ester 20:1cisΔ11

glycerol ether 18:0

glycerol diether 17:0, 17:0 with methyl branching

glycerol tetraether with methyl branching

glycerol:glycerol tetraether with cyclopropyl rings

glycerol:nonitol tetraether with methyl branching

POLAR HEAD

phosphatidic acid

phosphatidylserine

phosphatidylcholine

phosphatidylethanolamine

phosphatidylglycerol

phosphatidylinositol

Figure 1.10: Structures of glycerophospholipids. The central skeleton of the glycerol molecule (top) is derivatized with fatty acid hydrophobic chains (left) and with polar head groups (right) of various chemical nature. The lipidic polar heads in mesophilic and thermophilic organisms are the same. Instead, the typical aliphatic chains of mesophilic species (saturated or unsaturated esters) are often substituted by chains with other structures in thermophilic species. Such differences consist of methyl branching, cyclization, ether bonds, and longer chains spanning through the entire membrane (tetraether lipids). In thermophiles, the glycerol moiety may also be substituted by nonitol.

Of particular interest is the stabilization of the thermophilic enzymes, which need to be sufficiently rigid to prevent denaturation but flexible enough to undergo the conformational changes required by catalysis. An efficient compromise is achieved with the presence of large solvent filled cavities in the protein core, which provide the enzyme with "resilience", that is, the propensity of accumulating elastic energy upon deformation (Hough and Danson, 1999).

In conclusion, thermophilic macromolecules are particularly good candidates for structural studies, thanks to their pronounced stability. However, heterologous overproduction of membrane proteins from thermophiles can be problematic, in particular because of the strong differences in the lipid bilayer composition of the source and host organisms. For those membrane proteins naturally expressed in high abundance, ideal structural studies are performed after extracting the proteins from the native cells.

1.4.2. Aquifex aeolicus

Aquifex aeolicus (VF5) is an extremophilic organism because it requires high temperatures and exposure to hydrogen, oxygen and carbon dioxide for growth. It is a hyperthermophilic bacterium belonging to the order *Aquificales*. It grows optimally at 85 – 95 °C and was isolated from shallow hydrothermal volcanic areas in the Aeolian Islands, Italy (Huber and Eder, 2006)[24]. It can be artificially grown in liquid medium SME supplemented with thiosulfate (Huber and Eder, 2006). Its cells are rod-shaped, with a length of 2 – 6 μm and a diameter of approximately 0.5 μm (Huber and Eder, 2006).

A. aeolicus is a chemolithoautotroph. It uses inorganic carbon sources for biosynthesis[25] and inorganic chemical energy sources[26]. It is an obligate aerobic bacterium that cannot tolerate high oxygen concentrations but can survive in levels of oxygen as low as 7.5 ppm. It preferentially gains energy through the "Knallgas" reaction[27] that produces water[28]. Instead of hydrogen, it can also use thiosulfate as an energy source, producing sulfuric acid (H_2SO_4) and hydrogen sulfide (H_2S) (Huber and Eder, 2006). Its cellular membrane is predominantly composed of ester-linked fatty acids ($C_{18:0}$, $C_{20:1}$, C_{cy-21})[29], alkyl glycerol

---

[24] Therefore the species name "*aeolicus*". "VF5" indicates the strain that was isolated and whose genome has been sequenced (Deckert *et al.*, 1998).

[25] *A. aeolicus* fixes CO_2 through the reductive tricarboxylic acid cycle (Deckert *et al.*, 1998; Guiral *et al.*, 2005).

[26] i.e.: elemental sulfur, hydrogen, nitrate (Guiral *et al.*, 2005).

[27] $2H_2 + O_2 \rightarrow 2H_2O$

[28] This explains the genus name "*Aquifex*", which means "water maker" (Huber *et al.*, 1992).

[29] C_n indicates an alkyl chain with n carbon atoms; :0, saturated; :1, monounsaturated; *cy*, cyclopropyl ring.

monoethers ($C_{18:0}$, $C_{20:1}$) and glycerol diethers ($C_{18:0/18:0}$, $C_{18:0/21:1}$, $C_{20:0/20:0}$) (Huber and Eder, 2006).

Its gene pool constsits of the chromosomal genome and of an extrachromosomal element (ECE), both sequenced in 1998 with a calculated sequence redundancy of 4.83 (Deckert *et al.*, 1998). The genome size (1.551.335 base pairs in length) is only one-third that of *Escherichia coli* (Deckert *et al.*, 1998), indicating simpler metabolic pathways and a smaller proteome. Deckert and coworkers report that the genome is densely packed, that most genes are expressed in polycistronic operons and that many convergently transcribed genes overlap slightly. Interestingly, the genes that are functionally grouped within operons in other organisms or that encode subunits of the same enzyme are often separated in *A. aeolicus*. For example, it never recurs that two genes for amino acid biosynthesis are encoded in the same operon (Deckert *et al.*, 1998). In contrast, genes required for electron transport, hydrogenase subunits, transport systems, ribosomal subunits and flagella are often in functionally related operons. No introns or inteins were originally detected in the genome (Deckert *et al.*, 1998)[30]. The authors suggest that this might be related to autotrophy and a consequent lifestyle that decreases the need for selective regulation. A high G+C content in stable RNA, two genes for reverse gyrase and an overall abundance of charged amino acids with respect to mesophilic organisms are the genomic features that can be related to *A. aeolicus* thermal stability (Deckert *et al.*, 1998).

From the genome sequence some information can be obtained on the organisation of the *A. aeolicus* respiratory chain. The enzymes for oxygen-dependent energy production are similar to those of other bacteria. NADH:ubiquinone oxidoreductase (respiratory complex I), succinate:quinone oxidoreductase (respiratory complex II), ubiquinol:cytochrome *c* oxidoreductase (bc1 complex, respiratory complex III), cytochrome *c*, cytochrome *c* oxidase (respiratory complex IV) and F_1-F_o ATPase (respiratory complex V) are present. Additionally, the presence of a Rieske-like protein adjacent to a sulfide dehydrogenase subunit opens up the possibility that *A. aeolicus* respiration could be sulfur-dependent (Deckert *et al.*, 1998). Moreover, a number of other oxidoreductases have been identified in the genome and metabolize alternative electron donors and acceptors, also contributing to electron transfer in the membranes. These proteins are cytochrome *bd* ubiquinol oxidase, sulfide:quinone oxidoreductase (SQR), nitrate reductase, DMSO reductase[31], cytochrome c_{552}, hydrogenase I, hydrogenase II, formate dehydrogenase, flavocytochrome *c*:sulfide dehydrogenase (FCSD), ferredoxin, sulfur oxygenase reductase (SOR) and sulfur oxidation

---

[30] One intein was annotated recently (Huber and Eder, 2006).

[31] DMSO reductase was recently renamed sulfur reductase (Guiral *et al.*, 2005).
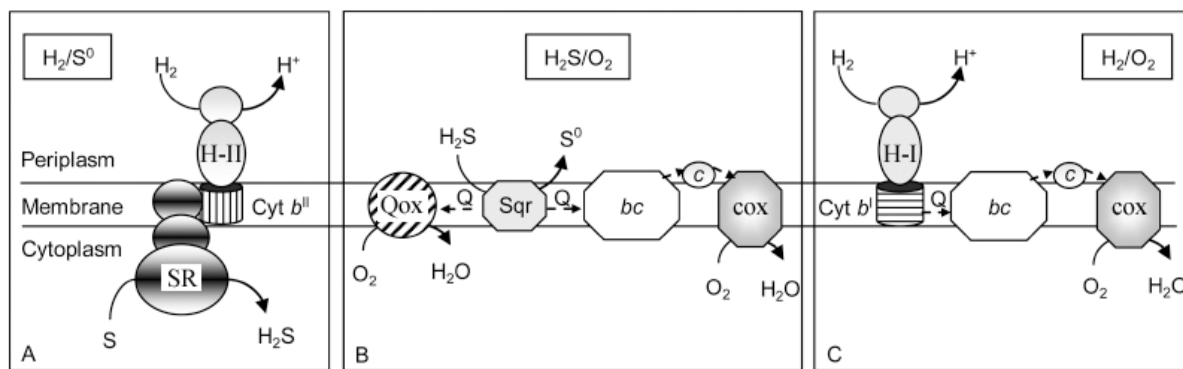
Figure 1.11: *A. aeolicus* **respiratory pathways.** Depending on the energy source, *A. aeolicus* has been proposed to use different pathways for cellular energy production. Panel A: in the presence of hydrogen (H_2) and elemental sulfur (S^0), it couples hydrogen oxidation by hydrogenase II (H-II, possessing a cytochrome b_{II} transmembrane subunit, Cyt b^{II}) with sulfur reduction by sulfur reductase (SR, previously annotated as DMSO reductase). Panel B: in the presence of sulfide (H_2S) and oxygen (O_2), it oxidizes sulfide to elemental sulfur by sulfide:quinone oxidoreductase (Sqr) and transfers electrons to the membrane quinone pool (Q) for oxygen-dependent respiration either *via* a quinol oxidase (Qox) or *via* the cytochrome bc_1 complex (bc), cytochrome *c* (c) and cytochrome *c* oxidase (cox). Panel C: in the presence of hydrogen and oxygen, it oxidizes hydrogen and reduces quinones through hydrogenase I (H-I, possessing a cytochrome b_I transmembrane subunit, Cyt b^I), thereby allowing oxygen-dependent respiration *via* the cytochrome bc_1 complex (bc), cytochrome *c* (c) and cytochrome *c* oxidase (cox). The Figure is adapted from Guiral *et al.*, 2009.

proteins (Sox). A number of schemes for possible *A. aeolicus* respiratory pathways in different growing conditions have been proposed (Nübel *et al.*, 2000; Guiral *et al.*, 2005; Guiral *et al.*, 2009) (Figure 1.11).

Finally, it must be considered that approximately 23 % of the *A. aeolicus* genome coding regions could not be assigned a known function by comparison to homologous sequences and were thus annotated as hypothetical proteins. Therefore, it will be impossible to have a comprehensive view of *A. aeolicus* metabolic pathways until functional and structural data becomes available for this large number of as yet uncharacterized proteins.

Currently, 160 unique structures[32] of proteins from *A. aeolicus* are deposited on the PDB, including 3 proteins associated with the membrane[33].

1.5.   The sulfur cycle and sulfur metabolism

Sulfur ($_{16}$S) is a compound of versatile chemistry. It is present in nature as 25 different isotopes, four of which are stable (^{32}S, ^{33}S, ^{34}S and ^{36}S) and one of which is widely used as a marker for its radioactivity (^{35}S). The electron configuration of sulfur, $[Ne]3s^2 3p^4$, allows it to acquire a broad range of valence states from -2 to +6 and thus be the most versatile source of oxidant and

---

[32] The numbers refer to a PDB database search performed on Jan, 15th 2010. On that day, the total number of proteins from *A. aeolicus* on the PDB was 309.

[33] Sulfide:quinone oxidoreductase (SQR), PDB id.: 3HYV, (Marcia *et al.*, 2009); a hybrid complex of *A. aeolicus* SecYEG and *Bacillus subtilis* SecA, PDB id.: 3DL8, (Zimmer *et al.*, 2008); and leucine transporter (LeuT), PDB id.: 2A65, (Yamashita *et al.*, 2005).

reductant species. Biologically relevant sulfur compounds are sulfide $[S^{2-}]$, polysulfides $[S_n^{2-}$, in the most stable chains n = 4 – 5], elemental sulfur [in different allotropic forms of which orthorhombic cyclooctasulfur (S_8) is the most stable], organic sulfoxides [R-SO-R', i.e. dimethyl sulfoxide, DMSO, $(CH_3)_2SO$], dithionite $[S_2O_4^{2-}]$, thiosulfate $[^-O_3S\text{-}S^-]$, sulfite $[SO_3^{2-}]$, polythionates $[^-O_3S\text{-}S_n\text{-}SO_3^-$, i.e. dithionate, n = 0, trithionate, n = 1, tetrathionate, n = 2, pentathionate, n = 3]

Figure 1.12: The sulfur cycle. It is calculated that 10^8 tonnes per year of sulfur are exchanged between the atmosphere and the soil and waters as sulfide (Kelly, 1980). The Figure is adapted from the Encyclopaedia Britannica online (http://www.britannica.com).

and sulfate $[SO_4^{2-}]$. The multiplicity of sulfur species, along with their pronounced reactivity, makes sulfur chemistry and biochemistry very rich.

Biological sulfur metabolism is part of the natural sulfur cycle (Kelly, 1980; Ehrlich and Newman, 2009) (Figure 1.12). Two metabolic lines include reactions involving sulfur compounds. Assimilatory sulfur metabolism consists of reactions introducing sulfur into biologically active organic molecules[34], while dissimilatory reactions are used to produce cellular energy. Organisms using a dissimilatory sulfur metabolism are predominantly in the bacterial world and can be classified according to whether they use oxidised or reduced sulfur species as electron donors/acceptors (Ehrlich and Newman, 2009).

Sulfur reducing bacteria (SRB) use sulfate, sulfite, thiosulfate or elemental sulfur as terminal electron acceptors. They are anaerobic species belonging to the group of Gram negative δ-proteobacteria, i.e. the genera *Desulfovibrio*, *Desulfomonas* and *Desulfobacter* and some species of *Clostridia*. Sulfate reduction to sulfide involves the enzymes ATP sulfurylase, inorganic pyrophosphatase (ppaC), APS reductase[35], sulfite reductase (SIR, Dsr) and a number of membrane associated complexes such as quinone-interacting membrane-bound oxidoreductase (Qmo), heterodisulfide reductase (Hdr) and Tmc (Pereira, 2008). These organisms are ecologically important because they release reduced sulfur species into the environment, i.e. in marine sediments and in the human colon.

In contrast, sulfur oxidising bacteria (SOB) use reduced sulfur compounds as electron donors. They are either phototrophic or chemotrophic bacteria. Phototrophic SOB include (i) green
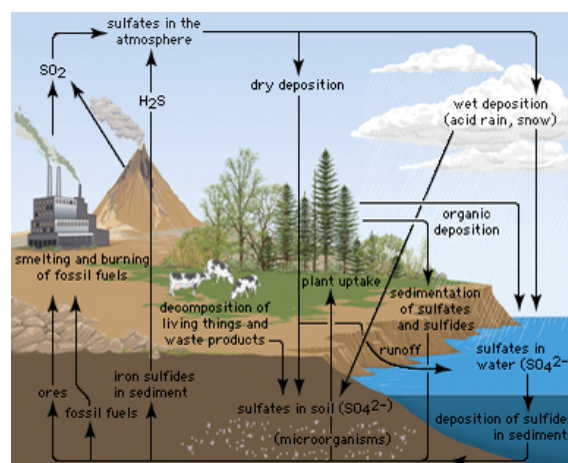
---

[34] Biologically relevant sulfur species are the amino acids cysteine and methionine, iron-sulfur clusters, coenzyme A, biotin, lipoic acid, molybdopterin, thionucleotides (Stacey, 1965) and thiamine.

[35] APS is adenosine 5'-phosphosulfate.

sulfur bacteria (GSB), i.e. of the genus *Chlorobium*, (ii) purple sulfur bacteria (PSB), i.e. γ-proteobacteria of the genus *Chromatium* (e.g. *Allochromatium vinosum*), (iii) purple non-sulfur bacteria, i.e. the genus *Rhodopseumonas*, and (iv) cyanobacteria, i.e. the genera *Oscillatoria*, *Aphanotece*, *Synechococcus* and *Nostoc*. Chemotrophic SOB belong to the class of proteobacteria, i.e. of the genera *Pseudomonas*, *Beggiatoa* and *Thiobacillus*. Many archaeal genera also oxidize sulfur compounds, i.e. *Acidianus* and *Sulfolobus*. Enzymes involved in sulfur oxidation are (i) sulfide:quinone oxidoreductase (SQR) and flavocytochrome *c*:sulfide dehydrogenase (FCSD) that oxidize sulfide to zero-valent sulfur, (ii) Dsr-like enzymes working in reverse that oxidize polysulfides to sulfite, (iii) enzymes of the rhodanese (thiosulfate:cyanide sulfur transferase) family that disproportionate thiosulfate to elemental sulfur and sulfite, (iv) the Sat/Apr/Qmo[36] system or the polysulfide reductase-like complex 3 that oxidize sulfite to sulfate and (v) thiosulfate:quinone oxidoreductase (TQO) and the sulfur-oxidizing enzyme (Sox) that oxidize thiosulfate to sulfate (Dahl and Friedrich, 2008).

In conclusion, it is worth mentioning that also eukaryotes can thrive in sulfur rich habitats and make use of sulfur compounds in dissimilatory reactions (Theissen and Martin, 2008a).

1.5.1. Biological roles of sulfide

Sulfide was first identified by the Swedish pharmacist Carl Wilhelm Scheele (1742 – 1786) (Dobbin, 1931). It is present in different physical states. The diprotonated form, H_2S, is a gas at standard pressure and temperature. In solution, hydrogen sulfide dissociates into the ions HS^- and S^{2-} with respective pK_A values of 6.89 and >14. At physiological pH in water, about one-fourth of free hydrogen sulfide exists as H_2S and three-fourths as HS^- (Furne *et al.*, 2008). Its reduction potential is -270 mV for oxidation to elemental sulfur and -200 mV for oxidation to disulfide (Brune, 1989). In this work, the term "sulfide" refers indistinctly to H_2S, HS^- or S^{2-}, if not mentioned otherwise.

The equilibrium between the gaseous and the water soluble phases confers on sulfide an important ecological role in the sulfur cycle. In fact, sulfur is predominantly exchanged between the atmospheric and the terrestrial or acquatic environments as sulfide (Kelly, 1980).

Sulfide is also a compound of notable biomedical interest.

Its toxicity is long known (Nicholls, 1975). Sulfide binds to iron in heme groups of cytochromes and globins therefore inhibiting respiration. Its LD_{50} in rodents is comparable to that of cyanide and corresponds to 1.1 mg/kg for intravenous administration or 100 – 200 mg/kg for skin or oral exposure (Wilson *et al.*, 2008).

---

[36] Sat stands for sulfate adenylyltransferase, Apr for APS reductase.

Nonetheless, life did originate in a highly sulfidic environment (Theissen *et al.*, 2003) and several organisms including eukaryotes tolerate sulfidic surroundings, such as deep sea hydrothermal vents, marine sediments (Grieshaber and Volkel, 1998; Theissen and Martin, 2008a) and the human colon (Levitt *et al.*, 1999; Wilson *et al.*, 2008). Additionally, sulfide is generated *in vivo* in the catabolism of the amino acids cysteine, methionine and homocysteine. Three enzymes release sulfide as a product, cystathionine β-synthase (CBS), cystathionine γ-lyase (CSE) and 3-mercaptopyruvate sulfurtransferase (3MST) (Stipanuk and Beck, 1982; Kuo *et al.*, 1983; Swaroop *et al.*, 1992). Other sources of sulfide are proteins containing FeS clusters, that release sulfide in acidic conditions or upon denaturation, and proteins containing bound sulfane sulfur, a divalent sulfur atom bound only to other sulfur (i.e. as persulfide or polysulfide) and released under reducing conditions (Kimura, 2010).

Cells are therefore exposed to sulfide. It has been determined that free sulfide concentrations in i.e. the human brain are in the range of 10 nM – 10 μM, depending on the method used. The physiological exposure to sulfide has certainly required the development of sulfide detoxifying pathways. In addition, research efforts of the last decade have demonstrated that, interestingly, sulfide also takes part in a number of important cellular processes. The various metabolic roles of sulfide have been reviewed recently (Kimura, 2010).

Firstly, in bacteria and archaea, sulfide acts as an electron donor in aerobic and anaerobic respiration and in anoxygenic photosynthesis (Arieli *et al.*, 1991; Nübel *et al.*, 2000).

Secondly, the homeostasis of sulfide levels in yeast is related to heavy metal tolerance processes since heavy metal sulfides precipitates at high sulfide levels and are cytotoxic (Shahak and Hauska, 2008).

Finally, the physiological relevance of sulfide in higher organisms, including mammals and in particular humans, is being extensively studied. In the first place, it has been questioned whether the physiological effects attributed to sulfide are occurring at physiological concentrations (Olson, 2009) and whether sulfide is the direct mediator of such effects or rather a "cofactor" potentiating the activity of other compounds (i.e. nitric oxide) (Grossi, 2009b; 2009a). However, it has been established that sulfide is indeed a substrate for human cells (Goubern *et al.*, 2007). Much experimental evidence attributed to sulfide the roles of antioxidant (Kimura, 2010) and nociceptive messenger (Smith, 2009). More interestingly, sulfide was recently classified as a "gasotransmitter", together with nitric oxide (NO) and carbon monoxide (CO). In fact, it possesses all the features of "gasotransmitters" (Wang, 2002; 2004), namely (i) it is a small compound gas at standard temperature and pressure (see above), (ii) it is freely permeable to membranes (Mathai *et*

al., 2009), (iii) it is endogenously generated (see above), (iv) it has specific functions at physiologically relevant concentrations[37], (v) it has specific cellular and molecular targets[38], and (vi) it has physiological functions that can be mimicked by exogenous counterparts, including potential pharmaceuticals already under clinical trials (Szabo, 2007). At a molecular level, sulfide may regulate the function of its targets by forming covalent S-sulfydryl cysteine adducts, in a similar way like nitric oxide (NO) induces the formation of S-nytrosyl cysteines (Mustafa *et al.*, 2009).

Last but not least, sulfide is also associated with many pathological states, including Alzheimer's disease (Kamoun, 2004), Down syndrome (Ichinohe *et al.*, 2005), ulcerative colitis (Roediger *et al.*, 1996; Moore *et al.*, 1997), penile erection disfunctions (d'Emmanuele di Villa Bianca *et al.*, 2009), inflammation (Pae *et al.*, 2009; Bhatia, 2010) and atherosclerosis (Muellner *et al.*, 2009; Schreier *et al.*, 2010).

1.6. Sulfide:quinone oxidoreductase

Sulfide:quinone oxidoreductases (SQRs) are membrane-associated enzymes of great evolutionary importance because they catalyze sulfide-dependent quinone reduction, an essential reaction used by early organisms to thrive in the sulfidic anoxic conditions of oceanic environments (Theissen *et al.*, 2003). At present, SQRs still play a number of physiological roles. They contribute to sulfide detoxification, to sulfide-dependent respiration (Nübel *et al.*, 2000), to anoxigenic photosynthesis (Arieli *et al.*, 1991), to heavy metal tolerance (HMT) in yeast (Vande Weghe and Ow, 1999) and possibly to sulfide homeostasis in mammals (Shahak and Hauska, 2008). Because of their importance, SQR genes have been conserved throughout evolution and the currently available annotated genome information reveals that they are present in a wide range of organisms, including all domains of life and also humans, though being absent in plants (Shahak and Hauska, 2008).

Sequence homology studies reveal that SQRs belong to the superfamily of flavoprotein disulfide oxidoreductases (FDRs).

1.6.1. Flavoprotein disulfide oxidoreductases (FDRs)

Flavoprotein disulfide oxidoreductases (FDRs) are flavoenzymes structurally characterized by the presence of two Rossmann-fold FAD- and NAD(P)H-binding domains and by a C-

---

[37] e.g.: it alters the induction of hippocampal long-term potentiation (LTP) (Eto *et al.*, 2002), it induces smooth muscle relaxation (Zhao *et al.*, 2001; Yang *et al.*, 2008; Chang *et al.*, 2010), it mediates sympathetic neurotransmission (Kulkarni *et al.*, 2009), it induces intracellular acidification in glia cells (Lu *et al.*, 2010), it inhibits insulin signalling (Yang *et al.*, 2005) and it is vasoactive (Benavides *et al.*, 2007; Olson and Whitfield, 2010).

[38] e.g.: NMDA receptors (Eto *et al.*, 2002), Cl^-/HCO_3^- exchanger and Na^+/H^+ exchanger (Lu *et al.*, 2010), Ca^{2+}-channels and Ca^{2+}-dependent K^+ channels (Kombian *et al.*, 1993).

terminal family-specific region, named "interface domain" (Figure 1.13) (Argyrou and Blanchard, 2004). They either catalyze the reduction of disulfide species to thiols or the opposite oxidative reaction.

Three groups of FDRs have been described according to the non-flavin redox active group that they possess (Argyrou and Blanchard, 2004).

Group I FDRs present a CysXXXXCys disulfide/thiol pair near the active site on the *si*-face of FAD (McNaught and Wilkinson,

Figure 1.13: The superfamily of flavoprotein disulfide reductases. FDRs are subdivided into three groups. They all possess two similar domains, one binding FAD ("FAD"), the other binding pyridine-nucleotides ("NAD(P)H"). Instead, their C-terminal region ("Central" and "Interface") is specific for each subfamily. Some proteins with significant sequence similarity to FDRs deviate from this common classification scheme ("Others"). The figure is reproduced from Argyrou and Blanchard, 2004.

1997). They bind FAD non-covalently. Their active oligomeric unit is a dimer and a HisXXXXGlu pair in the "interface domain" of one monomer participates in the reaction catalysed by the partner monomer. Members of group I FDRs are soluble pyridine-nucleotide:disulfide oxidoreductases, like glutathione reductase (E.C.: 1.8.1.7, PDB id.: 3DK4), lipoamide dehydrogenase (1.8.1.4, 3LAD), trypanothione reductase (1.8.1.12, 1FEB) and mycothione reductase (1.8.1.15) (Williams, 1992). Flavocytochrome *c*:sulfide dehydrogenase (FCSD) (1.8.2.-, 1FCD) can also be classified as a group I FDR (Shahak and Hauska, 2008). However, FCSD is exceptional because it oxidizes sulfide and transfers electrons to cytochrome *c*. Additionally, it binds FAD covalently through an 8-α-S-cysteinyl bond (Heuts *et al.*, 2009).

Group II FDRs use a second disulfide/thiol pair located in the "interface domain" on the protein surface, in addition to the one located on the *si*-face of FAD. The two cysteine couples are structurally close to each other, so that reducing equivalents are transferred from the protein core to the surface and become accessible by the substrates of these enzymes, which are much larger than the substrates of group I FDRs. Group II FDRs are high molecular weight thioredoxin reductase (1.8.1.9, 1ZKQ), mercuric ion reductase (1.16.1.1, 1ZK7) and alkylhydroperoxide (Ahp) reductase (1.8.1.-, 1N8J). Low molecular weight thioredoxin reductase (1.8.1.9, 1TRB) can also be classified in this group, but it shows atypical features. It does not possess the interface domain and it possesses only one cystine in a different position than that of group I FDRs. The two cysteines are separated

by two rather than four amino acids and they are located on the *re*-face of FAD. Low molecular weight thioredoxin reductase transfers reducing equivalents to the protein surface through a pronounced conformational change.

Finally, group III FDRs use a cysteine sulfenic acid (Cys-S-OH) or cysteine coenzyme A (Cys-S-S-CoA) group as non-flavin redox centres. They are NADH peroxidase (1.11.1.1, 1JOA), NADH oxidase (1.6.3.1) and coenzyme A disulfide reductase (1.8.1.14, 1YQZ).

1.6.2. Sulfide:quinone oxidoreductases (SQRs)

SQRs (1.8.5.-) belong to group I FDRs. At the beginning of this work, no structural information about any SQR was available. The closest structural model was represented by *Allochromatium vinosum* FCSD (PDB id.: 1FCD). In fact, SQRs oxidize sulfide similarly to FCSDs, while they are not known to reduce disulfide groups, as other group I FDRs do. Besides the limited structural knowledge, only a few SQR-like proteins have been functionally characterized. As a consequence, in some cases it is unclear whether SQR-like gene sequences do encode for true SQRs (Pham *et al.*, 2008; Chan *et al.*, 2009). Finally, the level of sequence similarity among putative SQRs is poor.

The scarce bioinformatic, functional and structural data make SQR classification complicated. Originally, three SQR groups were identified (Theissen *et al.*, 2003). Type I SQRs, found in bacteria like *Aquificales*, *α* and *β Proteobacteria* and *Firmicutes*, were described to have micromolar affinity to their substrates and contribute to cellular respiration (Nübel *et al.*, 2000) and anaerobic photosynthesis (Arieli *et al.*, 1991; Arieli *et al.*, 1994; Griesbeck *et al.*, 2000; Griesbeck *et al.*, 2002). Instead, type II SQRs, found in other bacteria and in mitochondria, were described to have a much lower affinity to sulfide and to require additional reaction partners in order to catalyse a still not well-understood biochemical reaction (Hildebrandt and Grieshaber, 2008; Theissen and Martin, 2008b). Finally, no type III SQR has been characterized yet. More recently, the classification has been extended to five SQR groups plus one uncharacterized group, based on improved sequence alignment calculations that made use of the increased number of available sequences (Pham *et al.*, 2008).

The bioinformatic and enzymatic studies that have been performed on SQRs identified sequence features involved in sulfide oxidation. Most importantly, type I SQRs present three conserved and essential cysteine residues in regions expected to be near the FAD (Griesbeck *et al.*, 2000; Griesbeck *et al.*, 2002), but none corresponds to the cysteine that covalently binds the flavin in FCSD. Therefore, it was believed that type I SQRs bound FAD non-covalently and that, in contrast to group I FDRs, their reaction cycle involved three and not two cysteines. In type II SQRs, one cysteine is substituted by a tyrosine,

which led to the hypothesis that a reaction partner might be needed by this SQR subfamily in order to compensate for the missing cysteine. Moreover, mutagenesis studies indicated that Val 294 (*A. aeolicus* SQR numbering) imparts high sulfide affinity to type I SQRs, because when it is mutated to or replaced by an aspartate, as in the type II SQRs, the K_M for sulfide increases by two orders of magnitude. Finally, Tyr 161, Glu 162 and Lys 312 were proposed to be involved in protonation / deprotonation events (Griesbeck *et al.*, 2002; Shahak and Hauska, 2008).

Upon sulfide oxidation, SQRs transfer two electrons to hydrophobic quinone acceptors. Therefore, SQRs need to interact with the membrane unlike all other FDRs, which are soluble enzymes. Both the membrane interaction and the mechanism of quinone reduction have also been difficult to describe in the absence of structural and functional data. For example, different SQR topology predictions have been proposed. SQRs were depicted either as transmembrane proteins (Guiral *et al.*, 2009) or as peripheral membrane proteins (Arieli *et al.*, 1991). Additionally, some SQR homologues have been reported to bind the membrane tightly and to require detergent treatment for solubilization, while others seem to bind the bilayer superficially and to be detachable by chaotropic agents (Schütz *et al.*, 1997). In addition, several residues were predicted to be involved in quinone reduction on the basis of mutagenesis and sequence conservation studies (Griesbeck *et al.*, 2002; Grüber, 2002). In particular, the hydrophobic amino acids Phe 39, Pro 46 and Leu 136 were proposed to bind quinone, while His 128 and His 193 (*A. aeolicus* SQR numbering) were thought to mediate proton transfer (Griesbeck *et al.*, 2002; Shahak and Hauska, 2008).

Besides the description of putative sites involved in sulfide oxidation and quinone reduction, the predictions about the SQR catalytic mechanism included hypotheses on the product of the reaction. On the basis of extensive but not conclusive biochemical experiments, it was proposed that type I SQRs polymerize sulfide to soluble polysulfide species. For *Rhodobacter caspulatus* SQR, it was suggested that the product of the reaction could be S_{10}^{2-} species, rapidly non-enzymatically hydrolyzed to the more stable S_{4-5}^{2-} (Griesbeck *et al.*, 2002). For type II SQRs, it has been observed that the reaction produces protein-bound persulfides to be converted to sulfite and eventually to thiosulfate by two mitochondrial-matrix protein partners. The latter are a sulfur dioxygenase (ETHE-1), structurally characterized in *Arabidopsis thaliana* (PDB id.: 2GCU) and pharmacologically relevant because its misfunction leads to ethylmalonic encephalopathy, a fatal autosomal recessive disorder, and a sulfurtransferase of the rhodanese family (McCoy *et al.*, 2006; Hildebrandt and Grieshaber, 2008; Tiranti *et al.*, 2009; Lagoutte *et al.*, 2010).

1.7. Aim of the work

The aim of this work is to determine the 3-D structure of the sulfide:quinone oxidoreductase from *Aquifex aeolicus* and to understand its reaction mechanism.

The work consists of (i) the identification and the isolation of the protein from the native membranes of *A. aeolicus* by conventional chromatography, electrophoresis and peptide mass fingerprinting, (ii) the biochemical characterization of the target by spectrophotometric techniques, analytical ultracentrifugation and electron microscopy, (iii) functional studies of the enzyme by means of a spectrophotometric activity assay and (iv) the structure determination of the purified protein complex by 3-D crystallization and X-ray diffraction experiments.

The high resolution 3-D structure will contribute to the clarification of a number of properties of the protein, particularly its oligomeric state, its membrane topology, its thermal stability, the mode of binding of its cofactor FAD and the geometry of its active sites. Ultimately, structure determination aims at defining an experimental basis for the understanding of the intriguing but complicated SQR catalytic mechanism.

The potential cellular roles of SQR and the physiological importance of both its substrates, sulfide and quinone, confer to the study a significant toxicological, microbiological, bioenergetical and biomedical relevance.

2. Results

Genome sequencing (Deckert *et al.*, 1998) indicated that *Aquifex aeolicus* possesses at least one gene with similarity to sulfide:quinone oxidoreductases, namely protein Aq_2186 (accession codes NP_214500, gi|15607118). Aq_2186 (hereafter referred to as SQR) is a protein composed of 430 amino acids, with a predicted molecular weight (MW) of 47.5 kDa, a theoretical isoelectric point (pI) of 6.5 and a GRAVY score of -0.032, indicating moderate hydrophobicity. Different programs predict a different number of transmembrane helices (0 to 3) from the primary structure (Marcia, 2005).

2.1.   Identification

2.1.1. Sulfide:quinone reductase activity in A. aeolicus *native membranes*

A. aeolicus membranes were prepared as described in the Material and Methods section and thawed on ice before use. The membranes possess SQR activity, as detected using a spectrophotometric activity test (see Table 2.2 in Chapter 2.2.1). This result confirmed a previous characterization of *A. aeolicus* membranes (Nübel *et al.*, 2000) and provided evidence for the presence of SQR in the membrane preparation.

2.1.2. Denaturing gel electrophoresis (SDS-PAGE)

The membranes solubilised in 3 % dodecyl-β-D-maltoside (DDM) were first fractionated on an anion exchanger shown in Figure 2.1.

The fractions containing SQR activity were analysed on SDS-PAGE gels. Two fluorescent

gel bands could be identified by illumination with ultraviolet light before staining, indicating the presence of flavoproteins. One band migrates at an apparent molecular weight of approximately 35 kDa and the other at 47 kDa, this latter value being in agreement with the expected molecular weight of the SQR monomer. The lower band is significantly more fluorescent than the upper one. Interestingly, the sample becomes homogeneous for the electrophoretic run after
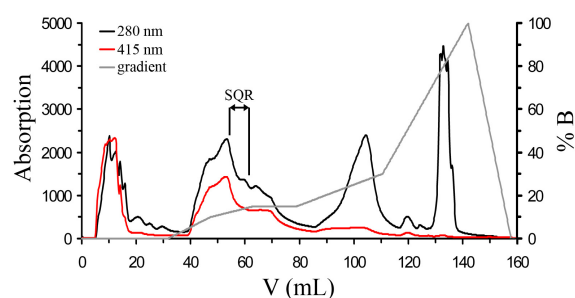
Figure 2.1: Fractionation of total solubilized membranes. The Figure shows the elution profile of total solubilized *A. aeolicus* membrane proteins loaded on the anion exchanger MonoQ. The black line represents the absorption at 280 nm and the red line the absorption at 415 nm. The grey line represents the elution gradient applied. The percentage of buffer B, containing 1 M NaCl (buffer A contains 50 mM NaCl), is indicated on the right axis. The position of the fractions possessing SQR activity and used for further purification is indicated by the black bars.
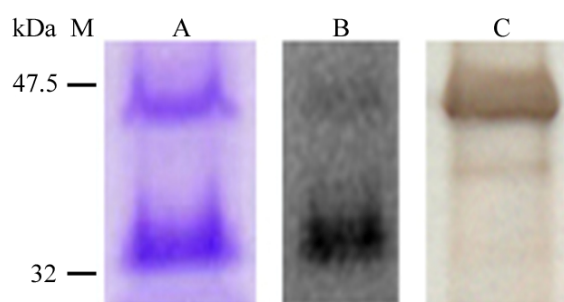
Figure 2.2: SDS-PAGE. Lane A is a Coomassie-brilliant-blue-stained gel. SQR can be identified in two bands migrating at approximately 35 and 47 kDa. Lane B is the same gel lane observed in a UV transilluminator before Coomassie staining. The SQR band at 35 kDa is significantly more fluorescent than the band at 47 kDa, indicating that it contains a flavin. Lane C is a silver stained SDS-PAGE of the same sample boiled for 1 h before the electrophoretic run. Under such conditions, SQR probably denatures completely, loses its fluorescence (not shown), and migrates as a single band at 47 kDa.

extensive boiling (1 h) or addition of urea before gel loading. Upon such treatment, only the band at 47 kDa is visible, but it has lost its fluorescence completely (Figure 2.2).

2.1.3. Peptide mass fingerprint (PMF)

After staining with Coomassie Brilliant Blue, both fluorescent bands were excised and subjected to trypsin digestion. Their peptide mass fingerprint (PMF) was characterized with MALDI-TOF mass spectrometry (T. Beckhaus, Goethe University, Frankfurt am Main, Germany) and easy-nLC-ESI-q-TOF mass spectrometry (Dr. J. Langer, Max Planck Institute of Biophysics, Frankfurt am Main, Germany). Both bands were identified to be protein Aq_2186 and show an identical sequence coverage of 70 % (Table 2.1). It is noteworthy that Met 1 has never been identified by PMF analysis and is also absent in the 3D structure. Therefore, it is likely that this residue is removed after protein synthesis and it is not present in the active form of SQR. Nonetheless, the numbering of the SQR sequence was not changed in order to maintain continuity with the previous literature.

2.2. Purification

2.2.1. Purification from the membrane fraction

SQR was purified in three chromatographic steps subsequent to the first anion exchange run. First, the protein mixture was fractionated through a size-exclusion chromatography (SEC) column (TSK1). Second, impurities were removed by detergent exchange. Finally, a second SEC column (TSK2) was used for a polishing run. The highest monodispersity in the final run is obtained at extreme ionic strength values, either in the absence of salt or with > 1 M NaCl (Marcia, 2005). As low ionic strength is advantageous for crystallization, in this work the pure protein was always kept in the absence of salt, unless otherwise indicated. The purification enrichment was approximately 20 fold with a yield of 32 % (Figure 2.3 and Table 2.2).

Table 2.1: SQR peptide mass fingerprinting (PMF)[a].

Start – End	Observed	MW (exp)	MW (cal)	Delta	Miss	Sequence	Modification	Ion score
2 – 22	1,026.5814	2,051.1483	2,051.164	-0,0157	1	M.AKHVVVIGGGVGGIATAYNLR.N		131
4 – 22	618.3462	1,852.0168	1,852.0319	-0,0151	0	K.HVVVIGGGVGGIATAYNLR.N		84
23 – 29	415.7227	829.4309	829.4368	-0,0059	0	R.NLMPDLK.I		48
23 – 29	846.4303	845.4230	845.4317	-0,0087	0	R.NLMPDLK.I	Oxidation (M)	33
30 – 53	937.1394	2,808.3964	2,808.4210	-0,0246	0	K.ITLISDRPYFGFTPAFPHLAMGWR.K	Oxidation (M)	17
54 – 68	833.9892	1,665.9638	1,665.9705	-0,0067	1	R.KFEDISVPLAPLLPK.F		93
55 – 68	513.6284	1,537.8634	1,537.8756	-0,0122	0	K.FEDISVPLAPLLPK.F		16
69 – 77	577.2939	1,152.5733	1,152.5815	-0,0082	0	K.FNIEFINEK.A		65
69 – 94	717.8476	2,867.3613	2,867.3825	-0,0212	1	K.FNIEFINEKAESIDPDANTVTTQSGK.K		40
78 – 94	867.4067	1,732.7988	1,732.8115	-0,0128	0	K.AESIDPDANTVTTQSGK.K		124
78 – 95	621.3044	1,860.8913	1,860.9065	-0,0152	1	K.AESIDPDANTVTTQSGKK.I		90
95 – 108	805.4394	1,608.8643	1,608.8763	-0,0119	1	K.KIEYDYLVIATGPK.L		110
96 – 108	741.3915	1,480.7684	1,480.7813	-0,0129	0	K.IEYDYLVIATGPK.L		93
178 – 203	919.4702	2,755.3888	2,755.4044	-0,0155	1	R.YKVPMTFITSEPYLGHFGVGGIGASK.R		68
180 – 203	822.4154	2,464.2245	2,464.2461	-0,0216	0	K.VPMTFITSEPYLGHFGVGGIGASK.R		73
180 – 204	656.0885	2,620.3249	2,620.3472	-0,0223	1	K.VPMTFITSEPYLGHFGVGGIGASKR.L		53
204 – 213	624.3362	1,246.6578	1,246.6670	-0,0092	1	K.RLVEDLFAER.N		84
205 – 213	546.2872	1,090.5598	1,090.5659	-0,0061	0	R.LVEDLFAER.N		61
205 – 224	772.4112	2,314.2117	2,314.2321	-0,0204	1	R.LVEDLFAERNIDWIANVAVK.A		66
214 – 224	414.8966	1,241.6679	1,241.6768	-0,0089	0	R.NIDWIANVAVK.A		59
225 – 246	818.0785	2,451.2136	2,451.2281	-0,0146	1	K.AIEPDKVIYEDLNGNTHEVPAK.F		80
247 – 266	1,073.4939	2,144.9732	2,144.9911	-0,0179	0	K.FTMFMPSFQGPEVVASAGDK.V		95
247 – 273	947.4540	2,839.3403	2,839.3673	-0,0271	1	K.FTMFMPSFQGPEVVASAGDKVANPANK.M		30
247 – 273	714.8409	2,855.3343	2,855.3622	-0,0279	1	K.FTMFMPSFQGPEVVASAGDKVANPANK.M	Oxidation (M)	3
247 – 273	958.1184	2,871.3333	2,871.3572	-0,0239	1	K.FTMFMPSFQGPEVVASAGDKVANPANK.M	2 Oxidation (M)	31
274 – 279	731.4171	730.4098	730.4160	-0,0061	0	K.MVIVNR.C		37
274 – 279	747.4102	746.4029	746.4109	-0,008	0	K.MVIVNR.C	Oxidation (M)	26
288 – 303	827.4771	1,652.9396	1,652.9501	-0,0105	0	K.NIFGVGVVTAIPPIEK.T		114
288 – 312	848.8234	2,543.4483	2,543.4727	-0,0244	1	K.NIFGVGVVTAIPPIEKTPIPTGVPK.T		64
304 – 312	455.2701	908.5256	908.5331	-0,0075	0	K.TPIPTGVPK.T		52
313 – 333	782.3728	2,344.0966	2,344.1160	-0,0194	0	K.TGMMIEQMAMAVAHNIVNDIR.N		99
313 – 333	787.7030	2,360.0873	2,360.1109	-0,0236	0	K.TGMMIEQMAMAVAHNIVNDIR.N	Oxidation (M)	83
313 – 333	793.0350	2,376.0833	2,376.1058	-0,0225	0	K.TGMMIEQMAMAVAHNIVNDIR.N	2 Oxidation (M)	24
334 – 342	537.7663	1,073.5180	1,073.5254	-0,0074	1	R.NNPDKYAPR.L		60
374 – 382	584.2863	1,166.5580	1,166.5695	-0,0115	1	K.MGKWAHYFK.T		60
383 – 392	666.8484	1,331.6823	1,331.6914	-0,0091	1	K.TAFEKYFLWK.V		68
393 – 405	730.8728	1,459.7310	1,459.7419	-0,0109	1	K.VRNGNIAPSFEEK.V		76
395 – 405	603.2895	1,204.5645	1,204.5724	-0,0079	0	R.NGNIAPSFEEK.V		78
395 – 412	1,024.5490	2,047.0834	2,047.0990	-0,0155	1	R.NGNIAPSFEEKVLEIFLK.V		19
406 – 412	431.2731	860.5316	860.5371	-0,0055	0	K.VLEIFLK.V		67

[a]: all of the peptides indicated in this table were identified in both of the SDS-PAGE gel bands, no additional peptides were detected.
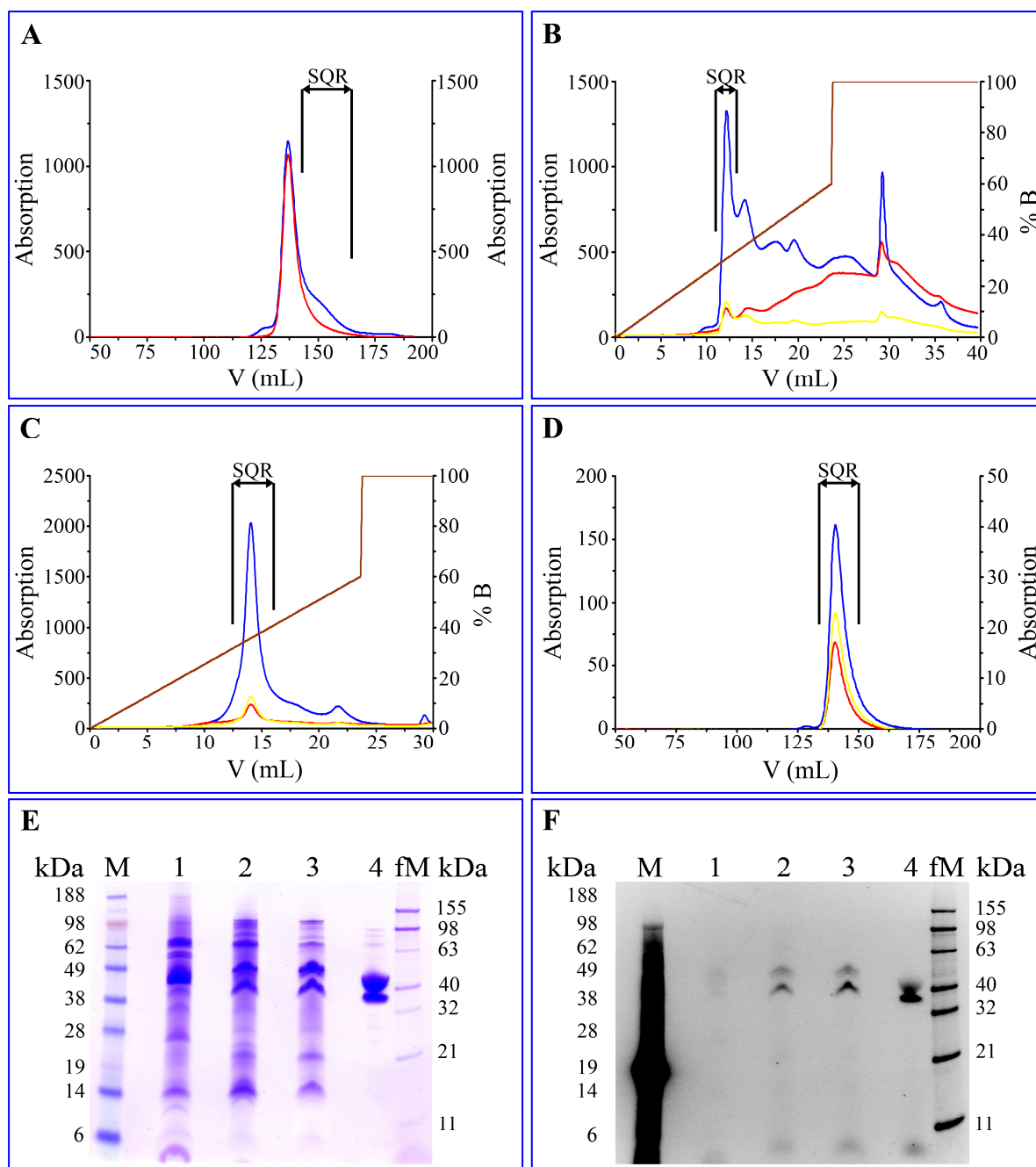
Figure 2.3: Purification steps. Panels A – D show the chromatograms of all purification steps that follow the anion exchange run depicted in Figure 2.1. All chromatograms show the profile of elution monitored at 280 nm (blue), 415 nm (red) and 456 nm (yellow). The gradients used for elution from the anion exchangers are indicated by the brown lines. % B indicates the percentage of buffer B, containing 1 M NaCl (buffer A contains no salt). The fractions containing SQR pooled at each step are indicated by the black bars. Panel A is the elution profile of the first SEC column (TSK1). Panel B is the elution profile after the detergent shift to Zwittergent 3-10. Panel C is the elution profile after the detergent shift back to DDM. Panel D is the elution profile of the second SEC column (TSK2). Panels E and F show the same SDS-PAGE gel illuminated under UV light before staining (F) and stained with Coomassie blue (E). "M" is the SeeBlue® Plus2 Prestained Standard and "fM" the BenchMark™ Fluorescence Standard. The MW of the standard proteins are indicated on the side in kDa. Lane 1 contains the solubilized membranes of *A. aeolicus*, lane 2 the fractions pooled from the anion exchanger column depicted in Figure 2.1, lane 3 the fractions pooled from TSK1 (panel A) and lane 4 those pooled from TSK2 (panel D).

Table 2.2: Purification enhancement[a].

Preparation	Protein amount	Specific activity	Total enzymatic units	Yield
membranes	92 mg	8.83 U/mg	812.4 U	100 %
solubilized membranes	140 mg	5.53 U/mg	774.2 U	95 %
MonoQ	22 mg	24.1 U/mg	530.2 U	65 %
TSK1	10 mg	46.9 U/mg	496.0 U	58 %
detergent exchange	4 mg	82.0 U/mg	328.0 U	40 %
TSK2	3 mg	87.6 U/mg	262.8 U	32 %

[a] the amount of protein was determined using the BCA assay. We find that this assay yields unrealistically low values when unsolubilized membranes are used (Peng *et al.*, 2003).

The purification protocol described in this work improves a previously assessed protocol (Marcia, 2005) to a significant extent, which proved crucial for the structural work. The protein purity was very high as judged by SDS-PAGE, clear native PAGE, analytical ultracentrifugation, *in solution* fingerprint mass spectrometry and single-particle electron microscopy (see following Chapters).

2.2.2. *Necessity of detergent solubilisation*

SQR extraction from the membranes, as well as all purification, kinetic and crystallization studies required the use of detergent. Dodecyl-β-D-maltoside (DDM) and Zwittergent 3-10 were used most extensively. In the absence of detergent micelles SQR is not stable. A time-course stability test was performed by storing the sample at 4 °C and testing its homogeneity by analytical gel filtration. The chromatograms show that the protein is extremely stable (for months) if solubilized in DDM, but quickly aggregates within 24 h in the absence of detergents (Figure 2.4).

Similarly, single-particle electron microscopy shows that the pure protein without detergent forms large aggregates, while the same sample in DDM is monodisperse (Figure 2.5) (Dr. J. Vonck, Max Planck Institute of Biophysics, Frankfurt am Main, Germany).
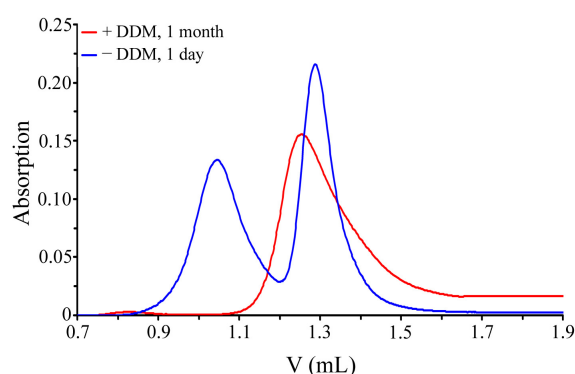
Figure 2.4: Analytical SEC profile in the presence and absence of detergent. SQR is homogeneous in the presence of DDM after 1 month at 4 °C (red line), while it forms larger aggregates in the absence of detergent after 1 day at 4 °C (blue line). The x-axis reports the elution volume (V) in mL. The y-axis reports the absorption monitored at 280 nm.
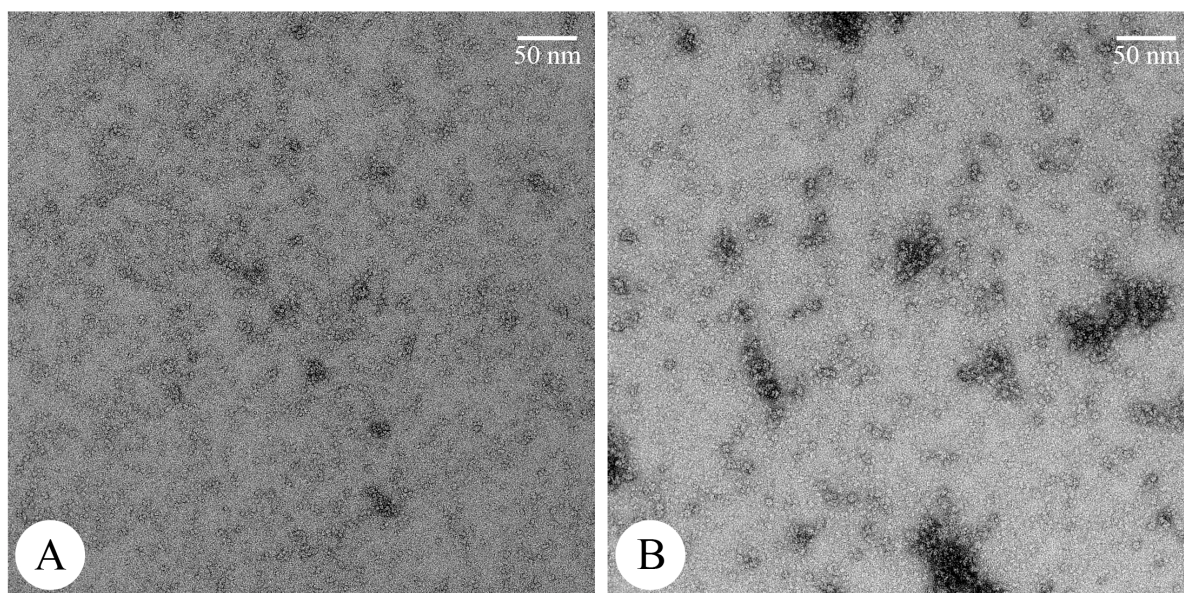
Figure 2.5: Single-particle EM in the presence and absence of detergent. SQR purified in DDM is monodisperse and stable for months at 4 °C (A), while it forms aggregates within one day when the detergent is removed (B). The image was kindly recorded by Dr. J. Vonck, Max Planck Institute of Biophysics, Frankfurt am Main, Germany.

2.2.3. Isoelectric focusing (IEF)

The pI of SQR was experimentally determined to be 4.2 – 4.5 on analytical isoelectric focusing gels incubated with detergent and ampholytes before the run (Figure 2.6). This value deviates significantly from that of 6.5 predicted from the protein sequence. This difference is likely to be related to the topology of SQR (see Chapter 3.2) and in particular to the masking effect that the detergent micelle exerts on its membrane-binding basic face (see Chapter 2.10.1 and Figure 2.34).
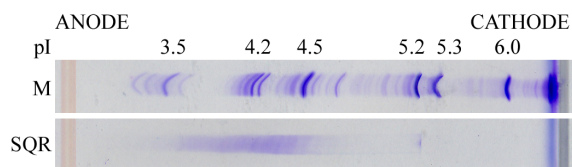
Figure 2.6: IEF. On an analytical isoelectric focusing gel SQR migrates as a smeary band at an approximate pI of 4.2 – 4.5. The gel was incubated with DDM and ampholytes in the pI range 3 – 7 before the run.

2.3. Determination of the oligomeric state

2.3.1. Size-exclusion chromatography (SEC)

A calibrated SEC system (TSK column, Tosoh Bioscience, Japan) was used to determine the apparent MW of SQR based on its retention time. Because of the significant shift observed in the retention time of SQR in buffers of different salt concentrations (Marcia, 2005), two calibration curves were calculated using buffers without salt and with 1 M NaCl. The apparent MW of SQR determined by SEC is about 120 kDa both in absence of salt and
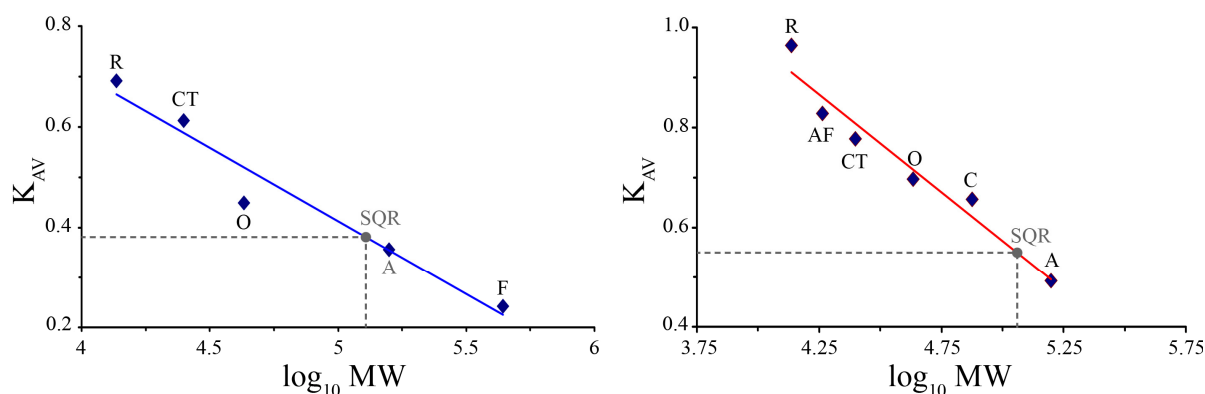
Figure 2.7: SEC calibration curves at low and high ionic strength. The gel permeation column TSK was calibrated in buffer A (TrisHCl pH 7.4 20 mM, DDM 0.05 %, left panel) and in buffer B (buffer A with 1 M NaCl, right panel). The standards used are ribonuclease A (R, MW = 13.7 kDa), apoferritin (AF, 18.3 kDa), chymotrypsinogen A (CT, 25.0 kDa), ovalbumin (O, 43.0 kDa), conalbumin (C, 75.0 kDa), aldolase (A, 158 kDa) and ferritin (F, 440 kDa). The x-axis reports the logarithm of the protein's MW ($\log_{10}$MW). The y-axis reports a function (K_{AV}) of the retention volume (V). $K_{AV} = (V - V_0)/(V_{tot} - V_0)$, V_{tot} corresponding to the total volume and V_0 to the void volume of the column. SQR (grey, dotted lines) results to have an apparent MW of about 120 kDa in both conditions.

in presence of 1 M NaCl (Figure 2.7). The detergent contribution to the total size of the complex is limited to about 15 kDa. This value can be estimated from the experiment described in Figure 2.4, which shows that the retention volume of the non-aggregated SQR population in the absence of detergent is only moderately smaller (1 % of the total column volume) than that of the protein-detergent complex.

2.3.2. Non-denaturing gel electrophoresis (native PAGE)

Four different protocols were used to determine the apparent molecular weight of SQR based on its electrophoretic mobility in a non-denatured state. Each experiment yielded a significantly different result. SQR runs homogeneously as a sharp band in gel systems run
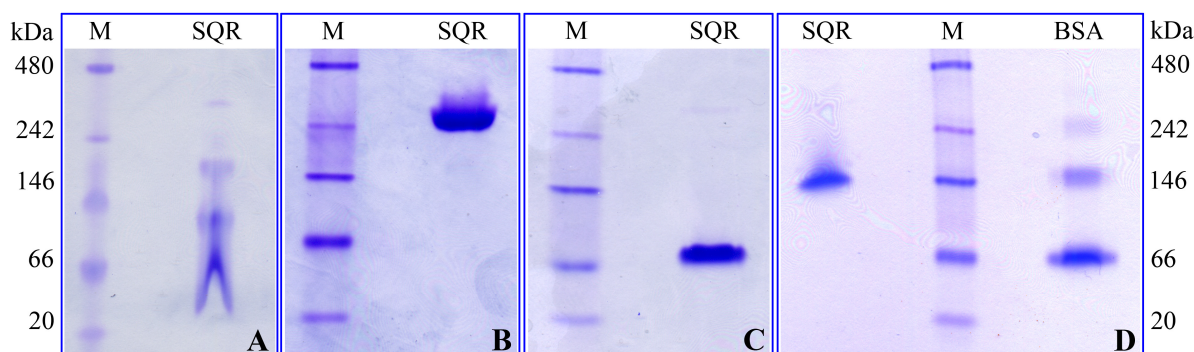
Figure 2.8: Native PAGE. A) BN-PAGE. B) CN-PAGE in DDM. C) CN-PAGE in DDM and DOC. D) CN-PAGE without detergents. M indicates the Native Marker™ Protein Standard, whose MW are indicated on both the left and the right side of the figure. BSA is bovine serum albumin run as an additional reference standard on every gel (shown only for D). BSA is a monomer of 66 kDa but tends to oligomerize to dimeric (132 kDa) and tetrameric (264 kDa) forms. While the migration profiles of the marker proteins and of BSA were not affected by the different experimental conditions, SQR electrophoretic behaviour changes significantly as described in the text.

in the absence of detergents (clear native PAGE), in the presence of DDM (clear native PAGE with detergent) or in the presence of the combination of DDM and DOC (high resolution clear native PAGE, (Wittig *et al.*, 2007)). However, it shows polydispersity on Blue-Native PAGE gels (Schägger and von Jagow, 1991). Additionally, the SQR mobility varies substantially depending on the running system. In the absence of detergent, SQR migrates at an apparent MW of about 150 kDa, with DDM at about 250 kDa, with DDM and DOC at 70 – 80 kDa and in BN-PAGE gels four bands can be detected at about 50 kDa, 90 kDa, 150 kDa and 280 kDa, respectively (Figure 2.8).

2.3.3.   *Laser-induced liquid-bead ion desorption (LILBID) mass spectrometry*

Laser-induced liquid-bead ion desorption mass spectrometry is a technique that analyses full-length protein solutions in their native form at low energy ionization, which not only preserves the full polypeptide sequence but also non-covalent interactions between protein subunits in a complex (Morgner *et al.*, 2006). The largest SQR species observed as single-charged oligomers have a size of 332.5 kDa, corresponding to seven SQR monomers. The series of double-charged species shows that a form with a theoretical mass of 522.5 kDa is also present, corresponding to the assembly of eleven SQR monomers (Figure 2.9). Measurements performed in ultrasoft mode could not determine confidently the size of the
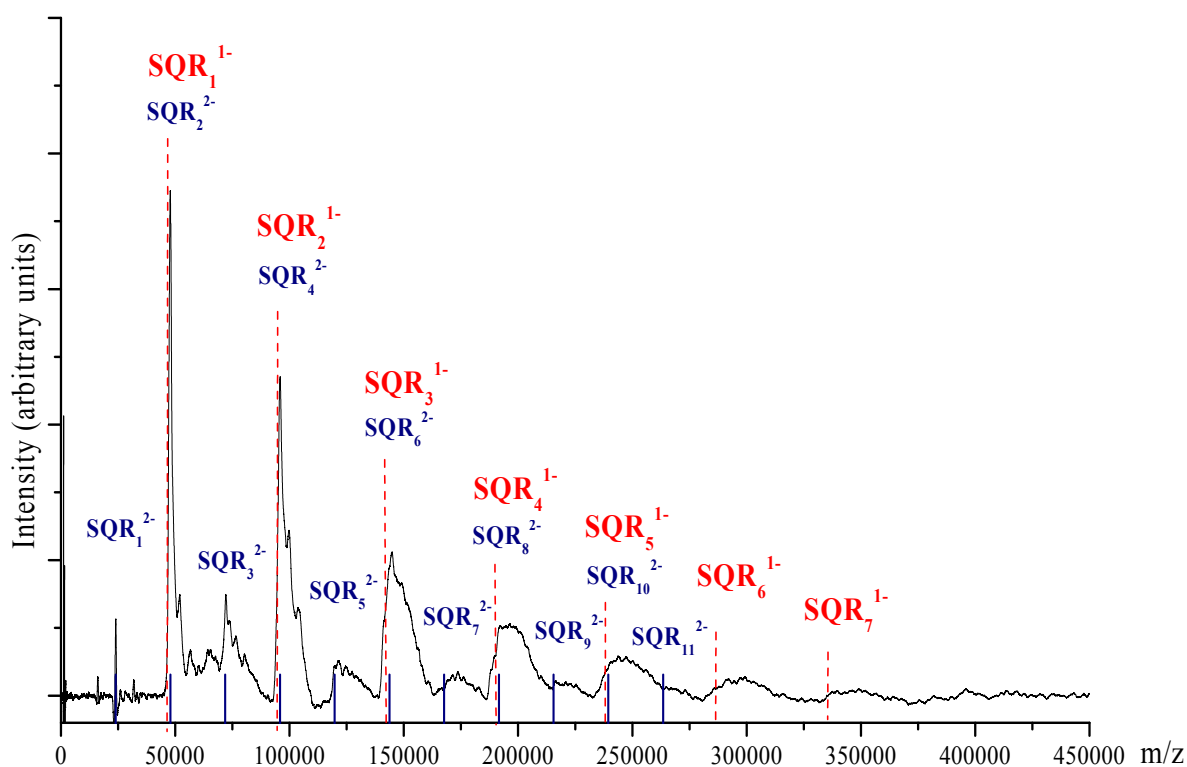
Figure 2.9: LILBID anion mass spectrum. Red bars indicate oligomers carrying a single charge, blue bars those with a double charge.

full SQR complex. Its complete charge distribution was not measurable, presumably because SQR aggregates and is heterogeneous under the experimental conditions (see Chapter 3.2) (L. Sokolova, Goethe University, Frankfurt am Main, Germany).

2.3.4. Crosslinking

SQR was crosslinked with glutaraldehyde and subsequently analysed by SDS-PAGE. Figure 2.10 shows that the highest oligomer that can be formed is the trimeric species.

2.3.5. Analytical ultracentrifugation (AUC)

Sedimentation velocity analytical ultracentrifugation (AUC) runs show that the SQR sample is homogeneous. The main component (89 % of the sample) sediments at $s_{20} = 5.52$ S. Calculation of the apparent MW from this value requires an approximate knowledge of the protein shape (in terms of hydrodynamic radius, also known as the Stokes radius) and is strongly dependent on the protein elliptical shape. Assuming the protein to be globular, the apparent MW is calculated to be about 90 kDa. However, using the dimensions of the protein complex from the crystal structure (55 x 130 Å, see Paragraph 2.10.1), the value shifts to an apparent MW of about 140 kDa, similar to that expected for the trimeric SQR complex (Figure 2.11) (Dr. V. Vogel, Goethe University, Frankfurt am Main, Germany).
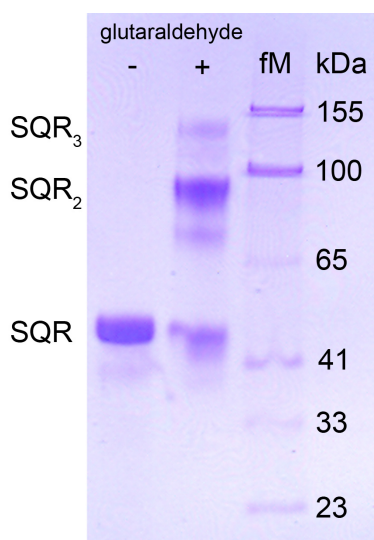
Figure 2.10: Crosslinking. SDS-PAGE gel of SQR treated (+) and untreated (-) with the crosslinker glutaraldehyde and boiled extensively before gel loading. The highest detectable oligomeric species for SQR is the trimer.
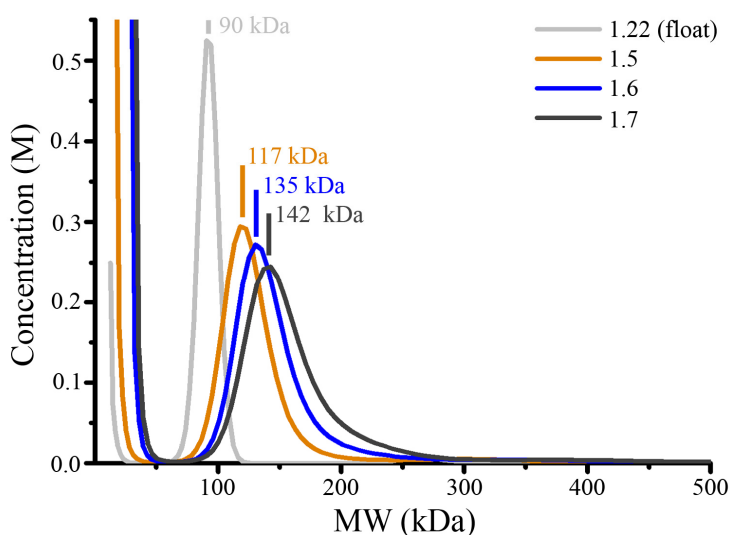
Figure 2.11: Sedimentation velocity AUC. The rough results of the experiment (not shown) were interpreted using different frictional ratio parameters (f/f_0), indicative of the protein shape. The light grey line was calculated using a floating value for f/f_0 eventually converging to 1.22 and corresponds to an apparent MW of 90 kDa. This fitting procedure is optimal for globular samples. The orange, blue and dark grey lines were calculated using fixed f/f_0 values of 1.5, 1.6 and 1.7, respectively, and correspond to MW of 117 kDa, 135 kDa and 142 kDa, respectively. Increasing f/f_0 values best approximate the behaviour of oblong samples, like SQR.

This result was additionally confirmed by sedimentation equilibrium runs in three different density-matching conditions, corresponding to a DDM/sucrose/D_2O mixture, a $C_{12}E_9$/sucrose mixture and a $C_{12}E_9$/D_2O mixture, respectively. The size of the protein complex determined by sedimentation equilibrium AUC is 160 kDa (± 10 %).

2.3.6.   Single-particle electron microscopy

Negative staining single-particle electron microscopy shows that SQR is monodisperse both in the presence and in the absence of salt. The particles were identified, filtered, averaged and aligned in self-organising maps from a neural network analysis. The result shows a similar protein shape in buffers without salt and with 1 M NaCl, respectively. The protein in the presence of salt is more compact than in the absence of salt, which correlates with the different retention factor from the SEC column (see Chapter 2.3.1). The majority of the particles are composed of three strong electron density regions. Dimeric and tetrameric particles are also observed, probably due to the different orientations of SQR on the carbon grid (Figure 2.12) (Dr. D. Parcej, Max Planck Institute of Biophysics, Frankfurt am Main, Germany).
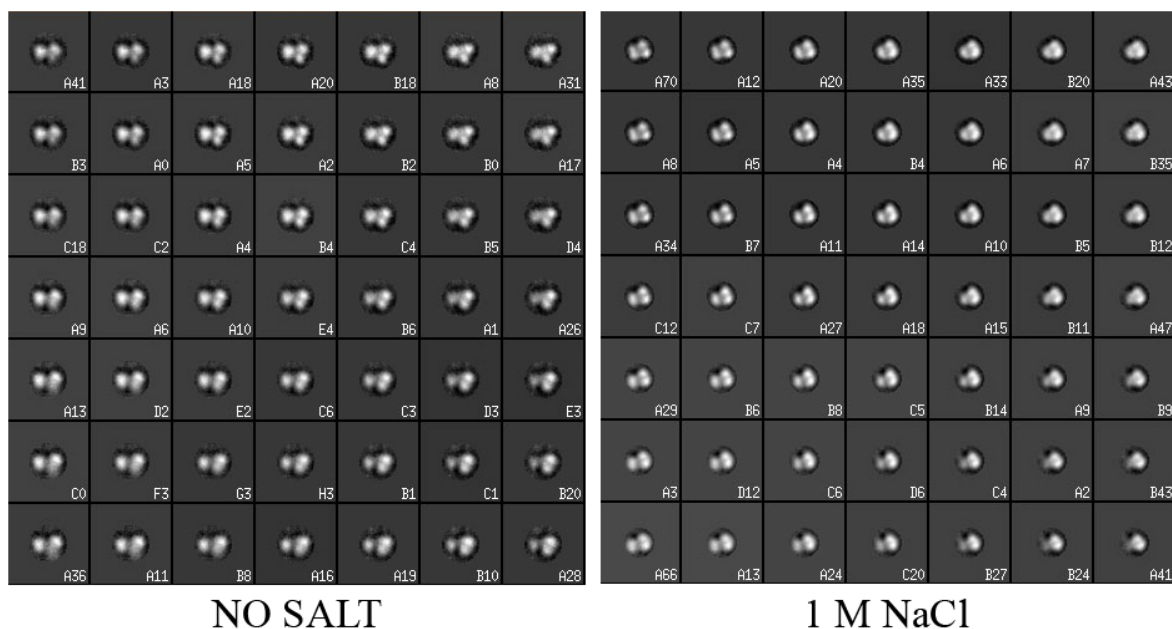
Figure 2.12: Single-particle EM. Self-organising maps from a neural network analysis of SQR single particles recorded by electron microscopy. The map on the left is produced from the sample purified in the absence of salt, while the map on the right from the sample purified in the presence of 1 M NaCl. The numbers indicate the number of particles corresponding to each node.

2.4. Characterization of the cofactor FAD

2.4.1. Full-length Matrix-Assisted Laser-Desorption Ionization Time-of-Flight mass spectrometry (MALDI-TOF MS)

Full length MALDI-TOF MS shows two m/Z signals for the purified SQR at values of 47,517 Da and 48,292 Da, respectively. The difference in mass approximately corresponds to the mass of FAD confirming that this is the flavin cofactor of SQR, as expected by homology with other FDRs (Shahak and Hauska, 2008) (Figure 2.13) (Dr. J. Langer, Max Planck Institute of Biophysics, Frankfurt am Main, Germany and Dr. U. Bahr, Goethe University, Frankfurt am Main, Germany).
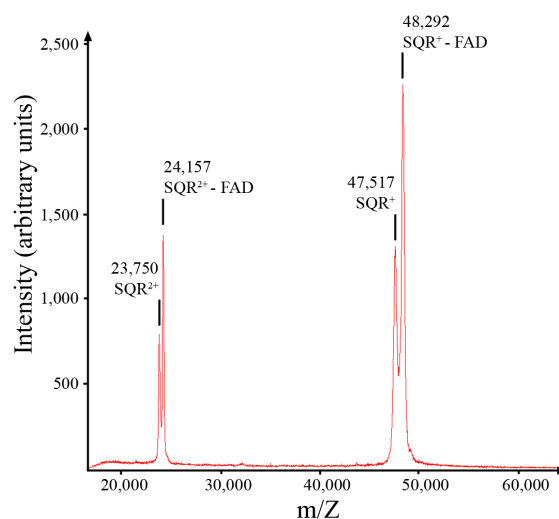
Figure 2.13: Full-length MALDI-TOF mass spectrum. SQR gives two signals at 47,517 Da and 48,292 Da. The difference between the two peaks (775 Da $\pm$ 50 Da) is in the range of the MW of the cofactor FAD.

2.4.2. VIS-UV absorption and fluorescence spectra

The SQR VIS-UV spectrum shows the typical flavin absorption bands at 356 nm and 456 nm in the air-oxidised sample, in addition to the absorption peak at 280 nm, typical of FAD and of the tryptophan residues in the protein (Figure 2.14). From the oxidised-minus-
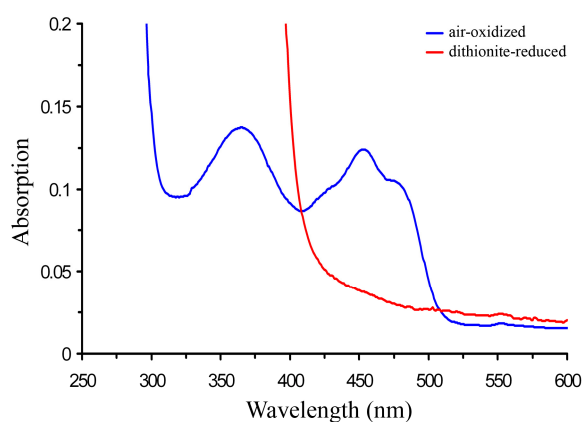
Figure 2.14: VIS-UV spectra. Air-oxidized SQR (blue) shows two characteristic absorption peaks at 365 and 456 nm respectively, which disappear in the dithionite-reduced (red) form.
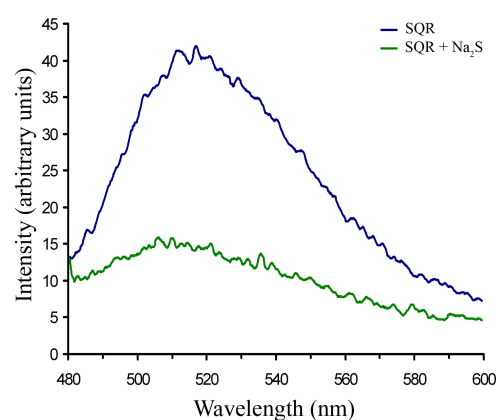
Figure 2.15: Fluorescence emission spectra. SQR was excited at 365 nm and its fluorescence recorded before (blue) and after (red) addition of Na_2S.
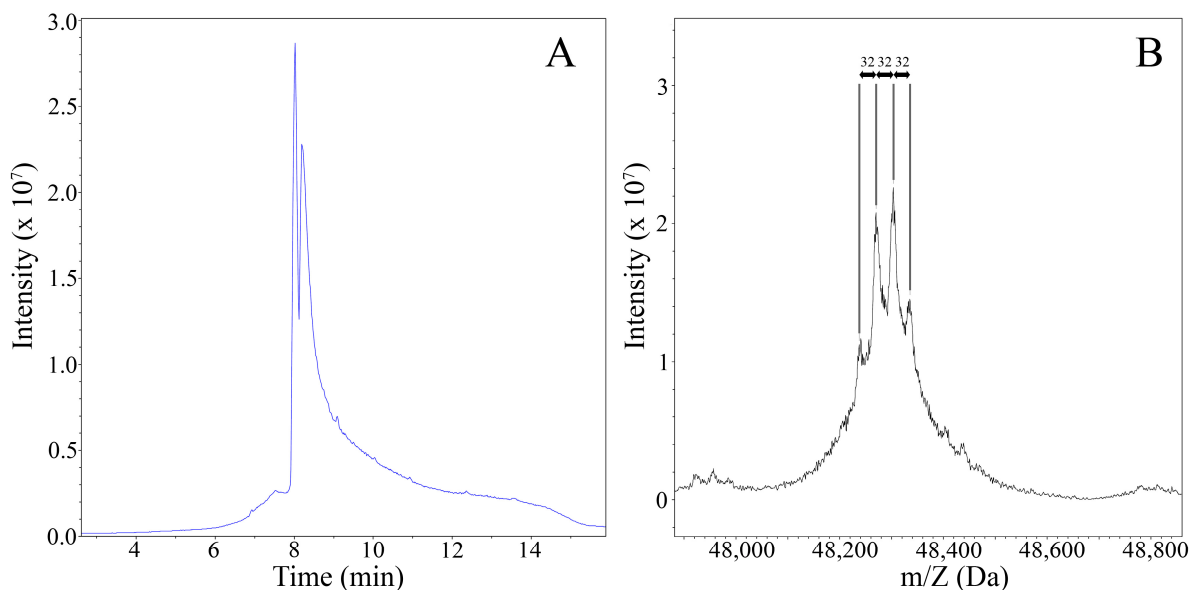
Figure 2.16: High resolution full-length easy-nLC-ESI-q-TOF mass spectrum. A) Chromatogram of the separation of full-length SQR on the reverse-phase nano-LC column. Two peaks can be separated, likely corresponding to two species with different oxidation states, as explained in the text. B) Deconvolution of the spectrum of the first eluting peak. Four species are detectable with a difference of 32 Da in MW, which corresponds to the atomic mass of sulfur.

reduced difference spectrum it can be calculated that there is one FAD molecule per protein monomer, as expected by homology with other FDRs (Shahak and Hauska, 2008). The protein also shows the typical FAD fluorescence emission band at 550 nm upon excitation at 365 nm. The intensity of the signal decreases upon addition of sulfide, as had been shown comprehensively in a previous work on the *R. capsulatus* SQR (Griesbeck *et al.*, 2002) (Figure 2.15).

2.5.   Characterization of additionally bound ligands

2.5.1.   *Total X-ray fluorescence (TXRF) spectroscopy*

TXRF spectroscopy was used to investigate the presence of possible additional metal cofactors in the sample. The results show that SQR does not contain any metal centres and that FAD is therefore the only cofactor. Unexpectedly, however, a careful analysis of the TXRF data showed that the protein contains a higher number of sulfur atoms than expected from the primary structure. Seven cysteines and thirteen methionines, that is 20 sulfur atoms, are expected from the sequence, but a calibrated TXRF experiment showed the presence of 30.8 sulfur atoms for every 2 phosphate atoms (one FAD molecule contains a diphosphate group). Although preliminary, this result suggests the presence of 10 – 11 sulfur atoms copurified with SQR (C. Rittmeyer and Dr. S. Metz, Goethe University, Frankfurt am Main, Germany).

2.5.2. Full-length Liquid Chromatography Electrospray Ionization quadrupole Time-of-Flight mass spectrometry (easy-nLC-ESI-q-TOF MS)

The sulfur content in the protein was investigated by high resolution full-length ESI-q-TOF mass spectrometry. The purified protein solution was separated by analytical HPLC into two fractions corresponding to two SQR species. Deconvolution of the spectra showed that each species displays peak series with 32 Da increments, corresponding to the atomic weight of sulfur. At least 4 peaks were detected for each series (Figure 2.16). The two species differed by 16 Da in size, indicating a single oxidation for the species with the biggest retention volume (Dr. J. Langer, Max Planck Institute of Biophysics, Frankfurt am Main, Germany).

2.5.3. Other techniques

The sulfur content of SQR was also investigated, unsuccessfully, by other techniques, such as peptide mass fingerprint mass spectrometry (PMF-MS), full length MALDI-TOF MS, high energy LILBID MS and X-ray diffraction after treating the crystals with reducing agents. In addition, other methodologies, such as sulfur K-edge XANES, radioactive labelling with $Na_2{}^{35}S$ and ^{33}S NMR have not been attempted yet because, although promising, they are either too costly or inaccessible. Table 2.3 summarizes all attempts performed to characterize the sulfur species in SQR.

Table 2.3: Characterization of the sulfur species bound to SQR.

Technique	Comment
TXRF	10 – 11 extra sulfur atoms every 2 phosphate atoms
ESI-QTOF MS	chain of minimum 4 sulfur atoms

Crystallization
   ▪ reduction before crystallization
 	
   ▪ reduction *in crystal*
	
	protein denatured or could not be crystallized
	when crystallized, the structure showed an unmodified putative polysulfur chain (see Chapter 2.10.4)
	crystals dissolved
	when stable, the structure showed an unmodified putative polysulfur chain (see Chapter 2.10.4)

PMF MS	fragments containing putative sulfur modification not identified
MALDI-TOF MS	below experimental resolution
LILBID MS	below experimental resolution
S K-edge XANES	no beam time allocated at ID 26, ESRF, Grenoble, France
$Na_2{}^{35}S$	too costly
^{33}S NMR	isotope not available as sulfide; too high protein amount required

2.6.   Characterization of the enzymatic activity

The enzymatic activity of purified SQR was determined via a spectrophotometric assay (Nübel *et al.*, 2000). The reaction buffer (50 mM Bis-TrisHCl, pH 7.0) was chosen due to the observation that the range of possible pH tests is limited by the chemical properties of the substrates used. Quinones are unstable at basic pH (Morton, 1965), whereas Na_2S develops $H_2S_{(g)}$ if the pH is acidic.
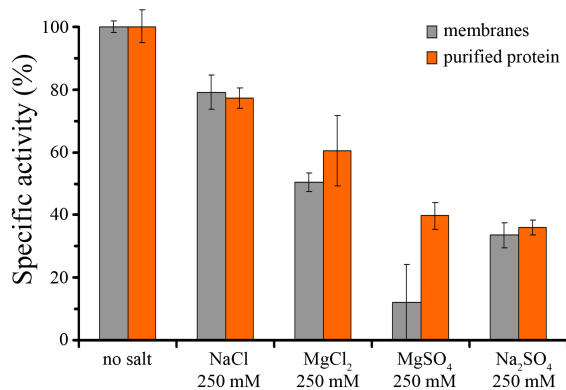
Figure 2.17: Salt dependency of the SQR enzymatic activity. The SQR activity was tested in buffers containing different salts at a concentration of 250 mM. The grey bars correspond to the activity measured in the membranes of *A. aeolicus*, while the red bars to the activity measured on the purified protein. The specific activity is indicated in percentage, with 100 % as the activity of the sample in the absence of salts.

The results confirm and extend a previous characterization of *A. aeolicus* SQR (Marcia, 2005). The Michaelis-Menten constants (K_M) for the substrates are in line with those determined for the non-solubilized enzyme in the membranes of *A. aeolicus* (Nübel *et al.*, 2000) and for the purified *A. aeolicus* SQR heterologously expressed in *Escherichia coli* (Schödl, 2003). Both sulfide and the quinones have micromolar affinity to the enzyme.

As expected for a hyperthermophilic enzyme, the activity is temperature-dependent and it is highest at 80 °C (the highest temperature tested), close to the optimal bacterial growth conditions. The activation energy of SQR is about 14 kJ/(K·mol). Particularly remarkable is the stability of the purified complex over time at high temperature. The protein incubated at 80 °C for longer than one day retains 50 % activity. Furthermore, SQR treatment with quinone

Table 2.4: Kinetic parameters.

K_M-decylUQ	2.16 ± 0.19 µM
K_M-UQ_1	5.40 ± 0.22 µM
K_M-UQ_4	1.6 ± 0.28 µM
K_M-UQ_9	6.43 ± 0.20 µM
K_M-Na_2S	5.94 ± 0.52 µM
specific activity at 40 °C	71.17 ± 4.28 U/mg
activation energy	14.08 ± 0.36 kJ/(K•mol)
$t_{1/2}$ at 80 °C	32 h
IC_{50} antimycin (in 10 µM decylUQ)	15 ± 2.5 µM
IC_{50} NaCl	250 mM
pH optimum	6.5

analogues (e.g. antimycin) shows a concentration-dependent, competitive mechanism of inhibition. The presence or absence of detergent in the reaction buffer does not affect the specific activity significantly. Under aerobic conditions 80 % of the activity is retained. Finally, the ionic strength – which influences the shape of the protein significantly (see Chapter 2.3.6) – affects its activity as well. Increasing salt concentrations were observed to reduce the activity of SQR, with 250 mM NaCl reducing it to 50 % with respect to an assay performed in absence of salt. Additionally, different salts show a different degree of inhibition (Figure 2.17). The calculated activity parameters are summarized in Table 2.4.

2.7. Three-dimensional crystallization

2.7.1. Crystallization by vapour diffusion

SQR could be crystallized in dodecyl-ß-D-maltoside (DDM) by hanging- and sitting-drop vapour diffusion using two different precipitating agents, MPD and ammonium sulfate.

In the MPD condition (Figure 2.18)[37], crystals were grown with 35 – 45 % (v/v) MPD in the absence of any additional salts or buffering agents. However, addition of 50 – 200 mM sodium phosphate makes the crystallization process faster and improves the crystal quality. The crystals grown in the MPD condition diffract X-rays and their resolution was improved from 10 Å (initial hit) to 3.5 Å.

In the ammonium sulfate condition (Figure 2.19)[38], crystals could only be grown at a concentration of 2.0 ± 0.1 M ammonium sulfate and exclusively in the presence of PEG as

additional precipitating agent (ideally 2 – 5 % PEG 400). The pH compatible with crystal formation ranges from 4.0 to 7.0 units and can be buffered with sodium citrate, sodium acetate or sodium 2-(N-morpholino)-ethanesulfonate (MES). In such conditions, the reservoir solution is already very close to saturation. If the crystallization drop is left open under the microscope,
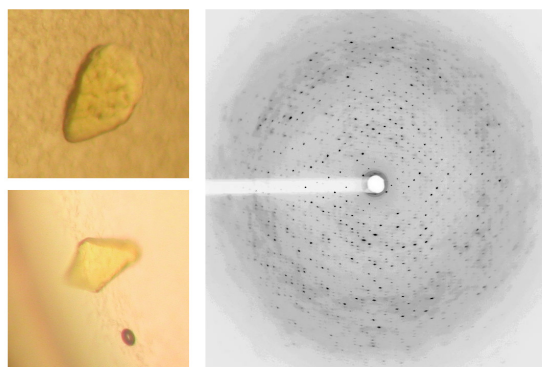
Figure 2.18: Crystallization in MPD. Left panels: two crystals grown by vapour diffusion in the MPD condition. Right panel: one of their representative diffraction patterns. The resolution at the edge is 3 Å.

---

[37] Initial hits were obtained from condition D1, Jena Bioscience Classic screen 7 (JB7_D1, 35 % MPD) and from condition 39, Quiagen MPD Suite screen (40 % MPD, 0.2 M sodium phosphate).

[38] Initial hits were obtained from condition C3, Jena Bioscience Classic screen 6 (JB6_C3, 0.1 M Na-MES pH 6.5, 2.0 M ammonium sulfate, 5 % PEG 400).
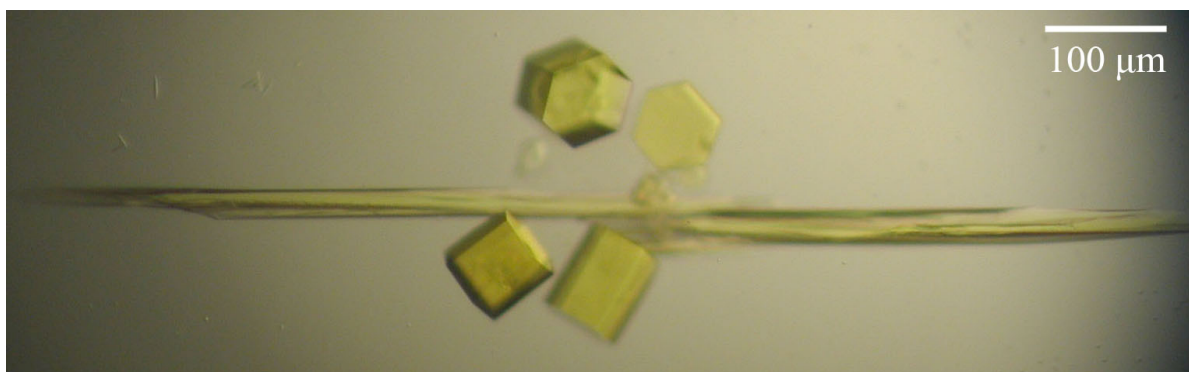
Figure 2.19: Crystallization in ammonium sulfate and PEG. Hexagonal and needle crystals grown in the same crystallization drop. The scale bar is indicated in the top right corner in micrometer.

phase separation occurs within 30 s and salt precipitation after about 120 s, making crystal handling particularly difficult. Two crystal forms can be grown in the ammonium sulfate condition and they often appear in the same crystallization drops. They have a hexagonal, prism shape and a thin, elongated shape. They will be hereafter referred to as the "hexagonal" and the "needle" crystals, respectively. The hexagonal crystals are thick and single, they are approximately $100 - 300$ µm big in all dimensions and grow in about $1 - 3$ weeks. They diffract X-rays to a resolution of $1.5 - 2.0$ Å. The needle crystals are thin and fragile, are often branched or broken along their longer axis, reach a size of approximately $10 \times 50 \times (200 - 1000)$ µm^3 and take longer to grow than the hexagonal crystals (up to 6 weeks). Their resolution could be improved from 4 Å to 1.8 Å. Both crystal forms tend to redissolve 2 weeks – 2 months after their formation. Further details about their biochemical and diffraction properties are reported in Chapters 2.7.3 and 2.8.1.

2.7.2. *Effect of physical and chemical factors on the crystallization by vapour diffusion*

Extensive screening of all physical and chemical parameters that affect the crystallization process was performed in the attempt to improve the crystal size, resolution and stability and the speed of crystal growth. These experiments included screening of temperature, pH, precipitating agents, the presence or absence of oxygen, detergents used to keep the protein in solution, detergents used as crystallization additives, other small molecules used as crystallization additives, crystallization plate types and the scale of the crystallization experiment. Table 2.5 presents a summary of the crystallization screening experiments.

Table 2.5: Crystallization screening.

Screening parameter	Crystals	Diffraction
Temperature		
10 °C	spherulites	–
16 °C	spherulites hexagonal, 4 weeks	– high mosaicity
18 °C	hexagonal, 1 – 3 weeks needle, 2 – 6 weeks	high mosaicity used for structure solution
25 °C	hexagonal, 1 – 3 days	high mosaicity
buffer, pH		
Na-citrate, 4.5 – 5.5	microcrystals	–
Na-acetate, 4.5 – 6.5	hexagonal, 1 – 3 weeks needle, 2 – 6 weeks	high mosaicity used for structure solution
Na-ADA, 6.0 – 7.0	microcrystals	–
Na-MES, 6.0 – 7.0	hexagonal, 1 – 3 weeks needle, 2 – 6 weeks	high mosaicity used for structure solution
TrisHCl, 7.5 – 8.5	hexagonal, 1 – 3 weeks	high mosaicity
Na-HEPES, 7.0 – 8.5	hexagonal, 1 – 3 weeks	high mosaicity
CAPSO, 9.5	–	–
Oxygen		
aerobic condition	hexagonal, 1 – 3 weeks needle, 2 – 6 weeks	high mosaicity used for structure solution
anaerobic condition	hexagonal, 1 – 3 weeks	high mosaicity
Precipitating agent (PEG)		
PEG 200 (< 5 % v/v)	hexagonal, 8 weeks	high mosaicity
PEG 400 (2 – 5 % v/v)	hexagonal, 1 – 3 weeks needle, 2 – 6 weeks	high mosaicity used for structure solution
PEG 350 MME (3 – 7 %)	–	–
PEG 550 MME (< 5 %)	hexagonal, 1 – 5 weeks	high mosaicity
PEG longer chain	phase separation	–
Detergents (after extensive detergent shift)		
UM	hexagonal	high mosaicity
α-DDM	hexagonal	high mosaicity
NG	hexagonal	high mosaicity
LDAO	phase separation	–

Zwittergent 3-10	phase separation	–

Small molecules additives (to the ammonium sulfate condition)

- ▪ **Small inorganic salts**

$FeSO_4$	microcrystals, > 8 weeks	–
$(NH_4)Fe(SO_4)_2$	microcrystals, > 8 weeks	–
NaCl	–	–
NaBr	needles	3 – 4 Å

- ▪ **Organic salts**

Na,K-tartrate	needles	3 – 4 Å
citrate trisodium salt	needles	3 – 4 Å
ATP disodium salt	needles	3 – 4 Å

- ▪ **Organic amines**

taurine	plates	4 – 6 Å
sarcosine	plates	4 – 6 Å
spermidine	plates	4 – 6 Å
D-pipecolic acid	plates	4 – 6 Å
glycyl-glycyl-glycine	microcrystals	–
urea	microcrystals	–
betaine	microcrystals	–
spermidine tetrachloride	microcrystals	–
proline	clear drop	–

- ▪ **Antioxidants**

glutathione	microcrystals	–
nicotinic acid	needles	too fragile
DTNB	microcrystals	–
ascorbate	microcrystals	–

- ▪ **Others**

DMSO	hexagonal	high mosaicity

Detergents (as additives to the ammonium sulfate condition)

- ▪ **Maltosides**

NM	phase separation	–
DM	microcrystals	–
UM	microcrystals	–

DDM	hexagonal, 1 – 3 weeks needle, 2 – 6 weeks	high mosaicity used for structure solution
13M	plates	3.5 – 4 Å
14M	plates	3.5 – 4 Å
16M	microcrystals	–

- **Thiomaltosides**

OTM	plates	3.5 – 4 Å
NTM	microcrystals	–
DTM	plates	3.5 – 4 Å

- **Cyclohexylmaltosides**

CYMAL-1	microcrystals	–
CYMAL-2	clear drop	–
CYMAL-3	plates	high mosaicity
CYMAL-4	microcrystals	–
CYMAL-5	plates	4 Å
CYMAL-6	plates, too fragile	–

- **Glucosides**

6G	clear drop	–
OG	microcrystals	–
NG	hexagonal plates	high mosaicity 3 Å

- **Thioglucosides**

7TG	precipitate	–
OTG	microcrystals	–
NTG	microcrystals	–

- **Alkylphoscholines**

FOS8	plates	10 Å
FOS9	plates	2.5 Å
FOS10	microcrystals	–
FOS12	microcrystals	–

- **Hydroxyethylglucamides (HEGA)**

HEGA8	hexagonal	high mosaicity
HEGA9	needles, too fragile	–

- **Cyclo-hydroxyethylglucamides (c-HEGA)**

c-HEGA8	clear drop	–
c-HEGA9	needles, too fragile	–
c-HEGA10	plates, < 2 weeks	3.5 Å
c-HEGA11	plates	–

- **Methylglucamides**

MEGA8	plates	–
MEGA9	plate	3 Å

- **Sucrose monoalkylesthers**

OS	hexagonal	high mosaicity
DS	hexagonal	–
LS	hexagonal	7 Å

- **Dimethylamine-N-oxide**

DDAO	microcrystals	–
LDAO	phase separation	–

- **Ammoniopropansulfonate**

Zwittergent 3-8	hexagonal	mosaicity
Zwittergent 3-10	prism plates	mosaicity 2.8 Å, space group $P2_1$
Zwittergent 3-12	plates	mosaicity
Zwittergent 3-14	microcrystals	–

- **Polyoxyethylenalkylethers**

$C_{12}E_8$	plates	3 Å
$C_{12}E_9$	microcrystals	–
Thesit	microcrystals	–

- **Tetramethylbutylphenylpolyoxyethylene**

Triton X-100	microcrystals	–

- **Others**

CTAB	microcrystals	–
DeoxyBigChap	plates	5 Å
HECAMEG	plates	3.5 - 4 Å
BAM	small plates	–
DDMAB	phase separation	–
CHAPSO	precipitate	–

2.7.3. *Biochemical characterization of the crystals obtained by vapour diffusion*

Both the hexagonal and the needle crystal forms obtained with ammonium sulfate and PEG by vapour diffusion were characterized biochemically. Their electrophoretic properties on SDS-PAGE are similar and they reproduce the behaviour of the pure protein in solution (Figure 2.20).

Their absorption spectrum was recorded in the visible range using a microspectrophotometer and differs slightly between the two forms, possibly due to the different polarization effect caused by the regular arrangement of the molecules in the two crystal lattices. Both FAD absorption bands are visible (Figure 2.21). Thirdly, both crystal forms were analysed by full-length MALDI-TOF mass spectrometry and show a similar profile, analogous to that of the pure protein in solution. Finally, an activity assay performed using dissolved crystals shows that both crystal forms retain similar enzymatic rates (± 20 %) as the purified protein solution.
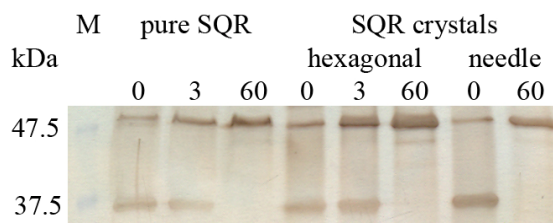
Figure 2.20: SDS-PAGE of the crystals. Silver-stained SDS-PAGE of the SQR solution and of the hexagonal and needle crystals. The samples were boiled for 0, 3 or 60 min before loading onto the gels as indicated in the third row. Due to lack of sample, the needle crystals were only boiled for 0 or 60 min. The lane marked with "M" is the Prestained Protein Marker, Broad Range. The MW of the two marker proteins visible in this Figure is indicated in kDa on the left.
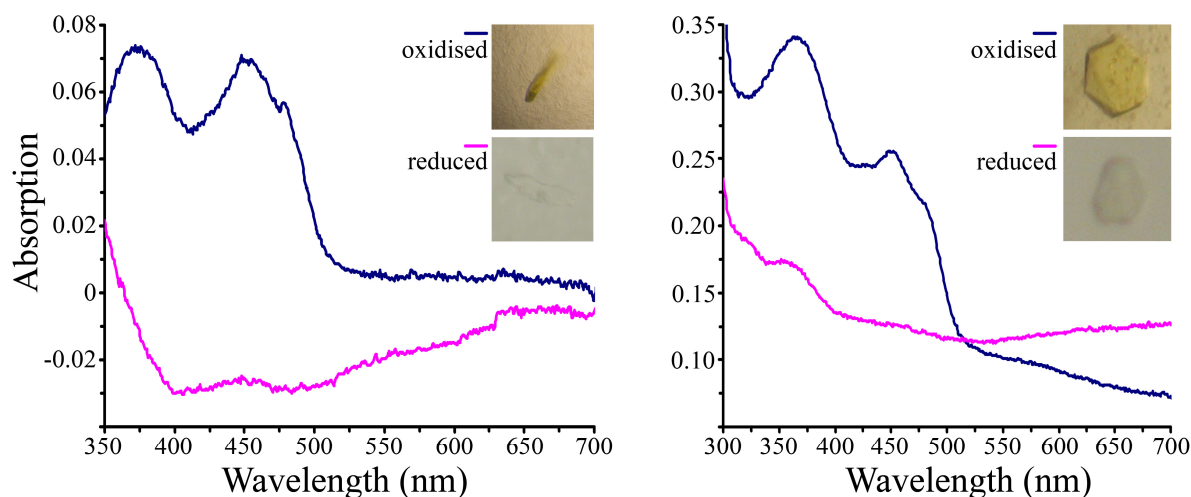
Figure 2.21: VIS spectra of the crystals. The absorption spectra of the needle crystals are shown in the left panel, those of the hexagonal crystals in the right panel. The spectra of the air-oxidised forms are shown in dark blue, those of the dithionite-reduced forms in light blue. The insets show the crystals used for the measurement, in the crystallization drop (oxidised state, upper inset) and in the quartz capillary (reduced state, lower inset), respectively.

2.7.4. Crystallization with substrates and inhibitors

Three strategies were followed in order to crystallize SQR with substrates, substrate analogues or inhibitors. (i) Copurification was performed by mixing SQR with a 10-fold excess of the compounds before the last chromatographic step (TSK2). (ii)
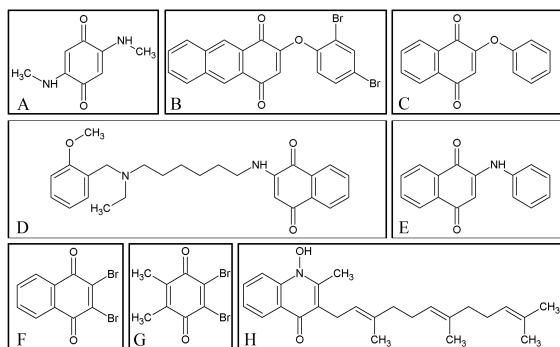
Figure 2.22: Quinones analogues used for crystallization screening. A – E) Quinone derivatives kindly provided by Prof. Bolognesi, Bologna University, Italy. F) 2,3-dibromo-1,4-naphtoquinone (Sigma-Aldrich). G) 2,3-dibromo-5,6-dimethyl-1,4-benzoquinone (Sigma-Aldrich). H) Aurachin C, kind gift of the Helmholtz Institut für Infektionsforschung, Braunschweig, Germany.

Cocrystallization was performed by mixing SQR with a 10-fold excess of the compounds immediately prior to crystallization (t_0). (iii) Soaking was performed by adding a concentrated solution of the compound either to the crystallization drop after the crystals had formed or directly to the cryo-protecting solution. The soaking experiments were carried out for 1 – 24 h. Table 2.6 summarizes the crystallization screening trials in the presence of substrates and inhibitors. The chemical structures of the most important quinone analogues used are shown in Figure 2.22.

2.7.5. Crystallization by other techniques

Three other crystallization techniques were also tested. In lipidic cubic phase setups (Landau and Rosenbusch, 1996) crystals could not be grown. By sponge phase (Wadsten *et al.*, 2006) crystals could be grown in a condition similar to the MPD condition already obtained by vapour diffusion. They were small and diffracted X-rays to about 8 Å resolution (Figure 2.23). Finally, in microbatch under paraffin oil, crystals were obtained in
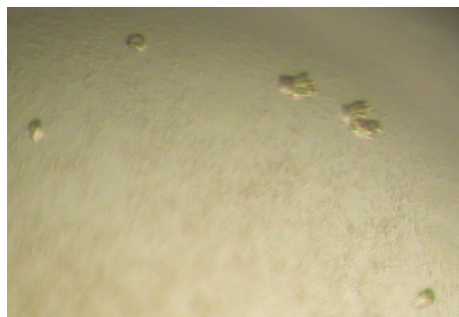
Figure 2.23: Crystals grown in sponge phase.
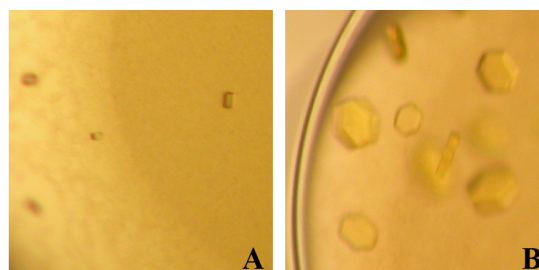
Figure 2.24: Crystals grown in microbatch under paraffin oil. A) Crystals grown in condition JBM1_A1. B) Crystals grown in condition JB6_C3.

condition A1 of the Jena Bioscience Membrane screen 1 (JBM1_A1)[39] and in condition C3 of the Jena Bioscience Classic screen 6 (JB6)[40]. The crystals grown in JBM1_A1 were too small for diffraction studies, while those grown in JB6 had the same characteristics as the hexagonal crystals grown by vapour diffusion (Figure 2.24). Table 2.7 summarizes the crystallization attempts using techniques different from the vapour diffusion method.

Table 2.6: Crystallization screening in the presence of substrates and inhibitors.

Additive	Crystals	Diffraction	Electron density
Copurification (addition of additive before TSK2)			
Decylubiquinone	hexagonal	high mosaicity	–
Antimycin	needle, too fragile	–	–
Na_2S	hexagonal	high mosaicity	–
Cocrystallization (addition of additive to the drop, at t_0)			
Na_2S (anaerobic)	hexagonal	high mosaicity	–
Stigmatellin	needles	–	–
Antimycin	needle, too fragile	–	–
Decylubiquinone	hexagonal needle	high mosaicity 2.5 Å	– NP[a]
UQ_1	hexagonal	high mosaicity	–
UQ_2	hexagonal	high mosaicity	–
UQ_4	hexagonal	high mosaicity	–
UQ_9	not grown	–	–
DMN	not grown	–	–
Soaking (addition of additive to the drop after crystal growth or to the cryo-protecting solution)			
Na_2S (anaerobic)	needle	2 – 3 Å	NP
Decylubiquinone	needle	2 – 3 Å	structure solved
Aurachin C	needle	2 – 3 Å	structure solved
Antimycin	needle	2 – 3 Å	NP
Stigmatellin	needle	2 – 3 Å	NP
Other quinones (Figure 2.22)	needle	2 – 3 Å	NP

[a] NP indicates a condition for which the structure was solved by rigid body refinement, but the ligand could not be identified in the electron density map.

---

[39] 15 % w/v PEG 400, 15 % w/v Glycerol, 100 mM Hepes sodium salt pH 7.5, 200 mM $CaCl_2$.

[40] 2 M ammonium sulfate, 5 % w/v PEG 400, 100 mM MES sodium salt pH 6.5.

Table 2.7: Crystallization screening using techniques different from vapour diffusion.

Technique	Condition	Resolution	Remarks
Lipidic cubic phase			
	no crystal growth	–	–
Sponge phase			
	MPD / AmPO$_4$	8 Å	small crystals
Microbatch			
	JBM1_A1	–	microcrystals
	JB6_C3	3 Å	high mosaicity

2.8. X-ray diffraction

2.8.1. X-ray diffraction patterns of the hexagonal and needle crystals

The two most promising crystal forms were the hexagonal and the needle crystals. Despite their biochemical properties being relatively similar (see Chapter 2.7.3), the diffraction of these two crystal forms is tremendously different.

The hexagonal crystals diffract X-rays to a higher resolution (on average 2.0 Å, at best 1.5 Å) but their diffraction pattern shows such high mosaicity – in terms of reflection size – that processing of the data is impossible. In most cases the diffraction does not show single
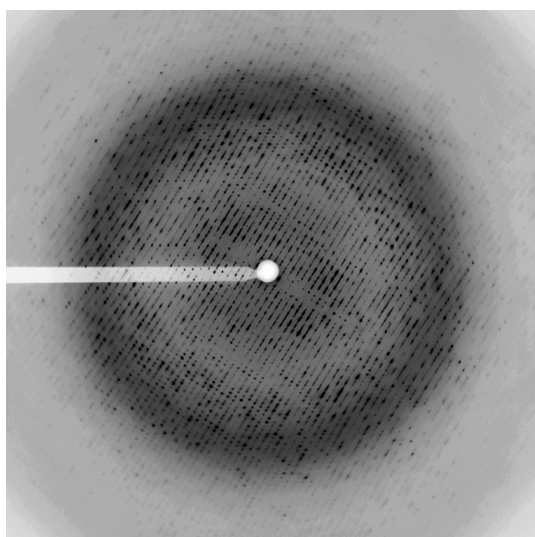
Figure 2.25: Diffraction of the hexagonal crystals. The diffraction pattern of the hexagonal crystals is characterized by high mosaicity and it is not possible to index it. The resolution at the edge is 1.5 Å.
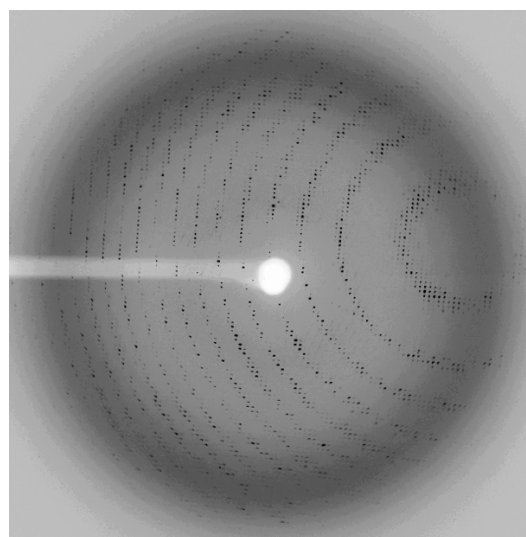
Figure 2.26: Diffraction of the needle crystals. The diffraction pattern of the needle crystals shows single spots and is suitable for structure determination. The resolution at the edge is 2.0 Å.

Table 2.8: Crystallographic data for the hexagonal and needle crystals.

Parameter	Hexagonal crystals	Needle crystals
best resolution	1.5 Å	1.8 Å
space group	–	$P2_12_12_1$
unit cell dimensions	–	112 x 154 x 178 Å^3
$N_{molecules}$ in AU	–	6
Matthews coefficient	–	2.67 Å^3/Da
Solvent content	–	53 %

spots, but rather smeared lines. Interestingly, the intensity of diffraction of the hexagonal crystals changes dramatically depending on the temperature at which data are collected. The same crystal showed a very weak diffraction pattern at room temperature (cooled with a jet of cold air), but the intensity became much stronger after freezing (with a jet of liquid nitrogen vapour). Nonetheless, at both temperatures, the diffraction is characterized by high mosaicity (Figure 2.25).

The needle crystals are instead characterized by a more regular lattice, resulting in a diffraction pattern of good quality, but they diffract X-rays to a lower resolution (on average 3 – 4 Å, at best 1.8 Å, Figure 2.26). Therefore, they are suitable for data collection. They belong to the space group $P2_12_12_1$ with unit cell dimensions of about 112 x 154 x 178 Å^3 and they are generally isomorphous. The asymmetric unit contains six SQR molecules, which corresponds to a Matthews coefficient of 2.67 Å^3/Da and to a solvent content of 53 %. Table 2.8 summarizes the crystallographic data of the hexagonal and needle crystals.

2.8.2. *Attempts to improve X-ray diffraction of the hexagonal and needle crystals*

Both the hexagonal and the needle crystal forms required optimization for data collection and data processing.

The greatest difficulty with the hexagonal crystal form was the high mosaicity. Besides the screening of a number of parameters to improve crystallization (see Chapter 2.7) other attempts aimed at improving the diffraction quality on the crystals themselves. These included treatment of the crystals in a humidity chamber setup[41] (Dr. Christian Benda, Max Planck Institute of Biochemistry, Martinsried, Germany), extensive screening of

---

[41] The humidity chamber set-up consists of a jacket to surround the loop on the goniometer. The jacket is connected to a humidifier and guarantees a constant vapour jet parallel to the loop pin and directed towards the crystal. The humidity of the vapour jet can be controlled, thereby inducing swelling or shrinking of the crystal in the loop. Such alterations of the crystal packing have been in some cases useful for improving the diffraction quality (Sanchez-Weatherby *et al.*, 2009). The measurement on site has to be carried out at room temperature. Alternatively, the set-up can be used to treat and optimize the crystal, which is then immediately flash frozen and measured separately.

Table 2.9: Screening of cryoprotecting solution.

Cryoprotectant	Effect
mother liquor	crystals unstable
ammonium sulfate > 2 M	salt spots, ice rings
glycerol in the mother liquor	loss of diffraction
glycerol in buffer	loss of diffraction
PEG 400 in the mother liquor	immediate phase separation, crystals unstable
PEG 400 in buffer < 30 %	crystals dissolve
PEG 400 in buffer 40 – 50 % v/v	used for structure solution and soaking experiments, crystals stable for > 24 h
PEG 1000 or longer chains	loss of diffraction
perfluoropolyether (Hampton, 100 %)	salt spots, ice rings

cryoprotecting solutions (Table 2.9), seeding, recrystallization (Figure 2.27), annealing and optimization of the data collection parameters (detector distance, oscillation angle). None of these attempts improved the mosaicity of the hexagonal crystals.

The problems with the needle crystal form were the high fragility and the high sensitivity to radiation damage. The latter problem was tackled using antioxidants as crystallization additives (see Table 2.5), performing data collection with a highly filtered X-ray beam and merging multiple data sets. Improving the purification protocol was however the crucial determinant for obtaining complete and redundant data sets. The detergent exchange between the purification steps TSK1 and TSK2 improves the quality of the preparation, most likely because Zwittergent 3-10 separates SQR from traces of a NiFe hydrogenase B-type cytochrome subunit, a heme- and Fe-S cluster-containing protein (Brugna-Guiral *et al.*, 2003) responsible for a dominant spectral band in the heme absorption region detected in an early stage of the project (Marcia, 2005).

Figure 2.27: Recrystallized SQR. The crystals shown in this Figure were obtained by recrystallization of the hexagonal crystals. 50 hexagonal crystals were fished and washed in the cryo-protecting solution, dissolved in 2 µL water, mixed with an equal volume of the original mother liquor and equilibrated against the same solution in sitting drop vapour diffusion set-ups. The diffraction pattern of recrystallized SQR is still characterized by high mosaicity (not shown).

2.9. Determination of the structure

2.9.1. Heavy metal soaking

The SQR structure had to be solved by experimental phase determination because attempts to solve the phase problem by molecular replacement using *A. vinosum* FCSD coordinates (PDB id.: 1FCD) were unsuccessful. In the absence of a heterologous expression system to incorporate seleno-methionines into the protein sequence recombinantly, SQR was derivatized with heavy metal compounds in order to exploit multiple isomorphous replacement and anomalous scattering. A number of compounds were tested following three strategies. (i) Cocrystallization with the derivatives typically did not permit crystal growth. (ii) Soaking in the crystallization drop was also unsuccessful due to the instability of the drop (see Chapter 2.7.1) and probably because of the high concentration (2 M) of ammonium ions, which tend to form complexes with the heavy metals, preventing their binding to the protein (Sigler and Blow, 1965). (iii) Soaking in the cryoprotectant solution was successful. Table 2.10 summarises the results of the heavy metal screening experiments.

The presence of the metal in the crystals was detected by fluorescence X-ray scans at the synchrotron beam line. The fluorescence absorption spectrum of the crystals also determines the wavelengths with maximal f' and f'' to use for anomalous scattering data collection. Figure 2.28 shows the fluorescence absorption spectrum of a needle crystal derivatized with K_2OsCl_6. For certain compounds, metal soaking into the crystals could be conveniently monitored visually, as a change in the crystal colour (Figure 2.29).

Table 2.10: Screening of heavy atom derivatives.

Additive	Resolution	Absorption edge (peak energy in eV)		Sites
K_2OsCl_6	3 Å	Os L-III	(10,879.4)	12
$AuCl_3$	3 Å	Au L-III	(11,925.7)	33
$Hg(CH_3COO)$	3 Å	Hg L-III	(12,322.9)	not found
Na_3IrCl_6	3 Å	Ir L-III	(11,221.1)	not found
$Lu(CH_3COO)$	3 Å	Lu L-III	(9,283.4)	not found
K_2PtBr_4	3 Å	Br K	(13,509.8)	not found
K_2PtBr_4	3 Å	Pt L-III	(11,570.5)	not found
$Sm(CH_3COO)$	3 Å	Sm L-II	(7,313.6)	not found
K_2OsO_4	6 Å	–	–	–
$KAu(CN)_2$	10 Å	–	–	–
Ethyl mercury phosphate	no diffraction	–	–	–
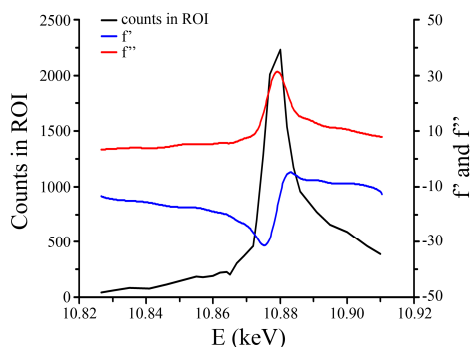
Figure 2.28: X-ray fluorescence spectrum. A needle crystal derivatized with K_2OsCl_6 was exposed to filtered X-ray radiation in the energy (in keV) range around the Os L-III absorption edge. The left Y-axis shows the fluorescence emission counts in the region of interest (ROI, black line). The right Y-axis shows the values of f' (blue) and f'' (red) derived from the fluorescence spectrum.
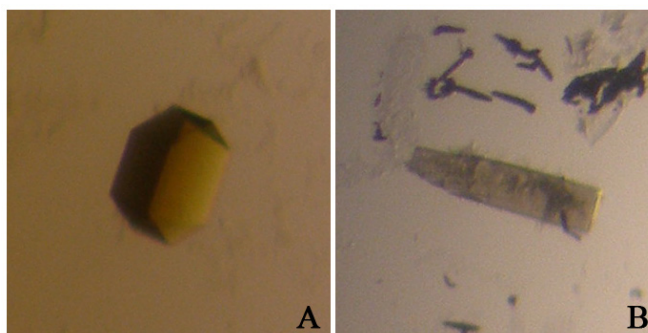
Figure 2.29: Crystals derivatized with heavy metals. Hexagonal (A) and needle (B) crystals soaked with K_2OsCl_6. After a few hours the crystals change colour from the original yellow to brown, indicating accumulation of the heavy metal in the crystal lattice.

2.9.2. Phase determination by Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS method)

The phases were determined by the MIRAS method using crystals derivatized with K_2OsCl_6 and $AuCl_3$. The metal binding sites were determined with SHELX (Sheldrick *et al.*, 2001) and refined with SHARP (Fortelle and Bricogne, 1997). 12 Os binding sites are particularly clear and give the predominant contribution to the phasing process. They are also beautifully symmetric, revealing the arrangement of the SQR molecules in the asymmetric unit (Figure 2.30). Automatic interpretation of the initial electron density map was poor, therefore the protein sequence was modelled manually starting from a 15-amino-acid peptide correctly positioned by ARP/wARP (Lamzin *et al.*, 2001). The initial model was refined to a final R_{free} of 23.5 % (R_{work} is 19.2 %) in the resolution range 20.0 – 2.30 Å, using TLS refinement and NCS averaging of the six

Figure 2.30: Heavy metal binding sites. These 12 Os binding sites in the asymmetric unit of the needle crystals were used together with 33 Au binding sites (not shown) to calculate the initial phases of the SQR structure.

monomers in the asymmetric unit. Final rounds of refinements were done without NCS. Table 2.11 summarizes the data collection statistics for the derivatized crystals.

Table 2.11: Crystallographic statistics for the derivative data sets used for MIRAS phasing[a].

Parameter	Derivative (Os)			Derivative (Au)
Space group	$P2_12_12_1$			$P2_12_12_1$
Unit cell dimensions (Å)	a = 112.30 b = 155.86 c = 178.26			a = 112.10 b = 154.30 c = 176.70
	Peak	**Inflection**	**Remote**	**Peak**
Wavelength (Å)	1.14000	1.14050	1.13010	1.03864
Resolution (Å)	50.0 – 3.5 (3.6 – 3.5)	50.0 – 3.5 (3.6 – 3.5)	50.0 – 4.0 (4.1 – 4.0)	50.0 – 2.3 (2.4 – 2.3)
R_{merge} (%)	12.3 (38.2)	12.8 (42.4)	8.2 (19.2)	6.2 (8.9)
$I/\sigma(I)$	12.69 (4.30)	12.26 (3.96)	15.77 (8.41)	15.38 (12.38)
Completeness (%)	98.9 (98.9)	98.9 (99.1)	98.7 (99.2)	99.7 (99.9)
Redundancy	3.36 (3.34)	3.37 (3.42)	3.33 (3.42)	3.83 (3.42)

[a] the numbers in parenthesis refer to the highest resolution shell.

2.9.3. Structure determination by rigid body refinement of quinone- and aurachin C-bound SQR and at high X-ray energy

The crystals of decylubiquinone-bound and aurachin C-bound SQR were isomorphous with the ones of the protein "as-purified". One data set of the protein bound to decylubiquinone was collected at high energy wavelength (near the Fe absorption edge, 6.5 keV) in order to study the residual anomalous scattering of sulfur. Also this data set was isomorphous to the data set used for structure solution. The corresponding structures were determined by rigid body refinement introducing the appropriate ligands modeled with the program Sketcher of the CCP4 suite (Collaborative computational project number 4, 1994). They have an RMSD of about 0.4 – 0.5 Å with respect to the "as-purified" structure. Table 2.12 summarizes the corresponding crystallographic data.

Table 2.12: Data collection and refinement statistics for the SQR structures solved[a].

Parameter	"As purified"	Decylubiquinone	Aurachin C	6.5 keV
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å)	a = 112.85	a = 110.78	a = 111.18	a = 110.78
	b = 154.92	b = 154.01	b = 154.23	b = 154.01
	c = 178.04	c = 175.55	c = 176.53	c = 175.55
Wavelength (Å)	0.91984	1.00150	1.00148	1.90744
Resolution (Å)	50.0 – 2.30	50.0 – 2.00	50.0 – 2.90	50.0 – 2.30
	(2.40 – 2.30)	(2.10 – 2.00)	(3.00 – 2.90)	(2.40 – 2.30)
R_{merge} (%)	8.3 (34.4)	10.5 (76.5)	8.3 (39.6)	12.4 (58.6)
$I/\sigma(I)$	14.87 (3.44)	13.17 (1.86)	16.74 (4.05)	9.97 (2.40)
Completeness (%)	97.9 (87.9)	96.2 (76.7)	98.8 (95.7)	91.1 (75.9)
Redundancy	3.47 (2.47)	4.74 (1.86)	3.72 (3.41)	2.54 (1.63)
Refinement				
Resolution (Å)	2.30	2.00	2.90	2.30
Total reflections	471,639	921,962	249,689	596,959
R_{work} / R_{free} (%)	19.2 / 23.5	20.0 / 23.7	19.4 / 24.1	20.3 / 24.7
RMSD				
▪ bond length (Å)	0.010	0.008	0.006	0.009
▪ bond angles (°)	1.420	1.179	0.989	1.308
Number of atoms				
▪ protein	20,136	20,029	20,094	20,029
▪ cofactor (FAD)	318	318	318	318
▪ detergent (DDM)	210	210	210	210
▪ water	571	1,142	194	546
▪ ions (SO_4^{2-})	120	120	120	120
▪ ligand	–	138	168	138
▪ sulfur	54	50	59	50
▪ buffer (MES)	72	–	–	–

[a] the numbers in parenthesis refer to the highest resolution shell, R_{work} and R_{free} were calculated according to eq. 1.2.6.1 (see Chapter 1.2.6).

2.10. Description of the structure

2.10.1. Overall structure

The six monomers in the asymmetric unit of *A. aeolicus* SQR are associated in two homotrimeric units. Each trimer is elliptical in shape, with dimensions of 130 Å in diameter and 50 – 60 Å in thickness (Figure 2.31). Additionally, the surface of the trimer is characterized by a pronounced polarity, as visualized on the basis of electrostatic potential calculations performed using GRASP (Nicholls *et al.*, 1993). One side of the trimer, formed by the two Rossmann-fold

Figure 2.31: Biological unit. SQR forms a trimer with a thickness of approximately 55 Å (left). When seen from its soluble face (right), the complex can be inscribed in a circle with a radius of about 65 Å. Each monomer is shown in a cartoon representation and coloured in a different tone of blue. All Figures showing the structure were generated with PyMOL (www.pymol.org).

domains, shows an overall negative surface potential, while the other side is characterized by an excess of positive charges and a pronounced hydrophobic patch in each monomer (Figure 2.32). Oligomerization contacts are mediated by two protein regions. In helices 202 – 213, near the central three-fold axis, Arg 204 of each monomer forms salt bridges with
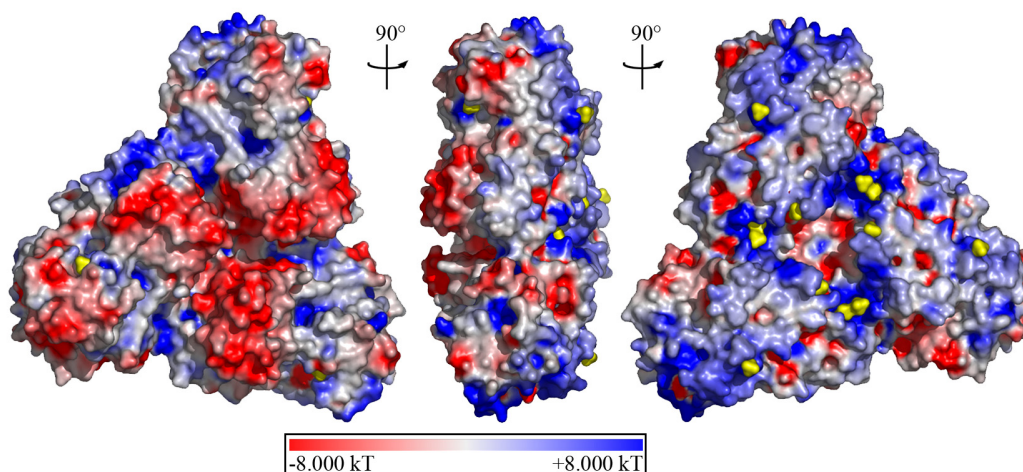
Figure 2.32: Electrostatic surface potential. Left panel: the SQR trimer has one overall negative (red) surface corresponding to the Rossmann domains facing the solvent. Right panel: rotation of the trimer by 180° shows the domain that mediates the interaction with the membrane. It has an overall intense positive charge (blue) with hydrophobic (white) patches corresponding to helices 376 – 395 and 400 – 412. Sulfate ions and MES molecules are shown in yellow. The electrostatic surface potential was calculated using the software GRASP excluding solvent molecules. The scale bar is indicated in the inset at the bottom of the Figure.
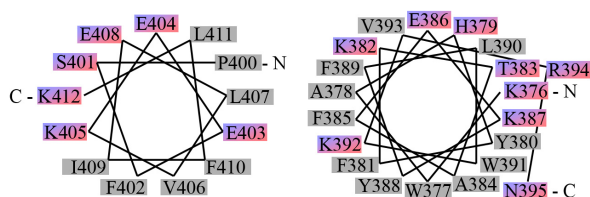
Figure 2.33: Amphipathic helices. Helical wheels of the amphipathic helices 376 – 395 (right) and 400 – 412 (left). The hydrophobic amino acids are in grey boxes, the polar residues are in blue/red boxes. "N" and "C" indicate the N and C termini of the helices, respectively.
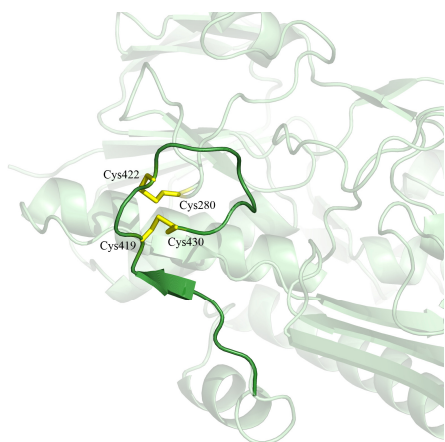
Figure 2.34: Disulfide bonds. Snapshot of the disulfide bridges between Cys 419 – Cys 430 and Cys 422 – Cys 280 in the C-terminal domain of one SQR monomer. The protein chain is represented as semitransparent cartoon in light green, the C_α trace of the C-terminal 18 amino acids is highlighted in dark green and the side chains of the cysteines forming the disulfide bonds are represented as yellow sticks.

Glu 207 and Asp 208 of the neighboring monomers. On the outer side, the C-terminal 18-amino-acid loops and residues 364 – 371 form contacts with residues 174 – 181 and 211 – 215 of their neighbouring monomers.

The six monomers are structurally nearly identical. They are composed of two FAD-binding Rossmann domains typical of FDRs and of a C-terminal domain distinctive of each FDR subfamily. In *A. aeolicus* SQR, the C-terminal domain consists of two parts. First, residues 376 – 412 form an amphipathic helix-turn-helix motif. The two helices (Lys 376 – Asn 395 and Pro 400 – Lys 412) are graphically illustrated in Figure 2.33. Their positive net charge (+4) and their relative hydrophobicity (1.21 kcal/mol) are comparable with values of other amphipathic helical domains of monotopic membrane proteins (Johnson and Cornell, 1999). Second, residues 413 – 430 form an 18-amino-acid long loop characterized by the presence of two disulfide bridges (Figure 2.34).

2.10.2. Interaction with the membrane

The structure suggests that the SQR trimer interacts with the membrane through its hydrophobic patches and its positively charged face and the membrane interaction domain can be subdivided into four structural motifs (Figure 2.35).

First, the N-terminal surface residues (Ala 2, Lys 3, Arg 22, Lys 29) and the spatially adjacent Arg 333 and Asn 334 of each subunit are distributed on a layer perpendicular to the central trimeric three-fold axis.

Second, within the same layer, but on the inner central region of the trimer, the subunits form a positively charged ring with a diameter of about 25 Å circling the three-fold axis. Its positive potential is due to three lysines (Lys 172, Lys 173, Lys 412) and three arginines (Arg 177, Arg 213, Arg 369) and a few main chain N atoms. Two putative sulfate groups

form ion pairs with protein atoms compensating the excess of acidic residues that would weaken the interactions between the monomers. Both are present as strong electron density peaks in all six monomers of the asymmetric unit and appear to originate from one sulfate ion and one MES molecule (Figure 2.36).

Third, four lysine residues (Lys 54, Lys 68, Lys 387 and Lys 405), conserved in type I SQRs, are oriented towards the bulk solvent.

Finally, the base of the trimer body is the part of the protein most deeply inserted into the membrane, approximately 12 Å below the layer of the N-terminus and of the sulfate groups. It consists of the helix-turn-helix amphipathic motif (residues 376 – 412). At the positive N-terminal end of helix 376 – 395, the indole group of conserved Trp 377 "snorkels up" (Lomize *et al.*, 2007) in a characteristic manner and, at the C-terminal end, Trp 391, Arg 394 and Asn 395 bind a putative sulfate ion (Figure 2.37). A dodecyl-β-D-maltoside (DDM) molecule can be identified in the electron density lying in between the helices. The detergent mimics both the physiological lipids as well as the substrate quinone that presumably enters SQR at this site (see Chapter 2.10.5). The maltose polar head group of DDM presents its oxygen atoms at hydrogen bond distances (3 – 3.5 Å) from residues Lys
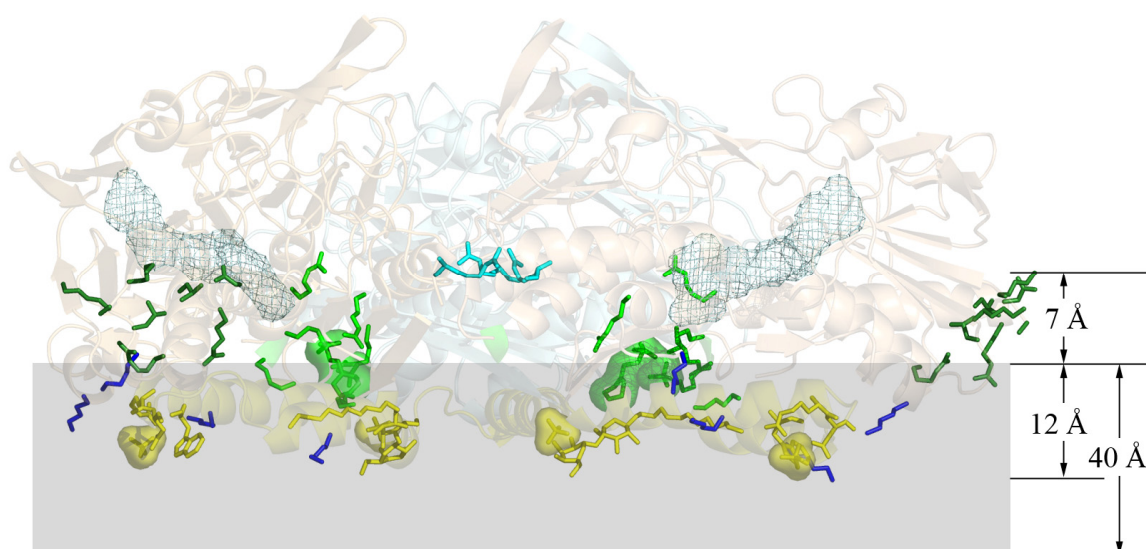
Figure 2.35: Membrane-binding motifs. The SQR trimer interacts with the membrane through four distinctive motifs, here represented in different colours. The N-terminal domain is dark green, the inner domain binding one sulfate ion and one MES molecule is light green, the four conserved lysines are dark blue and the base of the trimer body is yellow. The membrane is indicated in grey. The overall trimer is in a semi-transparent cartoon representation. The side chains of Arg 204 of all three monomers are represented in cyan sticks to highlight the central trimerization contacts. Other residues and molecules are shown only for the two subunits in the foreground for a clearer visualization. The FAD is in cyan mesh. The side chains of the residues and the molecules belonging to the different membrane interacting motifs are shown as sticks. For sulfate groups a semi-transparent surface representation is also shown. The distances reported in angstrom on the right were calculated from the plane of the sulfur atoms of the MES molecules respectively to the C_α atom of Arg 204 (7 Å) and the C3B atom of the maltose head of DDM (12 Å) in chain A. An approximate value (40 Å) indicates the membrane thickness according to White and Wimley, 1994.

373 and Trp 377 and from one putative sulfate ion, whereas its acyl chain makes hydrophobic contacts (3.5 – 4.2 Å) with residues Phe 357, Phe 381, Leu 407, Phe 410 and Leu 411. Other unexplained electron density peaks can also be identified in the amphipathic domain and more precisely between the two helices near the DDM molecule, in the quinone binding cavity (see Chapter 2.10.5) and in contact with loop 396 – 399.

2.10.3. Binding site of the cofactor FAD

SQR binds FAD in the same position as other FDRs do. The isoalloxazine ring binds the central Rossmann domain and it is planar, in contrast with the more bent conformation found in the related FCSD structure. The ribitol-adenosine moiety extends into the N-terminal Rossmann domain. The electron density of FAD is connected from its C8M group (the methyl group in position C8 of the FAD isoalloxazine ring) to the side chain of Cys 124. Most interestingly, their distance is too long for a direct covalent linkage and therefore this flavin-protein linkage is of a type that has not been observed before (Heuts *et al.*,
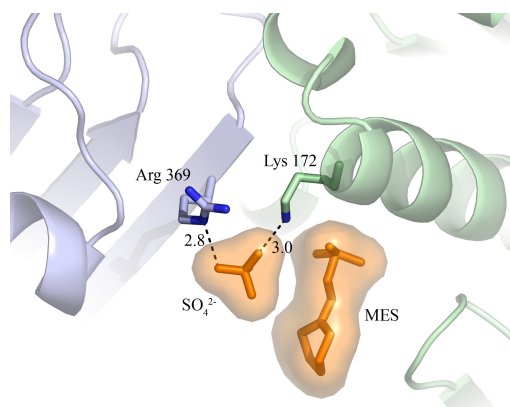
Figure 2.36: Monomer-monomer contacts mediated by sulfate groups. In the inner membrane-facing side of the trimer, one sulfate ion (SO_4^{2-}) and one MES molecule bind at the monomer-monomer interface. The sulfate ion is forming ion pairs between Arg 369 of one monomer (light blue) and Lys 172 of a neighbouring chain (light green). The protein is in a semi-transparent cartoon representation, relevant amino acids and protein-bound molecules are shown in sticks. A semitransparent surface representation of the sulfate ion and the MES molecule is shown in orange. Interatomic distances are indicated in angstrom and depicted as black dotted lines.
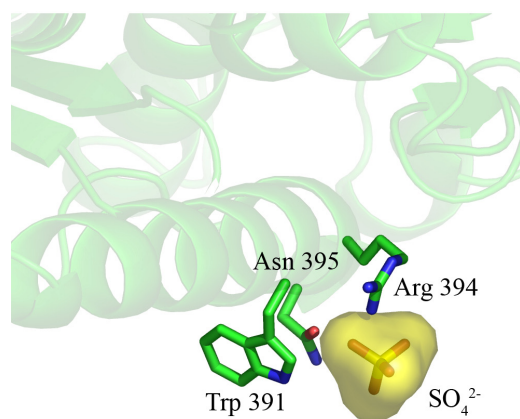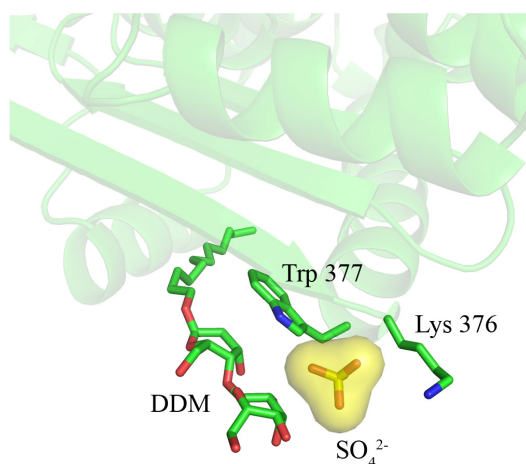
Figure 2.37: Putative lipid clamps. Sulfate ions (SO_4^{2-}) bind to the SQR amphipathic domain in two positions, possibly mimicking sites of specific binding of zwitterionic lipids. Left panel: the environment of conserved Trp 377, whose side chain interacts with the maltose head group of DDM and with a sulfate ion. Right panel: the environment of Trp 391, whose indole group interacts with a sulfate ion together with Arg 394 and Asn 395.
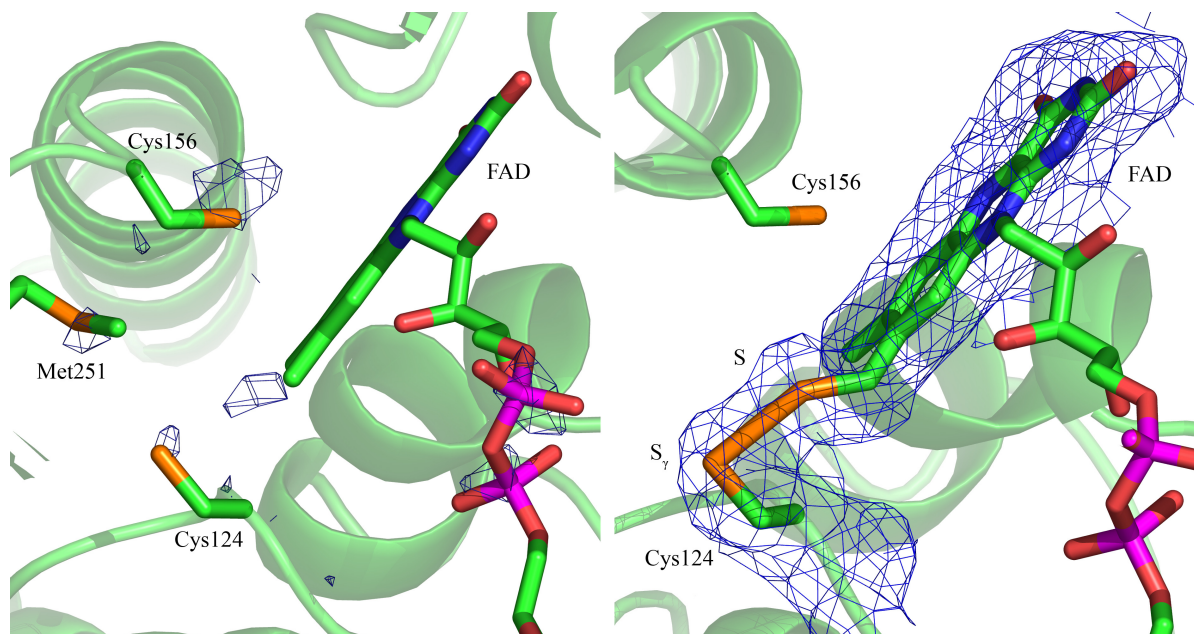
Figure 2.38: Cys 124-FAD linkage. Left panel: anomalous difference map around the catalytic site calculated from a data set collected at 6.5 keV and shown at 3.0 σ contour level in blue mesh. Anomalous peaks are visible for S atoms of cysteines and methionines and for the phosphate groups of the FAD. One additional peak is present between the sulfhydryl group of Cys 124 and the C8M group of FAD, indicating that a relatively heavy atom with residual anomalous scattering at 6.5 keV, possibly S, is present in the structure. Right panel: $2F_o$-F_c electron density map at 1 σ contour level in blue mesh of the protein-FAD linkage. The S_γ atom of Cys 124 and the putative S atom providing the linkage are indicated. The protein monomer is shown in the same orientation in the two panels. Its C_α trace is shown in a semitransparent cartoon representation in green. The FAD and the side chains of the relevant residues are in sticks, colour-coded according to the atom type (C green, N blue, O red, S orange and P magenta).

2009). The connection is provided by a relatively heavy atom as identified in an anomalous difference map calculated on the basis of a 2.30 Å resolution data set collected at a wavelength of 1.907 Å (Figure 2.38). We interpreted it as a sulfur atom possibly contributed by the substrate, sulfide (see Chapter 3.4). Thus, according to our interpretation of the electron density, SQR binds FAD through a labile persulfide bond (Cys-S_γ-S-CH_2-FAD) and not through a stable thioether bond (Cys-S_γ-CH_2-FAD).

2.10.4. Sulfide oxidation site

The sulfide oxidation site is located on the *re*-face of FAD. The aromatic FAD rings and the protein segment Ser 155 – Pro 159 delimit a cavity that extends approximately 15 Å parallel and 4 Å perpendicular to the isoalloxazine plane. The cavity is accessible from the bulk solvent through a channel characterized by a positive electrostatic surface potential. Four solvent molecules occupy the channel at consecutive sites from the periplasm to Ser 155, conserved only in type I SQRs, and might mimic the trace of the substrate, H_2S. At the end of the channel sits Val 294, also conserved only in type I SQRs and essential for activity (Griesbeck *et al.*, 2002) (Figure 2.39). Interestingly, the loop Val 294 – Lys 312
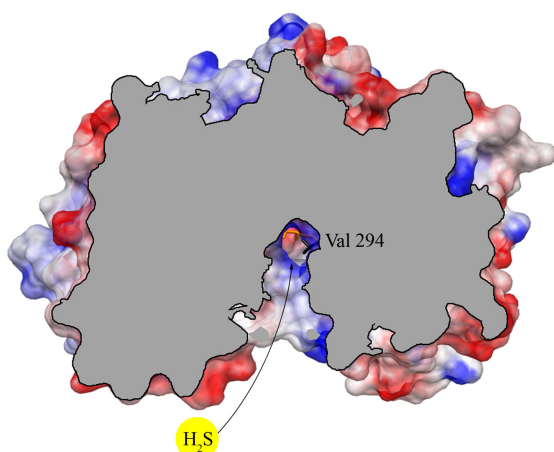
Figure 2.39: Putative sulfide access channel. Sulfide (here indicated as H_2S in a yellow circle) reaches the SQR active site through a positively charged channel (black arrow). The side chain of Val 294, represented in black sticks, resides at the inner end of such channel. The surface of the protein is colored according to its electrostatic surface potential (see Figure 2.32).
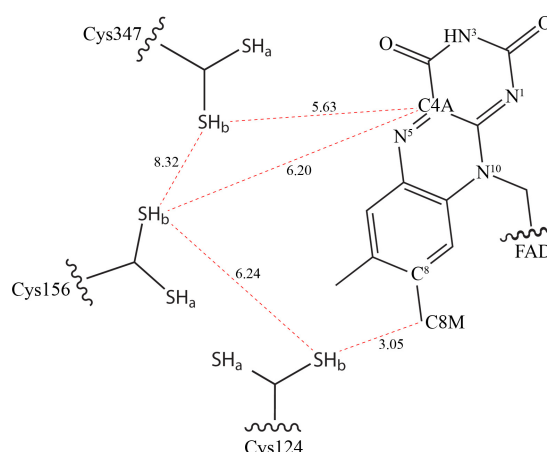
Figure 2.41: Distances between key atoms in the active site. The distances between the essential SQR cysteine residues (Cys 124, Cys 156, and Cys 347) and FAD positions C8M and C4A are represented as red dotted lines and indicated in angstrom. The side chains of Cys 124, Cys 156 and Cys 347 are in a double conformation ($-SH_a/-SH_b$) as modelled in the structure.
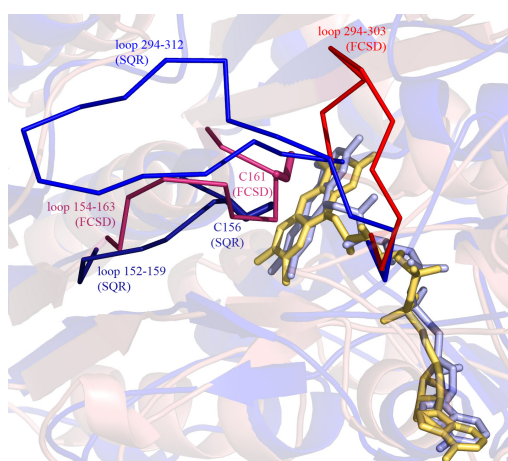
Figure 2.40: Superposition of SQR on FCSD. SQR (blue semitransparent cartoon representation) has an overall fold similar to FCSD (red semitransparent cartoon representation). However, two loops around the sulfide oxidation site have different conformations. SQR loop 294 – 312 (bright blue) is longer than FCSD loop 294 – 303 (bright red) and extends towards the left side of the Figure. SQR loop 152 – 159 (dark purple) and FCSD loop 154 – 163 (dark pink) also show a different orientation, such that FCSD Cys 161 is closer to FAD than SQR Cys 156. The FAD molecules are in cyan and yellow for the SQR and FCSD respectively. The side chains of the Cys residues are shown in sticks.

delimiting the channel on one side and the segment Pro 152 – Pro 159 opposite to the isoalloxazine ring are linked by hydrogen bonds and exhibit a conformation different from that of the FCSD structure (Chen *et al.*, 1994) (Figure 2.40). Three cysteines essential for activity (Griesbeck *et al.*, 2002) face the cavity as schematically depicted in Figure 2.41. Cys 124 binds the FAD dimethylbenzyl ring as described in the previous Chapter 2.10.3. Cys 347 is part of a rigid β-sheet on the pyrimidine-2,4-dione side. Cys 156 is localized in the loop Pro 152 – Pro 159 and, among the three cysteines, it is the residue with the highest potential degree of mobility.

The sulfide oxidation pocket is unexpectedly not empty in the structure. The density of the S_γ atom of Cys 156 in one conformation extends to form a long chain on one side. The

chain is oriented towards Cys 347, it is surrounded by conserved hydrophobic amino acids (Phe 194, Val 196, Ile 199, Ala 349 and Phe 358) and it is near the couple Tyr 161 – Glu 162 implicated in proton exchange during the reaction (Nübel *et al.*, 2000; Shahak and Hauska, 2008). In some monomers (i.e. monomer A and F in the structure of the "as purified" protein), the density closes to form a circle (Figure 2.42). In some other monomers (i.e. monomer D) the density connected to Cys 156 is shorter but a chain is also observable connected to Cys 347 (Figure 2.43).

We currently have no evidence for the crystallographic and possibly functional reasons for such differences between the densities of the active site of the single monomers. Since polysulfur is known to be the product of the reaction (Griesbeck *et al.*, 2002) and because sulfur has the property of forming chains, we have interpreted $2F_o$-F_c density regions higher than 1 σ with chains of S atoms adjusting the occupancy (30 %) to have F_o-F_c difference maps with positive and negative signals lower than 3 σ and B factors close to the average values of the whole protein chain ($B_{overall}$ = 38 Å^2). The geometry of the S chains modelled into the electron density is compatible with that of sulfur species known from
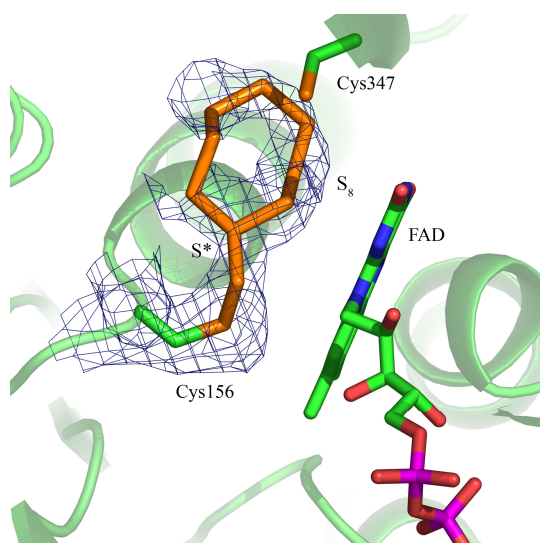
Figure 2.42: Putative polysulfur chain attached to Cys 156. In chain A of the "as-purified" structure, the electron density map of the sulfhydryl group of Cys 156 extends to form a circle and was thus interpreted as a covalently bound cyclooctasulfur ring (S_8). The density is visible also in the simulated annealing $2F_o$-F_c omit map represented here in blue mesh at 1.0 σ contour level. The second sulfur atom of the chain (S*) would be trivalent, as explained in the text. The colour code is the same as in Figure 2.38.
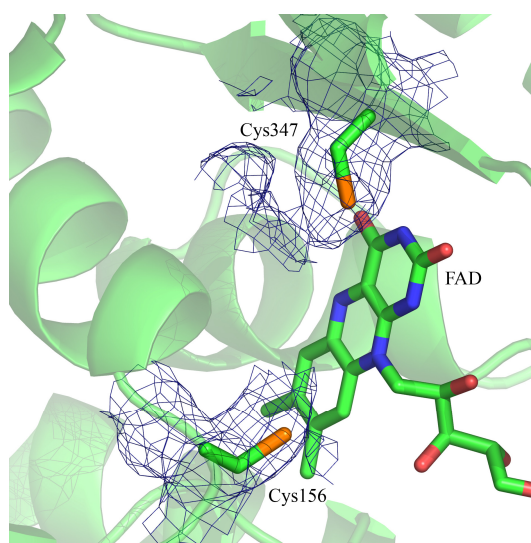
Figure 2.43: Putative polysulfur chain attached to Cys 347. The sulfide oxidation site of monomer D in the "as-purified" structure shows an elongated electron density peak extending from the sulfhydryl group of Cys 347, and a weaker density prolonging the side chain of Cys 156. Here the simulated annealing $2F_o$-F_c omit map around Cys 156 and Cys 347 is represented in blue mesh at 1.0 σ contour level. The colour code is the same as in Figure 2.38.

inorganic chemistry[42]. A few additional electron density peaks still remain unexplained in the sulfide oxidation cavity.

2.10.5. Quinone reduction site

The quinone binding site is on the FAD *si*-face and it is accessible through a channel formed between the amphipathic helices described above. The quinone was localised in the $2F_o$-F_c simulated annealing omit map at 0.8 σ (Figure 2.44). In the same region, the structure of the protein "as-purified" shows a density peak of different shape, possibly corresponding to a co-purified hydrophobic molecule. In the quinone-bound structure, the binding site is not fully occupied, which correlates to the low affinity (in the micromolar range) measured in kinetic experiments. However, it is possible to locate the aromatic ring and the hydrophobic side chain unequivocally. The aromatic ring is bound between the benzene ring of Phe 385 and the side chain of Ile 346 (both are conserved residues among SQRs but not in FCSDs). There is a network of solvent molecules and charged amino acids including Glu 318 and Lys 382 (conserved among SQRs) and Gln 319, Lys 392 and Glu 403 around the benzoquinone ring. The aliphatic side chain extends towards the membrane attachment domain and interacts with the hydrophobic side chains of amino acids Met 315, Ile 348, Phe 381, Tyr 388, Phe 402, Val 406, Leu 407. The position of the quinone binding site is confirmed also by the structure of SQR in an inhibitor-bound state. Aurachin C, a quinone-analogue reported to be a potent competitive inhibitor of *A. aeolicus* SQR (Nübel *et al.*, 2000), was soaked into the crystals and the structure was solved at 2.9 Å resolution. It also does not differ significantly from the structure of the protein "as-purified". The structure shows an electron density peak for the inhibitor in the same pocket that was identified to bind decylubiquinone.

Figure 2.44: The quinone binding site. Decylubiquinone (UQ, light blue stick representation) was localised in the electron density map between the side chains of Ile 346 and Phe 385 on FAD *si*-face. Here, the simulated annealing $2F_o$-F_c omit map is shown in blue mesh at 0.8 σ contour level. The colour code is the same as in Figure 2.38.

[42] see for instance the description of the structure of the S_8 ring in Rettig *et al.* (Rettig and Trotter, 1973), of the allotropic forms of elemental sulfur in Meyer (Meyer, 1964) and of trivalent S atoms in Wong *et al.* (Wong *et al.*, 2002).

3. Discussion

3.1. The structural membrane proteomics approach

This work was undertaken as part of a structural membrane proteomic project on *A. aeolicus* described previously (Wedemeyer, 2007). It is worth here presenting and discussing the working strategy because it is both an uncommon approach and has had important consequences on the choice of the experiments, on the protocols established and on the results achieved.

Most importantly, it has to be considered that the project consists of the isolation of membrane proteins from native cells. Such an approach, directed to structural studies, has advantages and disadvantages. First of all, since *A. aeolicus* cells require extreme and costly growing media, the same batches of membranes need to be used in our laboratories by a whole research group. Consequently, it is not possible to improve the production yield of a single protein specifically by optimizing i.e. the cell growing conditions or the membrane solubilisation protocol or the first chromatographic fractionation of the solubilized material. Therefore, this approach is only applicable to proteins that are already highly expressed in the source organism. Secondly, the work on a native system excludes the possibility of exploiting protein tagging for purification. Homologous recombination to enhance expression, introduce tags or mutate the target are also discouraged by the difficult growing conditions of the organism. Nonetheless, SQR is one of the most abundant enzymes in *A. aeolicus* membranes and an efficient protocol could be established to prepare a very pure sample in sufficient amounts for crystallization studies by means of conventional chromatographic techniques.

Despite some limitations, isolating SQR from native *A. aeolicus* cells had numerous advantages. In general, as a hyperthermophilic organism, *A. aeolicus* is a source of stable, compact protein complexes, ideal for structural studies. At the same time, membrane protein production in a native lipidic environment is advantageous in comparison to heterologous expression in another bacterial system like *E. coli*, given the substantial differences in membrane composition between thermophilic and mesophilic organisms. Furthermore, previous attempts to purify *A. aeolicus* SQR from the heterologous expression host *E. coli* were hampered by unsatisfactory levels of FAD incorporation (Schödl, 2003), while spectral characterization suggests that in our preparation the presence of apoprotein is negligible. Finally, certain features of the SQR structure, namely the unusual FAD-to-protein linkage and the putative polysulfur chain in the active site, are presumably a consequence of the activity of SQR *in vivo* before cell harvesting and might have not been present if the enzyme were produced heterologously.
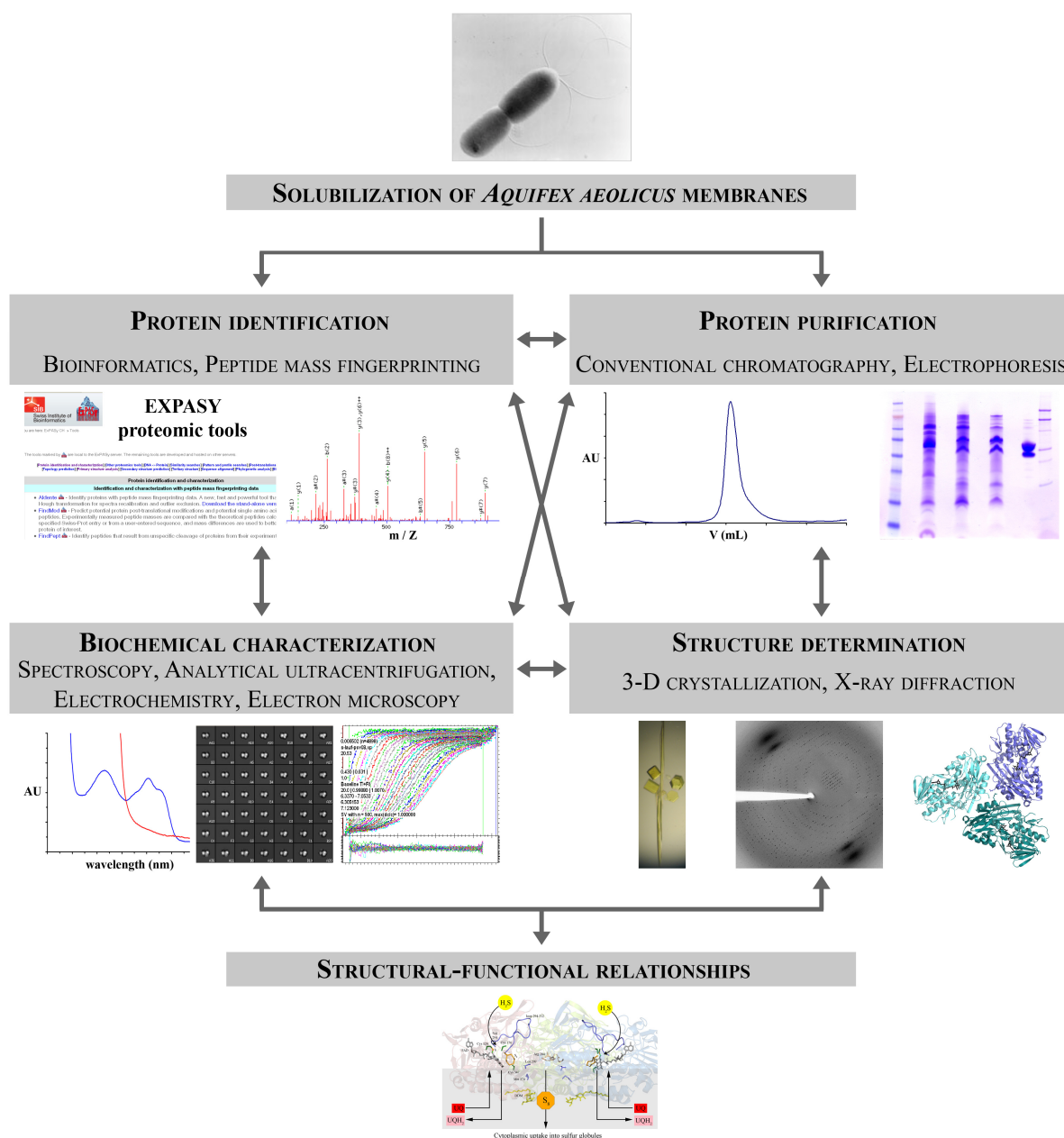
Figure 3.1: The structural membrane proteomics approach.

Figure 3.1 summarizes the working strategy in a flowchart and highlights the iterative nature of the process. In contrast to a genomic approach, which focuses on a single (set of) specific gene(s), a proteomic project concerns itself with the analysis of the full proteome of the target organism. Each experiment aims in first place to the identification and to the choice of the target, even before contributing to its direct biochemical characterization. In other words, parallel studies on the target need to be carried out to determine its sequence, its abundance, its purity and its interactions with other partners, besides its biochemical and its structural properties. It is therefore important that each result is used to improve and interpret all the others at every stage of the project. The iterative nature of such a working flowchart is well-

exemplified by two experiences from this work, the first related to the establishment of the purification protocol, the second to the interpretation of the electron density map. The purification protocol was optimized in two successive steps. An initial protocol was useful for obtaining the first crystals. Then, during crystallization trials, one additive, the detergent Zwittergent 3-10, was identified as modifying the crystal properties. The biochemical investigation on the effect that Zwittergent 3-10 has on the protein preparation led to a crucial improvement in the original purification protocol, by including a detergent exchange step. Only the crystals grown from this second, more pure protein preparation were suitable for structure determination. The second example is related to the unexpected structural features that, as mentioned above, were observed in the SQR active site and that could have hardly been predicted before the determination of the structure. As a consequence, after structure solution, the biochemical characterization of the protein had to be extended in order to support the crystallographic observations and to suggest an experiment-based interpretation of the electron density maps.

3.2.   Biochemical and crystallographic studies on a monotopic membrane protein

As observed in the crystal structure and in agreement with the expectation from previous biochemical and bioinformatic investigation (Shahak and Hauska, 2008), SQR belongs to the class of integral monotopic membrane proteins (MMPs). Here it is commented on what consequences the SQR topology has had on the biochemical and crystallographic studies, while its structural and functional implications are discussed in Chapter 3.3.2.

Integral MMPs are, like transmembrane proteins (TMPs), amphipathic. However, the electrostatic surface potential distribution in MMPs and TMPs strongly differs. TMPs consist of two polar regions on opposite sides separated by a hydrophobic belt. Instead, in MMPs one face has hydrophilic character and is exposed to the bulk solvent, while the opposite face is enriched in basic residues (positively charged) and in hydrophobic patches. The latter face binds the membrane *in vivo*, while *in vitro* it requires the presence of detergent for correct folding and to prevent aggregation. The surface dipolarity typical of MMPs is particularly pronounced in SQR.

In addition, the dimensions of the SQR complex reveal that the protein is rather flat, resembling the shape of an elliptical disc rather than that of a globular sphere. The SQR shape is a consequence of its necessity to bind the membrane as a trimer, as discussed later.

Because of its dipolar surface, its binding to detergent molecules and its ellipticity, all features inherent to its membrane topology, SQR behaves in a way that strongly deviates from standard soluble proteins. In particular, all biochemical techniques that are affected by surface charge

distribution and shape are biased by the membrane-bound nature of SQR. This is specifically true for techniques used to estimate the molecular weight of the protein, such as size-exclusion chromatography and native electrophoresis, and possibly also for the LILBID MS setup, where SQR aggregation may result from non-specific electrostatic or hydrophobic interactions in the small energised droplets generated by the ion source.

The crystallographic behaviour of SQR is also significantly affected by the asymmetry of its surface, and it could not be otherwise, since crystal contacts form through surface residues. The disordered detergent molecules, the steric impediment of the detergent micelle and the non-
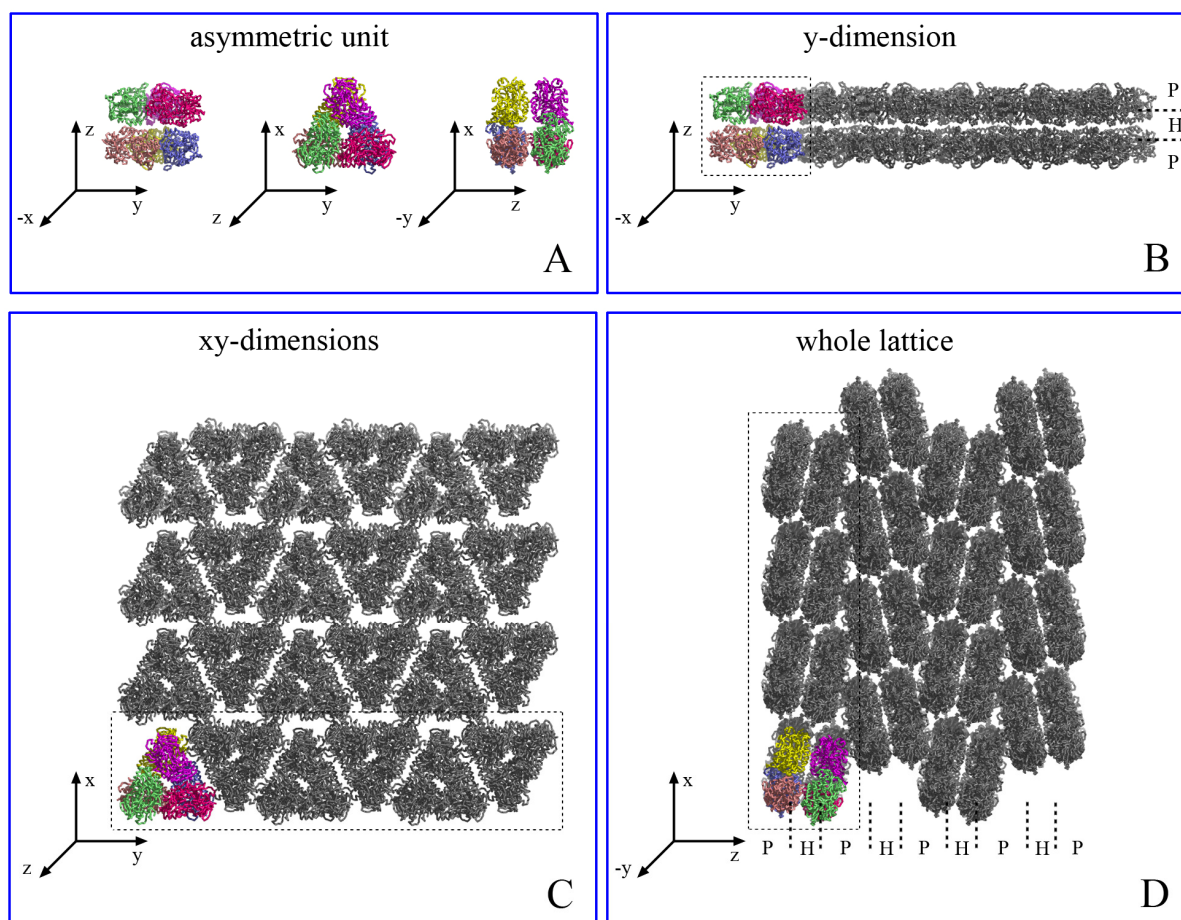
Figure 3.2: The crystal lattice of SQR. Panel A shows the asymmetric unit of the crystals in the three orientations, indicated by the x-, y- and z-axis, almost parallel to the crystallographic unit cell axis. It is composed by a dimer of trimers, here represented as ribbons, with each protein chain coloured differently. Panel B shows the disposition of the SQR hexamers along the y-axis. The protein complexes form a row confining the hydrophobic portion (H) of the structure in between two layers of trimers and exposing the polar regions (P) on the two sides. The dotted black box indicates one asymmetric unit (see panel A, left). The molecules obtained by symmetry operations are depicted in grey. Panel C shows the disposition of the SQR complexes in the x- and y- dimensions. The protein complexes form a bidimensional plane. One asymmetric unit is coloured as in panel A, the symmetry-related units are coloured grey. The dotted black box indicates one row of SQRs in the y-dimension (see panel B, rotated by 90° over the y-axis). Panel D shows the full SQR lattice. The xy-planes described in panel B (one is indicated here by the dotted black box) are repeated in the z-dimension. Although the detergent-filled region is not exactly continuous along the x-axis, it is possible to identify successive layers of polar (P) and hydrophobic (H) regions, as it is defined in the text for "type I" monotopic membrane protein crystals.

specific hydrophobic contacts formed by the membrane-inserted regions make it difficult for membrane proteins to form regular three-dimensional arrays. TMPs overcome these problems in two ways (Michel, 1983; 2006). One possibility is that two-dimensional crystals form in the plane of the membrane and stack over each other in an ordered manner (type I membrane protein crystals). Another possibility is that crystal contacts are formed exclusively among the polar regions of the protein and that the hydrophobic cores of the proteins are solvated by the detergent micelle (type II membrane protein crystals). Considering the structures available to date, MMPs also seem to adopt two alternative crystalline patterns which closely recall those described for TMPs. In the "type I" MMP crystals, the molecules also form a two-dimensional array mimicking the disposition over the membrane. However, TMPs in the type I crystals can form crystal contacts with both their polar regions on the two sides of the membrane, while MMPs cannot, because they possess only one polar face. Consequently, each two-dimensional plane of MMPs needs to come in contact with an identical one oriented "upside-down", so that the hydrophobic faces are facing each other and are excluded from the bulk solvent. This pattern can then be translated up- and downwards in the third dimension. It is noteworthy to remark that, in this type I crystalline arrangement, the two dimensional layers of the protein are alternatively in contact through polar and non-polar faces. The hydrophobic contacts, which are generally less specific, confer fragility and irregularity to the crystals. Possibly, this is the reason why it has been so difficult and so rare to obtain SQR crystals not characterized by high mosaicity. On the other hand, in the "type II" MMP crystals, the membrane binding domain surrounded by the detergent is segregated in cavities of the crystal lattice and the contacts between the protein units are mediated only by the polar faces. Oxidosqualene cyclase (PDB id.: 1W6K) and prostaglandin H_2 synthase 1 (1Q4G) crystallize in the "type I" form, while fatty acid amide hydrolase (1MT5), 11-β-hydroxysteroid dehydrogenase (1XSE), cytochrome P450 (2BDM), prostaglandin H_2 synthase 2 (1CVU), apocarotenoid cleavage oxygenase (2BIX), carnitine acyltransferase (1T7Q), electron transfer flavoprotein:quinone oxidoreductase (2GMH) and squalene hopene cyclase (2SQC) form "type II" crystals.

The lattice of SQR needle crystals (Figure 3.2) shows that in one dimension (along the crystallographic y-axis) the molecules are arranged on two sides of a layer filled by detergent molecules and corresponding to a hypothetical membrane plane. These rows of SQR complexes visible along the y-axis extend to form planes along the x-axis. Although the detergent-filled region is not continuous along the x-axis, it is confined to the core of the planes. Instead, the surface of the planes is composed on both sides by the hydrophilic portions of the protein complexes. Finally, the xy planes are repeated in the z dimension exclusively through polar crystal contacts. This distribution of the molecules in the SQR lattice is similar to that defined above for "type I" MMP crystals. It would be curious to study the tendency of SQR to

form two-dimensional crystals by detergent-lipid exchange and, because of the pronounced polarity of the SQR surface potential, to test whether crystallization in a stationary magnetic field (Moreno *et al.*, 2007) would improve the speed of crystal growth or the resolution of diffraction.

3.3. Overall three-dimensional structure

SQR is composed of three structural regions, like other type I FDRs. The N-terminal and the central domains are two Rossmann-fold domains, similar to other nucleotide-binding proteins, as inferred from a structural alignment performed with DALI (Table 3.1). Instead, the C-terminal or "interface domain" (Argyrou and Blanchard, 2004) is a structural region specific for each FDR subfamily. In SQRs, it is composed of two amphipathic helices (residues 376 – 412) and a C-terminal 18-amino-acid loop (residues 413 – 430). Its function is to form oligomeric contacts, to interact with the membrane and to interact with the quinone, as discussed in Chapters 3.3.1, 3.3.2 and 3.6, respectively.

Although not possessing a completely new structural fold, *A. aeolicus* SQR shows a number of interesting and in some cases unexpected features. In particular, it shows surprising differences

Table 3.1: Structural alignment of SQR against known protein structures[a].

Target[b] (PDB code / chain id)	RMSD[c]	N° aligned residues[d]	Identity[e]
1FCD / A	2.4	362	21
3H8L / A	2.8	356	24
2YVF / A	2.8	321	18
2GMH / A	3.1	143	17
1DX1 / C	3.1	329	14
1NDA / B	3.2	325	16
2QCU /B	3.2	148	15
4GR1	3.2	328	15
1ZK7 / A	3.3	322	14
2CFY / F	3.4	327	15
1QLA / A	4.5	190	13

[a] The alignment was performed with the software DALI (www.ebi.ac.uk/dali) against the PDB database.

[b] PDB codes correspond to the following proteins: 1FCD, flavocytochrome *c*:sulfide dehydrogenase; 3H8L, sulfide:quinone oxidoreductase (from *Acidianus ambivalens*); 2YVF, ferredoxin reductase; 2GMH, electron transfer flavoprotein (ETF):ubiquinone reductase; 1DXl, dihydrolipoamide dehydrogenase; 1NDA, trypanothione oxidoreductase; 2QCU, glycerol-3-phosphate dehydrogenase; 4GR1, glutathione reductase; 1ZK7, mercuric reductase; 2CFY, thioredoxin reductase; 1QLA (2BS2), quinol:fumarate reductase.

[c] Average root-mean-square deviations (RMSD) for all atoms indicated in angstrom.

[d] Total number of structurally aligned residues.

[e] Percentage of structurally aligned residues that are identical in SQR and in the respective aligned partner.

from the structure of *Allochromatium vinosum* flavocytochrome *c*:sulfide dehydrogenase (PDB id.: 1FCD), which had been for almost two decades the only structural reference for SQRs, and to another structure of an SQR solved recently, that of *Acidianus ambivalens* SQR (3H8L) (Brito *et al.*, 2009). The differences pertain to the oligomeric state, to the structural basis of membrane attachment, to the details of the substrate binding sites and to the hypothesis on the reaction mechanism.

3.3.1. Oligomeric state

All structures of MMPs, and in particular of monotopic membrane enzymes, reported to date are either monomeric, like ETF:UQ oxidoreductase (PDB id.: 2GMH), oxidosqualene cyclase (1W6K), apocarotenoid cleavage oxygenase (2BIX), cytochrome P450 (2BDM), 11-β-hydroxysteroid dehydrogenase (1XSE) and carnitine acyltransferase (1T7Q), or dimeric, like fatty acid amide hydrolase (1MT5), prostaglandin H_2 synthase 1 (1Q4G) and 2 (1CVU), monoamine oxidase B (1OJA) and squalene hopene cyclase (2SQC) (Balali-Mood *et al.*, 2009).

Also all FDRs characterized to date were shown to be functional in a dimeric state (Argyrou and Blanchard, 2004). Only the oligomeric state of SQRs was difficult to characterize unequivocally because of the uncertainties related to the fact that SQRs are the only FDRs bound to the membrane. Solubilized *Oscillatoria limnetica* SQR was proposed to be either a monomer surrounded by a detergent belt or a dimer with less detergent associated to it (Arieli *et al.*, 1994), while *Rhodobacter capsulatus* SQR was later shown to be functional in a multimeric form (Schütz *et al.*, 1997).

The *A. aeolicus* SQR structure shows that the protein forms a trimeric complex. Crosslinking and density-matched sedimentation equilibrium analytical ultracentrifugation, techniques which are least sensitive to protein shape and electrostatic effects and which allow masking the contribution from the detergent micelle, suggest that SQR is trimeric also in solution. Also the reconstruction by single-particle electron microscopy reveals that SQR exists as a trimer, probably with a tendency to form hexamers, as observed in the asymmetric unit of the crystal structure.

Comparison of SQR sequences shows poor conservation of the residues that mediate oligomerization, so that trimerization might not be a common feature of SQRs. For example, the *A. ambivalens* SQR (Brito *et al.*, 2009) is dimeric, even though it must be said that the structural disorder of the 50-amino-acid C-terminal domain absent in the structure may be responsible for a non-physiological oligomerization in that protein. It would not be exceptional that SQRs adopt different oligomerization states in different organisms. For example, another monotopic membrane enzyme, lanosterol cyclase, is dimeric in bacteria

(squalene hopene cyclase), but monomeric in eukaryotes (oxidosqualene cyclase) (Ruf *et al.*, 2004). However, it is certain that, in *A. aeolicus* SQR, the trimeric state guarantees thermal stability and an optimal membrane attachment surface and therefore has a physiological relevance.

3.3.2. Membrane-binding mode

MMP structures often do not reveal the details of their protein-to-lipid interactions (Balali-Mood *et al.*, 2009). The SQR structure presented in this work is exceptional in this respect because it provides information on the degree of penetration into the bilayer and on specific binding sites for phospholipids.

Membrane insertion of SQR follows two of the common strategies described for peripheral membrane proteins (see Chapter 1.3.1), namely the presence of in-plane amphipathic helices and of lipid clamps. The amphipathic helix-loop-helix C-terminal domain (Figure 2.33) presumably induces SQR to insert into the lipid bilayer to a depth of about 12 Å, which implicates complete penetration through the lipidic polar head group region (White and Wimley, 1994; Johnson and Cornell, 1999) and contacts between the exposed hydrophobic side chains and the membrane hydrocarbon core. In the same domain, two tryptophan residues (Trp 377 and Trp 391, Figure 2.37) are involved in specific binding to sulfate groups in the crystal structure. Sulfate groups are an important determinant for assessing protein-lipid interactions because sulfate chemically mimics the phosphate heads of cellular lipids. Therefore, position 377 and 391 in *A. aeolicus* SQR represent two putative lipid clamps and may possess affinity for zwitterionic phospholipids, preferred by tryptophan (Cho, 2001).

Furthermore, the observed binding sites of some sulfates between the subunits (i.e. one sulfate ion bridges Lys 172 of one monomer and Arg 369 of a neighbouring monomer, Figure 2.36) suggest that membrane insertion stabilizes oligomerization and, *vice versa*, the trimer creates an appropriate surface for binding lipid phosphate groups. The presence of oligomeric interactions mediated by salt ions also explains the fact that SQR is more compact in buffers with high salinity, as observed by size exclusion chromatography (Chapter 2.3.1) and electron microscopy (Chapter 2.3.6). The reduction in the rate of enzymatic activity measured at high salt concentrations might instead result from a competition of salt anions with sulfide and is unlikely to be physiological. In addition, it must be mentioned that, conversely, an increase of activity at high salt concentrations was observed for *A. aeolicus* SQR overexpressed in *E. coli* (Schödl, 2003). Therefore further investigation is needed to characterize the SQR reaction at high values of ionic strength. Certainly, the coupling between membrane insertion, trimerization and enzymatic activity

might ensure that SQR exclusively reduces hydrophobic quinones and, perhaps, might also facilitate the release of the water-labile polysulfur product into the membrane via a non-polar environment (see following Chapter 3.7).

SQR shares common features of membrane association, such as characteristic tryptophans, exposed hydrophobic residues, amphipathic segments and positively charged surface areas, with some other monotopic membrane proteins, i.e. the prostaglandin H_2 synthase 1 (PDB id.: 1Q4G) and the squalene hopene cyclase (2SQC). However, the fold of the inserted segments differs substantially. Even more divergent are the membrane insertion motifs of the structurally and functionally related flavoproteins ETF-UQ oxidoreductase (PDB id. 2GMH) and glycerol-3-phosphate dehydrogenase (PDB id. 2QCU). Interestingly, the comparison with *A. ambivalens* SQR, which might have one single amphipathic membrane anchor, seems to suggest that the membrane binding motifs are different even within the SQR family itself. Membrane attachment of monotopic membrane proteins has been probably developed at a later stage of protein evolution than enzyme differentiation for a specific catalytic activity. Within SQRs, the differences may correlate with the different sensitivity to quinone analogues (Griesbeck *et al.*, 2000).

A final consideration needs to be made for the description of how SQR insertion perturbs the lipidic bilayer. The effect of amphitropic and monotopic membrane proteins on the membrane has been extensively studied by coarse-grained molecular dynamics simulations (Fowler *et al.*, 2007; Balali-Mood *et al.*, 2009) (Figure 3.3). It was found that some protein complexes insert into the membrane in a "shallow" manner, perturb only one membrane leaflet and interact with the membrane "symmetrically", meaning that each monomer
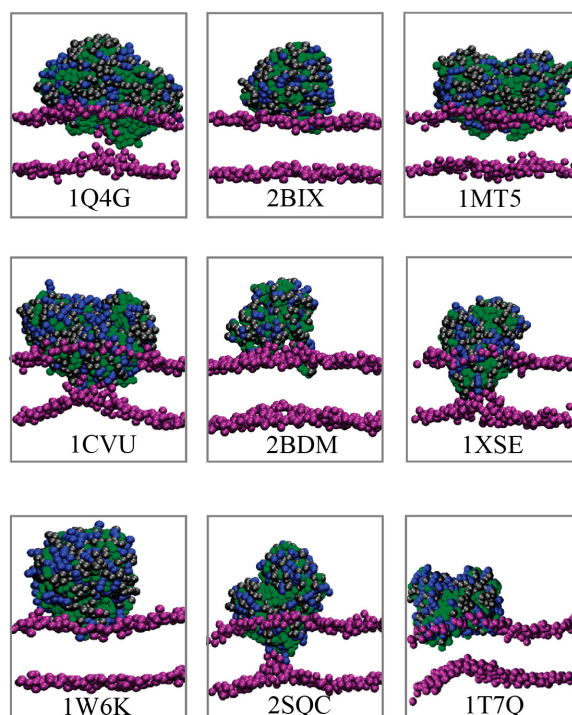
Figure 3.3: Lipid bilayer perturbation by monotopic membrane proteins. Final snapshots (t = 700 ns) generated by coarse-grained molecular dynamics simulations are shown for nine monotopic membrane proteins (the corresponding PDB codes are indicated in the insets). Phosphate atoms of the lipids are coloured purple, basic protein residues blue, hydrophobic residues green, and others gray. The Figure is reproduced from Balali-Mood *et al.*, 2009.

penetrates it to an identical extent. Other proteins, in contrast, insert "asymmetrically" and more deeply, so that some residues may even come in contact with the polar heads of the lipids belonging to the opposite membrane leaflet. The latter proteins induce a pronounced distortion of the entire bilayer. Based on those computational studies and considering the symmetry and the electrostatic properties of the SQR trimer, it seems more likely that SQR interacts with only one membrane leaflet through a shallow insertion.

3.3.3. Thermal stability

Functional assays show that *A. aeolicus* SQR has an exceptionally high thermal stability (Table 2.4). The three-dimensional structure shows that the protein adopts at least five different strategies to adapt to high temperatures. (i) As discussed above, SQR is trimeric and not monomeric or dimeric like other monotopic membrane proteins and other FDRs. "Increased oligomerization" (Yano and Poulos, 2003) was proposed to be a way of enhancing protein complex stability (Tanaka et al., 2004), although this issue is debated (Kumar et al., 2000). (ii) The six direct hydrogen bonds plus the two mediated by solvent molecules that constitute the monomer to monomer interactions and the involvement of an arginine (Arg 204) in the central contact region of the trimer are typical features of thermostable proteins (Kumar *et al.*, 2000; Yano and Poulos, 2003; Tanaka *et al.*, 2004). (iii) It is interesting to notice that seven cysteines are present in the *A. aeolicus* SQR. This is unusual because highly reactive cysteines are normally less frequent in hyperthermophilic proteins (Vieille and Zeikus, 2001). However, in SQR, four of these cysteines are reciprocally bound to form two disulfide bridges (Figure 2.34). Such disulfides stabilize the C-terminal domain, a sensitive region of the structure that mediates trimer formation, contributes to cover the active site (see Chapter 3.5) and ensures a stable and site-specific membrane interaction. The C-terminal disulfide bonds also corroborate the idea that the protein is localized in a periplasmic oxidising environment (Schütz *et al.*, 1999). (iv) The cofactor FAD is bound tightly through a covalent bond, as discussed in Chapter 3.4, and (v) large solvent-filled cavities surround its isoalloxazine moiety and constitute the active sites of the enzyme. These features confer an advantageous resilience (Hough and Danson, 1999) to SQR and so also contribute to its structural stability.

3.4. FAD binding mode

SQR binds FAD covalently, thus preventing dissociation and consequent protein destabilization (Heuts *et al.*, 2009). An helix dipole at the N^1-C^2=O^2 locus, a glutamate near the N^5 FAD atom and a cysteine tethering the flavin, three recently identified structural features needed for covalent flavinylation (Heuts *et al.*, 2009), are present in the structure. The

connection consists of an unusual 8-α-X-S$_\gamma$-cysteinyl bond with Cys 124, that might affect the reduction potential of FAD (Heuts *et al.*, 2009) and that we interpreted as a persulfide bond on the basis of an experimental anomalous difference electron density map (Chapter 2.10.3). Should this interpretation be confirmed, it would represent the first case of an 8α-S-S$_\gamma$-cysteinyl bond to FAD (Heuts *et al.*, 2009). Cysteine persulfides were however already found in the active site of other enzymes, i.e.: sulfurtransferase (PDB id: 1E0C), NifS/CsdB (1KMJ), rhodanese (1RHS) and sulfur oxygenase reductase (2CB2). Covalent FAD binding is unusual in group I FDRs. It is known to occur only in the other sulfide oxidising enzyme, FCSD, but the structural origin of the linking cysteine is different in this latter protein. In FCSD, Cys 42 binds FAD from its *si*-face inducing a distortion in the planarity of the isoalloxazine ring, while in SQR Cys 124 is situated near the dimethylbenzyl side and the FAD ring is planar. Surprisingly, a difference is present also among SQRs. In fact, *A. ambivalens* SQR binds FAD through a direct thiol 8α-S$_\gamma$-cysteinyl bond with Cys 129, whose S$_\gamma$ atom is only 1.53 Å away from the FAD position C8M. In other SQRs, then, the corresponding cysteine is replaced by a tyrosine, so that these enzymes probably form an 8-α-O-tyrosyl bond with FAD (Heuts *et al.*, 2009) (see Chapter 3.8).

The location of the catalytically active residues around FAD also differs significantly between FDRs and SQRs and even among SQRs themselves. The redox active disulfide pair typical of group I FDRs normally forms a disulfide bridge in the oxidised state. The cysteines lie on the *si*-side of FAD, 2 – 4 Å apart from its C4A atom, and one forms a covalent bond with it during the reaction cycle. The cysteines are proximal in the protein sequences too, with a consensus motif C$_1$XXXXC$_2$. In both SQRs and FCSDs, instead, the two cysteines are located on the *re*-face of FAD, a situation curiously recalling the case of low molecular weight thioredoxin reductase (PDB id: 1CL0), a group II and not a group I FDR (Argyrou and Blanchard, 2004). Additionally, they are far away from each other in the sequence (Cys 161 and Cys 337 in *A. vinosum* FCSD, Cys 156 and Cys 347 in *A. aeolicus* SQR and Cys 178 and Cys 350 in *A. ambivalens* SQR). Only in the FCSD structure do they form a disulfide bond and Cys 161 is sufficiently close (3.34 Å) to the C4A atom of FAD to form a covalent connection. In *A. aeolicus* SQRs and in *A. ambivalens* SQR, instead, the corresponding cysteines are far apart from each other (the S$_\gamma$ – S$_\gamma$ distance is 8.3 Å and 6.5 Å, respectively). Moreover, in the conformations observed in the structures, the distances between Cys 156 and the C4A atom of FAD in *A. aeolicus* SQR (6.24 Å) and between Cys 178 and the C4A atom in *A. ambivalens* SQR (4.90 Å) are too long for a covalent attachment (Table 3.2).

Table 3.2: Location of key atoms in the active site of FDRs.

Protein	PDB code	Position[a]	Distance[b]	FDR group
glutathione reductase	3DK4	*si*	3.37	I
lipoamide dehydrogenase	3LAD	*si*	3.24	I
trypanothione reductase	1FEB	*si*	3.44	I
A. vinosum FCSD	1FCD	*re*	3.34	I
A. aeolicus SQR	3HYV	*re*	6.24 (C156); 5.57 (C347)	I
A. ambivalens SQR	3H8L	*re*	4.90 (C178); 7.57 (C350)	I
low MW thioredoxin reductase	1CL0	*re*	3.40	II

[a] Position of the redox active disulfide centre relative to FAD isoalloxazine ring (*re*- or *si*-face).

[b] Distances of the Cys-Sγ atoms relative to FAD-C4A atom in angstrom.

3.5. Sulfide oxidation site

In SQRs, sulfide oxidation occurs on the *re*-face of FAD. In FDRs, this region is in general accessible from the bulk solvent and binds the pyridine-nucleotide substrate NAD(P)H. In SQRs, however, NAD(P)H binding is hindered by the presence of specific polypeptide segments, hereafter referred to as "capping loops", that guarantee exclusive access of sulfide to the active site. In *A. aeolicus* SQR, the "capping loops" are residues Val 294 – Lys 312, Pro 189 – Arg 204 and part of the 18-amino-acid C-terminal segment. The latter two loops interestingly mediate the oligomerization contacts in the quaternary structure. Loop 294 – 312 contains the crucial invariant residues Val 294, shown to be responsible for sulfide affinity (Griesbeck *et al.*, 2002), and Lys 312, proposed to be involved in proton exchange during the reaction cycle (Brito *et al.*, 2009). Residues Pro 311 (through the main chain oxygen atom) and Lys 312, together with Ser 155, coat a channel that connect the active site to the bulk solvent. Electron density peaks for several solvent molecules, suggest that this channel might be occupied by H_2S (a chemical analogue of H_2O) during the catalytic reaction.

Surprisingly, the geometry of the "capping loops" is not conserved in the other two sulfide oxidising enzymes of known structure, namely *A. vinosum* FCSD and *A. ambivalens* SQR. As a consequence, sulfide reaches the active site through completely different pathways in those enzymes. In FCSD the "capping loops" are substituted by short segments so that the redox active disulfide bridge is exposed to the solvent. No additional residues have been described to be involved in attracting or guiding sulfide to its binding site. In *A. ambivalens* SQR, instead, the active site is shielded by the "capping loop" Gly 154 – Cys 178. This segment extends in the opposite direction as loop 294 – 312 of *A. aeolicus* SQR (Figure 3.4). Moreover, the sulfide access channel, defined by residues Ser 176, Ser 214, Asp 215 and Asp 353 in *A. ambivalens*

SQR, is located in a different structural position than the access channel of *A. aeolicus* SQR. As a reference, the distance between the carboxyl group of Asp 215 (*A. ambivalens* SQR) and the hydroxyl group of Ser 155 (*A. aeolicus* SQR) is 12.5 Å.

Finally, both *A. aeolicus* SQR (this work) and *A. ambivalens* SQR (Brito *et al.*, 2009) present an active site occupied by a chain covalently attached to the redox active cysteines. In both cases, the electron density has been interpreted with polysulfur species, assuming that the structures represent an intermediate of the reaction cycle, with the forming product still attached to the protein. The geometry of such polysulfur chains

Figure 3.4: SQR "capping loops". *A. aeolicus* SQR (semitransparent cartoon representation in blue) was superposed on *A. ambivalens* SQR (in red) with the server DALI (see Table 3.1). Their "capping loops", represented as ribbons, possess completely different orientations. Loops 294 – 312 and 412 – 430 in *A. aeolicus* SQR extend from right to left in the Figure, while loop 154 – 178 in *A. ambivalens* SQR from left to right. Consequently, in the two proteins the sulfide oxidation site varies and the whole reaction mechanism may be different, as described in the text. The FAD molecules are shown as sticks to facilitate the localization of the active sites of the enzymes.

differs in the two structures. In *A. aeolicus* SQR, the chain never connects Cys 156 to Cys 347, it is 3 – 9 putative sulfur atoms long, it extends either the side chain of Cys 156 or the side chain of Cys 347 and it closes to form a putative S_8 ring in some monomers. In *A. ambivalens* SQR, instead, it consists of a 3-sulfur-atom connection between Cys 178 and Cys 350. Such structural differences lead to different catalytic hypothesis for the two SQRs (see Chapter 3.7).

3.6.   Quinone reduction site

Quinone reduction takes place on the *si*-face of FAD. Quinone reaches the active site through a channel in between the two amphipathic helices in the C-terminal membrane-attachment domain. Based on a classification of the different strategies by which membrane-bound enzymes interact with their lipophilic substrates (Forneris and Mattevi, 2008), it can be concluded that SQRs "desorb" the quinones from the lipid bilayer.

Considering the disposition of the residues around quinone-binding pockets of other quinone reductases and quinol oxidases, it must be said that no general rule has yet been identified. In SQR, the pocket extends approximately 8.2 Å (distance between Ile 346 and Phe 385) x 11.7 Å (distance between Glu 318 and Lys 382). The benzoquinone ring is sandwiched between the two apolar residues Ile 346 and Phe 385 as in other membrane proteins (Murray *et al.*, 1999). Histidines, which are often observed to form hydrogen bonds with the quinone oxygen atoms

(Fisher and Rich, 2000) and which were expected to bind the quinone also in SQR (Griesbeck *et al.*, 2000; Griesbeck *et al.*, 2002), are not found in the binding pocket. The closest histidines are His 47 and His 379, located 10 – 13 Å apart from the binding site. On the contrary, a number of other polar residues and of putative solvent molecules surround the quinone and form a hydrogen-bonding network. Glu 318 and / or Lys 382, conserved among SQRs, seem to be the most likely candidates mediating proton transfer. Finally, the quinone-bound SQR structure determines the shortest distance between quinone and FAD to be about 3.5 Å, which allows fast electron transfer.

3.7. Enzymatic activity

Kinetic data are reported for SQRs from *A. aeolicus* [(Nübel *et al.*, 2000; Schödl, 2003) and this work], *R. capsulatus* (Schütz *et al.*, 1997; Griesbeck *et al.*, 2002), *O. limnetica* (Arieli *et al.*, 1994), *Aphanothece halophytica* (Bronstein *et al.*, 2000), *Acidithiobacillus ferrooxidans* (Wakai *et al.*, 2004), *Paracoccus denitrificans* (Schütz *et al.*, 1998) and *A. ambivalens* (Brito *et al.*, 2009) and also for the FCSD from *A. vinosum* (Fukumori and Yamanaka, 1979)[41]. In general, SQRs and FCSDs exhibit similar substrate affinities in the micromolar range. Exceptionally, a number of bacterial SQRs [from *Pseudomonas putida* (Shibata and Kobayashi, 2006) and *Bacillus stearotheromophilus* (Shibata *et al.*, 2007)] and all reported mitochondrial homologues [from *Saccharomyces pombe* (Vande Weghe and Ow, 1999) and *Arenicola marina* (Hildebrandt and Grieshaber, 2008; Theissen and Martin, 2008b)] show millimolar affinity to sulfide and an increased catalytic efficiency only in the presence of reaction partners such as cyanide or thioredoxin and sulfite.

Despite the observation of related kinetic values, it is an open question whether the complex polysulfur formation reaction proceeds by the same or a different enzymatic mechanism in different SQR types. In fact, the different structural features between *A. aeolicus* SQR, *A. ambivalens* SQR and *A. vinosum* FCSD described in the previous chapters argue for a different reaction cycle.

The overall SQR reaction couples a reductive step (sulfide oxidation) and an oxidative step (quinone reduction) to provide reducing equivalents for redox processes in the membrane. Electron flow between the two half-reactions is accomplished via FAD by exploiting its capability to accept an electron pair from a transiently bound sulfide adduct and to donate these two electrons one by one to the neighboured quinone via the space. Both half-reactions are spatially separated. Quinone reduction can be described only from the structure of *A. aeolicus* SQR (see Chapter 3.6). On the other hand, sulfide oxidation can be addressed in more detail by

---

[41] It was questioned whether the FCSD sulfide oxidising activity measured *in vitro* has physiological relevance (Reinartz *et al.*, 1998).

comparing *A. aeolicus* SQR, *A. ambivalens* SQR and *A. vinosum* FCSD. SQRs, and possibly also FCSDs, catalyze a reaction of sulfur polymerisation whereby polysulfur intermediates are covalently attached to the enzyme. This enigmatic multi-step process, composed of initiation, elongation and termination steps, requires more complicated structural and mechanistic features than for the established reactions of the other FDRs. The chemistry of sulfur in the binding pocket of a protein and – in the physiological state – at the high temperatures at which *A. aeolicus* and *A. ambivalens* grow is complex and its description certainly requires deeper investigation. However, the available structural data suggest hypotheses for plausible reaction schemes.

A typical reaction scenario (hereafter referred to as the C4A mechanism[42]) is established for FDRs and it involves a redox-active cysteine/cystine couple that is oxidized/reduced by FAD via a Cys-S_γ-C4A-FAD charge-transfer intermediate. This mechanism requires a distance between the sulfur atom of one of the cysteines and the C4A atom in the range of 3 – 4 Å (Table 3.2). A reaction scheme in analogy with the mechanism known for FDRs was indeed proposed for *A. ambivalens* SQR (Brito *et al.*, 2009). Surprisingly, though, in the presented structure of *A. aeolicus* SQR, the key atoms are with 5 – 8 Å too far away for a direct reaction (Table 3.2), which would imply substantial conformational changes for such a mechanism to take place. From the structure (B factors, quality of the electron density), however, there is no hint of flexibility for any protein segment. Two other reaction mechanisms are therefore proposed for *A. aeolicus* SQR, which require shorter protein movements.

The first proposal is a subtle variation of the FDR C4A mechanism. It implies the formation of a Cys-S_γ-S-C4A-FAD adduct, where the additional S atom would be contributed by the substrate of the reaction, sulfide. Therefore, it is hereafter referred to as the C4A-S mechanism (Figure 3.5). Forming a S_γ-S-C4A bond would require a movement of the protein chain approximately half as long as forming a direct S_γ-C4A bond. In this case, the thiol/disulfide pairs in the fully oxidised enzyme would be Cys 156 – Cys 347 (*A. aeolicus* SQR) or Cys 178 – Cys 350 (*A. ambivalens* SQR) and SQRs would possibly adopt a structure similar to that of FCSD in the fully oxidised state. The C4A-S mechanism would also include a Cys156-S_γ-S-S_γ-Cys347 trisulfide bridge proposed by Griesbeck *et al.* (Griesbeck *et al.*, 2000) as a central intermediate.

---

[42] "C4A" refers here to the general mechanism of FDRs and to the mechanism proposed for *A. ambivalens* SQR (Brito *et al.*, 2009), while the "C4A mechanism" described in (Marcia *et al.*, 2009) is referred to as "C4A-S".
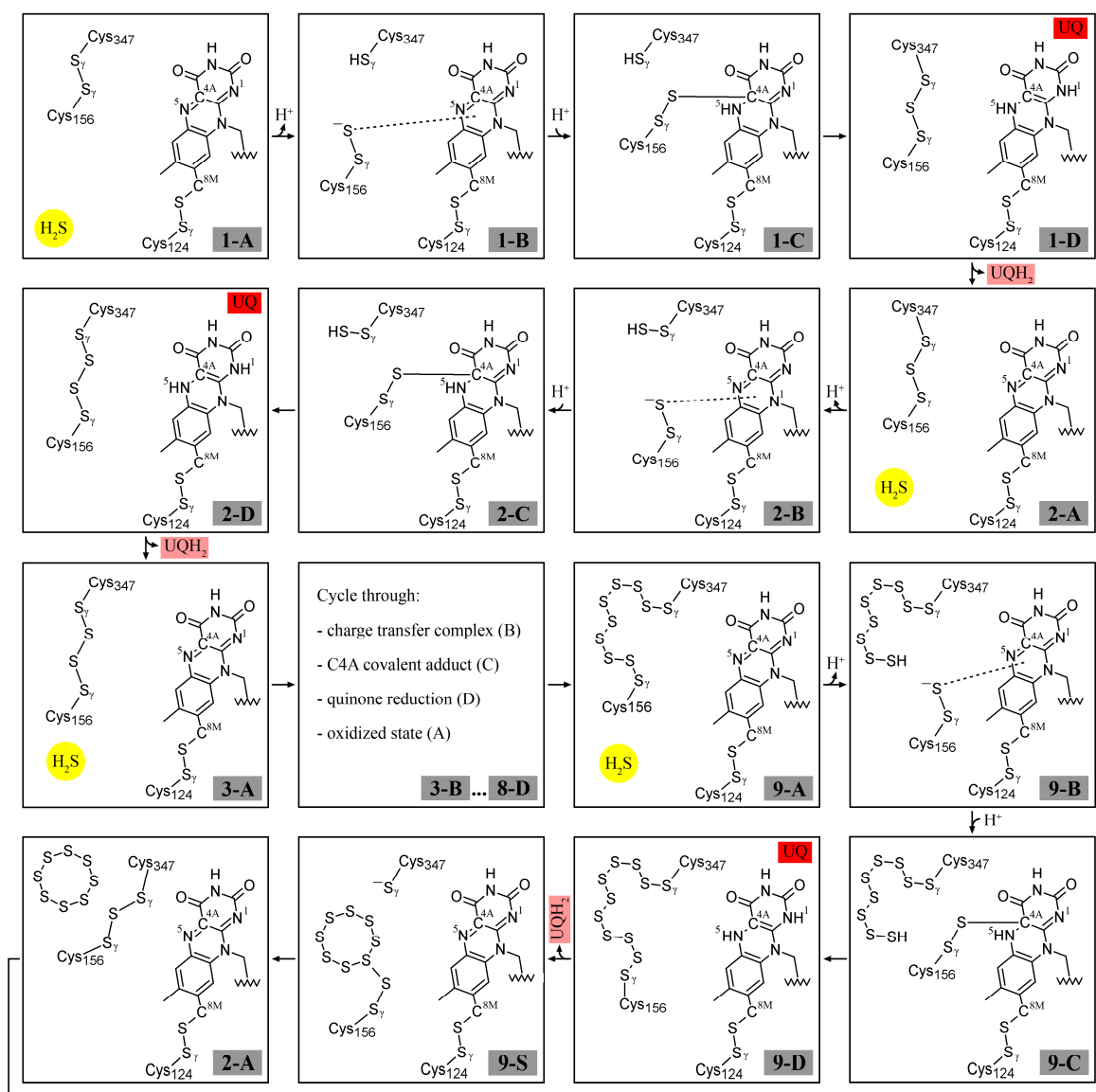
Figure 3.5: The C4A-S mechanism. The individual steps are numbered according to the following criteria. The number in the grey insets indicates the number of external S atoms accumulated during polymerization. For example, the steps marked with number 1 relate to the addition of the first sulfide atom, and so on. The letter indicates the oxidoreduction stage of the reaction. Letter A refers to the fully oxidized form, which presents a disulfide bridge in the initial step of catalysis and a polysulfide state in the following cycles. This state is open for sulfide binding (H_2S, yellow circle). Letter B refers to the charge-transfer complex between protein and FAD. Letter C refers to the Cys156-S_γ-S-C4A-FAD covalent adduct described in the text. The FAD is in a reduced state and the protein is open for binding the substrate quinone (UQ, red box). A significant protein conformational change seems to be required to bring Cys 156 closer to FAD position C4A. This possibly involves a rotation of loops 294 – 312 and 152 – 159 towards a conformation closer to that of the corresponding loops of FCSD (see Figure 2.40). Letter D indicates the state in which FAD is reoxidised after electrons are transferred to the quinone (the reduced quinol form is indicated as UQH_2, pink box). A polysulfide bridge connects Cys 156 and Cys 347. After quinol release the fully oxidised state (letter A) is formed again. After elongation to a 9-S-atom chain, the polysulfur can close on itself to form a cyclooctasulfur still covalently bound to the protein through one S atom in a trivalent state, as described in the text and observed in some monomers of the structure. This intermediate is indicated by the letter S. Protonation and deprotonation are mediated by charged residues present in the catalytic pocket of the enzyme but not depicted in this scheme. These might include i.e. Glu 162 on the *re*-side (Griesbeck *et al.*, 2002) and Glu 318 and/or water molecules on the *si*-side of FAD.
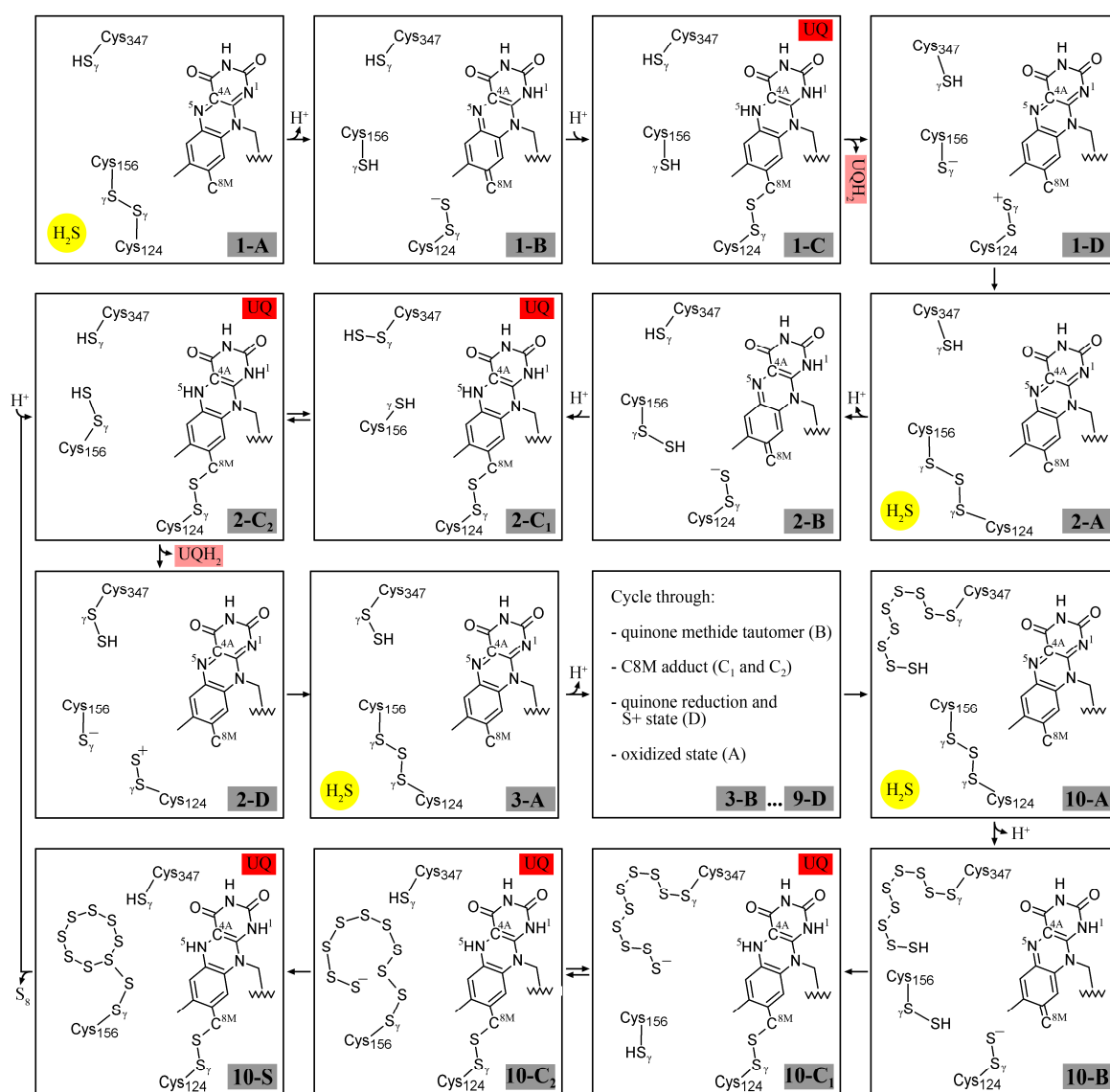
Figure 3.6: The C8M mechanism. The individual steps are numbered according to the criteria explained for Figure 3.5. Letter A refers to the fully oxidized form, which presents a disulfide bridge in the initial step of catalysis and a trisulfide bridge state in the following cycles. This state is open for sulfide binding (H_2S, yellow circle). Letter B refers to the state after sulfide binding. $Cys124-S_\gamma-S^-$ attacks FAD quinone-methide tautomer in position C8M. Letter C refers to the C8M adduct described in the text. The FAD is in a reduced state and the protein is open for binding the substrate quinone (indicated as UQ in a box with a red background). This stage presents an equilibrium with the growing polysulfur chain exchanged between Cys 156 and Cys 347 after the second cycle. Letter C_1 refers to the conformation in which the chain is attached to Cys 347, letter C_2 to the one in which the chain binds Cys 156. A significant protein conformational change seems to be required to favour the S chain exchange between Cys 156 and Cys 347. This would possibly involve a rotation of loop 294 – 312 and of the loop including Cys 156 towards a conformation closer to that of the corresponding loops of FCSD (see Figure 2.40). Letter D indicates the state in which FAD is reoxidised after electrons are transferred to the quinone (the reduced quinol form is indicated as UQH_2 in a box with pink background) and a transient Cys124-S^+ state is formed. After quinol release the fully oxidised state (letter A) is formed again. The reaction terminates as described for Figure 3.5.

The third crucial cysteine Cys 124 which covalently binds to FAD via a disulfide bridge would certainly influence the FAD position and its electronic properties, i.e. its redox potential, but would not directly participate in polymerisation. While the C4A mechanism would result in a stoichiometry of the reaction of n:(n-1) sulfide molecules oxidised over quinone molecules reduced (Brito *et al.*, 2009), the stoichiometry would be 1:1 if a C4A-S mechanism would be followed. This stoichiometric difference is a fine but possibly measurable parameter that may help in discriminating between the two mechanisms in the future.

The *A. aeolicus* SQR structures, though, also suggest an alternative mechanism for conveying electron pairs from sulfide to FAD which, of course, also implies a different elongation cycle. The connection between the Cys 124 S_γ and FAD C8M atoms via a labile persulfide bridge inspires a fairly speculative reaction scheme. The mechanism is hereafter referred to as the C8M mechanism and it is schematically illustrated in Figure 3.6. It recalls the description (Edmondson and Newton-Vinson, 2001) of FAD incorporation into flavoproteins that covalently bind their cofactor. In the proposed reaction scheme, one sulfur atom of the persulfide is the S_γ atom of Cys 124 and the other one originates from the substrate, S^{2-}, after reduction of a Cys124-S-S-Cys156 disulfide fully oxidised state. The hypothesis that SQR presents a Cys124-S-S-Cys156 disulfide in the oxidised state is compatible with the observation that the sulfide access pathway in *A. aeolicus* SQR (but not in *A. ambivalens* SQR) leads to the region of the active site between Cys 124 and Cys 156. Successively, the species Cys124-S-S⁻ attacks the FAD quinone-methide tautomer to form the "C8M adduct" observed in the structure. After electron transfer, a transient Cys124-S-S⁺ state is generated and the trisulfide bridge (Griesbeck *et al.*, 2000) is formed by a nucleophilic attack by Cys156-S⁻ at Cys124-S-S⁺. Cys 347 is involved in taking over the growing sulfur chain and pulling it into a hydrophobic pocket, thereby setting Cys156-S⁻ free for the next reaction cycle. This alternative mechanism attributes a central role to Cys 124 but the hypothesis is weakened by the fact that Cys 124 is not conserved in mitochondrial SQR sequences, where it is replaced by a tyrosine, and by the fact that in *A. ambivalens* SQR the corresponding cysteine binds FAD directly. However, the active site and the mechanism of sulfide oxidation appear to be substantially different in *A. aeolicus*, *A. ambivalens* and in the mitochondrial SQRs. This conclusion is based (i) on the effect of site specific exchanges (V294D and C124Y) on the enzyme kinetics (Griesbeck *et al.*, 2002), (ii) on the different modes of FAD binding (Chapter 3.4), (iii) on the strong differences between entire segments ("capping loops", see Chapter 3.5 and 3.8) and (iv) on the necessity proven for type II SQRs to have reaction partners such as cyanide or thioredoxin and sulfite to oxidize sulfide efficiently (Hildebrandt and Grieshaber, 2008; Theissen and Martin, 2008b).

In all postulated mechanisms, the reaction is spontaneously initiated by sulfide, an oxidized

FAD, quinone and a disulfide bond. The specific conformational changes at the basis of the polymerization are difficult to postulate because in the structure the distances between Cys 124, Cys 156 and Cys 347 are too long for direct sulfur-sulfur coupling. It appears attractive to conclude that the distances between the cysteines increase stepwise during the elongation process. This conclusion is supported by the polysulfur chains visible in both SQR crystal structures, suggesting that the reaction is trapped in an advanced stage of the cycle. In *A. aeolicus* SQR, a concerted rotation of loops 294 – 312 and 152 – 157 towards a conformation closer to that observed for the equivalent FCSD segments could occur during elongation.

Figure 3.7: Putative product release channel. The sulfide oxidation site can be identified by the position of the FAD molecule, of Cys 156 and of the polysulfur chain, in the form of cyclooctasulfur S_8. The latter extends into a hydrophobic cavity towards surface residues Leu 209 and Met 374 (the surface of these residues is shown in dark blue). The cavity could potentially open into the inner membrane-facing region of the trimer near the MES molecule bound to one of the membrane-interacting domains in the crystal structure (see Result section and Figure 2.35 and 2.36). The protein surface is in light cyan. The colour code for the residues and molecules shown is the same as in Figure 2.38.

Independent of the mechanism, the termination of elongation is presumably triggered by the limited space of the active site. The true product of the SQR reaction is unknown. Nonetheless, the *A. aeolicus* SQR structure shows that the active site can only provide a catalytically productive environment for up to a 9-S-atom chain which is finally most likely cleaved to an S_8 ring, as found in the electron density of some monomers. The release of the stable S_8 ring might proceed via the sulfide channel. However, cyclooctasulfur is water labile and insoluble. Consequently, it becomes attractive to consider an alternative channel, very hydrophobic, conducting directly towards the inner membrane-facing side of the trimer and only closed by the contacting side chains of surface residues Leu 209 and Met 374 (Figure 3.7). Using this exit, S_8 would move directly towards the membrane interior and it could finally be incorporated into sulfur globules. In *A. aeolicus* these are accumulated in the cytoplasm (Guiral *et al.*, 2005) so that, despite being a monotopic and not a polytopic membrane protein, SQR would participate in cellular uptake of sulfur in an efficient manner (Figure 3.8). In other organisms, sulfur is stored in the periplasm (Reinartz *et al.*, 1998). In this case, either membrane-bound SQRs transfer S_8 to the membrane for periplasmic reuptake or they release hydrophilic polysulfides, as soluble FCSDs most likely do. More structural data on the C-terminal region of other SQRs would help in discriminating

Cytoplasmic uptake into sulfur globules

Figure 3.8: Proposal for the complete SQR reaction in the cell. SQR (here in a semi-transparent cartoon representation) may participate to cellular sulfur uptake despite being a monotopic membrane protein. Sulfide (H_2S) accesses the oxidation site (*re*-face of FAD) from a channel controlled by "capping loop" 294 – 312 (blue ribbon). FAD (grey sticks) gets reduced through either a charge-transfer complex in position C4A or through an unusual alternative covalent adduct at position C8M, suggested by the described Cys 124-S-FAD linkage. Polymerization occurs over Cys 156 and 347. The mature product (a putative cyclooctasulfur ring, orange sticks) is possibly released below the trimeric central contact (Arg 204, side chains in ochre sticks) directly into the lipid bilayer, through a channel controlled by the side chains of Leu 209 and Met 374 (blue sticks). In *A. aeolicus*, S_8 would then be uptaken from the cytoplasmic side to be incorporated into sulfur globules. At every cycle of sulfide oxidation a quinone molecule (UQ in a red box indicates the oxidized form, UQH_2 in a pink box the reduced form) is reduced in the quinone binding pocket (*si*-face of FAD). The membrane is here schematically represented as a light grey rectangular box. The DDM molecules bound to SQR are shown in yellow sticks in the membrane-embedded region. Details are highlighted only for the two monomers in the foreground for clearer visualization. Labels for the represented residues and compounds are shown only for the left monomer.

between these two possibilities.

3.8. New structure-based sequence fingerprints for the classification of SQRs

After about twenty years of research on SQRs, the first 3-D structures were solved in 2009 [this work and (Brito *et al.*, 2009)]. At the same time, new SQR sequences became available and the functional characterization has also been extended (Hildebrandt and Grieshaber, 2008; Theissen and Martin, 2008b; Brito *et al.*, 2009; Chan *et al.*, 2009). These data show a surprising and previously unexpected variability among SQRs, until now all considered homologous. Based on these new findings, it is now possible to revisit and improve the

Table 3.3: Structure-based SQR classification[a].

Type	Organisms	H₂S-affinity	Physiological role[b]	Model	References
I	Bacteria (*α/β Proteobacteria*, *Cyanobacteria*, *Aquificales*)	µM	respiration photosynthesis	3HYV	(Arieli *et al.*, 1994; Schütz *et al.*, 1997; Schütz *et al.*, 1998; Bronstein *et al.*, 2000; Griesbeck *et al.*, 2000; Nübel *et al.*, 2000; Griesbeck *et al.*, 2002; Schödl, 2003; Wakai *et al.*, 2004; Marcia *et al.*, 2009)
II	Bacteria (*Firmicutes*, *Proteobacteria*) Eukayotes (*Homo sapiens*)	mM, µM with partners	ND	1FCD	(Vande Weghe and Ow, 1999; Shibata and Kobayashi, 2006; Shibata *et al.*, 2007; Hildebrandt and Grieshaber, 2008; Theissen and Martin, 2008b)
III	Archaea (*Sulfulobales*) Bacteria (*Chlorobiaceae*)	ND	ND	1FCD	–
IV	Bacteria (*Chlorobiaceae*, *Chromatiaceae*)	yes (K_M ND)	respiration	3HYV	(Reinartz *et al.*, 1998; Chan *et al.*, 2009)
V	Archaea (*Sulfulobales*, *Acidianus ambivalens*)	µM	respiration	3H8L	(Brito *et al.*, 2009)
	Bacteria (*Chlorobiaceae* and others)	not found	not found	3H8L	(Chan *et al.*, 2009)
VI	Bacteria (*Chlorobiaceae*, *Aquificales*, and others)	yes (K_M ND)	respiration	ND	(Chan *et al.*, 2009)

[a]: ND = not determined.
[b]: besides sulfide detoxification.

classification of this protein family. The structure-based sequence fingerprints are hereafter listed for each newly defined SQR class (Table 3.3, Figure 3.9 – 10, Appendix tables 35 – 37). Type I SQRs are proteins of *Aquificaceae*, *α-* and *β-Proteobacteria* and *Cyanobacteria* (Theissen *et al.*, 2003). Their activity was studied for the SQRs of *O. limnetica* (Arieli *et al.*, 1994), *R. capsulatus* (Schütz *et al.*, 1997; Griesbeck *et al.*, 2002), *A. aeolicus* (Aq_2186) [(Nübel *et al.*, 2000; Schödl, 2003)], *A. halophytica* (Bronstein *et al.*, 2000), *A. ferrooxidans* (Wakai *et al.*, 2004) and *P. denitrificans* (Schütz *et al.*, 1998). They have affinity to their substrates in the micromolar range. In the cell, they are probably expressed in a constitutive manner in order to maintain efficient sulfide detoxification. For *A. aeolicus*, sulfide exposure primarily derives from the volcanic vents in its natural environment or from the catabolism of thiosulfate used as energy source (Chapter 1.4.2) and of cysteines, methionines and Fe-S clusters (Chapter 1.5.1) in the artificial growing medium. In addition, it was assessed that type I SQRs play a role in sulfide-dependent respiratory and photosynthetic processes (Arieli *et al.*,

1994; Nübel *et al.*, 2000). From the structure of *A. aeolicus* SQR, the following fingerprints can be defined for type I SQRs. Distinctive regions are a "capping loop" in position 294 – 312 (*A. aeolicus* SQR numbering) and a C-terminus composed of a two-helix amphipathic motif and of an elongated loop. Necessary key residues are (i) Cys 124, involved in FAD binding and possibly in electron transfer, (ii) Cys 156 and Cys 347, a redox active couple, (iii) Ser 155, Val 294 and Pro 311, mediating sulfide access to the active site, (iv) Glu 162 and Lys 312, probable proton donors/acceptors during the reductive semireaction, (v) Ile 346 and Phe 385, involved in quinone binding, and (vi) Glu 318 and Lys 382, involved in proton exchange during quinone reduction. Finally, the hydrophobicity of the residues shaping the cavity of polysulfur elongation is also a conserved feature in type I SQRs.

Type II SQRs are proteins of Bacteria (*Firmicutes* and *Proteobacteria*), including a few pathogens (i.e. *Staphylococcus aureus* and *Pseudomonas aeruginosa*) and of Eukarya, including yeast (i.e. *S. pombe*), worms (i.e. *A. marina*), insects (i.e. *Drosophila melanogaster*) and mammals (i.e. *Homo sapiens*), but not plants (Shahak and Hauska, 2008). SQR activity was proven for bacterial (*P. putida* and *B. stearotheromophilus*) and mitochondrial homologues (*S. pombe* and *A. marina* SQRs) but only low (millimolar) affinity to sulfide was detected in the absence of reaction partners (Vande Weghe and Ow, 1999; Shibata and Kobayashi, 2006; Shibata *et al.*, 2007; Hildebrandt and Grieshaber, 2008; Theissen and Martin, 2008b). It has been proposed that the physiological partners of type II SQRs are a sulfur dioxygenase (ETHE-1) and a sulfurtransferase (rhodanese) localized in the mitochondrial matrix (Hildebrandt and Grieshaber, 2008; Tiranti *et al.*, 2009; Lagoutte *et al.*, 2010), but the nature of the interactions occurring between the three proteins is unknown to date. It certainly acquires a special interest given the increased importance attributed to sulfide, a newly characterised neurotransmitter in mammals. We believe that the best structural model currently available for type II SQRs is the *A. vinosum* FCSD. In fact, type II SQRs and bacterial FCSDs cluster in the same branch when phylogenetic trees are calculated including both sets of protein sequences [(Chan *et al.*, 2009) and Figure 3.10]. The hallmark of type II SQRs is the absence of any conserved elongated loop, except for the elongated N-terminal signal peptide characteristic of the eukaryotic enzymes. In addition, a conserved tyrosine replaces the Cys 124 of type I SQRs and probably binds FAD through an 8-α-O-tyrosyl bond (Heuts *et al.*, 2009). The other two essential cysteine residues are instead conserved and are likely to be exposed to the bulk solvent, as in FCSD, where they would be accessible by the SQR reaction partners sulfur dioxygenase and sulfurtransferase. Val 294 is here substituted by an aspartate, while Glu 162 is replaced by a lysine. The residues involved in quinone reduction are not well conserved with respect to type I SQRs, and FCSD is obviously not a good model for that protein region. Further data are necessary to describe the quinone binding site and membrane attachment of type II SQRs.

Type III SQRs are the least characterized SQRs to date. They are proteins belonging to green sulfur bacteria (*Chlorobiaceae*, i.e. *C. tepidum* CT1025) and to Archaea (*Sulfulobales*). Neither their enzymatic activity nor their physiological role have been studied yet. From sequence comparison studies, it can be said that they are also missing the "capping loops", like type II SQRs, but that they constitute an independent group because the similarity is higher within type III SQRs than to type II SQR members. Currently, the best model for type III SQRs is also FCSD.

Type IV SQRs are predominantly found in green sulfur bacteria (*Chlorobiaceae*, i.e. *C. tepidum* CT0117), but not exclusively (one homologue is encoded in the purple sulfur γ-proteobacterium *A. vinosum*). Their SQR activity had been detected in *A. vinosum* (Reinartz *et al.*, 1998) and it was recently proven also for CT0117, a constitutively expressed protein, necessary but not sufficient for growth of C. *tepidum* on medium containing sulfide as the sole electron donor (Chan *et al.*, 2009). The type IV SQRs are characterized by a series of five elongated loops. Three of them extend surface regions of *A. aeolicus* SQR. Precisely, they are inserted at Thr 90 near the adenosine moiety of FAD, at His 193 near the Val 294 – Lys 312 "capping loop", and at Asn 278 near the C-terminal end. The other two loops correspond to the Val 294 – Lys 312 "capping loop" of *A. aeolicus* SQR, although this protein segment is longer in type IV SQRs than in type I SQRs, and to the C-terminal loop of *A. aeolicus* SQR (type I and IV SQRs approximately share the same length). Additionally, all three cysteine residues characteristic of type I SQRs are conserved. Interestingly, the chemical properties of the residues that define sulfide access in type I SQRs are also conserved in type IV SQRs (Val 294 is replaced by an isoleucine, Ser 155 by a threonine and Pro 311 is strictly conserved), indicating a similarity in the sulfide entry channel. Based on these observations and on the fact that type I and IV SQRs cluster in the same branch of the SQR phylogenetic tree [(Pham *et al.*, 2008), and Figure 3.10], the *A. aeolicus* SQR structure may be the best model for future studies on type IV SQRs.

Type V SQRs are a group of archaeal proteins, predominantly belonging to *Sulfulobales*. *A. ambivalens* SQR also belongs to this group (Pham *et al.*, 2008), unlike previously reported (Brito *et al.*, 2009). Type V SQRs possess SQR activity and show micromolar affinity to their substrates, and play a role in sulfide dependent respiration (Brito *et al.*, 2009). However, their structure differs significantly from that of type I SQRs (see Chapters 3.3 – 3.7). Important sequence fingerprints of type V SQRs are an elongated "capping loop" in position 154 – 178 (*A. ambivalens* SQR numbering) and a shorter C-terminus than in type I SQRs. Conserved key residues are Cys 129, which binds FAD covalently, Cys 178 and Cys 350, which constitute a thiol/disulfide pair, and Ser 214, Asp 215 and Asp 353, involved in sulfide binding and in proton exchange during sulfide oxidation along with Asp 307 and Lys 315.

```
FAD-binding fingerprints
SQR fingerprints                          Eukaryotic SQR signal peptide                                    ___I___
key residues (A. aeolicus SQR)
key residues (A. ambivalens SQR)
```

SQR type I
```
 1  gi15607118   Aaeolicus              ------------------------------------------------------------MAKHVVVIGGGVGG
 2  gi21464540   Tdenitrificans         ------------------------------------------------------------MA-HIVILGAGVGG
 3  gi2385380    Rcapsulatus            ------------------------------------------------------------MA-HIVVLGAGLGG
 4  gi8118249    Olimnetica             ------------------------------------------------------------MA-HVAVIGAGLAG
 5  gi103492069  Aferrooxidans          ------------------------------------------------------------MA-HVVILGAGTGG
 6  gi17232004   NostocPCC7120          ------------------------------------------------------------MA-HIVIVGAGLGG
 7  gi38505571   SynechocystisPCC6803   ------------------------------------------------------------MA-QIVVIGAGLGG
 8  gi8118252    Ahalophytica           ------------------------------------------------------------MA-HIVIVGGGFGG
```

SQR type II
```
 9  gi19112859   Spombe        M----LTLNST-I--KSV--------------------TG-SFQSASMLARFASTHHKVLVVGGGSAG
10  gi24657386   Dmelanogaster M----NRRLPG-TI-SQ-------CQKLLLATHNHAPTAFST-RFLSTS-RVNRERQECQVLVVGGGTGG
11  gi62896485   Hsapiens      M----VPLVAV--V-SGPRAQLFAC-LLRLGTQ--QVGPL---QLHTG-ASHAARNHYEVLVLGGGSGG
12  gi148696168  Mmusculus     SILVMAPLVAV--V-SSPRARLFAC-FLRLGTQ--QAGPL---QLHTG-ACCTAKNHYEVLVLGGGAGG
13  gi157092622  Amarina       M----ASRKVS-LVLTG---RVLMCQQLPAQAARAAVAPHVA-SFSSSAPLETAKKMNYKMVVVGGGTGG
14  gi82749809   Saureus       M----N--------------------------------------------------KHYQIVIGGGTAG
15  gi15597541   Paeruginosa   M----QSAN-----------------------------------------------ASELTTDIAVIGGGAAG
16  gi17546261   Rsolanacearum M----PPSPNA---------------------------------------------NPNPPHHPPTEVLIIGGGAAG
```

SQR type III
```
17  gi15920835   Stokodaii              -----------------------------------------------------------MK-RVIIAGGNIAG
18  gi21673852   Ctepidum               -----------------------------------------------------------MSKKIVVLGAGTAG
19  gi11498167   Afulgidus              -----------------------------------------------------------MK-NIVIVGGGVAG
20  gi23015306   Mmagnetotacticum       -----------------------------------------------------------MSKTILIVGGGLSG
21  gi15899356   Ssolfataricus          -----------------------------------------------------------MAKRVIIAGGNIAG
22  gi18313472   Paerophilum            -----------------------------------------------------------MK-SVVIIGGGVGG
```

SQR type IV
```
23  gi21672958   Ctepidum               -----------------------------------------------------------MA-KVVVLGAGVSG
24  gi193211736  Cparvum                -----------------------------------------------------------MA-KVVVLGAGVSG
25  gi241775880  Avinosum               -----------------------------------------------------------MA-RILILGAGISG
26  gi145218949  Pvibrioformis          -----------------------------------------------------------MA-KVVVLGAGVSG
27  gi78185961   Pluteolum              -----------------------------------------------------------MA-KVVVLGAGVSG
28  gi78776818   Sdenitrificans         -----------------------------------------------------------MA-RLVVLGGGVSG
```

SQR type V
```
29  gi238828305  Aambivalens            -------------------------------------------------------MTKVLVLGGGRFGA
30  gi229580793  Sislandicus            ---------------------------------------------------MKLIRMTKVLVLGARFGG
31  gi15899031   Ssolfataricus          ---------------------------------------------------MRLRSMTKVLVLGARFGG
32  gi15922815   Stokodaii              ----------------------------------------------------------MPKVLVLGARFGG
33  gi13542037   Tvolcanium             ------------------------------------------------------MMT-NPKVIILGGRFGG
34  gi16082149   Tacidophilum           ----------------------------------------------------M------KILILGGRFGG
35  gi21673704   Ctepidum               ------------------------------------------------------MKT-KPHVLILGGNFAG
36  gi193212535  Cparvum                ------------------------------------------------------MKT-KPHVLILGGNFAG
37  gi255021328  Acaldus                ------------------------------------------------------MSN-KPHVVVLGGNFAG
38  gi256759353  Tintermedia            ----------------------------------------------------MSAAKPRVLILGGNFAG
```

SQR type VI
```
39  gi21673913   Ctepidum               -------------------------------------------------------MKKVLILGGGIAG
40  gi15606162   Aaeolicus              --------------------------------------------------------MAKVLVLGGGIGG
41  gi78186989   Pluteolum              -------------------------------MPLRHAYYHDK-QHNAVKNIKSVLILGGGIAG
42  gi225850055  Pmarina                --------------------------------------------------------MKKVLVLGGGTAG
43  gi78364011   Tcrunogena             -----------------------------------------------------M-NITILGTGFAA
44  gi86748177   Rpalustris             -----------------------------------------------------MTNITIIGSGFAA
```

FCSD
```
45  gi241777285  Avinosum      ----------MTLNRRD-FIKTSGAA--VAAV-GIL----------GF--PHLAFGAGRKVVVGGGTGG
46  gi153872492  Beggiatoa     ----------MSNITRRNFIKIA-AG--TAAAG-MV----------SF--PFIAYGAKKQVVVVGGGAGG
47  gi19879585   Climicola     ----------MSISRRD-FNKLLVSG--V--AGSAF-----GLFGMSG-NTFAAGGSKKHVVVGGGFGG
48  gi192289979  Rpalustris    ----------MTSISRRSFGKLAA-------AG----------LATLAA-PSLLHAKDGARLVIIGGGPGG
49  gi15605781   Aaeolicus     ----------MGVNRRD-VFKLAGLG---VALG-ALQLPSIACAAKTS-NSLLSPSKGKRIVIVGGGWAG
50  gi21436536   Pdenitrificans ---------M ILTRRS-LIAGA-----TAATG-----------LLY--APVVLGQGKPKVVVIGGGAGG
```

Figure 3.9: Structure-based sequence alignment. 50 representative SQR and FCSD sequences were aligned with the program T-COFFEE in the mode Expresso, using the available structural information for the *A. aeolicus* and the *A. ambivalens* SQRs and for the *A. vinosum* FCSD. The sequences are indicated with increasing numbers (1 to 50, corresponding to the numbering used in Figure 3.10), with their *gi* identifier and with the abbreviated name of the organism which they belong to. The abbreviations correspond to: *Aquifex aeolicus* (Aaeolicus), *Thiobacillus denitrificans* (Tdenitrificans), *Rhodobacter capsulatus* (Rcapsulatus), *Oscillatoria limnetica* (Olimnetica), *Acidithiobacillus ferrooxidans* (Aferrooxidans), *Nostoc* species strain PCC 7120 (NostocPCC7120), *Synechocystis* species strain PCC 6803 (SynechocystisPCC6803), *Aphanothece halophytica* (Ahalophytica), *Saccharomyces pombe* (Spombe), *Drosophila melanogaster* (Dmelanogaster), *Homo sapiens* (Hsapiens), *Mus musculus* (Mmusculus), *Arenicola marina* (Amarina), *Staphylochoccus aureus* (Saureus), *Pseudomonas aeruginosa* (Paeruginosa), *Ralstonia solanacearum* (Rsolanacearum), *Sulfolobus tokodaii* (Stokodaii), *Chlorobium tepidum* (Ctepidum), (see next page)

```
FAD-binding fingerprints                              II
SQR fingerprints                                                              type V SQR "capping loop"
key residues (A. aeolicus SQR)                     C124                                           S155 C156      E162
key residues (A. ambivalens SQR)                   C129                                           C178

         1  gi15607118  Aaeolicus         96 IEYDYLV IATG 122 S IC -T 149 GA - - - - - - - - - - - - - - - - - - - - IPGVSCF -GPAYE
         2  gi21464540  Tdenitrificans    96 VDYDFLV IATG 126 S VC -S 153 GA - - - - - - - - - - - - - - - - - - - - VQGASCY -GPAYE
         3  gi2385380   Rcapsulatus       96 VSYDQ IV IATG 125 S IC -H 152 GA - - - - - - - - - - - - - - - - - - - - AQGASCF -GPAYE
SQR      4  gi8118249   Olimnetica        95 LEYDYVVVATG 125 SVC -N 152 GA - - - - - - - - - - - - - - - - - - - - VPGASCF -GPAYE
type I   5  gi103492069 Aferrooxidans     96 VHYDYLM IATG 126 S IC -T 153 GA - - - - - - - - - - - - - - - - - - - - MAGASCF -GPAYE
         6  gi17232004  NostocPCC7120     95 ISYDYL IIATG 124 SVC -N 151 GA - - - - - - - - - - - - - - - - - - - - LPKTSCL -GPAYE
         7  gi38505571  SynechocystisPCC6503 94 LNYDYL IIASG 121 S IC -T 148 GA - - - - - - - - - - - - - - - - - - - - APGTGCF -GPAYE
         8  gi8118252   Ahalophytica      96 FDYDYLV ITTG 126 SVC -N 153 GA - - - - - - - - - - - - - - - - - - - - VPGASCM -GPAYE

         9  gi19112859  Spombe           137 ISYDYLVMAAG 171 T IY -S 196 TQ - - - - - - - - - - - - - - - - - - PSGVLKCA -GAPQK
        10  gi24657386  Dmelanogaster    136 IKYDFL IIATG 170 S IY -S 195 TF - - - - - - - - - - - - - - - - - - PNCP IKCA -GAPQK
        11  gi62896485  Hsapiens         136 ISYRYL IIALG 168 SNY -S 193 TF - - - - - - - - - - - - - - - - - - PNTPVKCA -GAPQK
SQR     12  gi148696168 Mmusculus        140 ISYRYL IIALG 172 SNY -S 197 TF - - - - - - - - - - - - - - - - - - PNTPVKCA -GAPQK
type II 13  gi157092622 Amarina          142 VKYDFLVCAMG 175 SNY -W 200 TF - - - - - - - - - - - - - - - - - - PNSP IKCA -GAPQK
        14  gi82749809  Saureus           99 VYYDFLVVAPG 131 SNY -S 156 TH - - - - - - - - - - - - - - - - - - PNTP IKCG -GAPMK
        15  gi15597541  Paeruginosa      103 LEYRNL IVCPG 135 SNY -R 160 TQ - - - - - - - - - - - - - - - - - - PAMP IKCA -GAPQK
        16  gi17546261  Rsolanacearum    109 VRYSQL IVSPG 141 SNY -R 166 TQ - - - - - - - - - - - - - - - - - - PAMP IKCA -GAPQK

        17  gi15920835  Stokodaii        103 HSYDYLV ITTG 130 SPY -E 155 NV - - - - - - - - - - - - - - - - - - AKLPHRCP -VAPLE
SQR     18  gi21673852  Ctepidum          95 FTYDFLV ISTG 126 TFY -Y 151 N I - - - - - - - - - - - - - - - - - - AELPFKCP -VAP IE
type    19  gi11498167  Afulgidus        100 VNYDYLV IATG 127 WFY -N 152 SV - - - - - - - - - - - - - - - - - - MG IPHKCP -VAP IE
III     20  gi23015306  Mmagnetotacticum 101 FKYDYLVLATG 128 WFY -D 153 NV - - - - - - - - - - - - - - - - - - NAPHKCP -VAPLE
        21  gi15899356  Ssolfataricus    104 HSYDYLV ITTG 131 TVY -E 156 NT - - - - - - - - - - - - - - - - - - AKLPHRCP -VAPLE
        22  gi18313472  Paerophilum      100 LKYDYL IVATG 127 TLW -T 152 SV - - - - - - - - - - - - - - - - - - TSTPYKCP -VAPYE

        23  gi21672958  Ctepidum         110 VDYDYLVNATG 139 SVC -T 171 GT - - - - - - - - - - - - - - - - - - GHAMATCQ -GAAFE
SQR     24  gi193211736 Cparvum          110 VDYDYLVNATG 139 SVC -T 171 GT - - - - - - - - - - - - - - - - - - GHAMATCQ -GAAFE
type    25  gi241775880 Avinosum         110 IEYDYLVNATG 140 SVC -T 172 GT - - - - - - - - - - - - - - - - - - GHGMCTCQ -GAAFE
IV      26  gi145218949 Pvibrioformis    110 VEYDYLVNATG 139 SVC -T 171 GT - - - - - - - - - - - - - - - - - - GHAMATCQ -GAAFE
        27  gi78185961  Pluteolum        110 VEYDYLVNATG 139 SVC -T 171 GT - - - - - - - - - - - - - - - - - - GHAMATCQ -GAAFE
        28  gi78776818  Sdenitrificans   110 IEYDYL INATG 143 SVC -T 175 GT - - - - - - - - - - - - - - - - - - GHGLCTCQ -GAAFE

        29  gi238828305 Aambivalens      100 EEYDYV IVG IG 127 SVC -E 152 G SGPFYQGHNPKP - -KVPENFVP NADSACE -GPVFE
        30  gi229580793 Sislandicus      105 EDYDYL IIALG 132 SVC -E 157 G SGLFYQGKTPKP - -KVPENYVP VADAACE -GP IFE
        31  gi15899031  Ssolfataricus    105 EDYDYL IIALG 132 SVC -E 157 G SGLFYQGKSPKP - -KVPENYVP VADAACE -GP IFE
SQR     32  gi15922815  Stokodaii        100 EEYDYV IVA IG 127 SVC -E 152 G SGFFYQGHNPAP - -NVPKNFVP NADAACE -GPVFE
type V  33  gi13542037  Tvolcanium       103 EDYDYL IALG 130 SVC -E 155 G SAKF IQDNKNRPT -NNPDAYAP LLESACE -GPVYE
        34  gi16082149  Tacidophilum      99 EKYDFLV IALG 126 SVC -E 151 G SAKF IQDTKNRPT -NTPDRYAP IADSACE -GPVFE

        35  gi21673704  Ctepidum         108 LTYDYLV IALG 135 TVS -D 162 G SARFHQGTKGK - - - - - -LDFVP MAKAACE -GPPVE
        36  gi193212535 Cparvum          108 LTYDYLV IALG 135 TVS -D 162 G SARFHQGTKGK - - - - - -LDFVP MAKAACE -GPPVE
        37  gi255021328 Acaldus          107 MHYDYVVVAVG 134 TCT -D 160 G SARFHQGDGTKD IKLYGGHAFP SAEAACE -GPPVE
        38  gi256759353 Tintermedia      108 TTYDYLVVALG 135 TVT -D 161 G SARFTQGDGAKGLEPYPGGS IP AAEAACE -GPPVE

        39  gi21673913  Ctepidum          94 SDFDFLV IALG 118 S ICGA 147 GFG - - - - - - - - - - - - - - - - GNPKDPSAVRGG PGFE
SQR     40  gi15606162  Aaeolicus         94 EDFDYLVVALG 118 S ICGS 147 GFG - - - - - - - - - - - - - - - - GNPKDKTAVRG RPVFE
type    41  gi78186989  Pluteolum        112 SGTDAVVVALG 136 S ICAS 165 GFG - - - - - - - - - - - - - - - - GNPKDKSAVRGG PAFE
VI      42  gi225850055 Pmarina           94 SDFDYLV IALG 118 S ICGS 147 GFG - - - - - - - - - - - - - - - - GNPRAKEAVRGG PVFE
        43  gi78364011  Tcrunogena        93 FNNNGL IIASG 118 TPCEG 144 GFG - - - - - - - - - - - - - - - - GNPKEPSAMRGG PMFE
        44  gi86748177  Rpalustris        95 FSNDHL IVACG 119 IPCEG 145 GFS - - - - - - - - - - - - - - - - TNPEEPGAMRGG PMFE

        45  gi241777285 Avinosum         125 FGYDRCVVAPG 159 - -K -A 183 AP - - - - - - - - - - - - - - - - - - - PAAPFRCP -PGPYE
        46  gi153872492 Beggiatoa        125 FTYERL IVAPG 159 - -L -A 183 AV - - - - - - - - - - - - - - - - - - - PPKPFRCP -PGPYE
FCSD    47  gi19879585  Climicola        131 IGYDRLVVSPG 166 - -Q -A 190 CP - - - - - - - - - - - - - - - - - - - PDNPFRCP -PGPYE
        48  gi192289979 Rpalustris       124 IAYDKL IVSPG 158 VPE -R 184 A I - - - - - - - - - - - - - - - - - - - PANPFRCP -PGPYV
        49  gi15605781  Aaeolicus        135 VEYDYL ILAPG 173 - - I -P 196 VV - - - - - - - - - - - - - - - - - - - PPPPHRCP -PAPYE
        50  gi21436536  Pdenitrificans   123 LPYDRLVLSPG 157 - -K -A 181 VA - - - - - - - - - - - - - - - - - - - PPNPYRCP -PGPYE
```

Figure 3.9 (continued). *Archaeoglobus fulgidus* (Afulgidus), *Magnetospirillum magnetotacticum* (Mmagnetotacticum), *Sulfolobus solfataricus* (Ssolfataricus), *Pyrobaculum aerophilum* (Paerophilum), *Chlorobaculum parvum* (Cparvum), *Allochromatium vinosum* (Avinosum), *Prosthecochloris vibrioformis* (Pvibrioformis), *Pelodictyon luteolum* (Pluteolum), *Sulfurimonas denitrificans* (Sdenitrificans), *Acidianus ambivalens* (Aambivalens), *Sulfolobus islandicus* (Sislandicus), *Thermoplasma volcanium* (Tvolcanium), *Thermoplasma acidophilum* (Tacidophilum), *Acidithiobacillus caldus* (Acaldus), *Thiomonas intermedia* (Tintermedia), *Persephonella marina* (Pmarina), *Thiomicrospira crunogena* (Tcrunogena), *Rhodopseumonas palustris* (Rpalustris), *Beggiatoa* species PS (Beggiatoa), *Chlorobium limicola* (Climicola), *Paracoccus denitrificans* (Pdenitrificans). Sequences 35 to 38 correspond to the subbranch of type V SQR sequences, whose SQR activity has not been proven (see text). The fingerprint regions are highlighted in red and indicated at the top of the alignment. The alignment of the N- and C-terminal regions was manually adjusted with the software Jalview. The complete unmodified alignment is reported in Appendix table 37.

101

```
FAD-binding fingerprints                IIIa                        IIIb
SQR fingerprints                             type I SQR "capping loop"
key residues (A. aeolicus SQR)    V294                 P311 K312   E318              I346 C347
key residues (A. ambivalens SQR)  D307                 K315                          V349 C350 D353

SQR type I
 1 gi15607118  Aaeolicus            294 VVTA IPP IEKTP --------IPTG VPK TGMM IE   344 SA IC IADFGE
 2 gi21464540  Tdenitrificans       301 VCVA IPPVEATP --------VPTG TPK TG YM IE   351 NA IC LADFGD
 3 gi2385380   Rcapsulatus          300 VCVA IPPVG PTP --------VPCG VPK TG FM IE   350 NAVC LADFGD
 4 gi8118249   Olimmetica           291 V IVE INPPEKTP --------LPVG VPK TGQM TE   342 EA IC IADFGD
 5 gi103492069 Aferrooxidans        302 IATA IPPVETTP --------VPTGAPK TG YM IE   353 NAVC FADMGD
 6 gi17232004  NostocPCC7120        290 VVVE IKPSEVTP --------LPLG VPK TGQM TE   341 DA IC FADFGN
 7 gi38505571  SynechocystisPCC6803 286 VSVHLGQ INDTQ --------VS IGTPK SGQM TE   337 EALC FAEYGD
 8 gi8118252   Ahalophytica         292 V ITQ LAAPEETE --------VPLGAPK TGQM TE   343 EA ICMADFGD

SQR type II
 9 gi19112859  Spombe               340 DCSGLP -----------------TSKTYAA IT   380 YTSC PLLTGY
10 gi24657386  Dmelanogaster        339 DSASSP -----------------NSKTAAAAA   379 YSSC P IVTGY
11 gi62896485  Hsapiens             336 DCTNLP -----------------TSKTAAAVA   376 YTSC PLVTGY
12 gi148696168 Mmusculus            340 DCTNLP -----------------TSKTAAAVA   380 YTSC PLVTGY
13 gi157092622 Amarina              342 DNTD IP -----------------TAKTAAAVA   383 YTSC PLVTSK
14 gi82749809  Saureus              298 DASNVP -----------------TSKTGAA IR   338 YTSC P IVTGY
15 gi15597541  Paeruginosa          305 DVCGTA -----------------NAKTAAAAR   345 YGGC PLTVER
16 gi17546261  Rsolanacearum        311 DACSTP -----------------NAKTAAAVR   351 YGGC PLTVER

SQR type III
17 gi15920835  Stokodaii            292 DATDLP -----------------VSKAGSTAD   332 DVLC Y IATGT
18 gi21673852  Ctepidum             288 DATNVP -----------------TSKAGSVAH   328 HSTC F IVYSK
19 gi11498167  Afulgidus            289 DATNLP -----------------VSKAGSVAH   329 KAMC F IETGF
20 gi23015306  Mmagnetotacticum     288 DTTN IP -----------------ISKAGSTAH   328 KVFC FVETGL
21 gi15899356  Ssolfataricus        293 DATDLP -----------------VSKAGSTAD   333 DVLC Y IATGT
22 gi18313472  Paerophilum          288 DATNLP -----------------VPKAGSVAH   328 RV IC F ILTGF

SQR type IV
23 gi21672958  Ctepidum             347 IAFAPPHS ISKPMTSVNGRQ IFPTPPR TGMPSG   410 GAAC IVSAGF
24 gi193211736 Cparvum              347 IAFAPPHP ISKPLKSENGRQ IFPTPPR TGMPSG   410 GAAC IVSGGF
25 gi241775880 Avinosum             348 IAFAPPHP ISKVMKSPSGLQ ISPTPPR TGMPSA   410 GAAC VASTGM
26 gi145218949 Pvibrioformis        347 IAFAPPHA ISKPMTSEKGRA IFPTPPR TGMPSG   409 GAAC IVSGGF
27 gi78185961  Pluteolum            347 IAFAPPHA ISKPMTSVTGRP IFPTPPR TGMPSG   409 GAAC IVSGGF
28 gi78776818  Sdenitrificans       351 IAFAPPHP ISKPMSSPNGTP INPTPPR TGMPSG   414 GAAC VASAGK

SQR type V
29 gi238828305 Aambivalens          307 DANSMT -----------------VPKLG YLAV   347 T IVC VADNPY
30 gi229580793 Sislandicus          312 DANSLT -----------------VPKLG YLAV   352 T IIC VADNPF
31 gi15899031  Ssolfataricus        312 DANSLT -----------------VPKLG YLAV   352 T IIC VADNPF
32 gi15922815  Stokodaii            307 DANSLT -----------------VPKLG YLAV   347 T IVC VADNPY
33 gi13542037  Tvolcanium           309 D INAL I -----------------VPKMAALAV   349 K IVC IADDPY
34 gi16082149  Tacidophilum         305 DVNAL I -----------------VPKMAFLAV   345 K IVC VADSPY

35 gi21673704  Ctepidum             313 DAAACT -----------------VPKLGS IGH   357 EV IC YGDMGD
36 gi193212535 Cparvum              313 DAAACT -----------------VPKLGS IGH   357 EV IC YGDMGD
37 gi255021328 Acaldus              317 DAAA IT -----------------VPKLGG IGH   360 VVYC IGDMGA
38 gi256759353 Tintermedia          318 DCAAAT -----------------LPKLGA IGH   361 VVYC IGD IGN

SQR type VI
39 gi21673913  Ctepidum             288 DSAALEGPE -------------WKAKQGH IAE   335 NVLC VMDTGD
40 gi15606162  Aaeolicus            288 DSAA IEGPP -------------WTAKQGHLAE   335 N ILC LMDMGW
41 gi78186989  Pluteolum            306 DSAALEGPD -------------WKAKQGHLAE   353 N ILC VMD IGN
42 gi225850055 Pmarina              288 DVVS IEGPE -------------WKAKQGHLAE   335 N IFC LMDMGN
43 gi78364011  Tcrunogena           284 DSGSYPGPE -------------WQAKQAHMAD   328 ELVC VIDT -L
44 gi86748177  Rpalustris           286 DAGSFPGPD -------------WMPKQAHQAD   330 ELVC IVDT -L

FCSD
45 gi241777285 Avinosum             324 DAC IANP -----------------MPKSG YSAN   364 LNTC YS ILAP
46 gi153872492 Beggiatoa            326 DAA IASP -----------------MPKSAYAAN   366 INTC YSM IGE
47 gi19879585  Climicola            326 DAC IAGA -----------------MPKSG FAAS   366 VNTC YSL IGP
48 gi192289979 Rpalustris           321 DAN IGGP -----------------LPKSG F IAN   362 FNTC YSHVGD
49 gi15605781  Aaeolicus            331 DANSVKG -----------------FPKSGDMAN   374 TNTC YSMVNG
50 gi21436536  Pdenitrificans       317 DSSAQGD -----------------MPKSAFSAN   357 SNTC WSLLAP
```

Figure 3.9 (continued).

```
FAD-binding fingerprints
SQR fingerprints                              type I SQR helix-turn-helix amphipathic motif        C-terminal loop
key residues (A. aeolicus SOR)           K382   F385
key residues (A. ambivalens SQR)         K397   A401
```

```
                                              ↓      ↓
      1  gi15607118   Aaeolicus            376 K W A H Y F K - T A F E K Y F L W K V R N - G N I A P S F E E K V L E - I - F L K - V H P I - E L C K D C E G A P G S R C
      2  gi21464540   Tdenitrificans       383 K W V H L A K - V A F E K Y F M R K M K K - G S T E P L Y E K Y V L - - - - G L M G I K K L K - - - - - - - - - - - - - -
      3  gi2385380    Rcapsulatus          382 K W V H W A K - E G F E R Y F M H K L R R - G T S E T F Y E K A A M K - L L G I - - - - - - - - - - - - - D K L K A V K K G
      4  gi8118249    Olimnetica           379 K W V S W S K - T A F E T F F L S K M R F - G L A V P W F E R W G L R F - M G L S L V E P L D T T R E T G N Q A F A S K S
      5  gi103492069  Aferrooxidans        385 R W V H L A K - V A F E K Y F I R K M K M - G V S E P F Y E K V L F K - M M G I T R - - - - - - - - L K E E D T H R K A S
      6  gi17232004   NostocPCC7120        378 A W V T W A K - A A F E R Y F L A K M R F - G T A V P W F E K L A L K L - L G L S L V A P L A V - K - S S R N I S Q E N Y
      7  gi38505571   SynechocystisPCC6803 374 I W V N W A K - A A F E K Y F L A K M R L - G V G L P W F E K L G L R L L F G L S M S R P A - - - - M A E D I V Q P V N V
      8  gi8118252    Ahalophytica         380 L W V N W A K - N A F E W Y F L A K M R W - G T A V P W F E K L G L Y L - L R L T L V T P I S E T P T Q Q K D L T S I K G

      9  gi19112859   Spombe               435 R M F Y H L K K D F F P F V Y W N F A V R N G K W Y G S R G L I P P H F P V N - - - - - - - - - - - - - - - - - - - - - -
     10  gi24657386   Dmelanogaster        429 Y S M F F M K K E L M P V L Y W K L M M N - G Y W N G P A L M R K I F S V L K F N K K Q - - - - - - - - - - - - - - - - -
     11  gi62896485   Hsapiens             426 L S M Y L M K A D L M P F L Y W N M M L R - G Y W G G P A F L R K L F H L G M S - - - - - - - - - - - - - - - - - - - -
     12  gi148696168  Mmusculus            430 I T M Y L M K A D L M P F L Y W N M M L R - G Y W G G P A F L R K L F H L G M N - - - - - - - - - - - - - - - - - - - - -
     13  gi157092622  Amarina              433 R T M F H F K A D I F P Q M Y W H M L I K - G Y W H G P G L F R K A M H L G M S R - - - - - - - - - - - - - - - - - - - -
     14  gi82749809   Saureus              388 R S M Y I F K K D L L P K M Y W Y G M L K - G L I - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     15  gi15597541   Paeruginosa          395 R F A W L L K R H V L P W V Y W N A M L K - G R E W L - - - - A R P G R - - - - - - - - - - - - - - - - - - - - - - - -
     16  gi17546261   Rsolanacearum        401 Q S A W W L K T T V L P W L Y W N G L L K - G R E W L A R P S L P V E R - - - - - - - - - - - - - - - - - - - - - - - -

     17  gi15920835   Stokodaii            375 Y V H W W G K - I G Y N K L Y W T V T A K - A V V - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     18  gi21673852   Ctepidum             384 K M N W Y G K - L A F E P L Y W N V L L D - G K H L G M P P T L V M A G K E V G - - - - - - - - - - - - - - - - - - -
     19  gi11498167   Afulgidus            372 K F I H W M K - L A Y N R M Y W L - T A R - G M L - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     20  gi23015306   Mmagnetotacticum     371 Q M I H W F K - L A Y N R L Y W L - S A R - G L L - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     21  gi15899356   Ssolfataricus        376 Y V H W W G K - I M Y N K M Y W T V T A K - A V V - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     22  gi18313472   Paerophilum          372 R F F N R L K - D M T N Y S I W S V M R C - G L - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

     23  gi21672958   Ctepidum             443 D W - E - - K Y P E W G R D M T Y S V G E V G L A G - - H W L K F M L - H - Y L F - F H K A - K G Y P F W Y L I P E - - -
     24  gi193211736  Cparvum              443 D W - E - - K Y P E W G R D M N Y S V G E A G L A G - - H W L K F I L - H - Y L F - F H K A - K G Y P F W W L I P E - - -
     25  gi241775880  Avinosum             443 D Y - E - - T Y P E Y G R D M D L T F G E I G L A G - - H W M K Y L L - H - H V F - I Y Q A - K L R P G W S V L P D - - -
     26  gi145218949  Pvibrioformis        442 D W - D - - K Y P E W G R D M K Y S V G E A G L A G - - H W L K V L L - H - Y L F - F H K A - K G Y P F W W L I P E - - -
     27  gi78185961   Pluteolum            442 D W - D - - K Y P E W G R D M K Y S V G E A G L A G - - H W L K V L L - H - Y L F - F H K A - K G Y P F W W L I P E - - -
     28  gi78776818   Sdenitrificans       447 D F - D - - K Y P G T G R D T D Y T F G E I G L A G - - H W I K H I L - H - H L F - L W K A - K L K A G W T L I P E - - -

     29  gi238828305  Aambivalens          393 K Y Y M W T K G D M A - L E K F L A S W - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     30  gi229580793  Sislandicus          398 K Y F I W T K G D M A - L E K F L A S W - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     31  gi15899031   Ssolfataricus        398 K Y F I W T K G D M A - L E K F L A S W - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     32  gi15922815   Stokodaii            393 K Y F L W T K G D M A - L E K F L A S W - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     33  gi13542037   Tvolcanium           395 R Y F L W S K G D L A - M D K Y L T S W - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     34  gi16082149   Tacidophilum         391 R Y F L W S K G D L A - M D K Y L T S W - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

     35  gi21673704   Ctepidum             402 Y F A M G G K V P Y W Q W K M - - G S W M - - G D K I L - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     36  gi193212535  Cparvum              402 Y F A M G G K V P Y W Q W K M - - G S W M - - G D K I L - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     37  gi255021328  Acaldus              405 F F K T R G K M P E W G L D A - - S K L L A E K L F A A - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     38  gi256759353  Tintermedia          406 F F R S K G K M P P W G L D A - - A R I M A - E K M A A - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

     39  gi21673913   Ctepidum             376 V V G H W L K - K A W G Y Y Y K L S K - M K Y I P R I P G M - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     40  gi15606162   Aaeolicus            377 V L G H I M K - K A W G V Y Y K L Y K - L G K - - I P K I I - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     41  gi78186989   Pluteolum            394 L F G H W L K - H G W A H Y Y R F S K - L G R L L R T S G V - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     42  gi225850055  Pmarina              376 V I G H Y L K - K G W G V Y Y K L S K - L G K I P R I P G M - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     43  gi78364011   Tcrunogena           359 R L M H F A K - R F F E R W Y I K Q Y K - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     44  gi86748177   Rpalustris           361 K L F H W L K - R L F E R H Y L T T F R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

     45  gi241777285  Avinosum             420 E R - - E V Q Y A - - Y S W Y N N I V H D T F G - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     46  gi153872492  Beggiatoa            420 K R - - E V L Y A - - H S W F K N I T H D M F G - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     47  gi19879585   Climicola            420 A Q - - E A M F A - - E G W Y T N I S K D I W G - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     48  gi192289979  Rpalustris           419 Q R K L E A Q Y A - - D S W Y A S I T K D A F S - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     49  gi15605781   Aaeolicus            425 E R S Q R L A R A - T F E W A K A M Y R D M F S - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     50  gi21436536   Pdenitrificans       409 - R - - K L T Y E E S L G W Y A G I T A D I F G - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

SQR type I: 1–8
SQR type II: 9–16
SQR type III: 17–22
SQR type IV: 23–28
SQR type V: 29–38
SQR type VI: 39–44
FCSD: 45–50

Figure 3.9 (continued).

Figure 3.10: Structure-based phylogenetic tree. The phylogenetic tree was calculated from the alignment reported in Figure 3.9 and drawn with the software Geneious.

Additionally, Val 349 and Ala 401 may correspond to residues that bind quinones in type I SQRs while Lys 397 may mediate proton transfer, but the architecture of the quinone binding site in type V SQRs still has to be determined experimentally. The structural model for type V SQRs is *A. ambivalens* SQR..

Furthermore, a set of bacterial sequences clusters together with the archaeal type V SQRs. Their "capping loop" is present but shorter. Cys 129 is replaced by a serine or a threonine, residues that are not known to covalently bind FAD in position C8M (Hayashi *et al.*, 2001; Halada *et al.*, 2003). Finally, an alanine substitutes Ser 214, so that one putative hydrogen bonding group for sulfide in its access channel is missing. One member of this subgroup is *C. tepidum* CT0876, for which SQR activity was tested unsuccessfully (Chan *et al.*, 2009). In the absence of functional evidence, this cluster cannot be assigned to an independent SQR group.

Type VI SQRs are a group of bacterial proteins previously proposed as an "unclassified group" (Pham *et al.*, 2008), including CT1087 from *C. tepidum* and Aq_788 from *A. aeolicus* (which we have identified in our *A. aeolicus* membrane preparations, data not shown). The SQR activity of CT1087 was recently demonstrated (Chan *et al.*, 2009), so that these proteins can now be confidently classified as SQRs. Besides the absence of any obvious "capping loop", these proteins possess cysteine residues in the positions of *A. aeolicus* SQR Cys 124 and Cys 347 and an aspartate in place of Val 294. Strikingly, they also lack *A. aeolicus* SQR Cys 156, which is replaced by a valine or a methionine, according to our alignment. Interestingly, another cysteine (Cys 275 in CT1087) is conserved in type VI SQRs and corresponds to *A. aeolicus* SQR Phe 281, very far from the catalytic site. More experimental data are necessary in

order to understand whether type VI SQRs follow a completely different reaction mechanism not involving Cys 156 or whether they assume a different 3-D fold such that Cys 275 comes closer to the active site.

3.9.   Conclusions and future perspectives

In summary, this work has achieved (i) the identification and isolation of sulfide:quinone oxidoreductase (SQR) from the native membranes of the hyperthermophilic bacterium *Aquifex aeolicus*, (ii) the characterization of its biochemical properties, (iii) the characterization of its enzymatic activity and (iv) the determination of its three-dimensional X-ray structure in an "as-purified" form and in substrate- and inhibitor-bound forms. The structures reveal that SQR is a monotopic membrane protein, specifically interacting with the lipidic bilayer as a trimer and that the protein adopts various strategies to resist to high temperatures. Additionally, the architecture of both the sulfide oxidation and the quinone reduction active sites could be described. Most surprisingly, the cofactor flavin adenine dinucleotide (FAD) is bound to the protein through an unprecedented putative persulfide linkage in position C8M and SQR was isolated in an intermediate state of the reaction cycle, with the partially formed polysulfide product covalently bound to the active site. Based on the structural and biochemical characterization of *A. aeolicus* SQR and on the comparison with the structure of *A. ambivalens* SQR, also solved recently, the structural-functional relationships of SQR could be established, identifying the specific residues and the structural domains which are essential for catalysis. Consequently, a new structure-based classification of SQRs was proposed. On the basis of this structural, functional and bioinformatic characterization, the complicated sulfur polymerization SQR reaction can now be described in more detail. Two different mechanistic hypotheses are proposed in this work and it is suggested that not all SQRs share the same reaction scheme. Finally, the overall reaction of *A. aeolicus* SQR was described to a more speculative extent also in its cellular context. Interestingly, it seems plausible that despite being a monotopic membrane enzyme, SQR may play a role in facilitating sulfur transport across the cellular membrane.

Moreover, the project also opened new directions for future research on *A. aeolicus* SQR specifically and on SQRs in general.

First, the characterization of the *A. aeolicus* SQR active site needs to be extended in order to demonstrate conclusively the presence of a polysulfide bound to Cys 156 and to Cys 347 and the presence of a persulfide bridge between Cys 124 and FAD. Second, the protein structure must be solved in alternative conformations and intermediate stages for comparison with those predicted by the hypothesised reaction mechanisms. Attempts towards the characterization of other structural intermediates of the reaction cycle are particularly challenging, because of the

nature of the reaction itself. The polysulfide species are highly reactive and unstable under many working conditions, investigations on the protein-FAD connection is hampered by the fact that loss of the cofactor results in enzyme denaturation and, finally, reducing agents to remove the putative sulfur derivatizations from the active site have been so far unsuccessful, probably because they compromise the integrity of the two C-terminal disulfide bridges so crucial for the protein's stability. Nonetheless, further spectrophotometric and mass spectrometric experiments are already in progress. Reconstitution in artificial lipid vesicles is also a feasible biochemical experiment that would prove the overall SQR reaction. On one hand, this approach would allow studying the membrane-binding mode of SQR and determining i.e. whether the protein has or not the same affinity to bilayers of different composition. On the other hand, it would help in defining the nature of the reaction product and in particular whether the latter partitions in a hydrophilic or a hydrophobic environment. Finally, the establishment of an overexpression system is also planned as a tool to study SQR by molecular biological techniques.

To a wider extent then, the variety in the SQR family also requires further investigation. It cannot be excluded that SQRs with novel sequences will be found that define additional SQR types. In addition, the described heterogeneity among the SQR sequences and the still very limited structural information available raise intriguing questions about the mechanisms of catalysis and the functional roles of SQRs in different organisms. Of particular interest is the concomitant presence of SQRs of different types in some species. Their cellular roles and the regulation of their expression and activity must be in some way related to their structural and mechanistic differences. It is possible that new structural and functional data will become available soon, especially about *Acidithiobacillus ferrooxidans* SQR, whose structure has already been solved (Zhang *et al.*, 2009), but is not yet available.

Last but not least, the study of the eukaryotic SQRs has a high priority, as they are potentially important in humans for the homeostasis of sulfide, a mediator of sympathetic neurotransmission and a key metabolite in neurodegenerative diseases. Three parallel strategies are conceivable for the production of the human mitochondrial sulfide:quinone oxidoreductase dehydrogenase-like protein (SQRDL). First, heterologous expression in *E. coli* might yield high amounts of protein, considering the high production levels detected in previous attempts (Ackermann, 2007; Maier, 2007). Nonetheless the efficiency of FAD reincorporation might be a bottleneck for this approach. Second, cell-free expression either in the presence or in the absence of detergent might be advantageous in producing a soluble enzyme. In fact, SQR is a monotopic membrane protein and not a transmembrane protein and this work shows that for short periods of time (hours) *A. aeolicus* SQR is stable in the absence of detergents. In addition, the incorporation of the cofactor FAD need not be a limitation for the cell-free system.

Thioredoxin reductase and glutathione reductase, two members of the FDR superfamily to which SQR belongs, have been successfully produced in a cell-free system in an active form (Knapp and Swartz, 2004) by supplementing the reaction mixture with commercial FAD. Actually, the presence of FAD might be useful for evaluating the correct folding of the cell-free produced enzyme by simply recording its VIS-UV absorption spectrum. Third, production of SQRDL should be attempted in yeast, as a eukaryotic expression system might be advantageous for production of a mitochondrial target. Yeast expression has already been reported for *Arenicola marina* SQRDL in *Saccharomyces cerevisiae* (Theissen and Martin, 2008b). Once a convenient expression system is found, the study should attempt to clarify how SQRDL interacts with sulfur dioxygenase (ETHE-1) and sulfurtransferase functionally and structurally.

In conclusion, the data presented in this work allow drawing a first picture of the fascinating sulfide:quinone oxidoreductase reaction and they stimulate further genetic, biochemical and structural investigations.

4. Material and Methods

4.1. Materials

4.1.1. Suppliers

The suppliers of all devices, materials and chemicals used in this work are listed in Table 4.1.

Table 4.1: List of suppliers.

Company	Location	Web address
Adobe	San Jose, California, USA	http://www.adobe.com
Anatrace	Maumee, Ohio, USA	http://www.anatrace.com
Agilent	Böblingen, D	http://www.agilent.com
Applied Biosystems	Framingham, Massachusetts, USA	http://www.appliedbiosystems.com
Archaeenzentrum	Regensburg, D	http://www.biologie.uni-regensburg.de/ Mikrobio/Thomm/Arbeitsgebiete/ fermentation.htm
ATOMIKA Instruments	Tempe, Arizona, USA	http://www.cameca.com
Avanti Polar Lipids	Birmingham, Alabama, USA	http://avantilipids.com
Axygen Biosciences	Union City, California, USA	http://www.axygenbio.com
Bandelin Sonorex	Mörfelden-Walldorf, D	http://www.sonorex.com
Beckman Coulter	Krefeld, D	http://www.beckman.com
Biometra	Göttingen, D	http://www.biometra.de
Biomol	Hamburg, D	http://www.biomol.de
Bio-Rad Laboratories GmbH	Munich, D	http://www.bio-rad.com
Bio-Tek Instruments	Winooski, Vermont, USA	http://www.biotek.com
Bruker-Nonius	Rheinstetten, D	http://www.bruker.com
Calbiochem	Darmstadt, D	http://www.calbiochem.com
Cartesian Technologies	Huntingdon, UK	http://www.cartesiantech.com
Corning Life Sciences	Amsterdam, NL	http://www.corning.com
Coy Laboratories	Grass Lake, Michigan, USA	http://www.coylab.com
DigiLab	Huntingdon, UK	http://www.digilabglobal.com
Eppendorf	Hamburg, D	http://www.eppendorf.com
FEI-Philips	Hillsboro, Oregon, USA	http://www.fei.com
Fluka	Seelze, D	http://www.sigmaaldrich.com
GE Bayer Silicones	Bergen op Zoom, NL	http://www.gesilicones.com
GE Healthcare	Munich, D	http://www.gehealthcare.com
GERBU Biochemical Mart	Gaiberg, D	http://www.gerbu.de
Genomics Solutions	Huntingdon, UK	http://www.genomicsolutions.com

Genzyme	Haverhill, UK	http://www.genzyme.co.uk
Gilson	Bad Camberg, D	http://www.gilson.com
Glycon	Luckenwalde, D	http://www.glycon.de
Greiner Bio-one GmbH	Frickenhausen, D	http://www.gbo.com/bioscience
Hamilton Robotics	Bonaduz, CH	http://www.hamiltonrobotics.com
Hampton Research	Aliso Vejo, California, USA	http://www.hamptonresearch.com
Hansatech Instruments Ltd	Norfolk, UK	http://www.hansatech-instruments.com
Hartenstein A. Laborbedarf GmbH	Würzburg, D	http://www.laborversand.de
Hellma	Müllheim, D	http://www.hellma-worldwide.com
Helmholtz Institut für Infektionsforschung	Braunschweig, D	http://www.helmholtz-hzi.de
Heraeus	Hanau, D	http://www.wc-heraeus.com
Hitachi High-Technologies	Tokyo, J	http://www.hitachi-hitech.com
HKL Research	Charlottesville, Virginia, USA	http://www.hkl-xray.com
Hoffmann Dental Manufaktur GmbH	Berlin, D	http://www.hoffmann-dental.com
IKA Werke	Staufen, D	http://www.ika.de
Infors AG	Bottmingen-Basel, CH	http://www.infors-ht.com
Invitrogen	Karlsruhe, D	http://www.invitrogen.com
Jena Bioscience GmbH	Jena, D	http://www.jenabioscience.com
Kipp & Zonen	Delft, NL	http://www.kippzonen.com
Kindler O. GmbH	Freiburg, D	http://www.o-kindler.de
Knick	Egelsbach, D	http://www.knick.de
Kodak	Köln, D	http://www.kodak.com
Leica Microsystems	Wetzlar, D	http://www.leica-microsystems.com
MAGV	Rabenau, D	http://magv-gmbh.de
Matrix Science	Boston, Massachusetts, USA	http://www.matrixscience.com
Memmert	Schwalbach, D	http://www.memmert.com
Merck – VWR International	Darmstadt, D	http://de.vwr.com
Messer	Sulzbach, D	http://www.messergroup.com
Mettler Toledo GmbH	Gieβen, D	http://www.mt.com
Microfluidics Corporation	Newton, Massachusetts, USA	http://www.microfluidicscorp.com
Microsoft Corporation	Redmond, Washington, USA	http://www.microsoft.com
Millipore	Schwalbach, D	http://www.millipore.com
Molecular Dimensions Ltd	Soham, UK	http://www.moleculardimensions.com
Mozilla	Mountain View, California, USA	http://www.mozilla.org
Müller W.	Berlin, D	
NanoSeparations	Nieuwkoop, NL	http://www.nanoseparations.com
New England Biolabs	Ipswich, Massachusetts, USA	http://www.neb.com
Nu-Chek Prep, Inc.	Elysian, Minnesota, USA	http://www.nu-chekprep.com
NUNC GmbH	Wiesbaden, D	http://www.nunc.de

Olympus	Hamburg, D	http://www.olympus.de
OXYL	Bobingen, D	
Pall GmbH	Dreieich, D	http://www.pall.com
Perkin Elmer	Waltham, Massachusetts, USA	http://www.perkinelmer.com
Pierce Biotechnology	Rockford, Illinois, USA	http://www.piercenet.com
Promega	Madison, Wisconsin, USA	http://www.promega.com
Proteros	Martinsried, D	http://www.proteros.com
Qiagen	Hilden, D	http://www1.qiagen.com
RealVNC Ltd	Cambridge, UK	http://www.realvnc.com
Riedel De Haen AG	Seelze, D	http://www.riedeldehaen.com
Rigaku	Tokyo, J	http://www.rigaku.com
Roche Diagnostics GmbH	Mannheim, D	http://www.roche.de
Roth, Carl Roth GmbH + Co KG	Karlsruhe, D	http://www.carl-roth.de
Sartorius AG	Göttingen, D	http://www.sartorius.de
Scientific Industries	Bohemia, New York, USA	http://www.scientificindustries.com
SERVA Electrophoresis GmbH	Heidelberg, D	http://www.serva.de
Sharp Electronics GmbH	Hamburg, D	http://www.sharp.de
Sigma-Aldrich Laborchemikalien GmbH	Seelze, D	http://www.sigmaaldrich.com
Sophos	Abingdon, UK	http://www.sophos.com
Ted Pella	Redding, California, USA	http://www.tedpella.com
Thomson Reuters	Philadelphia, Pennsylvania, USA	http://thomsonreuters.com
TOSOH Bioscience	Montgomeryville, Pennsylvania, USA	http://www.tosohbioscience.com
TTP LabTech	Melbourn, UK	http://www.ttplabtech.com
Vivascience	Hannover, D	http://www.vivascience.com
VWR	Darmstadt, D	http://de.vwr.com
Zeiss, Carl Zeiss AG	Oberkochen, D	http://www.zeiss.de

4.1.2. Chemicals

The chemicals used in this work are listed in Table 4.2.

Table 4.2: List of chemicals.

Chemical name	Formula	Abbreviation	Supplier
acetic acid	CH_3COOH	–	Roth
N-(2-acetamido)iminodiacetic acid	$C_6H_{10}N_2O_5$	ADA	Sigma-Aldrich
acetonitrile	CH_3CN	ACN	Roth
acrylamide	$CH_2CHC(O)NH_2$	–	Roth
Al's oil	Paraffin oil:Silicon oil = 1:1	–	Hampton
adenosine diphosphate disodium salt	$C_{10}H_{13}N_5O_{10}P_2Na_2$	–	Sigma-Aldrich
ammonium carbonate	$(NH_4)_2CO_3$	–	Sigma-Aldrich
ammonium chloride	NH_4Cl	–	Merck
ammonium hydrogen carbonate	NH_4HCO_3	AHC	Sigma-Aldrich
ammonium persulfate	$(NH_4)_2S_2O_8$	APS	Genzyme
ammonium phosphate	$(NH_4)_3PO_4$	–	Sigma-Aldrich
ammonium sulfate	$(NH_4)_2SO_4$	–	Sigma-Aldrich
antimycin	$C_{13}H_{14}O_5$	–	Sigma-Aldrich
BCA^{TM} *Protein Assay Reagent* A + B	–	–	Pierce
benzyldimethyldodecylammonium bromide	$C_{21}H_{38}BrN$	BAM	Hampton
1,3-bis(tris(hydroxymethyl)methylamino)propane	$C_8H_{19}NO_5$	BisTris	Sigma-Aldrich
boric acid	H_3BO_3	–	Merck
bromophenol blue	–	–	Roth
cetrimonium (or hexadecyltrimethylammonium) bromide	$C_{19}H_{42}BrN$	CTAB	Hampton
chloroform	$CHCl_3$	–	Roth
3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate	$C_{32}H_{58}N_2O_7S$	CHAPS	Sigma-Aldrich
3-[(3-Cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate	$C_{32}H_{58}N_2O_5S$	CHAPSO	Sigma-Aldrich
Coomassie Brilliant Blue R250 and G250	–	–	Serva
α-cyano-4-hydroxy-cinnamic acid	$C_{10}H_6NO_3$	CHCA	Bruker
N-cyclohexyl-3-aminopropanesulfonic acid	$C_9H_{19}NO_3S$	CAPSO	Gerbu

cyclohexylethanoyl-N-hydroxyethylglucamide	$C_{16}H_{31}NO_7$	C-HEGA-8	Anatrace
cyclohexylpropanoyl-N-hydroxyethylglucamide	$C_{17}H_{33}NO_7$	C-HEGA-9	Anatrace
cyclohexylbutanoyl-N-hydroxyethylglucamide	$C_{18}H_{35}NO_7$	C-HEGA-10	Anatrace
cyclohexylpentanoyl-N-hydroxyethylglucamide	$C_{19}H_{37}NO_7$	C-HEGA-11	Anatrace
3-cyclohexylmethyl-4-O-(a-D-glucopyranosyl)-b-D-glucopyranoside	$C_{19}H_{34}O_{11}$	CYMAL-1	Anatrace
3-cyclohexylethyl-4-O-(a-D-glucopyranosyl)-b-D-glucopyranoside	$C_{20}H_{36}O_{11}$	CYMAL-2	Anatrace
3-cyclohexylpropyl-4-O-(a-D-glucopyranosyl)-b-D-glucopyranoside	$C_{21}H_{38}O_{11}$	CYMAL-3	Anatrace
3-cyclohexylbutyl-4-O-(a-D-glucopyranosyl)-b-D-glucopyranoside	$C_{22}H_{40}O_{11}$	CYMAL-4	Anatrace
3-cyclohexylpentyl-4-O-(a-D-glucopyranosyl)-b-D-glucopyranoside	$C_{23}H_{42}O_{11}$	CYMAL-5	Anatrace
3-cyclohexylhexyl-4-O-(a-D-glucopyranosyl)-b-D-glucopyranoside	$C_{24}H_{44}O_{11}$	CYMAL-6	Anatrace
n-decyl-β-D-maltoside	$C_{22}H_{42}O_{11}$	DM	Glycon
1-S-n-decyl-β-D-thiomaltoside	$C_{22}H_{42}O_{10}S$	DTM	Anatrace
n-decyl-N,N-dimethylamine-N-oxide	$C_{12}H_{27}NO$	DDAO	Sigma-Aldrich
n-decyl-N,N-dimethyl-3-ammonio-1-propanesulfonate	$C_{15}H_{33}NO_3S$	Zwittergent 3-10	Sigma-Aldrich
n-decyl-sucrose	$C_{22}H_{42}O_{11}$	DS	Calbiochem
decylubiquinone	$C_{19}H_{30}O_4$	decyl-UQ	Sigma-Aldrich
deuterium oxide	D_2O	–	Sigma-Aldrich
2,3-dibromo-1,4-naphthoquinone	$C_{10}H_4Br_2O_2$	–	Sigma-Aldrich
2,3-dibromo-5,6-dimethyl-1,4-benzoquinone	$C_8H_6Br_2O_2$	–	Sigma-Aldrich
2,5-dihydroxybenzoic acid and 5-methoxysalicylic acid, 1:9 mixture	$C_7H_6O_4$ and $C_8H_8O_4$	sDHB	Bruker
di-chloro-di-methyl-silane	$(CH_3)_2SiCl_2$	–	Sigma-Aldrich
2,3-dimethyl-1,4-naphthoquinone	$C_{12}H_{10}O_2$	DMN	Prof. Teixeira, Instituto de Tecnologia Química e Biológica, Lisboa, Portugal
dimethylsulfoxide	C_2H_6OS	DMSO	Sigma-Aldrich

n-dodecyl-N,N-dimethylamine-N-oxide	$C_{14}H_{31}NO$	LDAO	Sigma-Aldrich
n-dodecyl-β-D-maltoside	$C_{24}H_{46}O_{11}$	DDM (or LM)	Glycon
N-dodecyl-N,N-(dimethylammonio)butyrate	$C_{18}H_{37}NO_2$	DDMAB	Hampton
n-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate	$C_{17}H_{39}NO_3S$	Zwittergent 3-12	Sigma-Aldrich
n-dodecyl-sucrose	$C_{24}H_{46}O_{11}$	LS	Calbiochem
ethylene diamine tetra-acetic acid, di sodium salt	$C_{10}H_{16}N_2O_8$	EDTA	Gerbu
ethanol (p.a.)	CH_3CH_2OH	–	Roth
flavin adenine dinucleotide, di-sodium salt	$C_{27}H_{31}N_9O_{15}P_2Na_2$	FAD	Sigma-Aldrich
formaldehyde	H_2CO	–	Sigma-Aldrich
N,N-bis[3-(D-gluconamido)propyl]deoxycholamide	$C_{42}H_{75}N_3O_{15}$	DeoxyBigChap	Calbiochem
glutaraldehyde	$CH_2(CH_2CHO)_2$	–	Sigma-Aldrich
glycerol	$CH_2(OH)CH(OH)CH_2(OH)$	–	Gerbu
glycine	NH_2CH_2COOH	–	Gerbu
4-(2-hydroxyethyl)-piperazine-1-ethanesulfonic acid	$C_8H_{18}N_2O_4S$	HEPES	Gerbu
heptane-1,2,3-triol	$C_7H_{16}O_3$	–	Sigma-Aldrich
1-S-n-heptyl-β-D-thioglucopyranoside	$C_{13}H_{26}O_5S$	7TG	Glycon
n-hexadecyl-β-D-maltoside	$C_{28}H_{54}O_{11}$	16M	Glycon
n-hexane	C_6H_{14}	–	Roth
n-hexyl-β-D-glucopyranoside	$C_{12}H_{24}O_6$	6G	Glycon
hydrochloric acid	HCl	–	Roth
isopropanol (p.a.)	$CH_3CH(OH)CH_3$	–	Roth
2-mercapto-ethanol	$CH_2(OH)CH_2(SH)$	β-ME	Sigma-Aldrich
methanol (p.a.)	CH_3OH	–	Roth
2-methyl-2,4-pentanediol	$C_6H_{14}O_2$	MPD	Fluka and Hampton
methyl 6-O-(N-heptylcarbamoyl)-a-D-glucopyranoside	$C_{15}H_{29}NO_7$	HECAMEG	Biomol
monoolein	$C_{21}H_{40}O_4$	–	Nu-Chek Prep Inc.
2-(N-morpholine)-ethane sulphonic acid	$C_6H_{13}NO_4S$	MES	Gerbu
nitrogen (liquid)	N_2	–	Messer
nonaethylene glycol monododecyl ether	$C_{30}H_{62}O_{10}$	$C_{12}E_9$	Fluka
nonanoyl-N-hydroxyethylglucamide	$C_{17}H_{35}NO_7$	HEGA-9	Biomol
n-nonyl-β-D-glucopyranoside	$C_{15}H_{30}O_6$	NG	Biomol
n-nonyl-β-D-maltoside	$C_{21}H_{40}O_{11}$	NM	Glycon
N-nonyl-N-methyl-D-glucamide	$C_{20}H_{39}NO_7$	MEGA-9	Glycon

1-S-n-nonyl-β-D-thioglucopyranoside	$C_{15}H_{30}O_5S$	NTG	Glycon
1-S-n-nonyl-β-D-thiomaltoside	$C_{21}H_{40}O_{10}S$	NTM	Hampton
O,O'-bis(2-aminopropyl)-polyethylene glycol 500	–	Jeffamine ED600	Sigma-Aldrich
octaethylene glycol monododecyl ether	$C_{29}H_{60}O_{10}$	$C_{12}E_8$	Fluka
octanoyl-N-hydroxyethylglucamide	$C_{16}H_{33}NO_7$	HEGA-8	Biomol
n-octyl-β-D-glucopyranoside	$C_{14}H_{28}O_6$	OG	Glycon
n-octyl-N,N-dimethyl-3-ammonio-1-propanesulfonate	$C_{13}H_{27}NO_3S$	Zwittergent 3-8	Sigma-Aldrich
N-octyl-N-methyl-D-glucamide	$C_{19}H_{37}NO_7$	MEGA-8	OXYL
n-octyl-sucrose	$C_{20}H_{38}O_{11}$	OS	Hampton
1-S-n-octyl-β-D-thioglucopyranoside	$C_{14}H_{28}O_5S$	OTG	Glycon
1-S-n-octyl-β-D-thiomaltoside	$C_{20}H_{38}O_{10}S$	OTM	Hampton
paraffin oil	–	–	Hampton
perfluoropolyether	$C_{50}F_{102}O_{16}$	PFO-X175/08	Hampton
L-pipecolinic acid	$C_6H_{11}NO_2$	–	Sigma-Aldrich
N-octyl-phoscholine	$C_{13}H_{30}NO_3P$	FOS-8	Anatrace
N-nonyl-phoscholine	$C_{14}H_{32}NO_3P$	FOS-9	Anatrace
N-decyl-phoscholine	$C_{15}H_{34}NO_3P$	FOS-10	Anatrace
N-dodecyl-phoscholine	$C_{17}H_{38}NO_3P$	FOS-12	Anatrace
polyethylene glycol 200 – 20000	$CH_2OH(OCH_2CH_2)_nOH$	PEG 200 – 20000	Fluka
polyethylene glycol monomethylether	$CH_3(OCH_2CH_2)_nOH$	PEG MME 550, 2000, 5000	Fluka
polyethylene glycol 400 dodecyl ether	$C_{12}H_{23}O_2.(C_2H_4O)_n.H$	Thesit	Fluka
polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether	$C_{14}H_{22}O(C_2H_4O)_n$	Triton X-100	Gerbu
potassium chloride	KCl	–	Merck
potassium di-hydrogen phosphate	KH_2PO_4	–	Roth
di-potassium hydrogen phosphate	K_2HPO_4	–	Roth
potassium hydroxide	KOH	–	Roth
potassium sulfate	K_2SO_4	–	Sigma-Aldrich
propanol (p.a.)	$CH_3CH_2CH_2OH$	–	Roth
quinone derivatives	–	UQ_1, UQ_4, UQ_9	Sigma-Aldrich
quinone derivatives	–	–	Prof. Bolognesi, Bologna University, Italy
silicon oil	$C_8H_{24}O_2Si_3$	–	Hampton

silicon paste	–	Baysilone-Paste mittelviskös	GE Bayer Silicones
silver nitrate	$AgNO_3$	–	Sigma-Aldrich
sodium acetate	CH_3COONa	–	Merck and Roth
sodium azide	NaN_3	–	Sigma-Aldrich
sodium carbonate	Na_2CO_3	–	Sigma-Aldrich
sodium chloride	$NaCl$	–	VWR
tri-sodium citrate	$Na_3C_6H_5O_7$	–	Sigma-Aldrich
sodium deoxycholate	$C_{24}H_{39}O_4Na$	DOC	Sigma-Aldrich
sodium dithionite	$Na_2S_2O_4$	–	Sigma-Aldrich
sodium dodecyl sulfate	$C_{12}H_{25}SO_4Na$	SDS	Gerbu
sodium hydroxide	$NaOH$	–	Gerbu
sodium potassium tartrate	$C_4H_{12}KNaO_{10}$	–	Fluka
sodium sulfate	Na_2SO_4	–	Sigma-Aldrich
sodium sulfide	Na_2S	–	Sigma-Aldrich
sodium thiosulfate	$Na_2S_2O_3$	–	Sigma-Aldrich and Roth
stigmatellin	$C_{30}H_{42}O_7$	–	Sigma-Aldrich
sulfuric acid	H_2SO_4	–	Roth
n-tetradecyl-β-D-maltoside	$C_{26}H_{50}O_{11}$	14M	Glycon
n-tetradecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate	$C_{19}H_{45}NO_3S$	Zwittergent 3-14	Sigma-Aldrich
N,N,N,N-tetramethyl-*p*-ethylendiamide	$(CH_3)_2NCH_2CH_2N(CH_3)_2$	TEMED	Roth
n-tridecyl-β-D-maltoside	$C_{25}H_{48}O_{11}$	13M	Glycon
N-(Tri(hydroxymethyl)methyl)glycine	$C_6H_{13}NO_5$	Tricine	Gerbu
trifluoroacetic acid	CF_3COOH	TFA	Merck
tris-hydroxymethyl-9-aminomethane	$C_4H_{11}NO_3$	Tris	Roth
n-undecyl-β-D-maltoside	$C_{23}H_{44}O_{11}$	UM	Glycon
uranyl acetate	$UO_2(CH_3COO)_2$	–	Ted Pella
urea	$(NH_2)_2CO$	–	Merck

4.1.3. Sparse-matrix crystallization screens

The sparse-matrix screens used in this work are listed in Table 4.3.

Table 4.3: List of sparse-matrix crystallization screens.

Name of the screen	Number of conditions	Supplier
Basic Kit for Membrane Proteins	48	Sigma-Aldrich
JBScreen Classic 1 – 10	24 each	Jena Bioscience
JBScreen Membrane 1 – 3	24 each	Jena Bioscience
JBScreen PACT++	96	Jena Bioscience
The MPDs	96	Qiagen
The MbClass	96	Qiagen
The MbClass II	96	Qiagen
The pHClear	96	Qiagen
The pHClear II	96	Qiagen
The PEGs	96	Qiagen
The Cations	96	Qiagen
The Anions	96	Qiagen
Structure Screen 1 (MD1-01)	48	Molecular Dimensions Ltd
Structure Screen 2 (MD1-02)	48	Molecular Dimensions Ltd
MemStart (MD1-21)	48	Molecular Dimensions Ltd
Crystal Screen (HR2-110)	48	Hampton Research
Crystal Screen 2 (HR2-112)	48	Hampton Research
MembFac screen (HR2-114)	48	Hampton Research
PEG/Ion Screen	48	Hampton Research
Additive Screen (HR2-428)	96	Hampton Research
Detergent Additive Screens 1 – 3 (HR2-410, HR2-411, HR2-412)	24 each	Hampton Research
Heavy atom screen Pt (HR2-442)	12	Hampton Research
Heavy atom screen Au (HR2-444)	6	Hampton Research
Heavy atom screen Hg (HR2-446)	19	Hampton Research
Heavy atom screen M1 (HR2-448)	20	Hampton Research
Heavy atom screen M2 (HR2-450)	18	Hampton Research

4.1.4. Equipment and devices

Equipment and devices used in this work are listed in Table 4.4.

Table 4.4: List of equipment and devices.

Device	Type	Supplier
Gel imaging station	gel documentation system (UV) and PC (maxdata)	Bio-Rad
Ampholytes	SERVALYT^{TM} 3-10	Serva
Anaerobic tent		Coy Laboratories
Analytical balance	R180D-*D1	Sartorius AG
Balance	PM 4600 Delta Range	Mettler Toledo
Centrifuge	Sigma 4K10	Sigma-Aldrich
Centrifuge	5415 R	Eppendorf
Centrifuge	Optima XLA	Beckman Coulter
Chromatography system	Äkta Purifier 10	GE Healthcare
Chromatography system	Äkta Purifier 100	GE Healthcare
Chromatography system	SMART	GE Healthcare
Chromatography system	easy nano LC	Bruker / Proxeon
Chromatography system	1200 HPLC	Agilent
Concentrators	Amicon Ultra-15, Ultra-4, Ultra-500	Millipore
Concentrators	Centriprep	Millipore
Concentrators	Vivaspin 500 µL	Vivascience
Crystallography bridges	CrystalBridge^{TM}	Greiner Bio-one
Crystallography capillaries	Mark-Röhrchen für röntgenographische Aufnahmen, 0.5 – 1.0 mm	Müller W.
Crystallography glasses	Microscopic glasses	Kindler
Crystallography humidity chamber	Free Mounting System^{TM}	Proteros
Crystallography liquid nitrogen dewar	Agil-2	Air Liquide
Crystallography liquid nitrogen dewar	Spearlab Cryogenic Foam Dewar 2L (HR4-675)	Hampton Research
Crystallography loops	CrystalCap^{TM} (HR4-913)	Hampton Research
Crystallography loops	CrystalCap Magnetic^{TM} (HR4-733)	Hampton Research
Crystallography loops	MicroTube^{TM} 18 mm (HR4-921)	Hampton Research
Crystallography loops	MicroTube^{TM} 21 mm (HR4-923)	Hampton Research
Crystallography loops	CryoLoop^{TM} 0.05 – 1.0 mm	Hampton Research
Crystallography loops	Mounted CryoLoop^{TM} 0.025 – 1.0 mm	Hampton Research
Crystallography loops	CrystalCap HT^{TM} 0.025 – 1.0 mm	Hampton Research
Crystallography plate	AxyGem^{TM}	Axygen Bioscience
Crystallography plate	Crystal Quick^{TM}	Greiner Bio-one

Crystallography plate	Crystal Quick^{TM} hydrophob	Greiner Bio-one
Crystallography plate	IMP@CT^{TM}	Greiner Bio-one
Crystallography plate	Terasaki	Greiner Bio-one
Crystallography plate	View Seal	Greiner Bio-one
Crystallography plate	XRL Plates	Molecular Dimensions Ltd
Crystallography plate	Linbro	Hampton Research
Crystallography plate	Assay Deep Well Block, 2 mL	Corning
Crystallography plate	Square Well Storage Mat	Corning
Crystallography robot	MicroSys SQ Series	Cartesian Technologies
Crystallography robot	Mosquito	TTP LabTech
Crystallography robot	HoneyBee 963	DigiLab
Crystallography storage and imaging system	CRYSTAL FARM^{TM} 400	Bruker
Crystallography tools	Vial Clamp^{TM} curved (HR4-671)	Hampton Research
Crystallography tools	Vial Clamp^{TM} straight (HR4-670)	Hampton Research
Crystallography tools	Crystal Wand Magnetic^{TM} straight (HR4-729)	Hampton Research
Crystallography tools	Crystal Wand^{TM} long with tab (HR4-602)	Hampton Research
Crystallography tools	CryoTong^{TM} 18 mm (HR4-637)	Hampton Research
Crystallography tools	CryoTong^{TM} 21 mm (HR4-639)	Hampton Research
Crystallography tools	CryoCane 5 vial holder (HR4-709)	Hampton Research
Crystallography tools	CryoCane Color Coder (HR4-715)	Hampton Research
Crystallography tools	Cryo Sleeve (HR4-708)	Hampton Research
Crystallography tools	Micro-Tools^{TM} Set (HR4-811)	Hampton Research
Crystallography tools	ClearSeal Film^{TM} Applicator (HR4-525)	Hampton Research
Crystallography tools	VDX^{TM} Plate with Sealant (HR3-170)	Hampton Research
Crystallography tools	Manco Crystal Clear Sealing Tape	Jena Biosciences
Crystallography tools	Magnetic CryoVials with CryoCaps and DataMatrix (MD7-406)	Molecular Dimensions Ltd
Crystallography tools	EMBL/ESRF Sample Changer Kit	Molecular Dimensions Ltd
Crystallography tools	Hoffmann's Glaswachs	Hoffmann
Crystallography tools	Leather gloves	VWR
Developer	D19	Kodak
Electron microscope	Tecnai 12	FEI-Philips
Electron microscope	CM 120	FEI-Philips
Electrophoresis chamber	XCell Sure Lock Mini-Cell	Invitrogen
Electrophoresis chamber		home made

119

Electrophoresis power supplier	PowerPac Basic	Bio-Rad
Electrophoresis power supplier	PowerPac HC	Bio-Rad
Electrophoresis plates	Glass plates for Hoefer SE-250	GE Healthcare
ELISA reader	PowerWave X	Bio-Tek Instruments
Flat bed scanner	SCAI	Zeiss
Gel drier	Mididry D62	Biometra
Gels for isoelectric focusing	SERVALYT PRECOTES^{TM}	Serva
Heating block	TB1 Thermoblock	Biometra
Heating block – Mixer	Thermomixer Compact	Eppendorf
Heating block – Mixer	Thermomixer Comfort	Eppendorf
Incubator	BK-600	Heraeus
Films	SO-163	Kodak
Filter for producing deionised water	MilliQ Plus	Millipore
Filters	Nylon membranes	Pall
Fluorescence spectrophotometer	F-4500	Hitachi High-Technologies
Liquid Handling System	StarLET	Hamilton Robotics
Luminescence Spectrophotometer	LS 50 B	Perkin-Elmer
Magnetic stirrer	RET IKAMAG	IKA Werke
Mass spectrometer	micrOTOF-QII	Bruker
Mass spectrometer	maXis	Bruker
Mass spectrometer	MALDI-TOF Voyager STR	Applied Biosystems
Mass spectrometer	MALDI-TOF OmniFlex®	Bruker
Membrane vacuum pumps	MP40 / System III	Biometra
Microfluidizer	110LA	Microfluidics Corporation
Microspectrophotometer	Universal-Mikroskop-Spektral-Photometer, UMSP 30 / 50	Carl Zeiss
Microwave	R-222	Sharp
Multipipette	edp3	Mettler Toledo
Oxygen electrode	Oxygraph	Hansatech
Optical microscope	SZ40	Olympus
Optical microscope	M165C	Leica Microsystems
Peristaltic pump	Miniplus 2	Gilson
pH-meter	765 Calimatic	Knick
Photometer	Ultrospec 2100 pro	GE Healthcare
Photometer	Lambda 35	Perkin-Elmer
Photometer	Lambda 40	Perkin-Elmer
Photometer	Agilent 8453	Agilent
Pipettes	P2 – P5000	Gilson
Precision syringes		Hamilton
X-ray generator	R-Axis IV^{++}	Rigaku
Rotors		Beckman Coulter
SDS-Gels	NuPAGE Novex® 4-12% Bis-Tris	Invitrogen

SDS-Gels	NativePAGE Novex® 4-16% Bis-Tris	Invitrogen
Sonification bath	RK 255 S	Bandelin Sonorex
Speedvacuum	Concentrator 5301	Eppendorf
Thermostatic room 4 °C and 18 °C		
TXRF spectrometer	EXTRA IIA	ATOMIKA Instruments
Ultracentrifuge	Optima MAX UC	Beckman Coulter
UV-Monitor	Single Path Monitor UV-1	GE Healthcare
UV-Monitor Writer	BD 41	Kipp & Zonen
Vortexer	Vortex-Genie 2	Scientific Industries

4.1.5. Databases and servers

The databases and servers used in this work are listed in Table 4.5.

Table 4.5: List of databases and servers.

Database	Website
EBI	http://www.ebi.ac.uk
Electron Microscopy Data Bank	http://www.ebi.ac.uk/pdbe/emdb
Expasy	http://www.expasy.org
Google	http://www.google.com
NCBI	http://www.ncbi.nlm.nih.gov
OPM	http://opm.phar.umich.edu
PDB	http://www.rcsb.org
Membrane Protein Data Bank	http://www.mpdb.ul.ie
Membrane Protein of Known 3D Structure	http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html
Unix and Linux servers	in house computer facility
Wikipedia	http://en.wikipedia.org

4.1.6. Software

The software used in this work is listed in Table 4.6.

Table 4.6: List of software.

Software	Website or device	Reference or provider
3D-PSSM	http://www.sbg.bio.ic.ac.uk	(Kelley *et al.*, 2000)
Amphipaseek	http://npsa-pbil.ibcp.fr	(Sapay *et al.*, 2006)
Acrobat 7.0	http://www.adobe.com	Adobe
ARP/wARP	http://www.embl-hamburg.de/ARP	(Lamzin *et al.*, 2001)
BLAST	http://blast.ncbi.nih.gov/Blast.cgi	(Altschul *et al.*, 1997)
CCP4	http://www.ccp4.ac.uk	(Collaborative computational project number 4, 1994)
ChemSketch 10.0	http://www.acdlabs.com	ACDLabs
ClustalX2 and ClustalW2	http://www.ebi.ac.uk/Tools/clustalW2	(Thompson *et al.*, 1994; Thompson *et al.*, 1997)
CNS	http://cns.csb.yale.edu/v1.1	(Brunger *et al.*, 1998)
Coot	http://www.biop.ox.ac.uk/coot	(Emsley and Cowtan, 2004)
DALI	http://www2.ebi.ac.uk/dali	(Holm *et al.*, 2008)
DataAnalysis 4.0	ESI-q-TOF MS	Bruker
DM	http://www.ccp4.ac.uk/html/dm.html	(Cowtan, 1994)
D*Trek	http://www.rigaku.com/software/dtrek.html	Rigaku, (Pflugrath, 1999)
EndNote X3	http://www.endnote.com	Thomson Reuters
Firefox	http://www.mozilla.com/firefox	Mozilla
Flex Control	http://www.bdal.com/products/maldi-tof-toftof-ms.html	Bruker
Geneious	http://www.geneious.com	Drummond *et al.*
Ghostscript 8.53	http://www.ghostscript.com	
HKL Package	http://www.hkl-xray.com	HKL Research
HMMTOP	http://www.enzim.hu/hmmtop	(Tusnady and Simon, 2001)
IMP	http://xray.bmc.uu.se/usf/imp_man.html	(Kleywegt and Jones, 1994)
Interner Explorer	http://www.microsoft.com/windows/Internet-explorer/	Microsoft Corporation
Jalview	http://www.jalview.org	(Waterhouse *et al.*, 2009)
Mascot	http://www.matrixscience.com	Matrix Science
Matthew's coefficient Calculator	http://www.rigaku.co.jp/xrl/group1/tips/prg/vm.html	Rigaku

O	http://alpha2.bmc.uu.se/~alwyn	(Jones and Kjeldgaard, 1997)
Office 2003	http://office.microsoft.com	Microsoft Corporation
Photoshop 6.0	http://www.adobe.com/products/photoshop	Adobe
PyMOL	http://www.pymol.org	(DeLano, 2002)
ProtParam	http://www.expasy.ch/tools/protparam.html	
RealVNC 4	http://www.realvnc.com	RealVNC Ltd
Refmac 5	http://www.ccp4.ac.uk/dist/html/refmac5.html	(Murshudov *et al.*, 1997)
SEDFIT	http://www.analyticalultracentrifugation.com	(Schuck, 2000)
SEDNTERP	http://www.jphilo.mailway.com/download.htm	courtesy of Dr. J. Philo
SEDPHAT	http://www.analyticalultracentrifugation.com	
SHARP	http://www.globalphasing.com/sharp	(Fortelle and Bricogne, 1997)
SHELX	http://shelx.uni-ac.gwdg.de/SHELX	(Sheldrick *et al.*, 2001)
SignalP	http://www.cbs.dtu.dk/services/SignalP	(Emanuelsson *et al.*, 2007)
Sophos Anti-Virus	http://antivir.gwdg.de	Sophos
SPIDER	http://www.wadsworth.org/spider_doc/spider/docs/spider.html	(Frank *et al.*, 1996)
T-COFFEE	http://www.tcoffee.org	(Notredame *et al.*, 2000)
TMHMM 2.0	http://www.cbs.dtu.dk/services/TMHMM-2.0	(Sonnhammer *et al.*, 1998)
Unicorn v4.10 and v5.01	http://www5.gelifesciences.com/aptrix/upp01077.nsf/Content/aktadesign_platform~system_software	GE Healthcare
WEB	http://www.wadsworth.org/spider_doc/spider/docs/spider.html	(Frank *et al.*, 1996)
XDS	http://xds.mpimf-heidelberg.mpg.de	(Kabsch, 1993)
Xmipp	http://xmipp.cnb.csic.es	(Marabini *et al.*, 1996)

4.1.7.   Columns and chromatographic matrices

The columns and column materials used in this work are listed in Table 4.7.

Table 4.7: List of columns and chromatographic matrices.

Material	Separation type	Supplier
CHT Ceramic Hydroxyapatite	affinity, ionic interaction, hydrophobic interaction	Bio-Rad
Cibacron Blue 3GA	affinity	Sigma-Aldrich
Reactive dyes (blue 4, blue 72, brown 10, green 5, green 19, red 120, yellow 3, yellow 86)	affinity	Sigma-Aldrich
Mini Q 4.6/50 PE	ion exchange	GE Healthcare
MiniQ PC 3.2/3	ion exchange	GE Healthcare
MonoQ 10/100 GL	ion exchange	GE Healthcare
HiTrap^{TM} Q-Sepharose^{TM} *Fast Flow*	ion exchange	GE Healthcare
HiTrap^{TM} DEAE-Sepharose^{TM} *Fast Flow*	ion exchange	GE Healthcare
HiTrap^{TM} CM-Sepharose^{TM} *Fast Flow*	ion exchange	GE Healthcare
HiTrap^{TM} SP-Sepharose^{TM} *Fast Flow*	ion exchange	GE Healthcare
HiTrap^{TM} Phenyl^{TM} *Fast Flow*	hydrophobic exchange	GE Healthcare
HiTrap^{TM} Butyl^{TM} *Fast Flow*	hydrophobic exchange	GE Healthcare
HiTrap^{TM} Octyl^{TM} *Fast Flow*	hydrophobic exchange	GE Healthcare
Disposable PD-10 desalting column	size-exclusion	GE Healthcare
Superdex 200 10/300 GL	size-exclusion	GE Healthcare
Superdex 200 PC 3.2/30	size-exclusion	GE Healthcare
TSK-GEL G4000SW	size-exclusion	TOSOH Bioscience
NS-MP-10	reverse phase	NanoSeparations
NS-AC-10	reverse phase	NanoSeparations
Sorbax SB-C8 high resolution cartridge	reverse phase	Agilent

4.1.8.   *Enzymes, other proteins, kits and standards*

Enzymes, other proteins, kits and standards used in this work are listed in Table 4.8.

Table 4.8: List of enzymes, other proteins, kits and standards.

Standard	Application	Supplier
BenchMark^{TM} Fluorescence Standard	Gel electrophoresis	Invitrogen
Bovine Serum Albumin	Protein concentration	Pierce
Catalase, bovine liver	Activity assay	Sigma-Aldrich
Gel filtration calibration kit (low and high molecular weight standards)	Size-exclusion chromatography	GE Healthcare
Glucose oxidase, *Aspergillus niger*	Activity assay	Sigma-Aldrich
Lysozyme	Crystallization	Sigma-Aldrich
NativeMark^{TM} Protein Standard	Gel electrophoresis	Invitrogen
Peptide calibration standard	Mass spectrometry	Bruker
Prestained Protein Marker, Broad Range (6 – 175 kDa)	Gel electrophoresis	New England Biolabs
Trypsin digestion kit	ProteoExtract® All-in-one	Calbiochem
SeeBlue® Plus2 Prestained Standard	Gel electrophoresis	Invitrogen
Sequence grade modified trypsin	Mass spectrometry	Promega
Silver staining kit	SilverQuest^{TM}	Invitrogen
Tobacco mosaic virus	Electron microscopy	
Trypsin, sequence grade, modified	Mass spectrometry	Promega

4.2. Methods

4.2.1. *Protein biochemistry*

4.2.1.1. DENATURING GEL ELECTROPHORESIS (SDS-PAGE)

Denaturing gel electrophoresis was performed according to protocols modified from Laemmli (Laemmli, 1970). Protein samples mixed 1:5 with sample buffer (all solutions used are listed in Table 4.9) were typically incubated for 3 min at 90 °C. The samples were run in a purchased gel chamber (XCell SureLock Mini-Cell, Invitrogen) on 4 – 12 % BisTris gels or 10 % BisTris gels (NuPAGE™ Novex®, Invitrogen) using MES/SDS running buffer at 200 V constant for 45 min. Alternatively, self cast gels (5 % stacking gel and 15 % separating gel) were run in a homemade gel chamber using Tris/glycine running buffer and a constant voltage of 200 V for 1 h. After electrophoresis, the gels were fixed with 7 % (v/v) acetic acid solution for 10 min if required or directly stained with colloidal Coomassie blue staining solution (detection range 5 – 10 µg protein). After 30 min at room temperature, the gels were destained using destaining solution. For higher detection sensitivity (0.05 – 0.1 µg protein), silver staining was performed according to a protocol optimized for subsequent fingerprint mass spectrometry *in gel* analysis (Shevchenko *et al.*, 1996). In addition to the solutions listed in Table 4.9, the SilverQuest™ Silver Staining Kit (Invitrogen) was also used.

In gel detection of FAD was performed before staining by UV excitation (Unden *et al.*, 1980) in a gel-imaging station (Bio-Rad).

Table 4.9: List of solutions for SDS-PAGE.

Solution	Component	Amount
Stock solutions		
▪ **Tris-HCl pH 8.8, 4x**	Tris	36.34 g
	SDS	0.4 g
	H_2O	to 100 mL
▪ **Tris-HCl pH 6.8, 4x**	Tris	6.06 g
	SDS	0.4 g
	H_2O	to 100 mL
▪ **APS 10 %**	APS	1 g
	H_2O	to 10 mL

Sample buffer 5x

	glycerol	2.5 mL
	βME	1.25 mL
	SDS	0.75 g
	Tris-HCl pH 8.0, 1M	2.5 mL
	bromophenol blue	2.5 mL
	H_2O	to 10 mL

MES/SDS running buffer 20x

	MES	97.6 g
	Tris	60.6 g
	SDS	10 g
	EDTA	3 g
	H_2O	to 500 mL

Tris/glycine running buffer 10x

	Tris	60.7 g
	glycine	142.8 g
	SDS	10 g
	H_2O	to 1 L

Molecular weight markers

SeeBlue® Plus2 Prestained Standard
Prestained Protein Marker, Broad Range (6 – 175 kDa)
BenchMark^{TM} Fluorescence Standard

Stacking gel 5 % (for 5 gels)

	Acrylamide/bis (30 % T, 2.67 % C)	2.5 mL
	Tris-HCl pH 6.8, 4x	3.75 mL
	H_2O	8.75 mL
	TEMED	5 μL
	APS 10 %	36 μL

Separating gel 15 % (for 5 gels)

	Acrylamide/bis (30 % T, 2.67 % C)	12.375 mL
	Tris-HCl pH 8.8, 4x	6.25 mL
	H_2O	6.375 mL
	TEMED	7.75 μL
	APS 10 %	157.5 μL

Coomassie-blue staining solutions

Fixative solution	acetic acid	7 mL
	H$_2$O	to 100 mL
Staining solution	Coomassie Brilliant Blue R250	1 g
	ethanol	150 mL
	acetic acid	50 mL
	H$_2$O	to 500 mL
Destaining solution	ethanol	150 mL
	acetic acid	50 mL
	H$_2$O	to 500 mL

Silver staining solutions

Fixative solution	methanol	450 mL
	acetic acid	100 mL
	H$_2$O	to 1 L
Sensitizing solution	Na$_2$S$_2$O$_3$	0.02 g
	H$_2$O	to 100 mL
Staining solution	AgNO$_3$	0.1 g
	H$_2$O	to 100 mL
Developing solution	formaline (37 % formaldehyde in water)	0.04 g
	Na$_2$CO$_3$	2 g
	H$_2$O	to 100 mL
Quenching solution	acetic acid	1 mL
	H$_2$O	to 100 mL

4.2.1.2.   NON-DENATURING GEL ELECTROPHORESIS (NATIVE PAGE)

Non-denaturing electrophoresis was performed using self-cast Tris/glycine gels (3 % stacking gel and 8 % separating gel) in a homemade gel chamber using Tris/glycine native running buffer (all solutions are listed in the following Table 4.10). The gels and the running buffer were supplemented with no detergent (CN-PAGE), with DDM only (DDM-CN-PAGE or low resolution CN-PAGE) or with DDM and DOC (DOC-CN-PAGE or high resolution CN-PAGE) (Wittig *et al.*, 2007). Blue Native PAGE (BN-PAGE) was performed by adding Coomassie blue to the running buffer only on the cathode side (Schägger and von Jagow, 1991; Schägger *et al.*, 1994). Alternatively, the samples were run in commercial gel chambers (XCell SureLock Mini-Cell, Invitrogen) on 4 – 16 % BisTris gels (NativePAGE^{TM} Novex®, Invitrogen). In this case, the detergents were added, if appropriate, to the running buffer only. The runs were all performed at 125 V constant for the first 30 min, then the voltage was increased to 150 V for 30 min and finally to 175 V for the last 30 min.

Table 4.10: List of solutions for native PAGE.

Solution	Component	Amount
Stock solutions		
▪ **Tris-HCl pH 8.8, 4x**	Tris	36.34 g
	HCl	to pH 8.8
	H_2O	to 100 mL
▪ **Tris-HCl pH 6.8, 4x**	Tris	6.06 g
	HCl	to pH 6.8
	H_2O	to 100 mL
▪ **APS 10 %**	APS	1 g
	H_2O	to 10 mL
▪ **Tris-HCl pH 8.6, 1 M**	Tris	12.11 g
	HCl	to pH 8.6
	H_2O	to 100 mL

Tris/glycine running buffer 10x		
	Tris	29 g
	glycine	144g
	(DDM 20 %	2.5 mL)
	(DOC 10 %	5 mL)
	H_2O	to 1 L

BisTris running buffer 20x		
	BisTris	209.2 g
	Tricine	179.2 g
	H_2O	to 1 L

Sample buffer, Tris/glycine system, 2x		
	TrisHCl, pH 8.6, 1 M	1 mL
	glycerol	1 g
	bromophenol blue	crystals
	H_2O	to 10 mL

Sample buffer, BisTris system, 4x		
	BisTris	0.418 g
	HCl 6 N	0.107 mL
	glycerol	4 g
	NaCl	0.117 g
	H_2O	to 10 mL

Cathode buffer additive, for BN-PAGE, 20x

	Coomassie blue R250	1 g
	H$_2$O	to 250 mL

Stacking gel 3 % (for 5 gels)

	Acrylamide/bis (30 % T, 2.67 % C)	1.5 mL
	Tris-HCl pH 6.8, 4x	3.75 mL
	H$_2$O	9.7 mL
	TEMED	5 µL
	APS 10 %	36 µL
	(DDM 20 %	37.5 µL)
	(DOC 10 %	75 µL)

Separating gel 8 % (for 5 gels)

	Acrylamide/bis (30 % T, 2.67 % C)	6.25 mL
	Tris-HCl pH 8.8, 4x	6.25 mL
	H$_2$O	12 mL
	TEMED	7.75 µL
	APS 10 %	157.5 µL
	(DDM 20 %	62.5 µL)
	(DOC 10 %	125 µL)

Molecular weight markers

	NativeMark^{TM} Protein Standard	
	Self-made marker (BSA, 20 mg / mL)	1 µL

4.2.1.3. ISOELECTRIC FOCUSING (IEF)

Isoelectric focusing was performed using SERVALYT^{TM} PRECOTES^{TM} gels (150 µm thick) and a horizontal flat bed electrophoresis chamber (Servalyt). A broad range of ampholytes (Servalyt) was chosen to cover the pH range from 3 to 7. The gels were incubated for 30 min at room temperature with the ampholyte solution containing 0.05 % DDM before the run. The run was performed according to the manufacturer's guidelines.

4.2.1.4. MEMBRANE PREPARATION AND SOLUBILIZATION

All experiments were performed using native cells of *Aquifex aeolicus* VF5 purchased from the Archaeenzentrum (Regensburg, Germany). The cells were grown at 95 °C in SME medium supplied with thiosulfate (Huber and Eder, 2006) under an atmosphere

composed of 97 – 99 % of the mixture $N_2 : H_2 = 80 : 20$ and by 3 – 1 % of O_2[43]. Cells were flash-cooled and stored at -20 °C. 40-gram batches of thawed cells were resuspended in 200 mL buffer containing 50 mM Tris-HCl pH 7.4 and disrupted using a microfluidizer (110LA, Microfluidics Corporation) at 1,000 MPa. Cell debris was removed by centrifugation at 10,000x*g* for 15 min and membranes were sedimented from the supernatant by centrifugation at 100,000x*g* for 1 h. The membranes were washed and re-suspended in the same buffer at an approximate protein concentration of 12 mg/mL and stored in aliquots at -80 °C (Peng *et al.*, 2003). Membranes were solubilized for 1 h at 50 ºC after dilution with an equal volume of 6 % DDM in water if not otherwise specified. The insoluble fraction was removed by ultracentrifugation at 40,000 rpm at 4 °C for 1 h (Peng *et al.*, 2003).

4.2.1.5. ION-EXCHANGE CHROMATOGRAPHY

For preparative protein purification, the solubilized membrane proteins were loaded onto a MonoQ 10/100 GL column (GE Healthcare), preequilibrated with buffer A (20 mM Tris-HCl pH 7.4) containing NaN_3 0.05 %, NaCl 50 mM and DDM 0.05 %. The run was performed on an Äkta purifier 10. The proteins were eluted with a segmented gradient in the same buffer to 1 M NaCl. Fractions were analysed by SDS-PAGE and combined according to the protein content. The samples were concentrated in Amicon Ultra-15 concentrators (Millipore) with a membrane molecular weight cut-off of 10 kDa.

For analytical purposes, MiniQ or HiTrap-Sepharose anion exchange columns were used, either on an Äkta purifier 10, on an Äkta purifier 100 or on a SMART system according to the manufacturer's guidelines. The columns were preequilibrated with the running buffer at the appropriate pH and eluted with linear gradient of NaCl up to 1 M.

For detergent exchange from DDM to Zwittergent 3-10 and then back from Zwittergent 3-10 to DDM, a MonoQ 10/100 GL column was used. The protein in buffer A with 0.05 % DDM was desalted using PD-10 columns (if necessary) and loaded onto the MonoQ column preequilibrated with 5 column volumes in buffer A with 0.05 % DDM. The buffer was exchanged to buffer A with Zwittergent 3-10 (1.5x CMC) for at least 5 column volumes. The protein was then eluted with a linear gradient in the same buffer to 0.6 M NaCl in 3 column volumes. The fractions containing the protein in Zwittergent 3-10 were combined and desalted on the PD-10

---

[43] 0.01 % O_2 is necessary and sufficient for cell growth and the maximal growth rate is observed at 1 % O_2. Above 5 % O_2 the gas mixture becomes explosive (Dr. H. Huber, Archaeenzentrum Regensburg, Germany, personal communication).

columns in the Zwittergent 3-10 buffer and then reloaded onto the MonoQ column preequilibrated with buffer A and Zwittergent 3-10. The detergent was exchanged back to 0.05 % DDM in 10 column volumes to ensure the most extensive removal of Zwittergent 3-10, which has a strong effect on the crystallization. The protein was eluted in buffer A with 0.05 % DDM and 1 M NaCl with a sharp linear gradient in 3 column volumes. Finally the protein was desalted on PD-10 columns.

4.2.1.6. SIZE-EXCLUSION CHROMATOGRAPHY (SEC)

For preparative purposes, a Superdex 200 10/300 GL column (GE Healthcare) or the TSK-GEL G4000SW column (TOSOH Bioscience) were used either on an Äkta purifier 10 or Äkta purifier 100. For analytical purposes, a Superdex 200 PC 3.2/30 was used on a SMART system. After equilibration for 3 column volumes in the running buffer, the runs were performed according to the manufacturer's guidelines.

4.2.1.7. OTHER CHROMATOGRAPHY

For analytical purposes, other chromatographic resins (dye affinity resins, hydrophobic interaction resins, hydroxyapatite) were used according to the manufacturer's guidelines.

4.2.1.8. DETERMINATION OF THE PROTEIN CONCENTRATION

Protein concentration estimation was done by a colorimetric assay using the biuret method. The reagent was prepared as a 1:50 (v/v) mixture of BCA Assay reagents A and B (Pierce) respectively. BSA samples at concentration ranges from 0.01 to 0.1 mg/mL or from 0.1 to 1 mg/mL were used as standards. 10 μL of standard or of a bi-distilled water solution or of a convenient amount of protein sample were mixed with 200 μL reagent in a 96-well reaction plate and incubated for 30 minutes at 37 °C or at 60 °C, according to the standard used (the higher temperature was used to enhance the reaction of the less concentrated standards). The plate was then immediately analysed in an ELISA reader (Power Wave X, BioTek). The concentration of protein was calculated as an average of two measurements.

4.2.2. Protein characterization

4.2.2.1. BIOINFORMATICS

Analysis of the SQR sequence *in silico* (theoretical pI, calculated molecular weight, amino acid and elemental composition) was performed with the program ProtParam (Gasteiger *et al.*, 2003). Prediction of transmembrane helices was performed with the program TMHMM (Sonnhammer *et al.*, 1998).

Tertiary structure comparison was performed with the program 3D-PSSM (Kelley *et al.*, 2000). Comparison of the structural coordinates was performed using the server DALI [www.ebi.ac.uk/dali, (Holm *et al.*, 2008)].

A structure-based sequence alignment was computed using the program T-COFFEE, version 8.13 (Notredame *et al.*, 2000) on a set of 50 manually selected sequences in FASTA (Pearson) format (Pearson and Lipman, 1988). 44 sequences correspond to SQR members, equally distributed among the identified types (10 sequences for type V, 8 sequences for each type I and II, and 6 sequences each for types III, IV and VI). 6 sequences of bacterial FCSD homologues were also included. First, a fully automated run was performed online (http://tcoffee.vital-it.ch) using the Expresso/3DCoffee algorithm (Armougom *et al.*, 2006) [method sap_pair (Taylor and Orengo, 1989)] and the full set of 50 sequences in FASTA format (see Appendix tables 35 – 37). Second, the sequences were separated into the seven groups (six SQR types plus the FCSD group) and a T-COFFEE sequence alignment (regular mode) was calculated independently. The alignments were successively combined as profiles in T-COFFEE, running the program locally, providing the structural coordinates 1FCD, 3HYV and 3H8L and using the methods sap_pair (Taylor and Orengo, 1989) and TMalign_pair (Zhang and Skolnick, 2005). The two strategies yielded similar results. The final score of the alignment is 63, confidently above 40, the cutoff indicated by the developers of T-COFFEE as minimal for a reliable result. Manual adjustments in Jalview (Waterhouse *et al.*, 2009) slightly improved the original score to 64. The phylogenetic tree was visualized with Geneious (Drummond, A.J. *et al.*, available at www.geneious.com).

4.2.2.2. PEPTIDE MASS FINGERPRINT MASS SPECTROMETRY (PMF-MS)

Protein identification was performed using a MALDI-TOF mass spectrometer (OmniFlex®, Bruker). Protein samples were prepared from Coomassie-stained SDS-PAGE gels as described previously (Shevchenko *et al.*, 1996). The target bands were excised from the gels, destained and dried. A trypsin solution was soaked into the gel pieces overnight at 37 °C. The peptides were extracted from the gels by sonication and the solution was acidified with TFA and dried in a rotary evaporator (SpeedVac). Finally the peptides were resuspended and mixed with a saturated solution of matrix (α-cyano-4-hydroxy-cinnamic acid, CHCA, Bruker) and loaded on a 7x7-well target (Bruker). Spectra were acquired operating in reflector delayed-extraction negative mode. Spectra from at least 1000 individual shots were averaged with the software provided by the manufacturer and then smoothed and analyzed using Mascot (Perkins

et al., 1999) and MoverZ (bioinformatics.genomicsolutions.com).

PMF-MS experiments were also performed on a MALDI-TOF MS (Voyager STR, Bruker) in collaboration with Tobias Beckhaus (Goethe University, Frankfurt am Main, Germany).

Finally, additional PMF-MS experiments primarily aimed at the detection of covalent protein modifications were performed by easy-nLC-ESI-q-TOF tandem MS in collaboration with Dr. Julian Langer (Max Planck Institute of Biophysics, Frankfurt am Main, Germany). For this latter purpose, the protein samples were prepared from SDS-PAGE gels run in the presence or in the absence of reducing agent or directly on the protein solution. Excised gel bands containing the target proteins were proteolysed with the ProteoExtract All-in-one Trypsin Digestion kit (CalBiochem). The gel pieces were treated with reducing and blocking reagent or with a corresponding volume of water and digested with a protease (trypsin or chymotrypsin or a combination of the two) according to the manufacturer's guidelines. In the last step, an extra 10 µl of buffer was added to harvest the proteolytic fragments. The supernatant containing these peptides was loaded onto reverse phase columns in a Bruker/Proxeon easy-nLC and eluted with a water/acetonitrile gradient. Peptides eluting from the column were ionised using a Bruker ESI-source with a nanoSprayer emitter and analysed by tandem-MS in a q-TOF mass spectrometer (Bruker micrOTOF-Q-II and Bruker maXis). The data sets were processed using a standard proteomics script and the Bruker DataAnalysis 4.0 software. The resulting MS and MS/MS mass lists were used to search the NCBInr database using Mascot on a local server.

The solutions used in the PMF-MS experiments as well as the details on data processing and the Mascot search parameters are listed in Table 4.11.

4.2.2.3. FULL-LENGTH MATRIX-ASSISTED LASER-DESORPTION IONISATION TIME-OF-FLIGHT MASS SPECTROMETRY (MALDI-TOF MS)

Full-length MALDI-TOF mass spectrometry on protein from crystals was performed using a MALDI-TOF MS instrument (OmniFlex®, Bruker).

Additionally, full-length MALDI-TOF MS (Voyager STR, Bruker) was performed on the protein solution in collaboration with Dr. Ute Bahr (Goethe University, Frankfurt am Main, Germany) and Dr. Julian Langer (Max Planck Institute of Biophysics, Frankfurt am Main, Germany). The experimental conditions were similar to those used for MALDI-TOF PMF-MS (see Chapter 4.2.2.2 and Table 4.11) but a 50 g/L in 0.05 % TFA solution of sDHB (1:9 mixture of 2,5-dihydroxybenzoic acid and 5-methoxysalicylic acid) was used as a matrix.

Table 4.11: List of solutions and experimental conditions for PMF-MS.

MALDI-TOF PMF-MS

Solution	Component	Amount
▪ **Destainer**	AHC	100 mM
	ACN	50 %
	H_2O	
▪ **Wash**	ACN	100 %
▪ **Trypsin, stock solution (-70 °C)**	sequence grade trypsin (Sigma-Aldrich)	1 mg / mL
	acetic acid	50 mM
	H_2O	
▪ **Trypsin, working solution**	trypsin stock solution	5 μL
	AHC	40 mM
	H_2O	to 250 μL
▪ **Cleavage**	AHC	40 mM
	H_2O	
▪ **Acidification**	TFA (10 % in water)	2 μL / sample
▪ **Resuspension**	ACN	70 %
	TFA	0.1 %
	H_2O	
▪ **Matrix**	ACN	70 %
	TFA	0.1 %
	H_2O	
	matrix (CHCA)	to saturation

easy-nLC-ESI-q-TOF PMF-MS

▪ **HPLC**	Instruments	Bruker/Proxeon easy nano LC
	Solvent A	Water 0.1% FA
	Solvent B	ACN 0.1% FA
	nano-LC columns	NanoSeparations
− pre-column	NS-MP-10 5 μm C18, 360 μm OD, 100 μm ID, L = 20 mm
− analytical column	NS-AC-10 5um C18, 75μm OD, 75μm ID, L=10cm
	Bruker ESI-Source	nanoSprayer for nanoLC
	flow rate	300 nl / min
	gradient	25 min total run time, 95 : 5 => 65 : 35 (15 mins) => 0 : 100 (17 mins), for details, see chromatogram view of the individual runs
▪ **MS**	mode	"auto" MS/MS, positive ion mode
	mass range	50 - 2,200 m/z

	MS acquisition	2 Hz, cycle time 3 – 5.5 s
	parental ion selection	1 MS was followed by 3 MS / MS of the most intense ions; the precursor ions for MS/MS were chosen in the range from 400 to 1,400 m/Z; if an internal calibrant was used, its isotopes (1221.9-1225 m/Z) were excluded from fragmentation preferred charge states 2^+ - 4^+ single-charged ions were not excluded active exclusion after 2 spectra for 0.5 min
	collision gas	nitrogen
	MS / MS acquisition	threshold for MS / MS was set to an intensity of 1,000 ions with intensity of 10,000 counts in MS were fragmented for 0.5 s ions with intensity of 1,000 counts were fragmented for 1 s
▪ **Data analysis**	autoMS(n)	intensity threshold: 1,000 retention time window: 0
	deconvolution	sum peak S/N threshold:1 relative intensity threshold: 0.1% mass range: 100 – 3,000 m/z
	export	as .mgf file
▪ **Search**	search engine	Mascot 2.2.2
	type of search	MS/MS Ion Search
	enzyme	trypsin, chymotrypsin or combination of both
	variable modifications	oxidation on methionines
	fixed modification	carbamidomethylation of cysteines (if reducing agent used)
	mass values	monoisotopic
	protein mass	unrestricted
	peptide mass tolerance	± 0.05 Da
	fragment mass tolerance	± 0.05 Da
	max missed cleavages	1
	instrument type	ESI-QUAD-TOF (micrOTOF-Q-II and maXis)
	database	NCBI nr (downloaded 19/08/2009)
	minimum ion score	15
	significance threshold	$p < 0.05$
	minimum peptide length	5
	scoring	standard individual ion scores > 31 indicate identity or extensive homology ($p < 0.05$)

4.2.2.4. FULL-LENGTH LIQUID CHROMATOGRAPHY ELECTROSPRAY IONISATION QUADRUPOLE TIME-OF-FLIGHT MASS SPECTROMETRY (EASY-NLC-ESI-Q-TOF MS)

Full-length mass spectrometry was additionally performed in collaboration with Dr. Julian Langer (Max Planck Institute of Biophysics, Frankfurt am Main, Germany). Proteins were loaded on a Sorbax SB-C8 high resolution cartridge (2.1 x 30 mm, 3.5 µm) in an Agilent 1200 HPLC system. The samples were eluted using a water/acetonitrile gradient (see chromatogram views of individual chromatograms), ionised using a Bruker ESI-Sprayer emitter in an ESI-Source and analysed in a Bruker maXis q-TOF mass spectrometer. The ion optics were trimmed for detection of large proteins. Spectra in the eluting peaks were analysed using the Bruker DataAnalysis 4.0 software. The spectra were binned and deconvoluted using the Bruker Maximum Entropy Algorithm. Table 4.12 summarizes the experimental conditions.

Table 4.12: List of solutions and experimental conditions for full length easy-nLC-ESI-q-TOF MS.

Parameters	Value
HPLC conditions	
Instrument	Agilent 1200 HPLC
Solvent A	Water 0.1% FA
Solvent B	ACN 0.1% FA
nano-LC columns (Agilent technologies)	Sorbax SB-C8 high resolution cartridge (2.1 x 30 mm, 3.5 µm)
Bruker ESI-Source	ESI-Sprayer for Agilent LC
flow rate	500 µl / min
gradient	100 : 0 => 10 : 90 (15 min) => 10 : 90 (19 min) => 100 : 0 (30 min), for details, see chromatogram view of the individual runs

Trimming of the ion optics for detection of large m/z values	
ISCID	90 eV
funnel RF	400 Vpp
fultipole RF	400 Vpp
ion cooler RF	400 Vpp
collision cell transfer time	199 µs
collision cell RF	1,200 Vpp

Full length charge deconvolution (Maximum Entropy Algorithm)	
selected mass range	1,100 – 2,500 m/z
low mass	45,000 m/z
high mass	55,000 m/z
resolving power	15,000 m/z

4.2.2.5.   FULL-LENGTH LASER-INDUCED LIQUID BEAD ION DESORPTION MASS SPECTROMETRY (LILBID MS)

Details of the experimental set-up for LILBID MS have been published elsewhere (Morgner *et al.*, 2006). Briefly, micro-droplets of solution (diameter 50 μm) are produced on demand at 10 Hz by a piezo-driven droplet generator and introduced into vacuum via differential pumping stages. In vacuum the droplets are irradiated one by one by synchronized high-power mid-IR laser pulses of typically 5-ns pulse length. The pulses are generated in a home-made optical parametric oscillator (OPO) using $LiNbO_3$ crystals and a Nd:YAG laser as pump. The wavelength of the OPO radiation is tuned to the absorption maximum of water at around 3 μm, corresponding to an excitation of its stretching vibrations. At a threshold intensity the droplet explodes, resulting in the emission of ions from the liquid into the gas phase. The ions are mass-analysed in a time-of-flight (TOF) reflectron mass spectrometer. For detection of very large biomolecules, a Daly-type ion detector is used, working up to an m/Z range in the low MDa region. At low laser intensity, LILBID MS desorbs ions out of the liquid very gently (ultrasoft mode) enabling detection of non-covalently assembled protein complexes. At higher laser intensities the complex is thermolysed into its subunits (harsh mode). The signals from the detector are recorded by a transient recorder. For data acquisition and analysis a user-written labview program is used. The signal-to-noise ratio is improved by subtracting an unstructured background, caused by metastable loss of water and buffer molecules, from the original ion spectra. These difference spectra are smoothed by averaging the signal over a pre-set number of channels of the transient recorder, with the smoothing interval always lying within the time resolution of the TOF mass spectrometer. LILBID MS experiments were performed in collaboration with Lucie Sokolova (Goethe University, Frankfurt am Main, Germany).

4.2.2.6.   CROSSLINKING

Purified SQR at a concentration of 0.1 mg/mL was incubated at room temperature for 1 h with 5 mM glutaraldehyde in a buffer containing 10 mM HEPES, pH 8.2 and DDM 0.05 %. After this time, the reaction was quenched with Tris-HCl 100 mM, pH 7.4 and the protein sample was analyzed by SDS-PAGE (Simanshu *et al.*, 2006).

4.2.2.7.   ANALYTICAL ULTRACENTRIFUGATION

The apparent molecular weight of the SQR complex was estimated by sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation in an Optima XL-A centrifuge (Beckman Coulter). The data were collected at a wavelength of 280 nm.

Sedimentation velocity experiments were performed with 300 µL samples in Tris-HCl, pH 7.4, 20 mM, 0.05 % DDM at protein concentrations of 0.5 – 1 mg/mL ($OD_{280\,nm}$= 0.2 – 0.4). Absorbance data were acquired at rotor speeds of 30,000 – 40,000 rpm and at a temperature of 20 ºC. The buffer density, buffer viscosity and the protein partial specific volumes were calculated using the software SEDNTERP, kindly provided by Dr. J. Philo (www.rasmb.org). Sedimentation velocity data were analyzed using the c(s) continuous distribution of Lamm equation solutions with the software SEDFIT (Schuck, 2000). Density-matched sedimentation equilibrium experiments were done at 4 ºC at rotor speeds of 10,000, 15,000 and 20,000 rpm at a protein optical density ($OD_{280\,nm}$) of about 0.2 units in buffer 1 (Tris-HCl 20 mM, pH 7.4, DDM 0.05 % w/V, sucrose 27.6 % w/V, D_2O 95 % V/V), buffer 2 (Tris-HCl 20 mM, pH 7.4, $C_{12}E_9$ 0.01 % w/V, sucrose 6.4 % w/V) and buffer 3 (Tris-HCl 20 mM, pH 7.4, $C_{12}E_9$ 0.01 % w/V, D_2O 46 % V/V), respectively. Global non-linear regression of the experimental absorbance profiles was performed using the software SEDPHAT (www.analyticalultracentrifugation.com). The partial specific volume of the protein was calculated from the protein sequence and corrected for each buffer condition according to previous reports (Mayer *et al.*, 1999).

4.2.2.8. TOTAL X-RAY FLUORESCENCE (TXRF) SPECTROSCOPY

Total X-ray Fluorescence was performed in collaboration with Dr. Steffen Metz and Claudia Rittmeyer (Goethe University, Frankfurt am Main, Germany).

All measurements were performed on a TXRF spectrometer EXTRA IIA (ATOMIKA Instruments) equipped with molybdenum and tungsten excitation tubes (tungsten L_α and molybdenum K_α radiation), a Si(Li) solid state detector with an area of 80 mm^2, an automated sample changer and a computer-controlled multichannel analyser system. The tungsten tube was operated at 25 kV and 60 mA using a 30 µm Cu attenuation filter, whereas the molybdenum tube was operated at 50 kV and 40 mA using a 200 µm Mo, aluminium foil attenuation filter. The samples were measured for 1,000 s. Sulfur, phosphate and metal concentrations were determined on protein samples at a concentration of 20 µM, in Tris-acetate pH 7.4 20 mM with 0.05 % DDM. Cr (10 mg/L) was added (1:10 ratio) as an internal standard. 4 µL of the final solution was air dried at room temperature onto an unsiliconized quartz glass carrier under clean bench conditions. Each sample was measured four times. A reference measurement was performed on the buffer without protein sample and taken as blank.

4.2.2.9. ENZYMATIC ACTIVITY ASSAY

The enzymatic activity of the protein was performed according to Nübel and coworkers

(Nübel *et al.*, 2000). The assay was performed using a diode array UV/Vis spectrophotometer (Agilent) measuring the time course of the 285 nm minus 300 nm absorption difference for 300 s. The reaction buffer (1 mL, degassed and nitrogen saturated) contained 50 mM Tris-HCl, pH 7.4, 20 mM glucose, 1 unit of glucose oxidase (type II from *Aspergillus niger*, Sigma-Aldrich) and 10 units of catalase (from bovine liver, Sigma-Aldrich). Semi-anaerobic cells (1 cm light path) made of Quarz-Suprasil and sealed with a silicon lid (Hellma) were used. As a substrate, decylubiquinone was used at a final concentration of 10 µM. Commercial stocks of 25 mg decylubiquinone were dissolved in 100 % ethanol and 2 mM aliquots were stored at -20 °C in light-protected Eppendorf cups. The reaction was started by the addition of Na_2S to a final concentration of 100 µM. Stocks of Na_2S were prepared fresh in degassed Tris-HCl 1 M pH 8.0.

4.2.2.10. ABSORPTION AND FLUORESCENCE SPECTROSCOPY

Quantification of the FAD cofactor in the pure protein preparation was done using an extinction coefficient of 10.8 $cm^{-1}mM^{-1}$ (Weyler and Salach, 1985) for the air-oxidised minus dithionite-reduced difference spectrum measured at 456 nm in a dual-wavelength spectrophotometer (Perkin Elmer).

Fluorescence emission spectra were measured in a fluorescence spectrophotometer (F-4500, Hitachi) exciting the sample at a wavelength of 365 nm and recording the emission at 480 – 600 nm. The samples were measured in quartz cuvettes (Hellma) suitable for fluorescence data acquisition and adapted with a self-made holder to minimize sample volume to 200 µL.

4.2.2.11. SINGLE-PARTICLE ELECTRON MICROSCOPY

Single-particle electron microscopy was performed in the attempt to obtain a low resolution initial insight into the structure of the protein and information on its oligomeric state in solution. The experiments were performed in collaboration with Dr. D. Parcej (Max Planck Institute of Biophysics, Frankfurt am Main, Germany). 3 µL of pure sample at a concentration of approximately 0.04 mg/mL were applied to a glow-discharged carbon-coated copper electron-microscope grid (400 mesh size). After adsorption for 1 minute, the specimen was stained by repetitive addition of 1 % (w/v) uranyl acetate followed by rapid drying (Stoops et al., 1992). Images were collected under low dose conditions onto Kodak SO-163 film using a Philips CM120 microscope operating at 120 kV with a calibrated magnification of 58,000x. After development for 12 minutes in full-strength D19 developer, the films were digitized on a Zeiss SCAI flat bed scanner (Zeiss, Göttingen, Germany). Images were then

converted into SPIDER (Frank et al., 1996) format and binned to a final pixel size of 0.36 nm on the sample scale. Well separated and stained single particles (462 particles) were manually selected in WEB (Frank et al., 1996) and windowed. To avoid reference bias, the single-particle images were first aligned using reference-free alignment (Marco et al., 1993) and then by three further rounds of translational and rotational alignment. Variability in the images was analysed using a Kehonen self-organising map within the package Xmipp (Marabini et al., 1996). Data were distributed on 7 x 7 grid with hexagonal topology.

Additional single-particle electron microscopy studies were performed in collaboration with Dr. J. Vonck (Max Planck Institute of Biophysics, Frankfurt am Main, Germany) in order to study the stability of the sample in the absence of detergent.

4.2.3. *Protein crystallization and structure determination*

4.2.3.1. PROTEIN CRYSTALLIZATION

The homogeneous purified protein solution was concentrated and mixed with an equimolar amount of FAD (Sigma) dissolved in water to a final protein concentration of 8 – 10 mg/mL. The protein was crystallized by the hanging- or sitting-drop vapour diffusion method at 4, 10, 16, 18 or 25 °C. For random screening, 100 + 100 nL or 300 + 300 nL protein / reservoir drops were equilibrated against 100 µL reservoir solution using the crystallographic robots MicroSys SQ Series (Cartesian Technologies) or Mosquito (TTP LabTech) in Crystal Quick^{TM} (Greiner Bio-one) or AxyGem^{TM} (Axygen Bioscience) 96-well plates. The sparse matrix screens listed in Table 4.3 were used to obtain initial hits. For refinement, 1 + 1 µL protein / reservoir drops were equilibrated against 300 µL reservoir by hand in XRL 24-well plates (Molecular Dimensions Ltd) using silanized cover slides (silanization was performed using 2 % dichloro-dimethyl-silane in n-hexane and two successive washing steps in ethanol). All buffers were prepared with ddH$_2$O (18 Ω) and filtered (0.2 µm membrane cut-off) to ensure the highest homogeneity of the particles in solution. When additives were used (i.e. the additive screen, the detergent additive screens or the heavy metal screens, Hampton, or the quinone analogues stock solutions, see later), 1 µL of protein solution was mixed with 0.8 µL of the self-made crystallization buffer and 0.2 µL of a 10-fold concentrated additive solution. Initial hits were obtained in the conditions JB6_C3 (Jena Bioscience Classic screen 6, condition C3), JB7_D1 (Jena Bioscience Classic screen 7, condition D1) and MPD_39 (The MPDs, condition 39, Qiagen). The crystals used for structure determination were grown at 18 °C by hanging-drop vapour

diffusion in Na-acetate pH 5.6 0.1 M or Na-MES pH 6.5 0.1 M, ammonium sulfate 2.0 M and PEG 400 2 – 5 % v/v. The crystals become visible after 3 – 6 weeks and reach their full size within 1 – 2 days.

For phasing, the crystals were soaked with heavy metal solutions (prepared from the Hampton Research heavy atom screens). After crystal growth, a 10-fold concentrated water solution of the heavy metal was added to the crystallization drop to a final heavy atom concentration of 1 – 10 mM. After 1 – 24 h, the crystals were washed in cryo-buffer and frozen. Alternatively, the heavy metal solution was added directly to the cryo-buffer.

For substrate derivatization, the crystals were soaked with quinone analogues. Stock solutions in 100 % ethanol, methanol or DMSO were stored at -20 °C in light-protected cups (Eppendorf) and added to the crystallization drops as described for the heavy metal solutions. Alternatively, a minimal amount was mixed homogeneously with 100 % PEG 400 and diluted with water and sodium acetate to a final concentration of 0.1 M sodium acetate, pH 5.6, 40 % v/v PEG 400, 0.1 – 10 mM quinone. Then, the crystals were soaked for 24 h in this solution and directly frozen.

Alternative crystallization methods tested were the batch method under oil, the lipidic cubic phase (Landau and Rosenbusch, 1996) and the sponge phase (Wadsten *et al.*, 2006).

For crystallization in microbatch mode under oil, 1 µL protein solution was mixed with an equal volume of precipitant solutions from the commercial screens available in Terasaki plates (Greiner Bio-one). The drops were then covered with a thin layer of paraffin oil, Al's oil or silicon oil. Crystals grew in the conditions JBM1_A1 (Jena Bioscience Membrane screen 1, condition A1) and JB6_C3 under paraffin oil.

For crystallization in lipidic cubic phase, the phase was prepared according to the following protocol. Two gas-tight 250-µL syringes (here named A and B, Hampton Research) were connected through a home-made screw. Syringe A was filled with monoolein (solid, -80 °C stock, Nu-Chek Prep Inc.) and syringe B with the protein-containing buffer such that the monoolein-to-buffer w/w ratio was precisely 60:40. The syringes were connected avoiding air gaps between the substances. The phase was prepared by gentle mixing of the solution in syringe B into the lipid in syringe A. The quality of the phase was inspected under a light microscope. The phase was directly dispensed by means of a home-made 200-nL dispenser into 96-well plates. They were then mixed with reservoir solutions between two glass slides separated by a thin paper foil and allowed to equilibrate.

For crystallization in sponge phase, the lipidic cubic phase was first prepared using a

suitable detergent-containing buffer (Na-MES pH 6.5 20 mM, DDM 0.02 % w/v) instead of the protein solution. The phase was then aliquoted into small PCR tubes, centrifuged to the bottom of the tubes, covered with 4-fold volumes of a precipitating-agent-containing buffer and equilibrated at 18 °C for 24 h. A screen of suitable precipitating-agent-containing buffers was developed in house together with Sabine Buschmann and Dr. Chitra Rajendran (Max Planck Institute of Biophysics, Frankfurt am Main, Germany, see Appendix table 34). The sponge phase was then mixed with the protein solution in standard hanging-drop vapour diffusion set-ups. Small crystals grew in the sponge phase condition generated using reservoir MPD_39 (The MPDs screen, condition 39, Quiagen).

4.2.3.2. BIOCHEMICAL CHARACTERIZATION OF THE CRYSTALS

The VIS spectrum of the crystals was measured on a microspectrophotometer in house (Universal-Mikroskop-Spektral-Photometer, UMSP 30 / 50, Carl Zeiss). Single crystals were soaked from the cryo-protecting solution (see Chapter 4.2.3.3) into suitable quartz flat-faced capillaries (Müller W.) and the capillaries were sealed with wax. After measurement in the oxidized state, a few crystals of sodium dithionite were added to the solution and the capillary was sealed again for measurement of the same crystal in the reduced state.

For SDS-PAGE electrophoresis and for full-length MALDI-TOF MS on an in house spectrometer (OmniFlex®, Bruker, see Chapter 4.2.2.3), 10 crystals of about 0.1 mm in all dimensions (hexagonal crystals) or of about 0.3 mm in at least one dimension (needle crystals) were extracted from the mother liquor, washed three times in cryo-protecting solution and finally dissolved in water to be used directly.

4.2.3.3. X-RAY EXPOSURE AND DATA COLLECTION

The crystals were either harvested at room temperature in glass capillaries of suitable dimensions (Müller W.) or stored in liquid nitrogen after extracting them from the crystallization drop with suitable nylon loops (Hampton Research), washing them in a cryo-protecting buffer and flash-cooling them under a cryo-jet. The cryo-protectant solution used for structure determination was composed of the crystallization buffer (0.1 M Na-MES or Na-acetate), supplemented with PEG 400, 40 – 50 % v/v.

An attempt to improve diffraction by adjusting the humidity of the crystals was performed using the humidity chamber (Dr. Christian Benda, Max Planck Institute of Biochemistry, Martinsried, Germany).

Crystals were screened at an in house rotating anode generator (Rigaku), at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and at the Swiss

Light Source (SLS, Villigen, Switzerland). All data sets used for structure determination were collected at the SLS beam lines X10SA and X06SA.

4.2.3.4. STRUCTURE DETERMINATION

SQR formed needle crystals with a typical size of 0.02 x 0.05 x 0.2 mm^3 in the space group $P2_12_12_1$ with unit cell dimensions of 112 x 154 x 178 Å^3, a solvent content of 53 % and six molecules in the asymmetric unit. All data sets were reduced and scaled using the XDS suite (Kabsch, 1993). The phases were obtained by the MIRAS method using Os (K$_2$OsCl$_6$) and Au (AuCl$_3$) derivatives. Heavy atom positions were determined with SHELX (Sheldrick *et al.*, 2001) and refined with SHARP (Fortelle and Bricogne, 1997) to obtain initial phases. Solvent flattening was performed with the software DM (Cowtan, 1994) and non-crystallographic symmetry (NCS) averaging was performed in DM using the NSC matrix derived from LSQ in the program O (Jones and Kjeldgaard, 1997) and refined by IMP (Kleywegt and Jones, 1994). An initial model was obtained with the automated program ARP/wARP (Lamzin *et al.*, 2001) and subsequent model building was done manually in Coot (Emsley and Cowtan, 2004). The initial model had an R$_{work}$ of 47 % and an R$_{free}$ of 50 %. Initial refinement was performed with CNS (Brunger *et al.*, 1998) to an R$_{free}$ of 30 %. Further refinement was done with Refmac5, using TLS and imposing NCS restraints on the protein chains (Collaborative computational project number 4, 1994; Murshudov *et al.*, 1997). Final rounds of refinement were done without NCS restraints. The datasets of the decylubiquinone and aurachin C soaked crystals were directly refined with Refmac5 without the need of a molecular replacement run. Solvent, cofactor and substrate molecules were drawn with Sketcher and added to the coordinate files in Coot and Refmac5. Omit maps were calculated using CNS, excluding the atoms of interest and avoiding model bias by simulated annealing refinement. Details about the structure solution flowchart and the scripts used to run each program are reported in the Appendix tables 1 – 33.

References.

1. Ackermann M. (2007) Biochemische, molekularbiologische und histologische Charakterisierung des Sulfid-Chinon-Reduktase-ähnlichen Proteins (SQRDL) bei Ratte und Mensch. Biochemical, biomolecular and histolgical characterization of the sulfide-quinone reductase-like protein (SQRDL) in rats and humans. (Translated from German) MSc thesis (Regensburg, Germany).

2. Almen M. S., Nordstrom K. J., Fredriksson R., Schioth H. B. (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol* 7:50.

3. Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.

4. Andreeva A., Howorth D., Chandonia J. M., Brenner S. E., Hubbard T. J., Chothia C., Murzin A. G. (2008) Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res* 36:D419-425.

5. Argyrou A., Blanchard J. S. (2004) Flavoprotein disulfide reductases: advances in chemistry and function. *Prog Nucleic Acid Res Mol Biol* 78:89-142.

6. Arieli B., Padan E., Shahak Y. (1991) Sulfide-induced sulfide-quinone reductase activity in thylakoids of *Oscillatoria limnetica*. *J Biol Chem* 266:104-111.

7. Arieli B., Shahak Y., Taglicht D., Hauska G., Padan E. (1994) Purification and characterization of sulfide-quinone reductase, a novel enzyme driving anoxygenic photosynthesis in *Oscillatoria limnetica*. *J Biol Chem* 269:5705-5711.

8. Arinaminpathy Y., Khurana E., Engelman D. M., Gerstein M. B. (2009) Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov Today* 14:1130-1135.

9. Armougom F., Moretti S., Poirot O., Audic S., Dumas P., Schaeli B., Keduas V., Notredame C. (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* 34:W604-608.

10. Balali-Mood K., Bond P. J., Sansom M. S. P. (2009) Interaction of monotopic membrane enzymes with a lipid bilayer: A coarse-grained MD simulation study. *Biochemistry* 48:2135-2145.

11. Banner D. W., Bloomer A. C., Petsko G. A., Phillips D. C., Pogson C. I., Wilson I. A., Corran P. H., Furth A. J., Milman J. D., Offord R. E., Priddle J. D., Waley S. G. (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature* 255:609-614.

12. Benavides G. A., Squadrito G. L., Mills R. W., Patel H. D., Isbell T. S., Patel R. P., Darley-Usmar V. M., Doeller J. E., Kraus D. W. (2007) Hydrogen sulfide mediates the vasoactivity of garlic. *Proc Natl Acad Sci U S A* 104:17977-17982.

13. Bendtsen J. D., Binnewies T. T., Hallin P. F., Ussery D. W. (2005) Genome update: prediction of membrane proteins in prokaryotic genomes. *Microbiology* 151:2119-2121.

14. Bernal J. D., Crowfoot D. (1934) X-Ray photographs of crystalline pepsin. *Nature* 133:794-795.

15. Bhatia M. (2010) Hydrogen sulfide and substance P in inflammation. *Antioxidants & Redox Signaling* 12:1191-1202.

16. Bhatia V. K., Madsen K. L., Bolinger P. Y., Kunding A., Hedegard P., Gether U., Stamou D. (2009) Amphipathic motifs in BAR domains are essential for membrane curvature sensing. *EMBO J* 28:3303-3314.

17. Blake C. C., Koenig D. F., Mair G. A., North A. C., Phillips D. C., Sarma V. R. (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* 206:757-761.

18. Blakeley M. P., Ruiz F., Cachau R., Hazemann I., Meilleur F., Mitschler A., Ginell S., Afonine P., Ventura O. N., Cousido-Siah A., Haertlein M., Joachimiak A., Myles D., Podjarny A. (2008) Quantum model of catalysis based on a mobile proton revealed by subatomic x-ray and neutron diffraction studies of h-aldose reductase. *Proc Natl Acad Sci U S A* 105:1844-1848.

19. Blobel G. (1980) Intracellular protein topogenesis. *Proc Natl Acad Sci U S A* 77:1496-1500.

20. Blow D. (2002) *Outline of crystallography for biologists.* (Oxford University Press, New York, USA).

21. Branden C., Tooze J. (1998) *Introduction to protein structure.* (Garland Publishing, Inc., New York, USA) 2nd Ed.

22. Brito J. A., Sousa F. L., Stelter M., Bandeiras T. M., Vonrhein C., Teixeira M., Pereira M. M., Archer M. (2009) Structural and functional insights into sulfide:quinone oxidoreductase. *Biochemistry* 48:5613-5622.

23. Bronstein M., Schütz M., Hauska G., Padan E., Shahak Y. (2000) Cyanobacterial sulfide-quinone reductase: cloning and heterologous expression. *J Bacteriol* 182:3336-3344.

24. Brugna-Guiral M., Tron P., Nitschke W., Stetter K. O., Burlat B., Guigliarelli B., Bruschi M., Giudici-Orticoni M. T. (2003) [NiFe] hydrogenases from the hyperthermophilic bacterium *Aquifex aeolicus*: properties, function, and phylogenetics. *Extremophiles* 7:145-157.

25. Brune D. C. (1989) Sulfur oxidation by phototrophic bacteria. *Biochim Biophys Acta* 975:189-221.

26. Brunger A. T., Adams P. D., Clore G. M., DeLano W. L., Gros P., Grosse-Kunstleve R. W., Jiang J. S., Kuszewski J., Nilges M., Pannu N. S., Read R. J., Rice L. M., Simonson T., Warren G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905-921.

27. Chan L. K., Morgan-Kiss R. M., Hanson T. E. (2009) Functional analysis of three sulfide:quinone oxidoreductase homologs in *Chlorobaculum tepidum*. *J Bacteriol* 191:1026-1034.

28. Chang T., Untereiner A., Liu J., Wu L. (2010) Interaction of methylglyoxal and hydrogen sulfide in rat vascular smooth muscle cells. *Antioxid Redox Signal* 12:1093-1100.

29. Chen Z. W., Koh M., Van Driessche G., Van Beeumen J. J., Bartsch R. G., Meyer T. E., Cusanovich M. A., Mathews F. S. (1994) The structure of flavocytochrome *c* sulfide dehydrogenase from a purple phototrophic bacterium. *Science* 266:430-432.

30. Cherezov V., Rosenbaum D. M., Hanson M. A., Rasmussen S. G. F., Thian F. S., Kobilka T. S., Choi H. J., Kuhn P., Weis W. I., Kobilka B. K., Stevens R. C. (2007) High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor. *Science* 318:1258-1265.

31. Cho W. (2001) Membrane targeting by C1 and C2 domains. *Journal of Biological Chemistry* 276:32407-32410.

32. Collaborative computational project number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760-763.

33. Cowtan K. (1994) DM: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography.* 31:34-38.

34. d'Emmanuele di Villa Bianca R., Sorrentino R., Maffia P., Mirone V., Imbimbo C., Fusco F., De Palma R., Ignarro L. J., Cirino G. (2009) Hydrogen sulfide as a mediator of human corpus cavernosum smooth-muscle relaxation. *Proc Natl Acad Sci U S A* 106:4513-4518.

35. Dahl C., Friedrich C. G. (2008) *Microbial Sulfur Metabolism* (Springer, Berlin and Heidelberg).

36. De Colibus L., Mattevi A. (2006) New frontiers in structural flavoenzymology. *Curr Opin Struct Biol* 16:722-728.

37. Deckert G., Warren P. V., Gaasterland T., Young W. G., Lenox A. L., Graham D. E., Overbeek R., Snead M. A., Keller M., Aujay M., Huber R., Feldman R. A., Short J. M., Olsen G. J., Swanson R. V. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392:353-358.

38. Deisenhofer J., Epp O., Miki K., Huber R., Michel H. (1985) Structure of the protein subunits in the photosynthetic reaction center of *Rhodopseudomonas viridis* at 3 Å resolution. *Nature* 318:618-624.

39. DeLano W. L. (2002) *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, USA).

40. Dobbin L. (1931) *The Collected Papers of Carl Wilhelm Scheele* (York House, London (UK)) 1st Ed.

41. Edmondson D. E., Newton-Vinson P. (2001) The covalent FAD of monoamine oxidase: structural and functional role and mechanism of the flavinylation reaction. *Antioxid Redox Signal* 3:789-806.

42. Ehrlich H. L., Newman D. K. (2009) *Geomicrobiology* (CRC Press, Boca Raton, Florida (USA)) 5th Ed.

43. Eisenberg D., Weiss R. M., Terwilliger T. C. (1982) The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature* 299:371-374.

44. Emanuelsson O., Brunak S., von Heijne G., Nielsen H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2:953-971.

45. Emsley P., Cowtan K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126-2132.

46. Engelman D. M., Steitz T. A., Goldman A. (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu Rev Biophys Biophys Chem* 15:321-353.

47. Eto K., Ogasawara M., Umemura K., Nagai Y., Kimura H. (2002) Hydrogen sulfide is produced in response to neuronal excitation. *J Neurosci* 22:3386-3391.

48. Faham S., Watanabe A., Besserer G. M., Cascio D., Specht A., Hirayama B. A., Wright E. M., Abramson J. (2008) The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. *Science* 321:810-814.

49. Fang Y., Jayaram H., Shane T., Kolmakova-Partensky L., Wu F., Williams C., Xiong Y., Miller C. (2009) Structure of a prokaryotic virtual proton pump at 3.2 A resolution. *Nature* 460:1040-1043.

50. Fersht A. (1999) *Structure and mechanism in protein science. A guide to enzyme catalysis and protein folding.* (Freeman, W.H. and Company, New York, USA).

51. Fisher N., Rich P. R. (2000) A motif for quinone binding sites in respiratory and photosynthetic systems. *Journal of Molecular Biology* 296:1153-1162.

52. Forneris F., Mattevi A. (2008) Enzymes without borders: mobilizing substrates, delivering products. *Science* 321:213-216.

53. Fortelle E. d. l., Bricogne G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multi-wavelenght anomalous diffraction methods. *Methods Enzymol* 276:472-494.

54. Fowler P. W., Balali-Mood K., Deol S., Coveney P. V., Sansom M. S. P. (2007) Monotopic enzymes and lipid bilayers: A comparative study. *Biochemistry* 46:3108-3115.

55. Fraaije M. W., Mattevi A. (2000) Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem Sci* 25:126-132.

56. Frank J., Radermacher M., Penczek P., Zhu J., Li Y., Ladjadj M., Leith A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116:190-199.

57. Fukumori Y., Yamanaka T. (1979) Flavocytochrome *c* of *Chromatium vinosum*. Some enzymatic properties and subunit structure. *J Biochem* 85:1405-1414.

58. Furne J., Saeed A., Levitt M. D. (2008) Whole tissue hydrogen sulfide concentrations are orders of magnitude lower than presently accepted values. *Am J Physiol Regul Integr Comp Physiol* 295:R1479-1485.

59. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D., Bairoch A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784-3788.

60. Gerday C., Glansdorff N. (2007) *Physiology and biochemistry of extremophiles.* (ASM Press, Washington D.C., USA).

61. Goubern M., Andriamihaja M., Nübel T., Blachier F., Bouillaud F. (2007) Sulfide, the first inorganic substrate for human cells. *Faseb Journal* 21:1699-1706.

62. Granseth E., von Heijne G., Elofsson A. (2005) A study of the membrane-water interface region of membrane proteins. *J Mol Biol* 346:377-385.

63. Griesbeck C., Hauska G., Schütz M. (2000) Biological sulfide oxidation: sulfide-quinone reductase (SQR), the primary reaction. *Recent research developments in microbiology*, ed Pandalai S. G. (Research Signpost, Trivandrum, India), Vol 4, pp 179-203.

64. Griesbeck C., Schütz M., Schödl T., Bathe S., Nausch L., Mederer N., Vielreicher M., Hauska G. (2002) Mechanism of sulfide-quinone reductase investigated using site-directed mutagenesis and sulfur analysis. *Biochemistry* 41:11552-11565.

65. Grieshaber M. K., Volkel S. (1998) Animal adaptations for tolerance and exploitation of poisonous sulfide. *Annu Rev Physiol* 60:33-53.

66. Grossi L. (2009a) Hydrogen sulfide induces nitric oxide release from nitrite. *Bioorg Med Chem Lett* 19:6092-6094.

67. Grossi L. (2009b) Hydrogen sulfide: A gasotransmitter or just a cofactor of the nitrite in inducing nitric oxide release? *Nitric Oxide-Biology and Chemistry* 20:S34-S34.

68. Grüber L. (2002) Mechanistische Studien an der Sulfid-Chinon-Reductase (SQR) von *Rhodobacter capsulatus*. Mechanistic studies on the sulfide:quinone reductase (SQR) of *Rhodobacter capsulatus*. (Translated from German) MSc thesis (Regensburg, Germany).

69. Guidotti G. (1972) Membrane proteins. *Annu Rev Biochem* 41:731-752.

70. Guiral M., Prunetti L., Lignon S., Lebrun R., Moinier D., Giudici-Orticoni M. T. (2009) New insights into the respiratory chains of the chemolithoautotrophic and hyperthermophilic bacterium *Aquifex aeolicus*. *J Proteome Res* 8:1717-1730.

71. Guiral M., Tron P., Aubert C., Gloter A., Iobbi-Nivol C., Giudici-Orticoni M. T. (2005) A membrane-bound multienzyme, hydrogen-oxidizing, and sulfur-reducing complex from the hyperthermophilic bacterium *Aquifex aeolicus*. *J Biol Chem* 280:42004-42015.

72. Halada P., Leitner C., Sedmera P., Haltrich D., Volc J. (2003) Identification of the covalent flavin adenine dinucleotide-binding region in pyranose 2-oxidase from *Trametes multicolor*. *Anal Biochem* 314:235-242.

73. Harker D. (1956) The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. *Acta Crystallographica* 9:1-9.

74. Hayashi M., Nakayama Y., Yasui M., Maeda M., Furuishi K., Unemoto T. (2001) FMN is covalently attached to a threonine residue in the NqrB and NqrC subunits of Na(+)-translocating NADH-quinone reductase from *Vibrio alginolyticus*. *FEBS Lett* 488:5-8.

75. Heuts D. P., Scrutton N. S., McIntire W. S., Fraaije M. W. (2009) What's in a covalent bond? On the role and formation of covalently bound flavin cofactors. *FEBS J* 276:3405-3427.

76. Hildebrandt T. M., Grieshaber M. K. (2008) Three enzymatic activities catalyze the oxidation of sulfide to thiosulfate in mammalian and invertebrate mitochondria. *FEBS J* 275:3352-3361.

77. Holm L., Kaariainen S., Rosenstrom P., Schenkel A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24:2780-2781.

78. Horikoshi K., Grant W. eds (1998) *Extremophiles. Microbial life in extreme environments.* (Wiley-Liss, California, USA), Vol 20.

79. Hough D. W., Danson M. J. (1999) Extremozymes. *Curr Opin Chem Biol* 3:39-46.

80. Huber R., Eder W. (2006) Aquificales. *The Prokaryotes: ecophysiology and biochemistry.*, eds Dworkin M., Falkow S. (Springer Verlag, Heidelberg, Germany), Vol 7, pp 925 - 938.

81. Huber R., Wilharm T., Huber D., Trincone A., Burggraf S., Konig H., Rachel R., Rockinger I., Fricke H., Stetter K. O. (1992) *Aquifex pyrophilus* gen. nov. sp. nov. represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Systematic and Applied Microbiology* 15:340-351.

82. Hunte C., Koepke J., Lange C., Rossmanith T., Michel H. (2000) Structure at 2.3 angstrom resolution of the cytochrome *bc*(1) complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. *Structure* 8:669-684.

83. Hunte C., Richers S. (2008) Lipids and membrane protein structures. *Curr Opin Struct Biol* 18:406-411.

84. Ichinohe A., Kanaumi T., Takashima S., Enokido Y., Nagai Y., Kimura H. (2005) Cystathionine beta-synthase is enriched in the brains of Down's patients. *Biochem Biophys Res Commun* 338:1547-1550.

85. Itoh T., De Camilli P. (2006) BAR, F-BAR (EFC) and ENTH/ANTH domains in the regulation of membrane-cytosol interfaces and membrane curvature. *Biochim Biophys Acta* 1761:897-912.

86. Iwata S. ed (2003) *Methods and results in crystallization of membrane proteins.* (International University Line, La Jolla, USA).

87. Jaenicke R., Sterner R. (2006) Life at high temperatures. *The Prokaryotes: ecophysiology and biochemistry.*, eds Dworkin M., Falkow S. (Springer Verlag, Heidelberg, Germany), Vol 2, pp 925 - 938.

88. Jeffery C. J. (1999) Moonlighting proteins. *Trends Biochem Sci* 24:8-11.

89. Jelsch C., Teeter M. M., Lamzin V., Pichon-Pesme V., Blessing R. H., Lecomte C. (2000) Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc Natl Acad Sci U S A* 97:3171-3176.

90. Johnson J. E., Cornell R. B. (1999) Amphitropic proteins: regulation by reversible membrane interactions. *Mol Membr Biol* 16:217-235.

91. Jones T. A., Kjeldgaard M. (1997) Electron-density map interpretation. *Methods Enzymol* 277:173-208.

92. Kabsch W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* 26:795-800.

93. Kamoun P. (2004) H2S, un nouveau neuromodulateur. H2S, a new neuromodulator (Translated from French) (Paris, France).

94. Kelley L. A., MacCallum R. M., Sternberg M. J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* 299:499-520.

95. Kelly D. P. (1980) The sulphur cycle: definition, mechanism and dynamics. *Ciba Foundation Symposium - 72,* Sulphur in Biology, eds Elliot K., Whelan J. (Excerpta Medica, Amsterdam, The Netherlands), pp 3 - 18.

96. Kendrew J. C., Bodo G., Dintzis H. M., Parrish R. G., Wyckoff H., Phillips D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181:662-666.

97. Khersonsky O., Roodveldt C., Tawfik D. S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* 10:498-508.

98. Kimura H. (2010) Hydrogen sulfide: from brain to gut. *Antioxid Redox Signal* 12:1111-1123.

99. Kleywegt G. J., Jones T. A. (1994) Halloween...masks and bones. From First Map to Final Model. eds Bailey S., Hubbard R., Waller R. (SERC Daresbury Laboratory, Warrington, UK), pp 59-66.

100. Knapp K. G., Swartz J. R. (2004) Cell-free production of active *E. coli* thioredoxin reductase and glutathione reductase. *Febs Letters* 559:66-70.

101. Kombian S. B., Reiffenstein R. J., Colmers W. F. (1993) The actions of hydrogen sulfide on dorsal raphe serotonergic neurons in vitro. *J Neurophysiol* 70:81-96.

102. Krishnamurthy H., Piscitelli C. L., Gouaux E. (2009) Unlocking the molecular secrets of sodium-coupled transporters. *Nature* 459:347-355.

103. Kulkarni K. H., Monjok E. M., Zeyssig R., Kouamou G., Bongmba O. N., Opere C. A., Njie Y. F., Ohia S. E. (2009) Effect of hydrogen sulfide on sympathetic neurotransmission and catecholamine levels in isolated porcine iris-ciliary body. *Neurochem Res* 34:400-406.

104. Kumagai I., Takeda S., Miura K. (1992) Functional conversion of the homologous proteins alpha-lactalbumin and lysozyme by exon exchange. *Proc Natl Acad Sci U S A* 89:5887-5891.

105. Kumar S., Tsai C. J., Nussinov R. (2000) Factors enhancing protein thermostability. *Protein Eng* 13:179-191.

106. Kuo S. M., Lea T. C., Stipanuk M. H. (1983) Developmental pattern, tissue distribution, and subcellular distribution of cysteine: alpha-ketoglutarate aminotransferase and 3-mercaptopyruvate sulfurtransferase activities in the rat. *Biol Neonate* 43:23-32.

107. Laemmli U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685.

108. Lagoutte E., Mimoun S., Andriamihaja M., Chaumontet C., Blachier F., Bouillaud F. (2010) Oxidation of hydrogen sulfide remains a priority in mammalian cells and causes reverse electron transfer in colonocytes. *Biochim Biophys Acta* 1797:1500-1511.

109. Lamzin V. S., Perrakis A., Wilson K. S. (2001) The ARP/wARP suite for automated construction and refinement of protein models. *International Tables for Crystallography*, eds Rossmann M. G., Arnold E. (Kluwer Academic Publishers, Dordrecht, The Netherlands), Vol F, pp 720-722.

110. Landau E. M., Rosenbusch J. P. (1996) Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci U S A* 93:14532-14535.

111. Lenaz G., Degli Esposti M. (1994) Membrane-bound enzymes. *Biomembranes. Structural and functional aspects.*, ed Shinitzky M. (Balaban (VCH), Weinheim, Germany).

112. Lesk A. M., Branden C. I., Chothia C. (1989) Structural principles of alpha/beta barrel proteins: the packing of the interior of the sheet. *Proteins* 5:139-148.

113. Levitt M. D., Furne J., Springfield J., Suarez F., DeMaster E. (1999) Detoxification of hydrogen sulfide and methanethiol in the cecal mucosa. *J Clin Invest* 104:1107-1114.

114. Lomize A. L., Pogozheva I. D., Lomize M. A., Mosberg H. I. (2007) The role of hydrophobic interactions in positioning of peripheral proteins in membranes. *BMC Struct Biol* 7:44.

115. Lu M., Choo C. H., Hu L. F., Tan B. H., Hu G., Bian J. S. (2010) Hydrogen sulfide regulates intracellular pH in rat primary cultured glia cells. *Neurosci Res* 66:92-98.

116. Macelroy R. (1974) Some comments on the evolution of extremophiles. *Biosystems* 6:74-75.

117. Maier K. (2007) Molekularbiologische und biochemische Untersuchungen am menschlichen Sulfid-Chinon-Reduktase-ähnlichen Protein (SQRDL). Molecular biology and biochemical investigations on the human sulfide:quinone reductase-like protein (SQRDL). (Translated from German) MSc thesis (Regensburg, Germany).

118. Marabini R., Masegosa I. M., San Martin M. C., Marco S., Fernandez J. J., de la Fraga L. G., Vaquerizo C., Carazo J. M. (1996) Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol* 116:237-240.

119. Marcia M. (2005) Purificazione e caratterizzazione funzionale della solfuro:chinone ossidoreduttasi e della citocromo *bd*:chinolo ossidasi da *Aquifex aeolicus*. Purification and characterization of the sulfide:quinone oxidoreductase and the cytochrome *bd*:quinol oxidase from *Aquifex aeolicus*. (Translated from Italian) MSc thesis (Bologna , Italy).

120. Marcia M., Ermler U., Peng G., Michel H. (2009) The structure of *Aquifex aeolicus* sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration. *Proc Natl Acad Sci U S A* 106:9625-9630.

121. Marco S., Valpuesta J. M., Rivas G., Andres G., San Martin C., Carrascosa J. L. (1993) A structural model for the GroEL chaperonin. *FEMS Microbiol Lett* 106:301-308.

122. Martin A. C., Orengo C. A., Hutchinson E. G., Jones S., Karmirantzou M., Laskowski R. A., Mitchell J. B., Taroni C., Thornton J. M. (1998) Protein folds and functions. *Structure* 6:875-884.

123. Mathai J. C., Missner A., Kugler P., Saparov S. M., Zeidel M. L., Lee J. K., Pohl P. (2009) No facilitator required for membrane transport of hydrogen sulfide. *Proc Natl Acad Sci U S A* 106:16633-16638.

124. Mayer G., Ludwig B., Muller H. W., van den Broek J. A., Friesen R. H. E., Schubert D. (1999) Studying membrane proteins in detergent solution by analytical ultracentrifugation: different methods for density matching. *Analytical Ultracentrifugation V* 113:176-181.

125. McCoy J. G., Bingman C. A., Bitto E., Holdorf M. M., Makaroff C. A., Phillips G. N., Jr. (2006) Structure of an ETHE1-like protein from *Arabidopsis thaliana*. *Acta Crystallogr D Biol Crystallogr* 62:964-970.

126. McNaught A. D., Wilkinson A. (1997) *IUPAC Compendium of Chemical Terminology (Golden book)*. (Blackwell Synergy, Cambridge, UK) 2nd Ed.

127. McPherson A. ed (1999) *Crystallization of biological macromolecules* (Cold Spring Harbor Laboratory Press, New York, USA).

128. McPherson A. (2009) *Introduction to macromolecular crystallography.* (Wiley-Blackwell, Hoboken, USA) 2nd Ed.

129. Meyer B. (1964) Solid allotropes of sulfur. *Chem Rev* 64:429-451.

130. Michel H. (1982a) Characterization and crystal packing of three-dimensional bacteriorhodopsin crystals. *EMBO J* 1:1267-1271.

131. Michel H. (1982b) Three-dimensional crystals of a membrane protein complex. The photosynthetic reaction centre from *Rhodopseudomonas viridis*. *J Mol Biol* 158:567-572.

132. Michel H. (1983) Crystallization of membrane proteins. *Trends in Biochemical Sciences* 8:56-59.

133. Michel H. ed (1991) *Crystallization of membrane proteins* (CRC Press, Boca Raton, USA).

134. Michel H. (2006) Crystallization of membrane proteins. *International Tables for Crystallography* F:94-99.

135. Minor D. L., Jr., Kim P. S. (1996) Context-dependent secondary structure formation of a designed protein sequence. *Nature* 380:730-734.

136. Moore J. W., Babidge W. J., Millard S. H., Roediger W. E. (1997) Thiolmethyltransferase activity in the human colonic mucosa: implications for ulcerative colitis. *J Gastroenterol Hepatol* 12:678-684.

137. Moreno A., Quiroz-Garcia B., Yokaichiya F., Stojanoff V., Rudolph P. (2007) Protein crystal growth in gels and stationary magnetic fields. *Crystal Research and Technology* 42:231-236.

138. Morgner N., Barth H. D., Brutschy B. (2006) A new way to detect noncovalently bonded complexes of biomolecules from liquid micro-droplets by laser mass spectrometry. *Australian Journal of Chemistry* 59:109-114.

139. Morton R. A. (1965) Spectroscopy of quinones and related substances. I. Ultraviolet absorption spectra. *Biochemistry of quinones*, ed Morton R. A. (Academic Press, London, UK), pp 23-64.

140. Muellner M. K., Schreier S. M., Laggner H., Hermann M., Esterbauer H., Exner M., Gmeiner B. M., Kapiotis S. (2009) Hydrogen sulfide destroys lipid hydroperoxides in oxidized LDL. *Biochem J* 420:277-281.

141. Murray L., Pires R. H., Hastings S. F., Ingledew W. J. (1999) Models for structure and function in quinone-binding sites: the *Escherichia coli* quinol oxidase, cytochrome bo(3). *Biochemical Society Transactions* 27:581-585.

142. Murshudov G. N., Vagin A. A., Dodson E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.

143. Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247:536-540.

144. Mustafa A. K., Gadalla M. M., Sen N., Kim S., Mu W., Gazi S. K., Barrow R. K., Yang G., Wang R., Snyder S. H. (2009) H_2S signals through protein S-sulfhydration. *Sci Signal* 2:ra72.

145. Neumann P., Weidner A., Pech A., Stubbs M. T., Tittmann K. (2008) Structural basis for membrane binding and catalytic activation of the peripheral membrane enzyme pyruvate oxidase from *Escherichia coli*. *Proc Natl Acad Sci U S A* 105:17390-17395.

146. Newby Z. E., O'Connell J. D., 3rd, Gruswitz F., Hays F. A., Harries W. E., Harwood I. M., Ho J. D., Lee J. K., Savage D. F., Miercke L. J., Stroud R. M. (2009) A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nat Protoc* 4:619-637.

147. Nicholls A., Bharadwaj R., Honig B. (1993) GRASP: Graphical representation and analysis of surface properties. *Biophys J* 64:A166.

148. Nicholls P. (1975) The effect of sulphide on cytochrome *aa₃*. Isosteric and allosteric shifts of the reduced alpha-peak. *Biochim Biophys Acta* 396:24-35.

149. Notredame C., Higgins D. G., Heringa J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205-217.

150. Nübel T., Klughammer C., Huber R., Hauska G., Schütz M. (2000) Sulfide:quinone oxidoreductase in membranes of the hyperthermophilic bacterium *Aquifex aeolicus* (VF5). *Arch Microbiol* 173:233-244.

151. Olson K. R. (2009) Is hydrogen sulfide a circulating "gasotransmitter" in vertebrate blood? *Biochim Biophys Acta* 1787:856-863.

152. Olson K. R., Whitfield N. L. (2010) Hydrogen sulfide and oxygen sensing in the cardiovascular system. *Antioxid Redox Signal* 12:1219-1234.

153. Ostermeier C., Iwata S., Ludwig B., Michel H. (1995) F-v fragment mediated crystallization of the membrane-protein bacterial cytochrome-*c*-oxidase. *Nature Structural Biology* 2:842-846.

154. Ostermeier C., Michel H. (1997) Crystallization of membrane proteins. *Curr Opin Struct Biol* 7:697-701.

155. Pae H. O., Lee Y. C., Jo E. K., Chung H. T. (2009) Subtle interplay of endogenous bioactive gases (NO, CO and H(2)S) in inflammation. *Arch Pharm Res* 32:1155-1162.

156. Palsdottir H., Hunte C. (2004) Lipids in membrane protein structures. *Biochim Biophys Acta* 1666:2-18.

157. Pearson W. R., Lipman D. J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85:2444-2448.

158. Peng G., Fritzsch G., Zickermann V., Schagger H., Mentele R., Lottspeich F., Bostina M., Radermacher M., Huber R., Stetter K. O., Michel H. (2003) Isolation, characterization and

electron microscopic single particle analysis of the NADH:ubiquinone oxidoreductase (complex I) from the hyperthermophilic eubacterium *Aquifex aeolicus*. *Biochemistry* 42:3032-3039.

159. Pereira I. A. C. (2008) Respiratory membrane complexes of *Desulfovibrio*. *Microbial Sulfur Metabolism*, eds Dahl C., Friedrich C. G. (Springer, Heidelberg, D), pp 24-35.

160. Perkins D. N., Pappin D. J., Creasy D. M., Cottrell J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551-3567.

161. Perutz M. (1992) *Protein structure. New approaches to disease and therapy* (WH Freeman and Company, New York, USA).

162. Peter B. J., Kent H. M., Mills I. G., Vallis Y., Butler P. J., Evans P. R., McMahon H. T. (2004) BAR domains as sensors of membrane curvature: the amphiphysin BAR structure. *Science* 303:495-499.

163. Pflugrath J. W. (1999) The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* 55:1718-1725.

164. Pham V. H., Yong J. J., Park S. J., Yoon D. N., Chung W. H., Rhee S. K. (2008) Molecular analysis of the diversity of the sulfide:quinone reductase (sqr) gene in sediment environments. *Microbiology* 154:3112-3121.

165. Ramakrishnan V., Moore P. B. (2001) Atomic structures at last: the ribosome in 2000. *Curr Opin Struct Biol* 11:144-154.

166. Raman P., Cherezov V., Caffrey M. (2006) The Membrane Protein Data Bank. *Cell Mol Life Sci* 63:36-51.

167. Reinartz M., Tschape J., Bruser T., Truper H. G., Dahl C. (1998) Sulfide oxidation in the phototrophic sulfur bacterium *Chromatium vinosum*. *Arch Microbiol* 170:59-68.

168. Ressl S., Terwisscha van Scheltinga A. C., Vonrhein C., Ott V., Ziegler C. (2009) Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. *Nature* 458:47-52.

169. Rettig S. J., Trotter J. (1973) Comparison of super-secondary structures in proteins. *J Mol Biol* 76:241-256.

170. Robertus J. D., Alden R. A., Birktoft J. J., Kraut J., Powers J. C., Wilcox P. E. (1972) An x-ray crystallographic study of the binding of peptide chloromethyl ketone inhibitors to subtilisin BPN'. *Biochemistry* 11:2439-2449.

171. Roediger W. E., Babidge W., Millard S. (1996) Methionine derivatives diminish sulphide damage to colonocytes. Implications for ulcerative colitis. *Gut* 39:77-81.

172. Rossmann M. G., Moras D., Olsen K. W. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* 250:194-199.

173. Rothschild L. J., Mancinelli R. L. (2001) Life in extreme environments. *Nature* 409:1092-1101.

174. Ruf A., Muller F., D'Arcy B., Stihle M., Kusznir E., Handschin C., Morand O. H., Thoma R. (2004) The monotopic membrane protein human oxidosqualene cyclase is active as monomer. *Biochemical and Biophysical Research Communications* 315:247-254.

175. Russell R. J., Ferguson J. M., Hough D. W., Danson M. J., Taylor G. L. (1997) The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry* 36:9983-9994.

176. Sanchez-Weatherby J., Bowler M. W., Huet J., Gobbo A., Felisaz F., Lavault B., Moya R., Kadlec J., Ravelli R. B. G., Cipriani F. (2009) Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallographica Section D-Biological Crystallography* 65:1237-1246.

177. Sapay N., Guermeur Y., Deleage G. (2006) Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinformatics* 7:255.

178. Schägger H., Cramer W. A., von Jagow G. (1994) Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by 2-dimensional native electrophoresis. *Anal Biochem* 217:220-230.

179. Schägger H., von Jagow G. (1991) Blue Native Electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem* 199:223-231.

180. Schödl T. (2003) Sulfid-Chinon Reductase (SQR) aus *Aquifex aeolicus*: Gensynthese, Expression, Reinigung und biochemische Characterisierung. Sulfide-quinone reductase (SQR) from *Aquifex aeolicus*: gene synthesis, expression, purification and biochemical characterization. (Translated from German) PhD Thesis (Regensburg, Germany).

181. Schreier S. M., Muellner M. K., Steinkellner H., Hermann M., Esterbauer H., Exner M., Gmeiner B. M., Kapiotis S., Laggner H. (2010) Hydrogen sulfide scavenges the cytotoxic lipid oxidation product 4-HNE. *Neurotox Res* 17:249-256.

182. Schuck P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* 78:1606-1619.

183. Schütz M., Klughammer C., Griesbeck C., Quentmeier A., Friedrich C. G., Hauska G. (1998) Sulfide-quinone reductase activity in membranes of the chemotrophic bacterium *Paracoccus denitrificans* GB17. *Arch Microbiol* 170:353-360.

184. Schütz M., Maldener I., Griesbeck C., Hauska G. (1999) Sulfide-quinone reductase from *Rhodobacter capsulatus*: requirement for growth, periplasmic localization, and extension of gene sequence analysis. *J Bacteriol* 181:6516-6523.

185. Schütz M., Shahak Y., Padan E., Hauska G. (1997) Sulfide-quinone reductase from *Rhodobacter capsulatus*. Purification, cloning, and expression. *J Biol Chem* 272:9890-9894.

186. Shahak Y., Hauska G. (2008) Sulfide oxidation from cyanobacteria to humans: sulfide-quinone oxidoreductase (SQR). *Advances in Photosynthesis and Respiration*, eds Hell R., Dahl C., Knaff D. B., Leustek T. (Springer, Heidelberg, Germany), Vol 27, pp 319-335.

187. Sheldrick G. M., Hauptman H. A., Weeks C. M., Miller R., Uson I. (2001) *Ab initio* phasing. *International Tables for Crystallography*, eds Rossmann M. G., Arnold E. (Kluwer Academic Publishers, Dordrecht, The Netherlands), Vol F, pp 333-351.

188. Shevchenko A., Wilm M., Vorm O., Mann M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68:850-858.

189. Shibata H., Kobayashi S. (2006) Characterization of a HMT2-like enzyme for sulfide oxidation from *Pseudomonas putida*. *Can J Microbiol* 52:724-730.

190. Shibata H., Suzuki K., Kobayashi S. (2007) Menaquinone reduction by an HMT2-like sulfide dehydrogenase from *Bacillus stearothermophilus*. *Can J Microbiol* 53:1091-1100.

191. Sigler P. B., Blow D. M. (1965) A means of promoting heavy-atom binding in protein crystals. *J Mol Biol* 14:640-644.

192. Simanshu D. K., Savithri H. S., Murthy M. R. (2006) Crystal structures of *Salmonella typhimurium* biodegradative threonine deaminase and its complex with CMP provide structural insights into ligand-induced oligomerization and enzyme activation. *J Biol Chem* 281:39630-39641.

193. Smith H. S. (2009) Hydrogen sulfide's involvement in modulating nociception. *Pain Physician* 12:901-910.

194. Sonnhammer E. L., von Heijne G., Krogh A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6:175-182.

195. Stacey K. A. (1965) Intracellular modification of nucleic acids. *Br Med Bull* 21:211-216.

196. Stipanuk M. H., Beck P. W. (1982) Characterization of the enzymic capacity for cysteine desulphhydration in liver and kidney of the rat. *Biochem J* 206:267-277.

197. Stoops J. K., Kolodziej S. J., Schroeter J. P., Bretaudiere J. P., Wakil S. J. (1992) Structure-function relationships of the yeast fatty acid synthase: negative-stain, cryo-electron microscopy, and image analysis studies of the end views of the structure. *Proc Natl Acad Sci U S A* 89:6585-6589.

198. Swaroop M., Bradley K., Ohura T., Tahara T., Roper M. D., Rosenberg L. E., Kraus J. P. (1992) Rat cystathionine beta-synthase. Gene organization and alternative splicing. *J Biol Chem* 267:11455-11461.

199. Szabo C. (2007) Hydrogen sulphide and its therapeutic potential. *Nat Rev Drug Discov* 6:917-935.

200. Tanaka Y., Tsumoto K., Yasutake Y., Umetsu M., Yao M., Fukada H., Tanaka I., Kumagai I. (2004) How oligomerization contributes to the thermostability of an archaeon protein. Protein L-isoaspartyl-O-methyltransferase from *Sulfolobus tokodaii*. *J Biol Chem* 279:32957-32967.

201. Taylor G. (2003) The phase problem. *Acta Crystallogr D Biol Crystallogr* 59:1881-1890.

202. Taylor W. R., Orengo C. A. (1989) Protein structure alignment. *J Mol Biol* 208:1-22.

203. Theissen U., Hoffmeister M., Grieshaber M., Martin W. (2003) Single eubacterial origin of eukaryotic sulfide:quinone oxidoreductase, a mitochondrial enzyme conserved from the early evolution of eukaryotes during anoxic and sulfidic times. *Mol Biol Evol* 20:1564-1574.

204. Theissen U., Martin W. (2008a) Biochemical and evolutionary aspects of eukaryotes that inhabit sulfidic environments. *Microbial Sulfur Metabolism*, eds Dahl C., Friedrich C. G. (Springer, Heidelberg, Germany), pp 36-45.

205. Theissen U., Martin W. (2008b) Sulfide:quinone oxidoreductase (SQR) from the lugworm *Arenicola marina* shows cyanide- and thioredoxin-dependent activity. *FEBS J* 275:1131-1139.

206. Theobald D. L., Miller C. (2010) Membrane transport proteins: surprises in structural sameness. *Nat Struct Mol Biol* 17:2-3.

207. Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., Higgins D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876-4882.

208. Thompson J. D., Higgins D. G., Gibson T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.

209. Tiranti V., Viscomi C., Hildebrandt T., Di Meo I., Mineri R., Tiveron C., Levitt M. D., Prelle A., Fagiolari G., Rimoldi M., Zeviani M. (2009) Loss of ETHE1, a mitochondrial dioxygenase, causes fatal sulfide toxicity in ethylmalonic encephalopathy. *Nat Med* 15:200-205.

210. Turkenburg van Diepen M. G. W. (1996) Crystallographic studies on modified insulin. PhD (York, UK).

211. Tusnady G. E., Simon I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17:849-850.

212. Unden G., Hackenberg H., Kroger A. (1980) Isolation and functional aspects of the fumarate reductase involved in the phosphorylative electron transport of *Vibrio succinogenes*. *Biochim Biophys Acta* 591:275-288.

213. Vande Weghe J. G., Ow D. W. (1999) A fission yeast gene for mitochondrial sulfide oxidation. *J Biol Chem* 274:13250-13257.

214. Vieille C., Zeikus G. J. (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews* 65:1-43.

215. von Heijne G. (1994) Membrane proteins: from sequence to structure. *Annu Rev Biophys Biomol Struct* 23:167-192.

216. von Heijne G., Gavel Y. (1988) Topogenic signals in integral membrane proteins. *Eur J Biochem* 174:671-678.

217. Wadsten P., Wohri A. B., Snijder A., Katona G., Gardiner A. T., Cogdell R. J., Neutze R., Engstrom S. (2006) Lipidic sponge phase crystallization of membrane proteins. *J Mol Biol* 364:44-53.

218. Waight A. B., Love J., Wang D. N. (2010) Structure and mechanism of a pentameric formate channel. *Nat Struct Mol Biol* 17:31-37.

219. Wakai S., Kikumoto M., Kanao T., Kamimura K. (2004) Involvement of sulfide:quinone oxidoreductase in sulfur oxidation of an acidophilic iron-oxidizing bacterium, *Acidithiobacillus ferrooxidans* NASF-1. *Biosci Biotechnol Biochem* 68:2519-2528.

220. Wang R. (2002) Two's company, three's a crowd: can H_2S be the third endogenous gaseous transmitter? *Faseb J* 16:1792-1798.

221. Wang R. (2004) *Signal Transduction and the Gasotransmitters: NO, CO and H$_2$S in Biology and Medicine.* (Humana Press, Totowa, New Jersey, USA).

222. Waterhouse A. M., Procter J. B., Martin D. M., Clamp M., Barton G. J. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189-1191.

223. Wedemeyer U. (2007) Functional and structural membrane proteomics using *Aquifex aeolicus*. PhD thesis (Frankfurt am Main, Germany).

224. Weyand S., Shimamura T., Yajima S., Suzuki S., Mirza O., Krusong K., Carpenter E. P., Rutherford N. G., Hadden J. M., O'Reilly J., Ma P., Saidijam M., Patching S. G., Hope R. J., Norbertczak H. T., Roach P. C., Iwata S., Henderson P. J., Cameron A. D. (2008) Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* 322:709-713.

225. Weyler W., Salach J. I. (1985) Purification and properties of mitochondrial monoamine oxidase type A from human placenta. *J Biol Chem* 260:13199-13207.

226. White S. H., Wimley W. C. (1994) Peptides in lipid bilayers: structural and thermodynamic basis for partitioning and folding. *Curr Opin Struct Biol* 4:79-86.

227. Whitford D. (2005) *Proteins. Structure and function.* (John Wiley and Sons Ltd, West Sussex, UK).

228. Williams C. H. J. (1992) Lipoamide dehydrogenase, glutathione reductase, thioredoxin reductase and mercuric ion reductase: a family of flavoenzyme transhydrogenases. *Chemistry and Biochemistry of Flavoenzymes*, ed Müller F. (CRC Press, Boca Raton, USA), pp 121-211.

229. Wilson K., Mudra M., Furne J., Levitt M. (2008) Differentiation of the roles of sulfide oxidase and rhodanese in the detoxification of sulfide by the colonic mucosa. *Dig Dis Sci* 53:277-283.

230. Wittig I., Karas M., Schagger H. (2007) High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol Cell Proteomics* 6:1215-1225.

231. Wong M. W., Steudel Y., Steudel R. (2002) Novel species for the sulfur zoo: isomers of S-8. *Chemical Physics Letters* 364:387-392.

232. Yamashita A., Singh S. K., Kawate T., Jin Y., Gouaux E. (2005) Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature* 437:215-223.

233. Yang G., Wu L., Jiang B., Yang W., Qi J., Cao K., Meng Q., Mustafa A. K., Mu W., Zhang S., Snyder S. H., Wang R. (2008) H_2S as a physiologic vasorelaxant: hypertension in mice with deletion of cystathionine gamma-lyase. *Science* 322:587-590.

234. Yang W., Yang G., Jia X., Wu L., Wang R. (2005) Activation of K_{ATP} channels by H_2S in rat insulin-secreting cells and the underlying mechanisms. *J Physiol* 569:519-531.

235. Yano J. K., Poulos T. L. (2003) New understandings of thermostable and peizostable enzymes. *Curr Opin Biotechnol* 14:360-365.

236. Zhang Y., Skolnick J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302-2309.

237. Zhang Y. F., Cherney M. M., Solomonson M., Liu J. S., James M. N. G., Weiner J. H. (2009) Preliminary X-ray crystallographic analysis of sulfide: quinone oxidoreductase from *Acidithiobacillus ferrooxidans*. *Acta Crystallogr F Struct Biol Cryst Commun* 65:839-842.

238. Zhao W., Zhang J., Lu Y., Wang R. (2001) The vasorelaxant effect of H(2)S as a novel endogenous gaseous K(ATP) channel opener. *EMBO J* 20:6008-6016.

239. Zimmer J., Nam Y., Rapoport T. A. (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455:936-943.

Appendix

Appendix table 1: Flowchart for structure solution[a]

Program	Input	Output	Notes
1. Data collection			
at synchrotron	crystal	diffraction image files	data collection strategy calculated with XDS
2. Indexing, integration			
XDS	**XDS.INP**	XDS_ASCII.HKL	intensities
3. Scaling			
XDS	**XSCALE.INP**, XDS_ASCII.HKL	xscale.ahkl	intensities, scaled, merged from multiple data sets
4. Conversion to amplitudes			
XDS	**XDSCONV.INP**, xscale.ahkl	dataset.hkl	amplitudes (for SHELX)
		dataset.ccp4asc **F2MTZ.INP**	amplitudes (for CCP4)
F2MTZ (CCP4)	**XDSMTZ.COM**, F2MTZ.INP	dataset.mtz	
5. Identification of heavy atom positions			
SHELX	**SHELXC.INP**, dataset.hkl files	ha.pdb	heavy atom sites in PDB format (integral coordinates)
		ha.lst	heavy atom sites in .lst format (fractional coordinates) rename to **HA.HATOM** for SHARP
6. Refinement of heavy atom positions			
CAD (CCP4)	**CAD_ANO_XDS.COM**, dataset.mtz files	miras.data.mtz	combined amplitudes of native and derivative data sets
SHARP	**SIN**, ha.hatom, miras.data.mtz	eden-unique.mtz **HATOM.PDB**	structure factor file refined heavy atom file
7. Solvent flattening			
DM (in SHARP)	eden-unique.mtz	miras_50pc.mtz	structure factor file, solvent flattened
8. Density averaging using NCS operations derived from heavy atom positions			
	pdb file with symmetric heavy atom positions	ha_1.pdb, ha_2.pdb, ha_3.pdb	separated heavy atom groups, each belonging to a NCS region
FFT (CCP4)	**FFT.COM**, miras_50pc.mtz	miras_50pc.map	electron density map
MAPMAN	manual input, miras_50pc.map	miras_50pc.dsn6	electron density map in "O" format
MAPMAN	manual input, miras_50pc.dsn6	miras_50pc.odb	electron density bones
O	O macros (**M2**, **BN**, **BT**), miras_50pc.dsn6, miras_50pc.odb	bones_1.odb	electron density bones of region 1 (to "copy" with NCS)
MAMA	**BONES_MASK.MAMAC**, bones_1.odb	miras.mask	mask around bones of region 1

LSQ_EX (in O)	ha_1.pdb, ha_2.pdb, ha_3.pdb, miras.mask	1to2.lsq, 1to3.lsq	NCS operation matrices
IMP	**IMP.COM**, **P212121.SYM**, miras_50pc.map, miras.mask, 1to2.lsq, 1to3.lsq	**miras_1to2.lsq**, **miras_1to3.lsq**	refined NCS operation matrices
MAMA2CCP4	miras.mask	miras_ccp4.mask	mask around bones of region 1 (CCP4 format)
DM	eden-unique.mtz, miras_ccp4.mask, **DM_AVE.COM** (with NCS operators from miras_1to2.lsq, miras_1to3.lsq)	dm_miras.mtz	NCS-averaged structure factors

9. Model building

ARP/wARP	dm_miras.mtz, **SEQ.PIR**	first_model.pdb	partial model built automatically
COOT	first_model.pdb	first_model-coot.pdb	full initial model

10. Refinement

MTZ2VARIOUS	**MTZ2CNS.COM**, native.mtz	native.hkl	native amplitudes in CNS format
CNS	**CNS.COM**		
	MAKE_CV.INP, native.hkl	native.cv	native amplitudes, with R_{free} flag
	GENERATE.INP, first_model-coot.pdb	generate.pdb	coordinates of the full initial model in CNS format
	GET_NCS_MATRICES.INP, generate.pdb	**NCS.DEF**	NCS matrices from the model coordinates
	RIGID.INP, generate.pdb, ncs.def, native.cv	rigid.pdb	preliminary rigid body refinement
	MINI.INP, rigid.pdb, ncs.def, native.cv	mini.pdb	simulated annealing refinement
SKETCHER (in CCP4)		ligands.pdb, **LIBRARY.CIF**	design ligands, write library files
COOT	mini.pdb, ligands.pdb, library.cif	mini-coot.pdb	refined model with ligands, water molecules can be inserted with COOT, too
UNIQUEIFY (in CCP4)	native.mtz	native_unique.mtz	native amplitude in CCP4 format, with R_{free} flag
REFMAC5 (in CCP4)	native_unique.mtz, mini-coot.pdb, **TLSIN.TLS**, library.cif	TLSOUT.TLS, ref1.pdb	definition of parameters for TLS refinement
REFMAC5 (in CCP4)	ref1.pdb, native_unique.mtz, library.cif	final_model.pdb	successive cycles to convergence

11. Validation

COOT	final_model.pdb, native_unique.mtz	various statistics	**Ramachandran plot**, geometry analysis, peptide omega analysis, rotamer analysis, density fit analysis
SFCHECK (in CCP4)	final_model.pdb, native_unique.mtz	various statistics	**Ramachandran plot**, geometry analysis, peptide omega analysis, rotamer analysis, density fit analysis

[a] in bold the files reported in the following Appendix tables.

Appendix table 2: XDS.INP script

```
 JOB= ALL !XYCORR INIT COLSPOT IDXREF DEFPIX XPLAN INTEGRATE CORRECT
 ORGX=1535    ORGY=1488    !Detector origin (pixels). ORGX=NX/2; ORGY=NY/2
 DETECTOR_DISTANCE= 280.    !(mm)
 OSCILLATION_RANGE= 0.5          !degrees (>0)
 X-RAY_WAVELENGTH= 0.91984       !Angstroem
 MAXIMUM_NUMBER_OF_PROCESSORS=16!<25;ignored by single cpu version of xds
!MINUTE=0    !Maximum number of minutes to wait until data image must appears
!MAXIMUM_NUMBER_OF_JOBS=4
 NAME_TEMPLATE_OF_DATA_FRAMES= image_???.img DIRECT  TIFF
 DATA_RANGE=1 60     !Numbers of first and last data image collected
 BACKGROUND_RANGE= 1 9  !Numbers of first and last data image for background
 SPOT_RANGE= 1 9     !First and last data image number for finding spots
 SPACE_GROUP_NUMBER=19  !0 for unknown crystals; cell constants are ignored.
 UNIT_CELL_CONSTANTS=  112.85  154.92  178.04  90.0  90.0  90.0
!REIDX=    0  0 -1  0  0 -1  0  0 -1  0  0  0
 REFINE(IDXREF)=BEAM AXIS ORIENTATION CELL DISTANCE
 REFINE(INTEGRATE)=DISTANCE BEAM ORIENTATION CELL !AXIS
 REFINE(CORRECT)=DISTANCE BEAM ORIENTATION CELL AXIS
!FRIEDEL'S_LAW=FALSE !Default is TRUE.
!STARTING_ANGLE=  0.0       STARTING_FRAME=1
 RESOLUTION_SHELLS=10 6 5 4 3 2
!STARTING_ANGLES_OF_SPINDLE_ROTATION= 0 180 10
!TOTAL_SPINDLE_ROTATION_RANGES=30.0 120 15
!REFERENCE_DATA_SET= CK.HKL   !Name of a reference data set (optional)
 NX=3072 NY=3072 QX=0.073242  QY=0.073242   !MARCCD 225mm version
 ROTATION_AXIS= 1.0 0.0 0.0
 INCIDENT_BEAM_DIRECTION=0.0 0.0 1.0
 FRACTION_OF_POLARIZATION=0.99 ! SLS X06SA
 POLARIZATION_PLANE_NORMAL= 0.0 1.0 0.0
 AIR=0.001        !Air absorption coefficient of x-rays
 DETECTOR=CCDCHESS       MINIMUM_VALID_PIXEL_VALUE=0     OVERLOAD=65000
 DIRECTION_OF_DETECTOR_X-AXIS= 1.0 0.0 0.0
 DIRECTION_OF_DETECTOR_Y-AXIS= 0.0 1.0 0.0
 TRUSTED_REGION=0.0 1.35 !Relative radius limiting trusted region on detector
!DARK_CURRENT_IMAGE=../images/blank.tif   !hardly ever used
!TEST=1     !Test flag. 1,2 additional diagnostics and images
!INDEX_ORIGIN= 0 0 0          ! used by "IDXREF" to add an index offset
!INDEX_ERROR=0.05 INDEX_MAGNITUDE=8 INDEX_QUALITY=0.8
!SEPMIN=6.0 CLUSTER_RADIUS=3
!MAXIMUM_ERROR_OF_SPOT_POSITION=3.0
 VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS= 6000 30000 !Used by DEFPIX
 INCLUDE_RESOLUTION_RANGE=50.0 2.0 !Angstroem; used by DEFPIX,INTEGRATE,CORRECT
!EXCLUDE_RESOLUTION_RANGE= 3.93 3.87 !ice-ring at 3.897 Angstrom
!EXCLUDE_RESOLUTION_RANGE= 3.70 3.64 !ice-ring at 3.669 Angstrom
!EXCLUDE_RESOLUTION_RANGE= 3.47 3.41 !ice-ring at 3.441 Angstrom
!EXCLUDE_RESOLUTION_RANGE= 2.70 2.64 !ice-ring at 2.671 Angstrom
!EXCLUDE_RESOLUTION_RANGE= 2.28 2.22 !ice-ring at 2.249 Angstrom
!WFAC1=1.5 !This controls the number of rejected MISFITS in CORRECT;
!BEAM_DIVERGENCE=   0.80       !arctan(spot diameter/DETECTOR_DISTANCE)
!BEAM_DIVERGENCE_E.S.D.=   0.080 !half-width (Sigma) of BEAM_DIVERGENCE
!REFLECTING_RANGE=  0.780 !for crossing the Ewald sphere on shortest route
!REFLECTING_RANGE_E.S.D.=  0.113 !half-width (mosaicity) of REFLECTING_RANGE
!NUMBER_OF_PROFILE_GRID_POINTS_ALONG_ALPHA/BETA=9 !used by: INTEGRATE
!NUMBER_OF_PROFILE_GRID_POINTS_ALONG_GAMMA= 9    !used by: INTEGRATE
!CUT=2.0    !defines the integration region for profile fitting
!MINPK=75.0 !minimum required percentage of observed reflection intensity
!DELPHI= 5.0!controls the number of reference profiles and scaling factors
!PATCH_SHUTTER_PROBLEM=TRUE        !FALSE is default
!STRICT_ABSORPTION_CORRECTION=FALSE !TRUE  is default
! STRONG_PIXEL=5.0                         !used by: COLSPOT
!MAXIMUM_NUMBER_OF_STRONG_PIXELS=1500000        !used by: COLSPOT
!SPOT_MAXIMUM-CENTROID=3.0                !used by: COLSPOT
!MINIMUM_NUMBER_OF_PIXELS_IN_A_SPOT=6        !used by: COLSPOT
!NBX=3  NBY=3  !Define a rectangle of size (2*NBX+1)*(2*NBY+1)
!BACKGROUND_PIXEL=6.0                     !used by: COLSPOT,INTEGRATE
!SIGNAL_PIXEL=3.0                     !used by: INTEGRATE
```

Appendix table 3: XSCALE.INP script

```
 MAXIMUM_NUMBER_OF_PROCESSORS=16!<33;ignored by single cpu version of xscale
 RESOLUTION_SHELLS= 10 4 3.2 3.1 3.0 2.9 2.8 2.7 2.6 2.5 2.4 2.3
 SPACE_GROUP_NUMBER= 19
 UNIT_CELL_CONSTANTS=    112.85   154.92   178.040  90.000  90.000  90.000
!REIDX=-1 0 0 0     0 -1 0 0     0 0 -1 0
!0-DOSE_SIGNIFICANCE_LEVEL=0.10
 OUTPUT_FILE=xscale.ahkl  FRIEDEL'S_LAW= TRUE !FALSE !TRUE
 INPUT_FILE= ../../run2/proc2A/XDS_ASCII.HKL  XDS_ASCII 50 2.3
 INPUT_FILE= ../../run3/proc2A/XDS_ASCII.HKL  XDS_ASCII 50 2.3
 INPUT_FILE= ../../run4/proc2A/XDS_ASCII.HKL  XDS_ASCII 50 2.3
```

Appendix table 4: XDSCONV.INP script

```
For SHELX:

 SPACE_GROUP_NUMBER= 19
 UNIT_CELL_CONSTANTS=  112.85   154.92   178.040  90.000  90.000  90.000
 INPUT_FILE= xscale.ahkl  XDS_ASCII 50 2.3
 OUTPUT_FILE= native.hkl  SHELX    FRIEDEL'S_LAW=TRUE  !FALSE !TRUE
!GENERATE_FRACTION_OF_TEST_REFLECTIONS=0.05
!INHERIT_TEST_REFLECTIONS_FROM_FILE=old.hkl SHELX
!INHERIT_FRACTION_OF_TEST_REFLECTIONS=1.0 !Default is to keep them all (1.0)
 WILSON_STATISTICS=TRUE !TRUE   !Default is TRUE

For CCP4:

 SPACE_GROUP_NUMBER=19
 UNIT_CELL_CONSTANTS=  112.85   154.92   178.040  90.000  90.000  90.000
 INPUT_FILE= xscale.ahkl  XDS_ASCII 50 2.3
 OUTPUT_FILE= native.ccp4asc  CCP4    FRIEDEL'S_LAW=TRUE   !FALSE !TRUE
!GENERATE_FRACTION_OF_TEST_REFLECTIONS=0.05
!INHERIT_TEST_REFLECTIONS_FROM_FILE=old.hkl SHELX
!INHERIT_FRACTION_OF_TEST_REFLECTIONS=1.0 !Default is to keep them all (1.0)
 WILSON_STATISTICS=TRUE !TRUE   !Default is TRUE
```

Appendix table 5: F2MTZ.INP script

```
TITLE XDS to MTZ
FILE native.ccp4asc
SYMMETRY   19
CELL    112.850   154.920   178.040    90.000    90.000    90.000
LABOUT  H K L FP SIGFP
CTYPOUT H H H  F   Q
END
```

Appendix table 6: XDSMTZ.COM command file

```
f2mtz HKLOUT temp.mtz <F2MTZ.INP
cad HKLIN1 temp.mtz HKLOUT native.mtz<<EOF
LABIN FILE 1 ALL
END
EOF
```

Appendix table 7: SHELXC.INP script

```
CELL   112.30   155.86   178.26  90.000  90.000  90.000
SPAG P212121
FIND 20
NTRY 100
SFAC Os
PEAK ../run1/os_pk.hkl
INFL ../run2/os_ip.hkl
LREM ../run3/os_rm.hkl
SHEL 50 4.5
```

Appendix table 8: HA.HATOM file[a]

```
ATOM Os  0.45241 0.82864 0.22865  1.000  99.90
ATOM Os  0.84344 0.67213 0.23048  1.000  98.56
ATOM Os  0.82612 0.79963 0.17765  1.000  97.23
ATOM Os  0.59807 0.57184 0.20250  1.000  90.72
ATOM Os  0.46955 0.63644 0.28410  1.000  90.50
ATOM Os  0.62160 0.89828 0.26573  1.000  83.35
ATOM Os  0.72945 0.47730 0.26946  1.000  77.99
ATOM Os  0.27950 0.65623 0.22902  1.000  69.39
ATOM Os  0.70261 0.99218 0.17106  1.000  68.23
ATOM Os  0.91979 0.90998 0.25200  1.000  63.77
ATOM Os  0.87948 0.56358 0.14005  1.000  62.82
ATOM Os  0.31103 0.81094 0.33240  1.000  56.40
```

[a] Os sites in fractional coordinates obtained with SHELX and used as initial input in SHARP.

Appendix table 9: CAD_ANO_XDS.COM command file

```
cad    hklin1 native.mtz \
       hklin2 os_pk.mtz \
       hklin3 os_ip.mtz \
       hklin4 os_rm.mtz \
       hklin5 au_pk.mtz \
       hklout miras.data.mtz << eof-cad
TITLE  sqr
CELL    112.85   154.92   178.04  90.000  90.000  90.000
RESOLUTION OVERALL 50 2.30
SYMMETRY 19
LABIN FILE 1 E1=FP E2=SIGFP
CTYPE FILE 1 E1=F  E2=Q
LABO  FILE 1 E1=F1 E2=SF1
LABIN FILE 2 ALL
CTYPE FILE 2 E1=F  E2=Q   E3=D  E4=Q E5=Y
LABO  FILE 2 E1=F2 E2=SF2 E3=D2 E4=SD2 E5=IS2
LABIN FILE 3 ALL
CTYPE FILE 3 E1=F  E2=Q   E3=D  E4=Q E5=Y
LABO  FILE 3 E1=F3 E2=SF3 E3=D3 E4=SD3 E5=IS3
LABIN FILE 4 ALL
CTYPE FILE 4 E1=F  E2=Q   E3=D  E4=Q E5=Y
LABO  FILE 4 E1=F4 E2=SF4 E3=D4 E4=SD4 E5=IS4
LABIN FILE 5 ALL
CTYPE FILE 5 E1=F  E2=Q   E3=D  E4=Q E5=Y
LABO  FILE 5 E1=F5 E2=SF5 E3=D5 E4=SD5 E5=IS5
END
eof-cad
```

Appendix table 10: SIN, input file for SHARP

```
DATAFILES        /sharpfiles/datafiles
TITLE            nat 3A0 os 3A6 au 4A0
OBSFILE          miras.data.mtz
LEVEL            ADVANCED
MODE             REFINE RESIDUAL ELECTRON_DENSITY
CYCLES           20 21 21
REJECT           YES 5
NUM_BINS         20
STRATEGY         3 7 15 31 31 31 31 31 31 31 63 63 31 63 31 31 31 63 31 31
BOX_CONSTR_XYZ   1 0.71
BOX_CONSTR_F'    1 2.00 0.50
WEED             1 0.50 3.00
CELL             112.850 154.920 178.040 90.000 90.000 90.000
SYMMETRY         P212121
ATOMS {
 C 12173
 N 3378
 O 3730
 P 12
 S 127
}
G-SITES {
   G-SITE-01 X  0.528532   REFINE Y  0.638998   REFINE Z  0.783989   REFINE
   G-SITE-02 X  0.548631   REFINE Y  0.830452   REFINE Z  0.727468   REFINE
   G-SITE-03 X  0.155730   REFINE Y  0.670994   REFINE Z  0.731525   REFINE
   G-SITE-04 X  0.378970   REFINE Y  0.897592   REFINE Z  0.765533   REFINE
   G-SITE-05 X  0.400309   REFINE Y  0.572727   REFINE Z  0.703544   REFINE
   G-SITE-06 X  0.174353   REFINE Y  0.799422   REFINE Z  0.676523   REFINE
   G-SITE-07 X  0.270375   REFINE Y  0.476828   REFINE Z  0.768361   REFINE
   G-SITE-08 X  0.723138   REFINE Y  0.655017   REFINE Z  0.728915   REFINE
   G-SITE-09 X  0.118790   REFINE Y  0.562132   REFINE Z  0.639704   REFINE
   G-SITE-10 X  0.688070   REFINE Y  0.817195   REFINE Z  0.831721   REFINE
   G-SITE-11 X  0.301422   REFINE Y  0.995313   REFINE Z  0.672481   REFINE
   G-SITE-12 X  0.078067   REFINE Y  0.910678   REFINE Z  0.751955   REFINE
   G-SITE-13 X  0.779141   REFINE Y  0.053389   REFINE Z  0.217347   REFINE
   G-SITE-14 X  0.418918   REFINE Y  0.442703   REFINE Z  0.122543   REFINE
   G-SITE-15 X  0.182506   REFINE Y  0.976572   REFINE Z  0.162427   REFINE
   G-SITE-16 X  0.091036   REFINE Y  0.667825   REFINE Z  0.188532   REFINE
   G-SITE-17 X  0.536366   REFINE Y  0.581898   REFINE Z  0.239063   REFINE
   G-SITE-18 X  1.358789   REFINE Y  0.537464   REFINE Z  0.778046   REFINE
   G-SITE-19 X  1.181822   REFINE Y  0.602086   REFINE Z  0.837420   REFINE
   G-SITE-20 X  0.300749   REFINE Y  0.720930   REFINE Z  0.486300   REFINE
   G-SITE-21 X  0.542818   REFINE Y  0.710331   REFINE Z  0.887311   REFINE
   G-SITE-22 X  1.237249   REFINE Y  0.889942   REFINE Z  0.848405   REFINE
   G-SITE-23 X  1.336562   REFINE Y  0.632576   REFINE Z  0.968115   REFINE
   G-SITE-24 X  0.687219   REFINE Y  0.825471   REFINE Z  0.619063   REFINE
   G-SITE-25 X  0.503144   REFINE Y  0.336395   REFINE Z -0.010094   REFINE
   G-SITE-26 X  0.627612   REFINE Y  0.759746   REFINE Z  0.639095   REFINE
   G-SITE-27 X  0.322924   REFINE Y  0.581904   REFINE Z  0.593117   REFINE
   G-SITE-28 X  0.602359   REFINE Y  0.631177   REFINE Z  0.711207   REFINE
   G-SITE-29 X  0.175716   REFINE Y  0.864055   REFINE Z  0.549950   REFINE
   G-SITE-30 X  1.115541   REFINE Y  0.824034   REFINE Z  0.752826   REFINE
   G-SITE-31 X  0.270913   REFINE Y  0.868141   REFINE Z  0.590398   REFINE
   G-SITE-32 X  1.288995   REFINE Y  0.953406   REFINE Z  0.880922   REFINE
   G-SITE-33 X  0.579274   REFINE Y  0.644924   REFINE Z  0.918405   REFINE
   G-SITE-34 X  0.039610   REFINE Y  0.846836   REFINE Z  0.488607   REFINE
}
COMPOUND {
   COMPOUND-TEXT {
   }
   C-SITES {
   }
   CRYSTAL {
      CRYSTAL-TEXT {
      }
      T-SITES {
      }
      WAVELENGTH {
         WAVELENGTH-TEXT {
         }
         RESOLUTION    48.959      3.000
```

```
            BATCH {
               BATCH-TEXT {
               }
               SCAL_K              0.61227 NOREFINE
               SCAL_B             -0.00594 NOREFINE
               SCAL_B6_ADD         6.015   0.000   6.906   0.000   0.000 -12.922 NOREFINE
               NISO_BGLO           5.56207   REFINE
               NISO_CLOC           0.00000 NOREFINE
               NISO_BLOC           0.00000 NOREFINE
               NANO_BGLO           0.00000 NOREFINE
               NANO_CLOC           0.00000 NOREFINE
               NANO_BLOC           0.00000 NOREFINE
               COLUMNS H=H K=K L=L FMID=F1 SMID=SF1
            }
         }
      }
   }
}
COMPOUND {
   COMPOUND-TEXT {
   }
   C-SITES {
      C-SITE-01  G-SITE-01  Os
      C-SITE-02  G-SITE-02  Os
      C-SITE-03  G-SITE-03  Os
      C-SITE-04  G-SITE-04  Os
      C-SITE-05  G-SITE-05  Os
      C-SITE-06  G-SITE-06  Os
      C-SITE-07  G-SITE-07  Os
      C-SITE-08  G-SITE-08  Os
      C-SITE-09  G-SITE-09  Os
      C-SITE-10  G-SITE-10  Os
      C-SITE-11  G-SITE-11  Os
      C-SITE-12  G-SITE-12  Os
   }
   CRYSTAL {
      CRYSTAL-TEXT {
      }
      T-SITES {
         T-SITE-01 {
            C-SITE-01
            HAT_OCC             1.23854   REFINE
            HAT_B              38.06819   REFINE
            HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
         }
         T-SITE-02 {
            C-SITE-02
            HAT_OCC             1.33834   REFINE
            HAT_B              42.96159   REFINE
            HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
         }
         T-SITE-03 {
            C-SITE-03
            HAT_OCC             1.24144   REFINE
            HAT_B              40.09387   REFINE
            HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
         }
         T-SITE-04 {
            C-SITE-04
            HAT_OCC             1.08752   REFINE
            HAT_B              37.48332   REFINE
            HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
         }
         T-SITE-05 {
            C-SITE-05
            HAT_OCC             1.39571   REFINE
            HAT_B              70.70729   REFINE
            HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
         }
         T-SITE-06 {
            C-SITE-06
            HAT_OCC             1.73159   REFINE
            HAT_B              88.84402   REFINE
            HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
         }
```

```
          T-SITE-07 {
             C-SITE-07
             HAT_OCC             1.01912   REFINE
             HAT_B              43.04803   REFINE
             HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
          }
          T-SITE-08 {
             C-SITE-08
             HAT_OCC             1.21206   REFINE
             HAT_B              95.67684   REFINE
             HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
          }
          T-SITE-09 {
             C-SITE-09
             HAT_OCC             1.00403   REFINE
             HAT_B              58.90898   REFINE
             HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
          }
          T-SITE-10 {
             C-SITE-10
             HAT_OCC             0.90715   REFINE
             HAT_B              40.85367   REFINE
             HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
          }
          T-SITE-11 {
             C-SITE-11
             HAT_OCC             1.02623   REFINE
             HAT_B              48.74876   REFINE
             HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
          }
          T-SITE-12 {
             C-SITE-12
             HAT_OCC             0.96901   REFINE
             HAT_B              54.90962   REFINE
             HAT_B6_ADD          0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
          }
       }
       WAVELENGTH {
          WAVELENGTH-TEXT {
          }
          RESOLUTION   48.959     3.600
          BATCH {
             BATCH-TEXT {
             }
             SCAL_K              0.83847   REFINE
             SCAL_B             48.42561 NOREFINE
             SCAL_B6_ADD         8.376   0.000  -3.334   0.000   0.000  -5.042    REFINE
             NISO_BGLO           0.00000   REFINE
             NISO_CLOC           0.00824   REFINE
             NISO_BLOC           0.00000 NOREFINE
             NANO_BGLO           0.00000   REFINE
             NANO_CLOC           0.09352   REFINE
             NANO_BLOC           0.00000 NOREFINE
             HATOM_LABEL {
                Os   ATOM_f' -15.87000 NOREFINE ATOM_f"  19.90000 NOREFINE
             }
             COLUMNS H=H K=K L=L FMID=F2 SMID=SF2 DANO=D2 SANO=SD2 ISYM=IS2
          }
       }
       WAVELENGTH {
          WAVELENGTH-TEXT {
          }
          RESOLUTION   48.959     3.700
          BATCH {
             BATCH-TEXT {
             }
             SCAL_K              0.83189   REFINE
             SCAL_B             52.80913 NOREFINE
             SCAL_B6_ADD         7.836   0.000  -6.864   0.000   0.000  -0.972    REFINE
             NISO_BGLO           0.05022   REFINE
             NISO_CLOC           0.05333   REFINE
             NISO_BLOC           0.00000 NOREFINE
             NANO_BGLO           0.00000   REFINE
             NANO_CLOC           0.12022   REFINE
```

```
                    NANO_BLOC         0.00000 NOREFINE
                    HATOM_LABEL {
                        Os   ATOM_f' -20.00000 NOREFINE ATOM_f"  13.00000 NOREFINE
                    }
                    COLUMNS H=H K=K L=L FMID=F3 SMID=SF3 DANO=D3 SANO=SD3 ISYM=IS3
                }
            }
            WAVELENGTH {
                WAVELENGTH-TEXT {
                }
                RESOLUTION    48.959     3.900
                BATCH {
                    BATCH-TEXT {
                    }
                    SCAL_K             0.83411   REFINE
                    SCAL_B            51.79986 NOREFINE
                    SCAL_B6_ADD       8.794   0.000  -6.217   0.000   0.000  -2.577    REFINE
                    NISO_BGLO         0.00158   REFINE
                    NISO_CLOC         0.05253   REFINE
                    NISO_BLOC         0.00000 NOREFINE
                    NANO_BGLO         0.00000   REFINE
                    NANO_CLOC         0.14499   REFINE
                    NANO_BLOC         0.00000 NOREFINE
                    HATOM_LABEL {
                        Os   ATOM_f' -10.00000 NOREFINE ATOM_f"  10.00000 NOREFINE
                    }
                    COLUMNS H=H K=K L=L FMID=F4 SMID=SF4 DANO=D4 SANO=SD4 ISYM=IS4
                }
            }
        }
    }
}
COMPOUND {
    COMPOUND-TEXT {
    }
    C-SITES {
        C-SITE-01  G-SITE-01  Au
        C-SITE-02  G-SITE-02  Au
        C-SITE-03  G-SITE-03  Au
        C-SITE-04  G-SITE-04  Au
        C-SITE-05  G-SITE-05  Au
        C-SITE-06  G-SITE-06  Au
        C-SITE-07  G-SITE-07  Au
        C-SITE-08  G-SITE-08  Au
        C-SITE-09  G-SITE-10  Au
        C-SITE-10  G-SITE-11  Au
        C-SITE-11  G-SITE-12  Au
        C-SITE-12  G-SITE-13  Au
        C-SITE-13  G-SITE-14  Au
        C-SITE-14  G-SITE-15  Au
        C-SITE-15  G-SITE-16  Au
        C-SITE-16  G-SITE-17  Au
        C-SITE-17  G-SITE-18  Au
        C-SITE-18  G-SITE-19  Au
        C-SITE-19  G-SITE-20  Au
        C-SITE-20  G-SITE-21  Au
        C-SITE-21  G-SITE-22  Au
        C-SITE-22  G-SITE-23  Au
        C-SITE-23  G-SITE-24  Au
        C-SITE-24  G-SITE-25  Au
        C-SITE-25  G-SITE-26  Au
        C-SITE-26  G-SITE-27  Au
        C-SITE-27  G-SITE-28  Au
        C-SITE-28  G-SITE-29  Au
        C-SITE-29  G-SITE-30  Au
        C-SITE-30  G-SITE-31  Au
        C-SITE-31  G-SITE-32  Au
        C-SITE-32  G-SITE-33  Au
        C-SITE-33  G-SITE-34  Au
    }
    CRYSTAL {
        CRYSTAL-TEXT {
        }
        T-SITES {
            T-SITE-01 {
```

```
         C-SITE-01
         HAT_OCC              0.65116   REFINE
         HAT_B               40.35831 NOREFINE
         HAT_B6_ADD           2.202   6.719   3.304   0.665  -5.318  -5.507    REFINE
      }
      T-SITE-02 {
         C-SITE-02
         HAT_OCC              0.84302   REFINE
         HAT_B               45.46768 NOREFINE
         HAT_B6_ADD           2.112  -1.454   2.721   2.529  -7.678  -4.833    REFINE
      }
      T-SITE-03 {
         C-SITE-03
         HAT_OCC              0.79082   REFINE
         HAT_B               44.46211 NOREFINE
         HAT_B6_ADD           7.824  -1.415   0.907  -1.104   0.686  -8.732    REFINE
      }
      T-SITE-04 {
         C-SITE-04
         HAT_OCC              0.69823   REFINE
         HAT_B               55.23544 NOREFINE
         HAT_B6_ADD           1.836  -2.622   1.899  11.104  -1.914  -3.734    REFINE
      }
      T-SITE-05 {
         C-SITE-05
         HAT_OCC              0.82216   REFINE
         HAT_B               68.65910 NOREFINE
         HAT_B6_ADD           9.522  -4.414   2.937  -6.250  -0.563 -12.459    REFINE
      }
      T-SITE-06 {
         C-SITE-06
         HAT_OCC              0.84218   REFINE
         HAT_B               64.67327 NOREFINE
         HAT_B6_ADD           8.196   3.753  -2.825   4.039   7.907  -5.372    REFINE
      }
      T-SITE-07 {
         C-SITE-07
         HAT_OCC              0.51035   REFINE
         HAT_B               86.83897   REFINE
         HAT_B6_ADD           0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-08 {
         C-SITE-08
         HAT_OCC              0.37383   REFINE
         HAT_B               80.43972   REFINE
         HAT_B6_ADD           0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-09 {
         C-SITE-09
         HAT_OCC              0.41543   REFINE
         HAT_B               75.86452   REFINE
         HAT_B6_ADD           0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-10 {
         C-SITE-10
         HAT_OCC              0.40864   REFINE
         HAT_B               63.16261   REFINE
         HAT_B6_ADD           0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-11 {
         C-SITE-11
         HAT_OCC              0.48451   REFINE
         HAT_B               88.14870   REFINE
         HAT_B6_ADD           0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-12 {
         C-SITE-12
         HAT_OCC              0.63079   REFINE
         HAT_B               37.99104   REFINE
         HAT_B6_ADD           0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-13 {
         C-SITE-13
         HAT_OCC              0.69017   REFINE
```

```
       HAT_B           59.56096   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-14 {
       C-SITE-14
       HAT_OCC          0.72704   REFINE
       HAT_B           49.73992   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-15 {
       C-SITE-15
       HAT_OCC          0.08550   REFINE
       HAT_B            1.00000 NOREFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000   REFINE
    }
    T-SITE-16 {
       C-SITE-16
       HAT_OCC          0.56739   REFINE
       HAT_B           56.43816   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-17 {
       C-SITE-17
       HAT_OCC          0.11234   REFINE
       HAT_B            1.00000   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-18 {
       C-SITE-18
       HAT_OCC          0.53362   REFINE
       HAT_B           61.42216   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-19 {
       C-SITE-19
       HAT_OCC          0.38506   REFINE
       HAT_B           41.58083   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-20 {
       C-SITE-20
       HAT_OCC          0.52782   REFINE
       HAT_B           53.53659   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-21 {
       C-SITE-21
       HAT_OCC          0.52708   REFINE
       HAT_B           62.48635   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-22 {
       C-SITE-22
       HAT_OCC          0.37368   REFINE
       HAT_B           37.04106   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-23 {
       C-SITE-23
       HAT_OCC          0.33543   REFINE
       HAT_B           26.10094   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-24 {
       C-SITE-24
       HAT_OCC          0.34827   REFINE
       HAT_B           32.66870   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
    }
    T-SITE-25 {
       C-SITE-25
       HAT_OCC          0.51300   REFINE
       HAT_B           59.52833   REFINE
       HAT_B6_ADD       0.000  0.000  0.000  0.000  0.000  0.000 NOREFINE
```

```
         }
      T-SITE-26 {
         C-SITE-26
         HAT_OCC            0.62263   REFINE
         HAT_B            77.23315   REFINE
         HAT_B6_ADD        0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-27 {
         C-SITE-27
         HAT_OCC            0.43867   REFINE
         HAT_B            43.64398   REFINE
         HAT_B6_ADD        0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-28 {
         C-SITE-28
         HAT_OCC            0.17985   REFINE
         HAT_B             4.30505   REFINE
         HAT_B6_ADD        0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-29 {
         C-SITE-29
         HAT_OCC            0.31355   REFINE
         HAT_B            19.09131   REFINE
         HAT_B6_ADD        0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-30 {
         C-SITE-30
         HAT_OCC            0.40023   REFINE
         HAT_B            32.90198   REFINE
         HAT_B6_ADD        0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-31 {
         C-SITE-31
         HAT_OCC            0.35576   REFINE
         HAT_B            29.47239   REFINE
         HAT_B6_ADD        0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-32 {
         C-SITE-32
         HAT_OCC            0.02042   REFINE
         HAT_B             1.00000   REFINE
         HAT_B6_ADD        0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
      T-SITE-33 {
         C-SITE-33
         HAT_OCC            0.01762   REFINE
         HAT_B             1.00000   REFINE
         HAT_B6_ADD        0.000   0.000   0.000   0.000   0.000   0.000 NOREFINE
      }
   }
   WAVELENGTH {
      WAVELENGTH-TEXT {
      }
      RESOLUTION    50.000      4.000
      BATCH {
         BATCH-TEXT {
         }
         SCAL_K            0.82879   REFINE
         SCAL_B           15.61462 NOREFINE
         SCAL_B6_ADD       1.939   0.000  -0.151   0.000   0.000  -1.788    REFINE
         NISO_BGLO         1.27865   REFINE
         NISO_CLOC         0.50367   REFINE
         NISO_BLOC         0.00000 NOREFINE
         NANO_BGLO         0.00000   REFINE
         NANO_CLOC         0.09530   REFINE
         NANO_BLOC         0.00000 NOREFINE
         HATOM_LABEL {
            Au   ATOM_f' -22.50000    REFINE ATOM_f"   6.69440    REFINE
         }
         COLUMNS H=H K=K L=L FMID=F5 SMID=SF5 DANO=D5 SANO=SD5 ISYM=IS5
      }
END
```

Appendix table 11: HATOM.PDB[a]

```
CRYST1  112.850  154.920  178.040  90.00  90.00  90.00 P212121
SCALE1      0.008861  0.000000  0.000000        0.00000
SCALE2      0.000000  0.006455  0.000000        0.00000
SCALE3      0.000000  0.000000  0.005617        0.00000
ATOM      1 OS   HET 1    1     59.645  98.994 139.581  1.24 38.07
ATOM      2 OS   HET 1    2     61.913 128.654 129.518  1.34 42.96
ATOM      3 OS   HET 1    3     17.574 103.950 130.241  1.24 40.09
ATOM      4 OS   HET 1    4     42.767 139.055 136.295  1.09 37.48
ATOM      5 OS   HET 1    5     45.175  88.727 125.259  1.40 70.71
ATOM      6 OS   HET 1    6     19.676 123.847 120.448  1.73 88.84
ATOM      7 OS   HET 1    7     30.512  73.870 136.799  1.02 43.05
ATOM      8 OS   HET 1    8     81.606 101.475 129.776  1.21 95.68
ATOM      9 OS   HET 1    9     13.405  87.085 113.893  1.00 58.91
ATOM     10 OS   HET 1   10     77.649 126.600 148.080  0.91 40.85
ATOM     11 OS   HET 1   11     34.015 154.194 119.729  1.03 48.75
ATOM     12 OS   HET 1   12      8.810 141.082 133.878  0.97 54.91
ATOM     13 AU   HET 2    1     59.645  98.994 139.581  0.65 40.36
ATOM     14 AU   HET 2    2     61.913 128.654 129.518  0.84 45.47
ATOM     15 AU   HET 2    3     17.574 103.950 130.241  0.79 44.46
ATOM     16 AU   HET 2    4     42.767 139.055 136.295  0.70 55.24
ATOM     17 AU   HET 2    5     45.175  88.727 125.259  0.82 68.66
ATOM     18 AU   HET 2    6     19.676 123.847 120.448  0.84 64.67
ATOM     19 AU   HET 2    7     30.512  73.870 136.799  0.51 86.84
ATOM     20 AU   HET 2    8     81.606 101.475 129.776  0.37 80.44
ATOM     21 AU   HET 2   10     77.649 126.600 148.080  0.42 75.86
ATOM     22 AU   HET 2   11     34.015 154.194 119.729  0.41 63.16
ATOM     23 AU   HET 2   12      8.810 141.082 133.878  0.48 88.15
ATOM     24 AU   HET 2   13     87.926   8.271  38.697  0.63 37.99
ATOM     25 AU   HET 2   14     47.275  68.584  21.818  0.69 59.56
ATOM     26 AU   HET 2   15     20.596 151.291  28.918  0.73 49.74
ATOM     27 AU   HET 2   16     10.273 103.459  33.566  0.09  1.00
ATOM     28 AU   HET 2   17     60.529  90.148  42.563  0.57 56.44
ATOM     29 AU   HET 2   18    153.339  83.264 138.523  0.11  1.00
ATOM     30 AU   HET 2   19    133.369  93.275 149.094  0.53 61.42
ATOM     31 AU   HET 2   20     33.939 111.686  86.581  0.39 41.58
ATOM     32 AU   HET 2   21     61.257 110.044 157.977  0.53 53.54
ATOM     33 AU   HET 2   22    139.624 137.870 151.050  0.53 62.49
ATOM     34 AU   HET 2   23    150.831  97.999 172.363  0.37 37.04
ATOM     35 AU   HET 2   24     77.553 127.882 110.218  0.34 26.10
ATOM     36 AU   HET 2   25     56.780  52.114  -1.797  0.35 32.67
ATOM     37 AU   HET 2   26     70.826 117.700 113.785  0.51 59.53
ATOM     38 AU   HET 2   27     36.442  90.149 105.599  0.62 77.23
ATOM     39 AU   HET 2   28     67.976  97.782 126.623  0.44 43.64
ATOM     40 AU   HET 2   29     19.830 133.859  97.913  0.18  4.31
ATOM     41 AU   HET 2   30    125.889 127.659 134.033  0.31 19.09
ATOM     42 AU   HET 2   31     30.573 134.492 105.115  0.40 32.90
ATOM     43 AU   HET 2   32    145.463 147.702 156.839  0.36 29.47
ATOM     44 AU   HET 2   33     65.371  99.912 163.513  0.02  1.00
ATOM     45 AU   HET 2   34      4.470 131.192  86.992  0.02  1.00
```

[a] refined heavy atom positions used for initial phasing by SHARP.

Appendix table 12: input parameters for solvent flattening with DM in SHARP

```
Directory ............................. /sharpfiles/logfiles/miras.1
Date .................................. Thu Mar  6 09:53:09 CET 2008
Data .................................. eden-unique.mtz
FP, SIGFP ............................. FP, SIGFP ('native' amplitude and sigma)
FB, PHIB .............................. Fcent, PHIcent (for initial map)
HLA, HLB, HLC, HLD .................... HLA, HLB, HLC, HLD (for phase combination)
FH, PHIH .............................. FH, PHIH (heavy atom contribution)
Spacegroup ............................ P212121 (19)
Cell .................................. 112.850 154.920 178.040 90.000 90.000 90.000
Resolution limits ..................... 50.000 - 2.65 &Aring;
Solvent fraction ...................... 0.50
Flipping factor ....................... -1.0000
Radius of solvent sphere .............. 5 - 2.65 &Aring;
Protein to solvent density ratio ...... 1.34375
Truncating min / max density in protein . 0.4 0.99
no. of cycles ......................... 5 + 90 + 5
```

Appendix table 13: FFT.COM command file

```
fft hklin miras_50pc.mtz  mapout   miras_50pc.map << EOF
TITLE  sqr
GRID  112 156 180
RESOLUTION 50. 2.65
SYMMETRY 19
FFTSYMMETRY 1
XYZLIM 0 111 0 155 0 179
RHOLIM 100.
LABIN F1=FPsha PHI=PHIshasol W=FOMshasol
EOF
#
extend   mapin   miras_50pc.map  mapout  miras_50pc.exmap     << eof-extend
XYZLIM  5 68 20 109 25 95
SYMMETRY 19
eof-extend
```

Appendix table 14: O macro files

```
@M2

!contour a map at the current active center
 centre_id wait_id
 paint_case atom_name 4 c* n* o* s*  yellow blue red magenta
!mol sirm1 obj sph spher 40 end
 map_file ../dm/ miras_50pc.dsn6
 map_obj mp50
 map_par 30. 30. 30. 1.2 slate_blue 0.5 0.1 0  !11.
 map_act
 map_draw
 message ' '

##############################################################

@BN

bone_redefine
wait_id wait_id 5
5
message ' '

##############################################################

@BT

bone_redefine
wait_id wait_id 3
message ' '
```

Appendix table 15: BONES_MASK.MAMAC command file

```
new cell  112.8500  154.9200  178.0400   90.0000   90.0000   90.0000
new grid 112 156 180
new radius 5
new bones bonesmask
expand bonesmask
fill bonesmask
contract bonesmask
island bonesmask
WRite bonesmask miras.mask
```

Appendix table 16: NCS operations[a]

```
miras_1to2.lsq

.LSQ_RT_12                 R          12 (3(1x,f12.7))
   -0.5336170   -0.8399070    0.0990430
   -0.8402240    0.5131610   -0.1751830
    0.0963120   -0.1766990   -0.9795420
  146.5375820  113.2809370  273.5504760


#######################################

miras_1to3.lsq

.LSQ_RT_13                 R          12 (3(1x,f12.7))
    0.9082180    0.0026710    0.4184890
    0.0019220   -0.9999960    0.0022100
    0.4184930   -0.0012030   -0.9082190
  -50.7582440  227.9414830  231.1131440
```

[a] determined with LSQ_EX (in O) and IMP from the heavy metal binding positions.

Appendix table 17: IMP.COM command file

```
virus_imp << EOF
../maps/2402/miras_50pc.map
miras.mask
symm/p212121.sym
miras_1to2.lsq     # miras_1to3.lsq
A
EOF
```

Appendix table 18: P212121.SYM, symmetry operation file for space group $P2_12_12_1$[a]

```
.space_group_operators    r    48  (3F10.5)
   1.00000    0.00000    0.00000
   0.00000    1.00000    0.00000
   0.00000    0.00000    1.00000
   0.00000    0.00000    0.00000
  -1.00000    0.00000    0.00000
   0.00000   -1.00000    0.00000
   0.00000    0.00000    1.00000
   0.50000    0.00000    0.50000
   1.00000    0.00000    0.00000
   0.00000   -1.00000    0.00000
   0.00000    0.00000   -1.00000
   0.50000    0.50000    0.00000
  -1.00000    0.00000    0.00000
   0.00000    1.00000    0.00000
   0.00000    0.00000   -1.00000
   0.00000    0.50000    0.50000
```

[a] to be used in IMP.

Appendix table 19: DM_AVE.COM command file

```
dm                                    \
        hklin eden-unique.mtz \
        hklout dm_miras.mtz  \
        ncsin1 miras_ccp4.mask \
        histlib /soft/dec/xtal/ccp4/v4.2.1/lib/data/hist.lib            \
        << 'my-data'
SOLC 0.5
MODE SOLV HIST AVER
COMB PERT
NCYC 70
COMBINE PERT
SCHEME RES FROM 5.0
RESOLUTION 50. 2.5
AVER DOMAIN 1 REFI
OMAT
    1.00000    0.00000    0.00000
    0.00000    1.00000    0.00000
    0.00000    0.00000    1.00000
    0.00000    0.00000    0.00000
AVER DOMAIN 1 REFI
OMAT
   -0.5336170   -0.8399070    0.0990430
   -0.8402240    0.5131610   -0.1751830
    0.0963120   -0.1766990   -0.9795420
  146.5375820  113.2809370  273.5504760
AVER DOMAIN 1 REFI
OMAT
    0.9082180    0.0026710    0.4184890
    0.0019220   -0.9999960    0.0022100
    0.4184930   -0.0012030   -0.9082190
  -50.7582440  227.9414830  231.1131440
LABI FP=FP SIGFP=SIGFP PHIO=PHIB FOMO=FOM -
    HLA=HLA HLB=HLB HLC=HLC FDM=FB PHIDM=PHIB
LABO PHIDM=PHIDM FOMDM=WDM -
    HLADM=HLADM HLBDM=HLBDM HLCDM=HLCDM
'my-data'
```

Appendix table 20: SEQ.PIR file[a]

```
> sqr.pir

MAKHVVVIGGGVGGIATAYNLRNLMPDLKITLISDRPYFGFTPAFPHLAMGWRKFEDISVPLAPLLPKFN
IEFINEKAESIDPDANTVTTQSGKKIEYDYLVIATGPKLVFGAEGQEENSTSICTAEHALETQKKLQELY
ANPGPVVIGAIPGVSCFGPAYEFALMLHYELKKRGIRYKVPMTFITSEPYLGHFGVGGIGASKRLVEDLF
AERNIDWIANVAVKAIEPDKVIYEDLNGNTHEVPAKFTMFMPSFQGPEVVASAGDKVANPANKMVIVNRC
FQNPTYKNIFGVGVVTAIPPIEKTPIPTGVPKTGMMIEQMAMAVAHNIVNDIRNNPDKYAPRLSAICIAD
FGEDAGFFFADPVIPPRERVITKMGKWAHYFKTAFEKYFLWKVRNGNIAPSFEEKVLEIFLKVHPIELCK
DCEGAPGSRC
```

[a] sequence of *A. aeolicus* SQR in PIR format, for ARP/wARP.

Appendix table 21: MTZ2CNS.COM command file

```
mtz2various HKLIN native.mtz HKLOUT native.hkl << EOF
RESOLUTION 10000 2.3
OUTPUT CNS
LABIN FP=FP SIGFP=SIGFP
END
EOF
```

Appendix table 22: CNS.COM command file

```
 /soft/.../cns_solve-0208231106.exe  < mini.inp              > mini.log
#/soft/.../cns_solve-0208231106.exe  < generate.inp          > generate.log
#/soft/.../cns_solve-0208231106.exe  < rigid.inp             > rigid.log
#/soft/.../cns_solve-0208231106.exe  < make_cv.inp           > make_cv.log
#/soft/.../cns_solve-0208231106.exe  < get_ncs_matrices.inp  > get_ncs_matrices.log
```

Appendix table 23: MAKE_CV.INP script

```
{===================== crystallographic data =======================}
{===>} sg="P2(1)2(1)2(1)";
{===>} a=112.85;
{===>} b=154.92;
{===>} c=178.04;
{===>} alpha=90.0;
{===>} beta=90.0;
{===>} gamma=90.0;
{===>} test_set="test";
{===>} test_flag=1;
{===>} reflection_infile="native.hkl";
{===>} percentage=5.0;
{===>} low_res=500.0;
{===>} high_res=2.3;
{===>} ref_sel=(all);
{===>} reflection_outfile="native.cv";
{===========================================================================}
{        things below this line do not normally need to be changed          }
{===========================================================================}
...
```

Appendix table 24: GENERATE.INP script

```
{============================= important ================================}
{* Coordinates for molecules of the same type (eg. all protein, all
   nucleic acid etc) can be input in the same coordinate file if the
   different chains are separated by a TER card or each chain has
   a different segid or chainid. *}
{* A break in a chain can be detected automatically or should be delimited
   by a BREAK card. In this case no patch (head, tail or link) will be
   applied between the residues that bound the chain break. *}
{* If a segid is present in the coordinate file it will be read unless
   segid renaming is used below. If renaming is used then all chains in a
   coordinate file will be given the same segid. *}
{* If a PDB chain identifier is present in the coordinate file then this
   can be used for the segid *}
{* NB. All input PDB files must finish with an END statement *}
{========================== protein files ================================}
{===>} prot_coordinate_infile_1="initial_model.pdb";
{===>} prot_rename_1=false;
{===>} prot_segid_1="";
{===>} prot_convert_1=false;
{===>} prot_separate_1=true;
{======================= nucleic acid files =============================}
{===>} nucl_coordinate_infile_1="";
{===>} nucl_rename_1=false;
{===>} nucl_segid_1="";
{===>} nucl_convert_1=false;
{===>} nucl_separate_1=true;
{========================== water files ================================}
{===>} water_coordinate_infile_1="";
{===>} water_rename_1=false;
{===>} water_segid_1="";
{===>} water_convert_1=false;
{======================= carbohydrate files =============================}
{===>} carbo_coordinate_infile_1="";
{===>} carbo_rename_1=false;
{===>} carbo_segid_1="";
{===>} carbo_convert_1=false;
```

```
{======================== prosthetic group files ==========================}
{===>} prost_coordinate_infile_1="";
{===>} prost_rename_1=false;
{===>} prost_segid_1="";
{===>} prost_convert_1=false;
{=========================== ligand files =================================}
{===>} lig_coordinate_infile_1="";
{===>} lig_rename_1=false;
{===>} lig_segid_1="";
{===>} lig_convert_1=false;
{============================= ions files =================================}
{===>} ion_coordinate_infile_1="";
{===>} ion_rename_1=false;
{===>} ion_segid_1="";
{===>} ion_convert_1=false;
{========================= renaming atoms =================================}
{===>} ile_CD_becomes="CD1";
{===>} OT1_becomes="O";
{===>} OT2_becomes="OXT";
{====================== automatic mainchain breaks =======================}
{===>} auto_break=true;
{===>} break_cutoff=2.5;
{===>} prot_break_infile="CNS_TOPPAR:protein_break.top";
{====================== automatic disulphide bonds =======================}
{===>} auto_ss=true;
{===>} disulphide_dist=3.0;
{======================= manual disulphide bonds =========================}
{===>} ss_use_1=false;
{===>} ss_i_segid_1=""; ss_i_resid_1=0;
{===>} ss_j_segid_1=""; ss_j_resid_1=0;
{===>} ss_use_2=false;
{===>} ss_i_segid_2=""; ss_i_resid_2=0;
{===>} ss_j_segid_2=""; ss_j_resid_2=0;
{======================= RNA to DNA conversion  ==========================}
{===>} dna_sele=(none);
{========================= carbohydrate links  ===========================}
{===>} carbo_use_1=false;
{===>} carbo_patch_1="B1N";
{===>} carbo_i_segid_1="BBBB"; carbo_i_resid_1=401;
{===>} carbo_j_segid_1="AAAA"; carbo_j_resid_1=56;
{===>} carbo_use_2=false;
{===>} carbo_patch_2="B1N";
{===>} carbo_i_segid_2="BBBB"; carbo_i_resid_2=402;
{===>} carbo_j_segid_2="AAAA"; carbo_j_resid_2=182;
{===>} carbo_use_3=false;
{===>} carbo_patch_3="";
{===>} carbo_i_segid_3=""; carbo_i_resid_3=0;
{===>} carbo_j_segid_3=""; carbo_j_resid_3=0;
{===>} carbo_use_4=false;
{===>} carbo_patch_4="";
{===>} carbo_i_segid_4=""; carbo_i_resid_4=0;
{===>} carbo_j_segid_4=""; carbo_j_resid_4=0;
{===>} carbo_use_5=false;
{===>} carbo_patch_5="";
{===>} carbo_i_segid_5=""; carbo_i_resid_5=0;
{===>} carbo_j_segid_5=""; carbo_j_resid_5=0;
{===>} carbo_use_6=false;
{===>} carbo_patch_6="";
{===>} carbo_i_segid_6=""; carbo_i_resid_6=0;
{===>} carbo_j_segid_6=""; carbo_j_resid_6=0;
{======================= generate parameters =============================}
{===>} hydrogen_flag=false;
{===>} hydrogen_build="all";
{===>} atom_build=(not(known));
{===>} atom_delete=(none);
{===>} set_bfactor=false;
{===>} bfactor=15.0;
{===>} set_occupancy=false;
{===>} occupancy=1.0;
{============================ output files ===============================}
{===>} structure_outfile="generate.mtf";
{===>} coordinate_outfile="generate.pdb";
{===>} pdb_o_format=true;
{================== protein topology and parameter files ==================}
```

```
{===>} prot_topology_infile="CNS_TOPPAR:protein.top";
{===>} prot_link_infile="CNS_TOPPAR:protein.link";
{===>} prot_parameter_infile="CNS_TOPPAR:protein_rep.param";
{================ nucleic acid topology and parameter files ================}
{===>} nucl_topology_infile="CNS_TOPPAR:dna-rna.top";
{===>} nucl_link_infile="CNS_TOPPAR:dna-rna.link";
{===>} nucl_parameter_infile="CNS_TOPPAR:dna-rna_rep.param";
{=================== water topology and parameter files ===================}
{===>} water_topology_infile="CNS_TOPPAR:water.top";
{===>} water_parameter_infile="CNS_TOPPAR:water_rep.param";
{================ carbohydrate topology and parameter files ==============}
{===>} carbo_topology_infile="CNS_TOPPAR:carbohydrate.top";
{===>} carbo_parameter_infile="CNS_TOPPAR:carbohydrate.param";
{============= prosthetic group topology and parameter files =============}
{===>} prost_topology_infile="";
{===>} prost_parameter_infile="";
{================== ligand topology and parameter files ==================}
{===>} lig_topology_infile="";
{===>} lig_parameter_infile="";
{==================== ion topology and parameter files ===================}
{===>} ion_topology_infile="CNS_TOPPAR:ion.top";
{===>} ion_parameter_infile="CNS_TOPPAR:ion.param";
{========================================================================}
{         things below this line do not need to be changed unless        }
{         you need to apply patches - at the appropriate places marked    }
{========================================================================}
...
```

Appendix table 25: GET_NCS_MATRICES.INP script

```
{====================== molecular structures ======================}
{===>} structure_infile="generate.mtf";
{===>} coordinate_infile="generate.pdb";
{======================= atom selection ==========================}
{===>} atom_ref=(resid 4:418 and name CA);
{===>} atom_ncs_1=(resid 1004:1418 and name CA);
{===>} atom_ncs_2=(resid 2004:2418 and name CA);
{===>} atom_ncs_3=(resid 3004:3418 and name CA);
{===>} atom_ncs_4=(resid 4004:4418 and name CA);
{===>} atom_ncs_5=(resid 5004:5418 and name CA);
{===>} ignore_name=false;
{========================= output files ==========================}
{===>} list_outfile="get_ncs_matrices.list";
{========================================================================}
{         things below this line do not normally need to be changed      }
{========================================================================}
...
```

Appendix table 26: NCS.DEF script

```
{=========================== NCS type ============================}
{===>} ncs_type="restrain";
{==================== strict NCS (Skew matrix) ===================}
{===>} skew_use=false;
{===>} skew_matrix=( 1 0 0 )
                    ( 0 1 0 )
                    ( 0 0 1 );
{===>} skew_vector=(0 0 0);
{=============== strict NCS (X-ray and nonbonded terms) =============}
{===>} ncs_op_1=true;
{===>} ncs_matrix_1=( 1 0 0 )
                     ( 0 1 0 )
                     ( 0 0 1 );
{===>} ncs_vector_1=(0 0 0);
{===>} ncs_op_2=false;
{===>} ncs_matrix_2=( -0.999932 -0.011239  0.003138 )
                     (  0.010826 -0.793178  0.608894 )
                     ( -0.004354  0.608886  0.793246 );
{===>} ncs_vector_2=(174.4428 -1.5140 0.7045);
{===>} ncs_op_3=false;
```

```
{===>} ncs_matrix_3=( 1 0 0 )
               ( 0 1 0 )
               ( 0 0 1 );
{===>} ncs_vector_3=(0 0 0);
{===>} ncs_op_4=false;
{===>} ncs_matrix_4=( 1 0 0 )
               ( 0 1 0 )
               ( 0 0 1 );
{===>} ncs_vector_4=(0 0 0);
{===>} ncs_op_5=false;
{===>} ncs_matrix_5=( 1 0 0 )
               ( 0 1 0 )
               ( 0 0 1 );
{===>} ncs_vector_5=(0 0 0);
{===>} ncs_op_6=false;
{===>} ncs_matrix_6=( 1 0 0 )
               ( 0 1 0 )
               ( 0 0 1 );
{===>} ncs_vector_6=(0 0 0);
{===>} ncs_op_7=false;
{===>} ncs_matrix_7=( 1 0 0 )
               ( 0 1 0 )
               ( 0 0 1 );
{===>} ncs_vector_7=(0 0 0);
{================= strict NCS (nonbonded term only) =================}
{===>} nb_ncs_op_1=false;
{===>} nb_ncs_matrix_1=( 1 0 0 )
                  ( 0 1 0 )
                  ( 0 0 1 );
{===>} nb_ncs_vector_1=(0 0 0);
{===>} nb_ncs_op_2=false;
{===>} nb_ncs_matrix_2=( 1 0 0 )
                  ( 0 1 0 )
                  ( 0 0 1 );
{===>} nb_ncs_vector_2=(0 0 0);
{===>} nb_ncs_op_3=false;
{===>} nb_ncs_matrix_3=( 1 0 0 )
                  ( 0 1 0 )
                  ( 0 0 1 );
{===>} nb_ncs_vector_3=(0 0 0);
{===>} nb_ncs_op_4=false;
{===>} nb_ncs_matrix_4=( 1 0 0 )
                  ( 0 1 0 )
                  ( 0 0 1 );
{===>} nb_ncs_vector_4=(0 0 0);
{===>} nb_ncs_op_5=false;
{===>} nb_ncs_matrix_5=( 1 0 0 )
                  ( 0 1 0 )
                  ( 0 0 1 );
{===>} nb_ncs_vector_5=(0 0 0);
{===================== NCS restraint group 1 =======================}
{===>} group.1.weight=300;
{===>} group.1.sigb=2;
{===>} group.1.equiv.1=(resid 4:418 and name CA);
{===>} group.1.equiv.2=(resid 1004:1418 and name CA);
{===>} group.1.equiv.3=(resid 2004:2418 and name CA);
{===>} group.1.equiv.4=(resid 3004:3418 and name CA);
{===>} group.1.equiv.5=(resid 4004:4418 and name CA);
{===>} group.1.equiv.6=(resid 5004:5418 and name CA);
{===================== NCS restraint group 2 =======================}
{===>} group.2.weight=300;
{===>} group.2.sigb=2;
{===>} group.2.equiv.1=(none);
{===>} group.2.equiv.2=(none);
{===>} group.2.equiv.3=(none);
{===>} group.2.equiv.4=(none);
{===>} group.2.equiv.5=(none);
{===>} group.2.equiv.6=(none);
{==================================================================}
{       things below this line do not normally need to be changed       }
{==================================================================}
...
```

Appendix table 27: RIGID.INP script

```
{======================= molecular structure =========================}
{===>} structure_infile="generate.mtf";
{===>} parameter_infile_1="CNS_TOPPAR:protein_rep.param";
{===>} parameter_infile_2="";
{===>} parameter_infile_3="";
{===>} parameter_infile_4="";
{===>} parameter_infile_5="";
{===>} coordinate_infile="generate.pdb";
{====================== crystallographic data =======================}
{===>} sg="P2(1)2(1)2(1)";
{===>} a=112.85;
{===>} b=154.92;
{===>} c=178.04;
{===>} alpha=90;
{===>} beta=90;
{===>} gamma=90;
{===>} anom_library="";
{===>} reflection_infile_1="native.cv";
{===>} reflection_infile_2="";
{===>} reflection_infile_3="";
{===>} obs_f="fobs";
{===>} obs_sigf="sigma";
{===>} test_set="test";
{===>} test_flag=1;
{===>} obs_w="";
{===>} obs_i="";
{===>} obs_sigi="";
{===>} obs_pa="";
{===>} obs_pb="";
{===>} obs_pc="";
{===>} obs_pd="";
{===>} obs_phase="";
{===>} obs_fom="";
{===>} low_res=50.0;
{===>} high_res=4.0;
{===>} obs_type="amplitude";
{===>} sigma_cut=0.0;
{===>} obs_rms=10000;
{==================== non-crystallographic symmetry ==================}
{===>} ncs_infile="ncs.def";
{============ initial B-factor and bulk solvent corrections =========}
{===>} bscale="anisotropic";
{===>} low_res_bscale=6.0;
{===>} bulk_sol=true;
{===>} bulk_mask_infile="";
{===>} sol_k=-1;
{===>} sol_b=-1;
{======================== atom selection ==========================}
{===>} atom_select=(resid 4:5418);
{===>} r_group_1=(resid 4:418);
{===>} r_group_2=(resid 1004:1418);
{===>} r_group_3=(resid 2004:2418);
{===>} r_group_4=(resid 3004:3418);
{===>} r_group_5=(resid 4004:4418);
{===>} r_group_6=(resid 5004:5418);
{================ rigid body minimization parameters ================}
{===>} rigid_nstep=20;
{===>} num_cycles=1;
{===>} reftarget="mlf";
{===>} target_bins=-1;
{===>} fft_memory=-1;
{========================= output files ===========================}
{===>} coordinate_outfile="rigid.pdb";
{===>} pdb_o_format=true;
{==================================================================}
{         things below this line do not normally need to be changed         }
{         unless more than one group is required                            }
{==================================================================}
...
```

Appendix table 28: MINI.INP script

```
{======================= molecular structure ========================}
{===>} structure_infile="generate.mtf";
{===>} coordinate_infile="rigid.pdb";
{===>} topology_infile_1="CNS_TOPPAR:protein.top";
{===>} topology_infile_2="";
{===>} topology_infile_3="CNS_TOPPAR:water.top";
{===>} topology_infile_4="CNS_TOPPAR:ion.top";
{===>} topology_infile_5="";
{===>} prot_link_infile="CNS_TOPPAR:protein.link";
{===>} nucl_link_infile="CNS_TOPPAR:dna-rna.link";
{===>} parameter_infile_1="CNS_TOPPAR:protein_rep.param";
{===>} parameter_infile_2="";
{===>} parameter_infile_3="CNS_TOPPAR:water_rep.param";
{===>} parameter_infile_4="CNS_TOPPAR:ion.param";
{===>} parameter_infile_5="";
{===>} break_cutoff=2.5;
{===>} prot_break_infile="CNS_TOPPAR:protein_break.top";
{===>} disulphide_dist=3.0;
{===>} atom_delete=(none);
{========================== renaming atoms ==============================}
{===>} ile_CD_becomes="CD1";
{===>} OT1_becomes="O";
{===>} OT2_becomes="OXT";
{======================= crystallographic data ========================}
{===>} sg="P2(1)2(1)2(1)";
{===>} a=112.85;
{===>} b=154.92;
{===>} c=178.04;
{===>} alpha=90.;
{===>} beta=90.;
{===>} gamma=90.;
{===>} reflection_infile_1="native.cv";
{===>} reflection_infile_2="";
{===>} reflection_infile_3="";
{===>} anom_library="";
{===>} obs_f="fobs";
{===>} obs_sigf="sigma";
{===>} test_set="test";
{===>} test_flag=1;
{===>} obs_i="";
{===>} obs_sigi="";
{===>} obs_pa="";
{===>} obs_pb="";
{===>} obs_pc="";
{===>} obs_pd="";
{===>} obs_w="";
{===>} obs_phase="";
{===>} obs_fom="";
{===>} low_res=50;
{===>} high_res=2.30;
{===>} obs_type="amplitude";
{===>} sigma_cut=0.0;
{===>} obs_rms=10000;
{=================== non-crystallographic symmetry ==================}
{===>} ncs_infile="ncs.def";
{============ initial B-factor and bulk solvent corrections =========}
{===>} bscale="anisotropic";
{===>} low_res_bscale=6.0;
{===>} bulk_sol=true;
{===>} bulk_mask_infile="";
{===>} sol_k=-1;
{===>} sol_b=-1;
{======================= atom selection ========================}
{===>} atom_select=(known and not hydrogen);
{===>} atom_fixed=(none);
{===>} atom_harm=(none);
{===>} k_harmonic=10;
{===>} atom_rigid=(none);
{===>} atom_main=(name ca or name n or name c or name o or name ot+);
{===>} conf_1=(none);
{===>} conf_2=(none);
```

```
{===>} conf_3=(none);
{===>} conf_4=(none);
{===>} restraints_infile="";
{===================== annealing parameters ========================}
{===>} anneal=true;
{===>} md_type="cartesian";
{===>} geometry_min=20;
{===>} md_scheme="slowcool";
{===>} temperature=3000;
{===>} cool_rate=50;
{===>} constant_steps=50;
{===>} seed=82364;
{===>} torsion_maxlength=50;
{===>} torsion_maxtree=20;
{===>} torsion_maxchain=2000;
{===>} torsion_maxbond=6;
{==================== minimization parameters =====================}
{===>} minimize_nstep=30;
{===>} bfactor_nstep=20;
{===>} rweight=-1;
{===>} bmin=1;
{===>} bmax=200;
{===>} bsig_main=1.5;
{===>} asig_main=2.0;
{===>} bsig_side=2.0;
{===>} asig_side=2.5;
{==================== refinement parameters =======================}
{===>} num_cycles=3;
{===>} reftarget="mlf";
{===>} wa=-1.4;
{===>} target_bins=-1;
{===>} fft_memory=-1;
{=================== map generation parameters ====================}
{===>} write_map=true;
{===>} map_type="sigmaa";
{===>} map_coeff_1="fofc";
{===>} map_coeff_2="2fofc";
{===>} grid=0.35;
{===>} fill_in=false;
{===>} map_scale=true;
{===>} map_format="cns";
{===>} map_mode="molecule";
{===>} xmin=0;
{===>} xmax=0;
{===>} ymin=0;
{===>} ymax=0;
{===>} zmin=0;
{===>} zmax=0;
{===>} atom_map=(known and not hydrogen);
{===>} map_cushion=3.0;
{========================= output files ===========================}
{===>} output_root="mini";
{===>} pdb_o_format=true;
{=========================== defaults =============================}
{===>} defaults_file="";
{=================================================================}
{       things below this line do not normally need to be changed       }
{=================================================================}
...
```

Appendix table 29: UNIQUEIFY.INP script (in CCP4i)

```
NREF 0
SYMMETRY
END
CELL 112.8500 154.9200 178.0400 90.0000 90.0000 90.0000
SYMMETRY 'P 21 21 21'
LABOUT F=FUNI SIGF=SIGFUNI
RESOLUTION 2.300
#############################################
FREERFRAC 0.05
END
#############################################
LABI FILE 2  E1=FreeR_flag
LABI FILE 1  ALLIN
END
#############################################
COMPLETE FREE=FreeR_flag
END
```

Appendix table 30: REFMAC5.INP script (in CCP4i)

```
refi resolution 2.300 20.041
make check NONE
make -
    hydrogen YES -
    hout NO -
    peptide NO -
    cispeptide YES -
    ssbridge YES -
    symmetry YES -
    sugar YES -
    connectivity NO -
    link NO
ncsr nchains 6 chains A B C D E F -
    nspans 1 -
    2 430 4
refi -
    type REST -
    resi MLKF -
    meth CGMAT -
    bref ISOT
ncyc 10
scal -
    type SIMP -
    reso 2.3 20.041 -
    LSSC -
    ANISO -
    EXPE
solvent YES -
    VDWProb 1.2 -
    IONProb 0.8 -
    RSHRink 0.8
weight -
    MATRIX 0.05
monitor MEDIUM -
    torsion 10.0 -
    distance 10.0 -
    angle 10.0 -
    plane 10.0 -
    chiral 10.0 -
    bfactor 10.0 -
    bsphere 10.0 -
    rbond 10.0 -
    ncsr 10.0
labin  FP=FP SIGFP=SIGFP -
   FREE=FreeR_flag
labout  FC=FC FWT=FWT PHIC=PHIC PHWT=PHWT DELFWT=DELFWT PHDELWT=PHDELWT FOM=FOM
PNAME unknown
DNAME unknown220408
RSIZE 80
END
```

Appendix table 31: TLSIN.TLS script

```
TLS    Chain A
RANGE  'A   1' 'A 430' ALL
TLS    Chain B
RANGE  'B   1' 'B 430' ALL
TLS    Chain C
RANGE  'C   1' 'C 430' ALL
TLS    Chain D
RANGE  'D   1' 'D 430' ALL
TLS    Chain E
RANGE  'E   1' 'E 430' ALL
TLS    Chain F
RANGE  'F   1' 'F 430' ALL
```

Appendix table 32: LIBRARY.CIF for the SQR "as-purified" structure (PDB id.: 3HYV) to be used in REFMAC5

```
global_
_lib_name          ?
_lib_version       ?
_lib_update        ?
# ------------------------------------------------
#
# ---   LIST OF MONOMERS ---
#
data_comp_list
loop_
_chem_comp.id
_chem_comp.three_letter_code
_chem_comp.name
_chem_comp.group
_chem_comp.number_atoms_all
_chem_comp.number_atoms_nh
_chem_comp.desc_level
H2S     .    'HYDROSULFURIC ACID                 ' non-polymer       3   1 .
MES     .    'N-(EHTYLSULFITE)MORPHOLINE         ' non-polymer      24  12 .
PS9     PS9 'OCTATHIOCANE                        ' non-polymer       8   8 .
LMT     LMT 'DODECYL-BETA-D-MALTOSIDE            ' non-polymer      81  35 .
FAD     FAD 'FLAVIN-ADENINE-DINUCLEOTIDE         ' non-polymer      83  53 .
CSS     CSS 'S-MERCAPTOCYSTEINE                  ' non-polymer      14   8 .
#
# ---   LIST OF LINKS ---
#
data_link_list
loop_
_chem_link.id
_chem_link.comp_id_1
_chem_link.mod_id_1
_chem_link.group_comp_1
_chem_link.comp_id_2
_chem_link.mod_id_2
_chem_link.group_comp_2
_chem_link.name
 1.61   FAD      .      .        H2S       .        .
 bond_FAD-C8M_=_H2S-S
 1.62   FAD      .      .        H2S       .        .
 bond_FAD-C8M_=_H2S-S
 2.04   CYS      .      .        H2S       .        .
 bond_CYS-SG_=_H2S-S
 1.94   CSS      .      .        PS9       .        .
 bond_CSS-SD_=_PS9-S2
 1.93   CSS      .      .        PS9       .        .
 bond_CSS-SD_=_PS9-S2
 1.33   SER      .      .        CSS       .        .
 bond_SER-C_=_CSS-N
 1.34   CSS      .      .        PHE       .        .
 bond_CSS-C_=_PHE-N
#
```

```
# --- DESCRIPTION OF MONOMERS ---
#
data_comp_H2S
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.partial_charge
 H2S          HS1     H    H          0.000
 H2S          S       S    S          0.000
 H2S          HS2     H    H          0.000
loop_
_chem_comp_tree.comp_id
_chem_comp_tree.atom_id
_chem_comp_tree.atom_back
_chem_comp_tree.atom_forward
_chem_comp_tree.connect_type
 H2S       HS1   S      .      END
 H2S       S     n/a    HS1    START
 H2S       HS2   S      .      .
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
 H2S       S     HS1     coval       0.960    0.020
 H2S       HS2   S       coval       0.960    0.020
loop_
_chem_comp_angle.comp_id
_chem_comp_angle.atom_id_1
_chem_comp_angle.atom_id_2
_chem_comp_angle.atom_id_3
_chem_comp_angle.value_angle
_chem_comp_angle.value_angle_esd
 H2S       HS1   S    HS2     109.470    3.000
#
data_comp_MES
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.partial_charge
_chem_comp_atom.x
_chem_comp_atom.y
_chem_comp_atom.z
 MES          O3S     O    OS       -0.330     0.069    0.043   -0.007
 MES          S       S    ST        0.000     1.156    1.058   -0.016
 MES          O1S     O    OS       -0.330     1.953    0.844   -1.240
 MES          O2S     O    OS       -0.330     1.947    0.851    1.217
 MES          C8      C    CH2       0.000     0.337    2.494   -0.011
 MES          H81     H    H         0.000    -0.282    2.557    0.886
 MES          H82     H    H         0.000    -0.300    2.553   -0.896
 MES          C7      C    CH2       0.000     1.314    3.621   -0.024
 MES          H71     H    H         0.000     1.957    3.567   -0.904
 MES          H72     H    H         0.000     1.931    3.610    0.877
 MES          N4      N    N         0.000     0.548    4.826   -0.064
 MES          C5      C    CH2       0.000     1.340    6.058   -0.043
 MES          H51     H    H         0.000     1.935    6.143   -0.955
 MES          H52     H    H         0.000     2.004    6.069    0.824
 MES          C6      C    CH2       0.000     0.368    7.231    0.043
 MES          H62     H    H         0.000    -0.181    7.269   -0.900
 MES          H61     H    H         0.000     0.962    8.140    0.155
 MES          O1      O    O2        0.000    -0.546    7.129    1.119
 MES          C2      C    CH2       0.000    -1.067    5.834    1.210
 MES          H22     H    H         0.000    -2.129    5.927    1.446
 MES          H21     H    H         0.000    -0.555    5.335    2.036
 MES          C3      C    CH2       0.000    -0.898    5.013   -0.067
 MES          H32     H    H         0.000    -1.196    5.541   -0.975
```

```
 MES             H31    H    H          0.000    -1.392    4.039    -0.035
loop_
_chem_comp_tree.comp_id
_chem_comp_tree.atom_id
_chem_comp_tree.atom_back
_chem_comp_tree.atom_forward
_chem_comp_tree.connect_type
 MES        O3S    S       .      .
 MES        S      O2S     C8     .
 MES        O1S    S       .      .
 MES        O2S    n/a     S      START
 MES        C8     S       C7     .
 MES        H81    C8      .      .
 MES        H82    C8      .      .
 MES        C7     C8      N4     .
 MES        H71    C7      .      .
 MES        H72    C7      .      .
 MES        N4     C7      C3     .
 MES        C5     C6      H51    .
 MES        H51    C5      .      END
 MES        H52    C5      .      .
 MES        C6     O1      C5     .
 MES        H62    C6      .      .
 MES        H61    C6      .      .
 MES        O1     C2      C6     .
 MES        C2     C3      O1     .
 MES        H22    C2      .      .
 MES        H21    C2      .      .
 MES        C3     N4      C2     .
 MES        H32    C3      .      .
 MES        H31    C3      .      .
 MES        N4     C5      .      ADD
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
 MES        O1S    S       deloc    1.480    0.020
 MES        O2S    S       deloc    1.480    0.020
 MES        S      O3S     deloc    1.480    0.020
 MES        C8     S       single   1.662    0.020
 MES        H81    C8      single   1.092    0.020
 MES        H82    C8      single   1.092    0.020
 MES        C7     C8      single   1.524    0.020
 MES        H71    C7      single   1.092    0.020
 MES        H72    C7      single   1.092    0.020
 MES        N4     C7      single   1.455    0.020
 MES        C5     N4      single   1.455    0.020
 MES        H51    C5      single   1.092    0.020
 MES        H52    C5      single   1.092    0.020
 MES        C6     C5      single   1.524    0.020
 MES        H62    C6      single   1.092    0.020
 MES        H61    C6      single   1.092    0.020
 MES        N4     C3      single   1.455    0.020
 MES        H31    C3      single   1.092    0.020
 MES        H32    C3      single   1.092    0.020
 MES        C3     C2      single   1.524    0.020
 MES        H22    C2      single   1.092    0.020
 MES        H21    C2      single   1.092    0.020
 MES        C2     O1      single   1.426    0.020
 MES        O1     C6      single   1.426    0.020
loop_
_chem_comp_angle.comp_id
_chem_comp_angle.atom_id_1
_chem_comp_angle.atom_id_2
_chem_comp_angle.atom_id_3
_chem_comp_angle.value_angle
_chem_comp_angle.value_angle_esd
 MES        O3S    S       O1S     109.500    3.000
 MES        O3S    S       O2S     109.500    3.000
 MES        O3S    S       C8      109.500    3.000
 MES        O1S    S       O2S     109.500    3.000
```

```
MES       O1S      S        C8       109.500      3.000
MES       O2S      S        C8       109.500      3.000
MES       S        C8       H81      109.500      3.000
MES       S        C8       H82      109.500      3.000
MES       S        C8       C7       109.500      3.000
MES       H81      C8       H82      107.900      3.000
MES       H81      C8       C7       109.470      3.000
MES       H82      C8       C7       109.470      3.000
MES       C8       C7       H71      109.470      3.000
MES       C8       C7       H72      109.470      3.000
MES       C8       C7       N4       105.000      3.000
MES       H71      C7       H72      107.900      3.000
MES       H71      C7       N4       109.470      3.000
MES       H72      C7       N4       109.470      3.000
MES       C7       N4       C5       120.000      3.000
MES       C7       N4       C3       120.000      3.000
MES       C5       N4       C3       120.000      3.000
MES       N4       C5       H51      109.470      3.000
MES       N4       C5       H52      109.470      3.000
MES       N4       C5       C6       105.000      3.000
MES       H51      C5       H52      107.900      3.000
MES       H51      C5       C6       109.470      3.000
MES       H52      C5       C6       109.470      3.000
MES       C5       C6       H62      109.470      3.000
MES       C5       C6       H61      109.470      3.000
MES       C5       C6       O1       109.470      3.000
MES       H62      C6       H61      107.900      3.000
MES       H62      C6       O1       109.470      3.000
MES       H61      C6       O1       109.470      3.000
MES       C6       O1       C2       111.800      3.000
MES       O1       C2       H22      109.470      3.000
MES       O1       C2       H21      109.470      3.000
MES       O1       C2       C3       109.470      3.000
MES       H22      C2       H21      107.900      3.000
MES       H22      C2       C3       109.470      3.000
MES       H21      C2       C3       109.470      3.000
MES       C2       C3       H32      109.470      3.000
MES       C2       C3       H31      109.470      3.000
MES       C2       C3       N4       105.000      3.000
MES       H32      C3       H31      107.900      3.000
MES       H32      C3       N4       109.470      3.000
MES       H31      C3       N4       109.470      3.000
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
MES       var_1    O3S      S        C8       C7       180.000    20.000    1
MES       var_2    S        C8       C7       N4       180.000    20.000    3
MES       var_3    C8       C7       N4       C5       180.000    20.000    1
MES       var_4    C7       N4       C3       C2       120.000    20.000    1
MES       var_5    C7       N4       C5       C6       180.000    20.000    1
MES       var_6    N4       C5       C6       O1        60.000    20.000    3
MES       var_7    C5       C6       O1       C2       -60.000    20.000    1
MES       var_8    C6       O1       C2       C3         0.000    20.000    1
MES       var_9    O1       C2       C3       N4        60.000    20.000    3
loop_
_chem_comp_chir.comp_id
_chem_comp_chir.id
_chem_comp_chir.atom_id_centre
_chem_comp_chir.atom_id_1
_chem_comp_chir.atom_id_2
_chem_comp_chir.atom_id_3
_chem_comp_chir.volume_sign
MES       chir_01  S        O1S      O2S      O3S        positiv
loop_
_chem_comp_plane_atom.comp_id
_chem_comp_plane_atom.plane_id
_chem_comp_plane_atom.atom_id
```

```
_chem_comp_plane_atom.dist_esd
 MES      plan-1   N4        0.020
 MES      plan-1   C7        0.020
 MES      plan-1   C5        0.020
 MES      plan-1   C3        0.020
#
data_comp_PS9
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.partial_charge
_chem_comp_atom.x
_chem_comp_atom.y
_chem_comp_atom.z
 PS9          S3     S   S2       0.000    0.000    0.000    0.000
 PS9          S4     S   S2       0.000    1.414    1.414    0.000
 PS9          S5     S   S2       0.000    0.552    3.219   -0.001
 PS9          S6     S   S2       0.000    1.960    4.630    0.164
 PS9          S7     S   S2       0.000    1.116    6.436   -0.001
 PS9          S8     S   S2       0.000    2.494    7.848    0.325
 PS9          S9     S   S2       0.000    1.691    9.651    0.001
 PS9          S2     S   S2       0.000    3.018   11.068    0.483
loop_
_chem_comp_tree.comp_id
_chem_comp_tree.atom_id
_chem_comp_tree.atom_back
_chem_comp_tree.atom_forward
_chem_comp_tree.connect_type
 PS9     S3     S2      .      END
 PS9     S4     n/a     S5     START
 PS9     S5     S4      S6     .
 PS9     S6     S5      S7     .
 PS9     S7     S6      S8     .
 PS9     S8     S7      S9     .
 PS9     S9     S8      S2     .
 PS9     S2     S9      S3     .
 PS9     S3     S4      .      ADD
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
 PS9     S2     S3      single    2.000    0.020
 PS9     S2     S9      single    2.000    0.020
 PS9     S4     S3      single    2.000    0.020
 PS9     S5     S4      single    2.000    0.020
 PS9     S6     S5      single    2.000    0.020
 PS9     S7     S6      single    2.000    0.020
 PS9     S8     S7      single    2.000    0.020
 PS9     S9     S8      single    2.000    0.020
loop_
_chem_comp_angle.comp_id
_chem_comp_angle.atom_id_1
_chem_comp_angle.atom_id_2
_chem_comp_angle.atom_id_3
_chem_comp_angle.value_angle
_chem_comp_angle.value_angle_esd
 PS9     S4     S3     S2      109.470    3.000
 PS9     S3     S4     S5      109.470    3.000
 PS9     S4     S5     S6      109.470    3.000
 PS9     S5     S6     S7      109.470    3.000
 PS9     S6     S7     S8      109.470    3.000
 PS9     S7     S8     S9      109.470    3.000
 PS9     S8     S9     S2      109.470    3.000
 PS9     S9     S2     S3      109.470    3.000
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
```

```
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 PS9      var_1     S3    S4    S5    S6      175.000   20.000   1
 PS9      var_2     S4    S5    S6    S7      175.000   20.000   1
 PS9      var_3     S5    S6    S7    S8      175.000   20.000   1
 PS9      var_4     S6    S7    S8    S9      175.000   20.000   1
 PS9      var_5     S7    S8    S9    S2      175.000   20.000   1
 PS9      var_6     S8    S9    S2    S3      175.000   20.000   1
 PS9      var_7     S9    S2    S3    S4      175.000   20.000   1
 PS9      var_1     S5    S4    S3    S2      175.000   20.000   1
#
data_comp_LMT
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.partial_charge
_chem_comp_atom.x
_chem_comp_atom.y
_chem_comp_atom.z
 LMT          'O6''   O    OH1     0.000     0.000    0.000    0.000
 LMT          'HO6''  H    H       0.000     0.414   -0.870   -0.079
 LMT          'C6''   C    CH2     0.000     1.013    1.013    0.000
 LMT          'H6'1'  H    H       0.000     1.634    0.905   -0.891
 LMT          'H6'2'  H    H       0.000     1.634    0.905    0.891
 LMT          'C5''   C    CH1     0.000     0.356    2.388    0.000
 LMT          'H5''   H    H       0.000    -0.189    2.528   -0.944
 LMT          'C4''   C    CH1     0.000    -0.616    2.529    1.165
 LMT          'H4''   H    H       0.000    -1.386    1.748    1.091
 LMT          'C3''   C    CH1     0.000    -1.286    3.897    1.155
 LMT          'H3''   H    H       0.000    -2.370    3.768    1.026
 LMT          'O3''   O    OH1     0.000    -1.036    4.561    2.399
 LMT          'HO3''  H    H       0.000    -1.394    5.459    2.363
 LMT          'C2''   C    CH1     0.000    -0.746    4.762    0.022
 LMT          'H2''   H    H       0.000    -1.123    4.376   -0.935
 LMT          'O2''   O    OH1     0.000    -1.198    6.110    0.193
 LMT          'HO2''  H    H       0.000    -0.792    6.674   -0.480
 LMT          O1B     O    O2      0.000     0.090    2.375    2.395
 LMT          C1B     C    CH1     0.000    -0.750    2.612    3.522
 LMT          H1B     H    H       0.000    -1.026    3.675    3.527
 LMT          O5B     O    O2      0.000    -0.029    2.333    4.720
 LMT          C5B     C    CH1     0.000    -0.586    1.227    5.428
 LMT          H5B     H    H       0.000    -0.583    0.339    4.781
 LMT          C6B     C    CH2     0.000     0.247    0.951    6.674
 LMT          H6B1    H    H       0.000     0.319    1.861    7.273
 LMT          H6B2    H    H       0.000     1.248    0.631    6.378
 LMT          O6B     O    OH1     0.000    -0.377   -0.082    7.446
 LMT          HO6B    H    H       0.000     0.190   -0.312    8.194
 LMT          C4B     C    CH1     0.000    -2.017    1.554    5.836
 LMT          H4B     H    H       0.000    -2.439    0.703    6.389
 LMT          'O4''   O    OH1     0.000    -2.021    2.712    6.678
 LMT          'HO4''  H    H       0.000    -2.933    2.966    6.873
 LMT          C3B     C    CH1     0.000    -2.883    1.836    4.613
 LMT          H3B     H    H       0.000    -3.691    1.093    4.563
 LMT          O3B     O    OH1     0.000    -3.452    3.146    4.724
 LMT          HO3B    H    H       0.000    -3.936    3.356    3.914
 LMT          C2B     C    CH1     0.000    -2.058    1.768    3.333
 LMT          H2B     H    H       0.000    -1.788    0.722    3.133
 LMT          O2B     O    OH1     0.000    -2.833    2.273    2.240
 LMT          HO2B    H    H       0.000    -2.285    2.303    1.444
 LMT          'O5''   O    O2      0.000     1.359    3.394    0.117
 LMT          'C1''   C    CH1     0.000     0.809    4.705    0.015
 LMT          'H1''   H    H       0.000     1.182    5.305    0.857
 LMT          'O1''   O    O2      0.000     1.247    5.312   -1.199
 LMT          C1      C    CH2     0.000     0.634    6.582   -1.407
 LMT          H11     H    H       0.000     0.813    7.219   -0.538
 LMT          H12     H    H       0.000    -0.441    6.449   -1.544
 LMT          C2      C    CH2     0.000     1.229    7.236   -2.649
```

```
LMT              H21     H     H        0.000     1.130    6.545    -3.489
LMT              H22     H     H        0.000     2.287    7.435    -2.466
LMT              C3      C     CH2      0.000     0.513    8.541    -2.975
LMT              H31     H     H        0.000     0.536    9.183    -2.092
LMT              H32     H     H        0.000    -0.523    8.316    -3.233
LMT              C4      C     CH2      0.000     1.186    9.256    -4.141
LMT              H41     H     H        0.000     1.243    8.567    -4.987
LMT              H42     H     H        0.000     2.195    9.545    -3.839
LMT              C5      C     CH2      0.000     0.400   10.497    -4.547
LMT              H51     H     H        0.000     0.266   11.129    -3.667
LMT              H52     H     H        0.000    -0.577   10.185    -4.922
LMT              C6      C     CH2      0.000     1.135   11.278    -5.630
LMT              H61     H     H        0.000     1.348   10.606    -6.463
LMT              H62     H     H        0.000     2.074   11.651    -5.215
LMT              C7      C     CH2      0.000     0.294   12.450    -6.123
LMT              H71     H     H        0.000     0.007   13.060    -5.264
LMT              H72     H     H        0.000    -0.603   12.058    -6.606
LMT              C8      C     CH2      0.000     1.076   13.303    -7.115
LMT              H81     H     H        0.000     1.440   12.659    -7.919
LMT              H82     H     H        0.000     1.925   13.752    -6.597
LMT              C9      C     CH2      0.000     0.197   14.402    -7.702
LMT              H91     H     H        0.000    -0.238   14.976    -6.881
LMT              H92     H     H        0.000    -0.602   13.937    -8.283
LMT              C10     C     CH2      0.000     1.008   15.329    -8.599
LMT              H101    H     H        0.000     1.516   14.727    -9.355
LMT              H102    H     H        0.000     1.751   15.846    -7.987
LMT              C11     C     CH2      0.000     0.109   16.353    -9.282
LMT              H111    H     H        0.000    -0.465   16.879    -8.516
LMT              H112    H     H        0.000    -0.575   15.826    -9.950
LMT              C12     C     CH3      0.000     0.926   17.348   -10.076
LMT              H123    H     H        0.000     1.526   17.918    -9.415
LMT              H122    H     H        0.000     1.547   16.830   -10.759
LMT              H121    H     H        0.000     0.276   17.993   -10.608
loop_
_chem_comp_tree.comp_id
_chem_comp_tree.atom_id
_chem_comp_tree.atom_back
_chem_comp_tree.atom_forward
_chem_comp_tree.connect_type
LMT      'O6''  'C6''   'HO6''  .
LMT      'HO6'' 'O6''   .       .
LMT      'C6''  'C5''   'O6''   .
LMT      'H6'1' 'C6''   .       .
LMT      'H6'2' 'C6''   .       .
LMT      'C5''  'O5''   'C6''   .
LMT      'H5''  'C5''   .       .
LMT      'C4''  O1B     'C3''   .
LMT      'H4''  'C4''   .       .
LMT      'C3''  'C4''   'C2''   .
LMT      'H3''  'C3''   .       .
LMT      'O3''  'C3''   'HO3''  .
LMT      'HO3'' 'O3''   .       .
LMT      'C2''  'C3''   'C1''   .
LMT      'H2''  'C2''   .       .
LMT      'O2''  'C2''   'HO2''  .
LMT      'HO2'' 'O2''   .       .
LMT      O1B    C1B     'C4''   .
LMT      C1B    C2B     O1B     .
LMT      H1B    C1B     .       .
LMT      O5B    C5B     .       .
LMT      C5B    C4B     O5B     .
LMT      H5B    C5B     .       .
LMT      C6B    C5B     O6B     .
LMT      H6B1   C6B     .       .
LMT      H6B2   C6B     .       .
LMT      O6B    C6B     HO6B    .
LMT      HO6B   O6B     .       .
LMT      C4B    C3B     C5B     .
LMT      H4B    C4B     .       .
LMT      'O4''  C4B     'HO4''  .
LMT      'HO4'' 'O4''   .       .
LMT      C3B    C2B     C4B     .
LMT      H3B    C3B     .       .
LMT      O3B    C3B     HO3B    .
```

```
LMT      HO3B    O3B     .       .
LMT      C2B     O2B     C1B     .
LMT      H2B     C2B     .       .
LMT      O2B     n/a     C2B     START
LMT      HO2B    O2B     .       .
LMT      'O5''   'C1''   'C5''   .
LMT      'C1''   'C2''   'O1''   .
LMT      'H1''   'C1''   .       .
LMT      'O1''   'C1''   C1      .
LMT      C1      'O1''   C2      .
LMT      H11     C1      .       .
LMT      H12     C1      .       .
LMT      C2      C1      C3      .
LMT      H21     C2      .       .
LMT      H22     C2      .       .
LMT      C3      C2      C4      .
LMT      H31     C3      .       .
LMT      H32     C3      .       .
LMT      C4      C3      C5      .
LMT      H41     C4      .       .
LMT      H42     C4      .       .
LMT      C5      C4      C6      .
LMT      H51     C5      .       .
LMT      H52     C5      .       .
LMT      C6      C5      C7      .
LMT      H61     C6      .       .
LMT      H62     C6      .       .
LMT      C7      C6      C8      .
LMT      H71     C7      .       .
LMT      H72     C7      .       .
LMT      C8      C7      C9      .
LMT      H81     C8      .       .
LMT      H82     C8      .       .
LMT      C9      C8      C10     .
LMT      H91     C9      .       .
LMT      H92     C9      .       .
LMT      C10     C9      C11     .
LMT      H101    C10     .       .
LMT      H102    C10     .       .
LMT      C11     C10     C12     .
LMT      H111    C11     .       .
LMT      H112    C11     .       .
LMT      C12     C11     H123    .
LMT      H123    C12     .       END
LMT      H122    C12     .       .
LMT      H121    C12     .       .
LMT      'C5''   'C4''   .       ADD
LMT      C1B     O5B     .       ADD
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
LMT      C1B     C2B     single  1.524   0.020
LMT      C1B     O1B     single  1.426   0.020
LMT      O5B     C1B     single  1.426   0.020
LMT      C2B     C3B     single  1.524   0.020
LMT      O2B     C2B     single  1.432   0.020
LMT      C3B     C4B     single  1.524   0.020
LMT      O3B     C3B     single  1.432   0.020
LMT      C4B     C5B     single  1.524   0.020
LMT      'O4''   C4B     single  1.432   0.020
LMT      C6B     C5B     single  1.524   0.020
LMT      C5B     O5B     single  1.426   0.020
LMT      O6B     C6B     single  1.432   0.020
LMT      O1B     'C4''   single  1.426   0.020
LMT      'C1''   'C2''   single  1.524   0.020
LMT      'O1''   'C1''   single  1.426   0.020
LMT      'C1''   'O5''   single  1.426   0.020
LMT      'C2''   'C3''   single  1.524   0.020
LMT      'O2''   'C2''   single  1.432   0.020
LMT      'C3''   'C4''   single  1.524   0.020
```

```
LMT       'O3''  'C3''     single     1.432   0.020
LMT       'C4''  'C5''     single     1.524   0.020
LMT       'C5''  'C6''     single     1.524   0.020
LMT       'O5''  'C5''     single     1.426   0.020
LMT       'C6''  'O6''     single     1.432   0.020
LMT       C1     'O1''     single     1.426   0.020
LMT       C2     C1        single     1.524   0.020
LMT       C3     C2        single     1.524   0.020
LMT       C4     C3        single     1.524   0.020
LMT       C5     C4        single     1.524   0.020
LMT       C6     C5        single     1.524   0.020
LMT       C7     C6        single     1.524   0.020
LMT       C8     C7        single     1.524   0.020
LMT       C9     C8        single     1.524   0.020
LMT       C10    C9        single     1.524   0.020
LMT       C11    C10       single     1.524   0.020
LMT       C12    C11       single     1.513   0.020
LMT       H1B    C1B       single     1.099   0.020
LMT       H2B    C2B       single     1.099   0.020
LMT       H3B    C3B       single     1.099   0.020
LMT       H4B    C4B       single     1.099   0.020
LMT       H5B    C5B       single     1.099   0.020
LMT       H6B1   C6B       single     1.092   0.020
LMT       H6B2   C6B       single     1.092   0.020
LMT       HO2B   O2B       single     0.967   0.020
LMT       HO3B   O3B       single     0.967   0.020
LMT       'HO4'' 'O4''     single     0.967   0.020
LMT       HO6B   O6B       single     0.967   0.020
LMT       'H1''  'C1''     single     1.099   0.020
LMT       'H2''  'C2''     single     1.099   0.020
LMT       'H3''  'C3''     single     1.099   0.020
LMT       'H4''  'C4''     single     1.099   0.020
LMT       'H5''  'C5''     single     1.099   0.020
LMT       'H6'1' 'C6''     single     1.092   0.020
LMT       'H6'2' 'C6''     single     1.092   0.020
LMT       'HO2'' 'O2''     single     0.967   0.020
LMT       'HO3'' 'O3''     single     0.967   0.020
LMT       'HO6'' 'O6''     single     0.967   0.020
LMT       H11    C1        single     1.092   0.020
LMT       H12    C1        single     1.092   0.020
LMT       H21    C2        single     1.092   0.020
LMT       H22    C2        single     1.092   0.020
LMT       H31    C3        single     1.092   0.020
LMT       H32    C3        single     1.092   0.020
LMT       H41    C4        single     1.092   0.020
LMT       H42    C4        single     1.092   0.020
LMT       H51    C5        single     1.092   0.020
LMT       H52    C5        single     1.092   0.020
LMT       H61    C6        single     1.092   0.020
LMT       H62    C6        single     1.092   0.020
LMT       H71    C7        single     1.092   0.020
LMT       H72    C7        single     1.092   0.020
LMT       H81    C8        single     1.092   0.020
LMT       H82    C8        single     1.092   0.020
LMT       H91    C9        single     1.092   0.020
LMT       H92    C9        single     1.092   0.020
LMT       H101   C10       single     1.092   0.020
LMT       H102   C10       single     1.092   0.020
LMT       H111   C11       single     1.092   0.020
LMT       H112   C11       single     1.092   0.020
LMT       H121   C12       single     1.059   0.020
LMT       H122   C12       single     1.059   0.020
LMT       H123   C12       single     1.059   0.020
loop_
_chem_comp_angle.comp_id
_chem_comp_angle.atom_id_1
_chem_comp_angle.atom_id_2
_chem_comp_angle.atom_id_3
_chem_comp_angle.value_angle
_chem_comp_angle.value_angle_esd
LMT       'HO6'' 'O6''  'C6''    109.470    3.000
LMT       'O6''  'C6''  'H6'1'   109.470    3.000
LMT       'O6''  'C6''  'H6'2'   109.470    3.000
LMT       'O6''  'C6''  'C5''    109.470    3.000
```

LMT	'H6'1'	'C6''	'H6'2'	107.900	3.000
LMT	'H6'1'	'C6''	'C5''	109.470	3.000
LMT	'H6'2'	'C6''	'C5''	109.470	3.000
LMT	'C6''	'C5''	'H5''	108.340	3.000
LMT	'C6''	'C5''	'C4''	111.000	3.000
LMT	'C6''	'C5''	'O5''	109.470	3.000
LMT	'H5''	'C5''	'C4''	108.340	3.000
LMT	'H5''	'C5''	'O5''	109.470	3.000
LMT	'C4''	'C5''	'O5''	109.470	3.000
LMT	'C5''	'C4''	'H4''	108.340	3.000
LMT	'C5''	'C4''	'C3''	111.000	3.000
LMT	'C5''	'C4''	O1B	109.470	3.000
LMT	'H4''	'C4''	'C3''	108.340	3.000
LMT	'H4''	'C4''	O1B	109.470	3.000
LMT	'C3''	'C4''	O1B	109.470	3.000
LMT	'C4''	'C3''	'H3''	108.340	3.000
LMT	'C4''	'C3''	'O3''	109.470	3.000
LMT	'C4''	'C3''	'C2''	111.000	3.000
LMT	'H3''	'C3''	'O3''	109.470	3.000
LMT	'H3''	'C3''	'C2''	108.340	3.000
LMT	'O3''	'C3''	'C2''	109.470	3.000
LMT	'C3''	'O3''	'HO3''	109.470	3.000
LMT	'C3''	'C2''	'H2''	108.340	3.000
LMT	'C3''	'C2''	'O2''	109.470	3.000
LMT	'C3''	'C2''	'C1''	111.000	3.000
LMT	'H2''	'C2''	'O2''	109.470	3.000
LMT	'H2''	'C2''	'C1''	108.340	3.000
LMT	'O2''	'C2''	'C1''	109.470	3.000
LMT	'C2''	'O2''	'HO2''	109.470	3.000
LMT	'C4''	O1B	C1B	111.800	3.000
LMT	O1B	C1B	H1B	109.470	3.000
LMT	O1B	C1B	O5B	109.470	3.000
LMT	O1B	C1B	C2B	109.470	3.000
LMT	H1B	C1B	O5B	109.470	3.000
LMT	H1B	C1B	C2B	108.340	3.000
LMT	O5B	C1B	C2B	109.470	3.000
LMT	C1B	O5B	C5B	111.800	3.000
LMT	O5B	C5B	H5B	109.470	3.000
LMT	O5B	C5B	C6B	109.470	3.000
LMT	O5B	C5B	C4B	109.470	3.000
LMT	H5B	C5B	C6B	108.340	3.000
LMT	H5B	C5B	C4B	108.340	3.000
LMT	C6B	C5B	C4B	111.000	3.000
LMT	C5B	C6B	H6B1	109.470	3.000
LMT	C5B	C6B	H6B2	109.470	3.000
LMT	C5B	C6B	O6B	109.470	3.000
LMT	H6B1	C6B	H6B2	107.900	3.000
LMT	H6B1	C6B	O6B	109.470	3.000
LMT	H6B2	C6B	O6B	109.470	3.000
LMT	C6B	O6B	HO6B	109.470	3.000
LMT	C5B	C4B	H4B	108.340	3.000
LMT	C5B	C4B	'O4''	109.470	3.000
LMT	C5B	C4B	C3B	111.000	3.000
LMT	H4B	C4B	'O4''	109.470	3.000
LMT	H4B	C4B	C3B	108.340	3.000
LMT	'O4''	C4B	C3B	109.470	3.000
LMT	C4B	'O4''	'HO4''	109.470	3.000
LMT	C4B	C3B	H3B	108.340	3.000
LMT	C4B	C3B	O3B	109.470	3.000
LMT	C4B	C3B	C2B	111.000	3.000
LMT	H3B	C3B	O3B	109.470	3.000
LMT	H3B	C3B	C2B	108.340	3.000
LMT	O3B	C3B	C2B	109.470	3.000
LMT	C3B	O3B	HO3B	109.470	3.000
LMT	C3B	C2B	H2B	108.340	3.000
LMT	C3B	C2B	O2B	109.470	3.000
LMT	C3B	C2B	C1B	111.000	3.000
LMT	H2B	C2B	O2B	109.470	3.000
LMT	H2B	C2B	C1B	108.340	3.000
LMT	O2B	C2B	C1B	109.470	3.000
LMT	C2B	O2B	HO2B	109.470	3.000
LMT	'C5''	'O5''	'C1''	111.800	3.000
LMT	'O5''	'C1''	'H1''	109.470	3.000
LMT	'O5''	'C1''	'O1''	109.470	3.000

```
LMT    'O5''  'C1''  'C2''  109.470  3.000
LMT    'H1''  'C1''  'O1''  109.470  3.000
LMT    'H1''  'C1''  'C2''  108.340  3.000
LMT    'O1''  'C1''  'C2''  109.470  3.000
LMT    'C1''  'O1''  C1     111.800  3.000
LMT    'O1''  C1     H11    109.470  3.000
LMT    'O1''  C1     H12    109.470  3.000
LMT    'O1''  C1     C2     109.470  3.000
LMT    H11    C1     H12    107.900  3.000
LMT    H11    C1     C2     109.470  3.000
LMT    H12    C1     C2     109.470  3.000
LMT    C1     C2     H21    109.470  3.000
LMT    C1     C2     H22    109.470  3.000
LMT    C1     C2     C3     111.000  3.000
LMT    H21    C2     H22    107.900  3.000
LMT    H21    C2     C3     109.470  3.000
LMT    H22    C2     C3     109.470  3.000
LMT    C2     C3     H31    109.470  3.000
LMT    C2     C3     H32    109.470  3.000
LMT    C2     C3     C4     111.000  3.000
LMT    H31    C3     H32    107.900  3.000
LMT    H31    C3     C4     109.470  3.000
LMT    H32    C3     C4     109.470  3.000
LMT    C3     C4     H41    109.470  3.000
LMT    C3     C4     H42    109.470  3.000
LMT    C3     C4     C5     111.000  3.000
LMT    H41    C4     H42    107.900  3.000
LMT    H41    C4     C5     109.470  3.000
LMT    H42    C4     C5     109.470  3.000
LMT    C4     C5     H51    109.470  3.000
LMT    C4     C5     H52    109.470  3.000
LMT    C4     C5     C6     111.000  3.000
LMT    H51    C5     H52    107.900  3.000
LMT    H51    C5     C6     109.470  3.000
LMT    H52    C5     C6     109.470  3.000
LMT    C5     C6     H61    109.470  3.000
LMT    C5     C6     H62    109.470  3.000
LMT    C5     C6     C7     111.000  3.000
LMT    H61    C6     H62    107.900  3.000
LMT    H61    C6     C7     109.470  3.000
LMT    H62    C6     C7     109.470  3.000
LMT    C6     C7     H71    109.470  3.000
LMT    C6     C7     H72    109.470  3.000
LMT    C6     C7     C8     111.000  3.000
LMT    H71    C7     H72    107.900  3.000
LMT    H71    C7     C8     109.470  3.000
LMT    H72    C7     C8     109.470  3.000
LMT    C7     C8     H81    109.470  3.000
LMT    C7     C8     H82    109.470  3.000
LMT    C7     C8     C9     111.000  3.000
LMT    H81    C8     H82    107.900  3.000
LMT    H81    C8     C9     109.470  3.000
LMT    H82    C8     C9     109.470  3.000
LMT    C8     C9     H91    109.470  3.000
LMT    C8     C9     H92    109.470  3.000
LMT    C8     C9     C10    111.000  3.000
LMT    H91    C9     H92    107.900  3.000
LMT    H91    C9     C10    109.470  3.000
LMT    H92    C9     C10    109.470  3.000
LMT    C9     C10    H101   109.470  3.000
LMT    C9     C10    H102   109.470  3.000
LMT    C9     C10    C11    111.000  3.000
LMT    H101   C10    H102   107.900  3.000
LMT    H101   C10    C11    109.470  3.000
LMT    H102   C10    C11    109.470  3.000
LMT    C10    C11    H111   109.470  3.000
LMT    C10    C11    H112   109.470  3.000
LMT    C10    C11    C12    111.000  3.000
LMT    H111   C11    H112   107.900  3.000
LMT    H111   C11    C12    109.470  3.000
LMT    H112   C11    C12    109.470  3.000
LMT    C11    C12    H123   109.470  3.000
LMT    C11    C12    H122   109.470  3.000
LMT    C11    C12    H121   109.470  3.000
```

```
  LMT      H123    C12     H122    109.470    3.000
  LMT      H123    C12     H121    109.470    3.000
  LMT      H122    C12     H121    109.470    3.000
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
  LMT      var_1   'HO6''  'O6''   'C6''   'C5''    175.000   20.000   1
  LMT      var_2   'O6''   'C6''   'C5''   'O5''    175.000   20.000   3
  LMT      var_3   'C6''   'C5''   'C4''   O1B       60.000   20.000   3
  LMT      var_4   'C5''   'C4''   'C3''   'C2''      0.000   20.000   3
  LMT      var_5   'C4''   'C3''   'O3''   'HO3''   175.000   20.000   1
  LMT      var_6   'C4''   'C3''   'C2''   'O2''    180.000   20.000   3
  LMT      var_7   'C3''   'C2''   'O2''   'HO2''   175.000   20.000   1
  LMT      var_8   'C5''   'C4''   O1B     C1B      175.000   20.000   1
  LMT      var_9   'C4''   O1B     C1B     O5B      175.000   20.000   1
  LMT      var_10  O1B     C1B     C2B     C3B      180.000   20.000   3
  LMT      var_11  O1B     C1B     O5B     C5B     -120.000   20.000   1
  LMT      var_12  C1B     O5B     C5B     C4B      -60.000   20.000   1
  LMT      var_13  O5B     C5B     C6B     O6B      175.000   20.000   3
  LMT      var_14  C5B     C6B     O6B     HO6B     175.000   20.000   1
  LMT      var_15  O5B     C5B     C4B     C3B       60.000   20.000   3
  LMT      var_16  C5B     C4B     'O4''   'HO4''   175.000   20.000   1
  LMT      var_17  C5B     C4B     C3B     C2B        0.000   20.000   3
  LMT      var_18  C4B     C3B     O3B     HO3B     175.000   20.000   1
  LMT      var_19  C4B     C3B     C2B     O2B      180.000   20.000   3
  LMT      var_20  C3B     C2B     O2B     HO2B     175.000   20.000   1
  LMT      var_21  'C6''   'C5''   'O5''   'C1''    180.000   20.000   1
  LMT      var_22  'C5''   'O5''   'C1''   'O1''   -120.000   20.000   1
  LMT      var_23  'O5''   'C1''   'C2''   'C3''     60.000   20.000   3
  LMT      var_24  'O5''   'C1''   'O1''   C1       175.000   20.000   1
  LMT      var_25  'C1''   'O1''   C1      C2       175.000   20.000   1
  LMT      var_26  'O1''   C1      C2      C3       175.000   20.000   3
  LMT      var_27  C1      C2      C3      C4       175.000   20.000   3
  LMT      var_28  C2      C3      C4      C5       175.000   20.000   3
  LMT      var_29  C3      C4      C5      C6       175.000   20.000   3
  LMT      var_30  C4      C5      C6      C7       175.000   20.000   3
  LMT      var_31  C5      C6      C7      C8       175.000   20.000   3
  LMT      var_32  C6      C7      C8      C9       175.000   20.000   3
  LMT      var_33  C7      C8      C9      C10      175.000   20.000   3
  LMT      var_34  C8      C9      C10     C11      175.000   20.000   3
  LMT      var_35  C9      C10     C11     C12      175.000   20.000   3
  LMT      var_36  C10     C11     C12     H121     175.000   20.000   3
loop_
_chem_comp_chir.comp_id
_chem_comp_chir.id
_chem_comp_chir.atom_id_centre
_chem_comp_chir.atom_id_1
_chem_comp_chir.atom_id_2
_chem_comp_chir.atom_id_3
_chem_comp_chir.volume_sign
  LMT      chir_01  C1B    C2B     O1B     O5B       negativ
  LMT      chir_02  C2B    C1B     C3B     O2B       negativ
  LMT      chir_03  C3B    C2B     C4B     O3B       positiv
  LMT      chir_04  C4B    C3B     C5B     'O4''     negativ
  LMT      chir_05  C5B    C4B     C6B     O5B       negativ
  LMT      chir_06  'C1''  'C2''   'O1''   'O5''     positiv
  LMT      chir_07  'C2''  'C1''   'C3''   'O2''     negativ
  LMT      chir_08  'C3''  'C2''   'C4''   'O3''     positiv
  LMT      chir_09  'C4''  O1B     'C3''   'C5''     negativ
  LMT      chir_10  'C5''  'C4''   'C6''   'O5''     negativ
#
data_comp_FAD
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
```

```
_chem_comp_atom.type_energy
_chem_comp_atom.partial_charge
_chem_comp_atom.x
_chem_comp_atom.y
_chem_comp_atom.z
FAD           O2P      O    OP      -0.500      0.000      0.000      0.000
FAD           P        P    P        0.000      1.068      1.068      0.000
FAD           O1P      O    OP      -0.500      2.063      1.137      1.134
FAD           'O5''    O    O2       0.000      1.964      0.883     -1.325
FAD           'C5''    C    CH2      0.000      2.988      1.822     -1.642
FAD           'H5'1'   H    H        0.000      3.673      1.914     -0.797
FAD           'H5'2'   H    H        0.000      2.538      2.795     -1.851
FAD           'C4''    C    CH1      0.000      3.755      1.342     -2.869
FAD           'H4''    H    H        0.000      4.273      0.404     -2.628
FAD           'O4''    O    OH1      0.000      2.838      1.111     -3.945
FAD           'HO4''   H    H        0.000      3.331      0.883     -4.745
FAD           'C3''    C    CH1      0.000      4.779      2.380     -3.314
FAD           'H3''    H    H        0.000      4.254      3.289     -3.641
FAD           'O3''    O    OH1      0.000      5.640      2.702     -2.216
FAD           'HO3''   H    H        0.000      6.337      3.300     -2.518
FAD           'C2''    C    CH1      0.000      5.629      1.851     -4.463
FAD           'H2''    H    H        0.000      4.970      1.517     -5.276
FAD           'O2''    O    OH1      0.000      6.412      0.743     -4.006
FAD           'HO2''   H    H        0.000      6.892      0.358     -4.752
FAD           'C1''    C    CH2      0.000      6.568      2.930     -4.989
FAD           'H1'1'   H    H        0.000      7.168      3.322     -4.165
FAD           'H1'2'   H    H        0.000      5.982      3.740     -5.428
FAD           N10      N    NR6      0.000      7.450      2.355     -6.008
FAD           C9A      C    CR66     0.000      7.411      1.043     -6.263
FAD           C9       C    CR16     0.000      6.534      0.225     -5.561
FAD           H9       H    H        0.000      5.882      0.650     -4.808
FAD           C8       C    CR6      0.000      6.493     -1.139     -5.826
FAD           C8M      C    CH3      0.000      5.543     -2.026     -5.065
FAD           H8M3     H    H        0.000      5.815     -2.040     -4.042
FAD           H8M2     H    H        0.000      4.557     -1.652     -5.163
FAD           H8M1     H    H        0.000      5.588     -3.009     -5.456
FAD           C7       C    CR6      0.000      7.326     -1.682     -6.788
FAD           C7M      C    CH3      0.000      7.282     -3.160     -7.075
FAD           H7M3     H    H        0.000      7.607     -3.695     -6.221
FAD           H7M2     H    H        0.000      6.290     -3.444     -7.314
FAD           H7M1     H    H        0.000      7.919     -3.380     -7.893
FAD           C6       C    CR16     0.000      8.203     -0.864     -7.490
FAD           H6       H    H        0.000      8.854     -1.290     -8.243
FAD           C5X      C    CR66     0.000      8.247      0.498     -7.229
FAD           N5       N    NRD6     0.000      9.099      1.293     -7.911
FAD           C10      C    CR66     0.000      8.293      3.143     -6.683
FAD           C4X      C    CR66     0.000      9.130      2.597     -7.650
FAD           C4       C    CR6      0.000     10.007      3.416     -8.352
FAD           O4       O    OH1      0.000     10.827      2.881     -9.299
FAD           HO4      H    H        0.000     11.312      3.591     -9.743
FAD           N3       N    NRD6     0.000     10.047      4.741     -8.094
FAD           C2       C    CR6      0.000      9.234      5.270     -7.155
FAD           O2       O    O        0.000      9.271      6.497     -6.917
FAD           N1       N    NRD6     0.000      8.333      4.467     -6.426
FAD           O3P      O    O2       0.000      0.342      2.505      0.000
FAD           PA       P    P        0.000      1.207      3.857      0.121
FAD           O1A      O    OP      -0.500      1.868      3.874      1.478
FAD           O2A      O    OP      -0.500      2.243      3.854     -0.978
FAD           O5B      O    O2       0.000      0.092      5.005     -0.058
FAD           C5B      C    CH2      0.000      0.438      6.380      0.092
FAD           H5B1     H    H        0.000      1.264      6.622     -0.580
FAD           H5B2     H    H        0.000      0.742      6.568      1.123
FAD           C4B      C    CH1      0.000     -0.768      7.247     -0.251
FAD           H4B      H    H        0.000     -1.651      6.921      0.316
FAD           C3B      C    CH1      0.000     -0.475      8.719      0.016
FAD           H3B      H    H        0.000     -1.162      9.090      0.788
FAD           O3B      O    OH1      0.000      0.873      8.862      0.476
FAD           HO3B     H    H        0.000      1.084      9.801      0.570
FAD           C2B      C    CH1      0.000     -0.649      9.548     -1.251
FAD           H2B      H    H        0.000      0.290     10.076     -1.470
FAD           O2B      O    OH1      0.000     -1.694     10.508     -1.053
FAD           HO2B     H    H        0.000     -1.751     11.085     -1.827
FAD           C1B      C    CH1      0.000     -1.016      8.664     -2.437
FAD           H1B      H    H        0.000     -0.233      8.737     -3.204
```

```
FAD          O4B    O    O2        0.000    -1.132    7.311   -2.003
FAD          N9A    N    NR5       0.000    -2.311    9.114   -3.008
FAD          C4A    C    CR56      0.000    -2.894    8.632   -4.110
FAD          C5A    C    CR56      0.000    -4.087    9.318   -4.301
FAD          N7A    N    NRD5      0.000    -4.216   10.217   -3.301
FAD          C8A    C    CR15      0.000    -3.136   10.106   -2.499
FAD          H8A    H    H         0.000    -2.946   10.692   -1.608
FAD          N3A    N    NRD6      0.000    -2.522    7.665   -4.983
FAD          C2A    C    CR16      0.000    -3.303    7.370   -6.027
FAD          H2A    H    H         0.000    -3.006    6.597   -6.725
FAD          N1A    N    NRD6      0.000    -4.451    8.030   -6.210
FAD          C6A    C    CR6       0.000    -4.822    8.994   -5.341
FAD          N6A    N    N         0.000    -5.986    9.663   -5.526
FAD          HN6A   H    H         0.000    -6.247   10.344   -4.912
loop_
_chem_comp_tree.comp_id
_chem_comp_tree.atom_id
_chem_comp_tree.atom_back
_chem_comp_tree.atom_forward
_chem_comp_tree.connect_type
FAD       O2P    P        .       .
FAD       P      O3P      'O5''   .
FAD       O1P    P        .       .
FAD       'O5''  P        'C5''   .
FAD       'C5''  'O5''    'C4''   .
FAD       'H5'1' 'C5''    .       .
FAD       'H5'2' 'C5''    .       .
FAD       'C4''  'C5''    'C3''   .
FAD       'H4''  'C4''    .       .
FAD       'O4''  'C4''    'HO4''  .
FAD       'HO4'' 'O4''    .       .
FAD       'C3''  'C4''    'C2''   .
FAD       'H3''  'C3''    .       .
FAD       'O3''  'C3''    'HO3''  .
FAD       'HO3'' 'O3''    .       .
FAD       'C2''  'C3''    'C1''   .
FAD       'H2''  'C2''    .       .
FAD       'O2''  'C2''    'HO2''  .
FAD       'HO2'' 'O2''    .       .
FAD       'C1''  'C2''    N10     .
FAD       'H1'1' 'C1''    .       .
FAD       'H1'2' 'C1''    .       .
FAD       N10    'C1''    C10     .
FAD       C9A    C5X      .       .
FAD       C9     C8       H9      .
FAD       H9     C9       .       END
FAD       C8     C7       C9      .
FAD       C8M    C8       H8M3    .
FAD       H8M3   C8M      .       .
FAD       H8M2   C8M      .       .
FAD       H8M1   C8M      .       .
FAD       C7     C6       C8      .
FAD       C7M    C7       H7M3    .
FAD       H7M3   C7M      .       .
FAD       H7M2   C7M      .       .
FAD       H7M1   C7M      .       .
FAD       C6     C5X      C7      .
FAD       H6     C6       .       .
FAD       C5X    N5       C6      .
FAD       N5     C4X      C5X     .
FAD       C10    N10      N1      .
FAD       C4X    C4       N5      .
FAD       C4     N3       C4X     .
FAD       O4     C4       HO4     .
FAD       HO4    O4       .       .
FAD       N3     C2       C4      .
FAD       C2     N1       N3      .
FAD       O2     C2       .       .
FAD       N1     C10      C2      .
FAD       O3P    PA       P       .
FAD       PA     O2A      O3P     .
FAD       O1A    PA       .       .
FAD       O2A    n/a      PA      START
FAD       O5B    PA       C5B     .
```

```
FAD       C5B       O5B       C4B       .
FAD       H5B1      C5B       .         .
FAD       H5B2      C5B       .         .
FAD       C4B       C5B       O4B       .
FAD       H4B       C4B       .         .
FAD       C3B       C2B       O3B       .
FAD       H3B       C3B       .         .
FAD       O3B       C3B       HO3B      .
FAD       HO3B      O3B       .         .
FAD       C2B       C1B       C3B       .
FAD       H2B       C2B       .         .
FAD       O2B       C2B       HO2B      .
FAD       HO2B      O2B       .         .
FAD       C1B       O4B       C2B       .
FAD       H1B       C1B       .         .
FAD       O4B       C4B       C1B       .
FAD       N9A       C1B       C8A       .
FAD       C4A       N3A       .         .
FAD       C5A       N7A       C6A       .
FAD       N7A       C8A       C5A       .
FAD       C8A       N9A       N7A       .
FAD       H8A       C8A       .         .
FAD       N3A       C2A       C4A       .
FAD       C2A       N1A       N3A       .
FAD       H2A       C2A       .         .
FAD       N1A       C6A       C2A       .
FAD       C6A       C5A       N1A       .
FAD       N6A       C6A       HN6A      .
FAD       HN6A      N6A       .         .
FAD       N10       C9A       .         ADD
FAD       C9A       C9        .         ADD
FAD       C10       C4X       .         ADD
FAD       C4B       C3B       .         ADD
FAD       N9A       C4A       .         ADD
FAD       C4A       C5A       .         ADD
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
FAD       O1A       PA        deloc     1.510   0.020
FAD       O2A       PA        deloc     1.510   0.020
FAD       O5B       PA        single    1.610   0.020
FAD       PA        O3P       single    1.610   0.020
FAD       C5B       O5B       single    1.426   0.020
FAD       C4B       C5B       single    1.524   0.020
FAD       C4B       O4B       single    1.426   0.020
FAD       C3B       C4B       single    1.524   0.020
FAD       O4B       C1B       single    1.426   0.020
FAD       O3B       C3B       single    1.432   0.020
FAD       C2B       C3B       single    1.524   0.020
FAD       O2B       C2B       single    1.432   0.020
FAD       C1B       C2B       single    1.524   0.020
FAD       N9A       C1B       single    1.485   0.020
FAD       N9A       C8A       aromatic  1.337   0.020
FAD       C4A       N9A       aromatic  1.337   0.020
FAD       C8A       N7A       aromatic  1.350   0.020
FAD       N7A       C5A       aromatic  1.350   0.020
FAD       C5A       C6A       aromatic  1.390   0.020
FAD       C5A       C4A       aromatic  1.390   0.020
FAD       N6A       C6A       double    1.355   0.020
FAD       C6A       N1A       aromatic  1.350   0.020
FAD       N1A       C2A       aromatic  1.337   0.020
FAD       C2A       N3A       aromatic  1.337   0.020
FAD       N3A       C4A       aromatic  1.355   0.020
FAD       N1        C2        aromatic  1.350   0.020
FAD       N1        C10       aromatic  1.350   0.020
FAD       O2        C2        double    1.250   0.020
FAD       C2        N3        aromatic  1.350   0.020
FAD       N3        C4        aromatic  1.350   0.020
FAD       O4        C4        single    1.362   0.020
FAD       C4        C4X       aromatic  1.390   0.020
```

```
FAD       C4X    N5        aromatic   1.350   0.020
FAD       C4X    C10       aromatic   1.390   0.020
FAD       N5     C5X       deloc      1.350   0.020
FAD       C5X    C6        aromatic   1.390   0.020
FAD       C5X    C9A       aromatic   1.390   0.020
FAD       C6     C7        aromatic   1.390   0.020
FAD       C7M    C7        single     1.506   0.020
FAD       C7     C8        aromatic   1.384   0.020
FAD       C8M    C8        single     1.506   0.020
FAD       C8     C9        aromatic   1.390   0.020
FAD       C9     C9A       aromatic   1.390   0.020
FAD       C9A    N10       aromatic   1.337   0.020
FAD       C10    N10       deloc      1.337   0.020
FAD       N10    'C1''     single     1.465   0.020
FAD       'C1''  'C2''     single     1.524   0.020
FAD       'O2''  'C2''     single     1.432   0.020
FAD       'C2''  'C3''     single     1.524   0.020
FAD       'O3''  'C3''     single     1.432   0.020
FAD       'C3''  'C4''     single     1.524   0.020
FAD       'O4''  'C4''     single     1.432   0.020
FAD       'C4''  'C5''     single     1.524   0.020
FAD       'C5''  'O5''     single     1.426   0.020
FAD       'O5''  P         single     1.610   0.020
FAD       O1P    P         deloc      1.510   0.020
FAD       P      O2P       deloc      1.510   0.020
FAD       O3P    P         single     1.610   0.020
FAD       H5B1   C5B       single     1.092   0.020
FAD       H5B2   C5B       single     1.092   0.020
FAD       H4B    C4B       single     1.099   0.020
FAD       H3B    C3B       single     1.099   0.020
FAD       HO3B   O3B       single     0.967   0.020
FAD       H2B    C2B       single     1.099   0.020
FAD       HO2B   O2B       single     0.967   0.020
FAD       H1B    C1B       single     1.099   0.020
FAD       H8A    C8A       single     1.083   0.020
FAD       HN6A   N6A       single     0.954   0.020
FAD       H2A    C2A       single     1.083   0.020
FAD       HO4    O4        single     0.967   0.020
FAD       H6     C6        single     1.083   0.020
FAD       H7M1   C7M       single     1.059   0.020
FAD       H7M2   C7M       single     1.059   0.020
FAD       H7M3   C7M       single     1.059   0.020
FAD       H8M1   C8M       single     1.059   0.020
FAD       H8M2   C8M       single     1.059   0.020
FAD       H8M3   C8M       single     1.059   0.020
FAD       H9     C9        single     1.083   0.020
FAD       'H1'1' 'C1''     single     1.092   0.020
FAD       'H1'2' 'C1''     single     1.092   0.020
FAD       'H2''  'C2''     single     1.099   0.020
FAD       'HO2'' 'O2''     single     0.967   0.020
FAD       'H3''  'C3''     single     1.099   0.020
FAD       'HO3'' 'O3''     single     0.967   0.020
FAD       'H4''  'C4''     single     1.099   0.020
FAD       'HO4'' 'O4''     single     0.967   0.020
FAD       'H5'1' 'C5''     single     1.092   0.020
FAD       'H5'2' 'C5''     single     1.092   0.020
loop_
_chem_comp_angle.comp_id
_chem_comp_angle.atom_id_1
_chem_comp_angle.atom_id_2
_chem_comp_angle.atom_id_3
_chem_comp_angle.value_angle
_chem_comp_angle.value_angle_esd
FAD       O2P    P      O1P     119.900   3.000
FAD       O2P    P      'O5''   108.200   3.000
FAD       O2P    P      O3P     108.200   3.000
FAD       O1P    P      'O5''   108.200   3.000
FAD       O1P    P      O3P     108.200   3.000
FAD       'O5''  P      O3P     102.600   3.000
FAD       P      'O5''  'C5''   120.500   3.000
FAD       'O5''  'C5''  'H5'1'  109.470   3.000
FAD       'O5''  'C5''  'H5'2'  109.470   3.000
FAD       'O5''  'C5''  'C4''   109.470   3.000
FAD       'H5'1' 'C5''  'H5'2'  107.900   3.000
```

FAD	'H5'1'	'C5''	'C4''	109.470	3.000
FAD	'H5'2'	'C5''	'C4''	109.470	3.000
FAD	'C5''	'C4''	'H4''	108.340	3.000
FAD	'C5''	'C4''	'O4''	109.470	3.000
FAD	'C5''	'C4''	'C3''	111.000	3.000
FAD	'H4''	'C4''	'O4''	109.470	3.000
FAD	'H4''	'C4''	'C3''	108.340	3.000
FAD	'O4''	'C4''	'C3''	109.470	3.000
FAD	'C4''	'O4''	'HO4''	109.470	3.000
FAD	'C4''	'C3''	'H3''	108.340	3.000
FAD	'C4''	'C3''	'O3''	109.470	3.000
FAD	'C4''	'C3''	'C2''	111.000	3.000
FAD	'H3''	'C3''	'O3''	109.470	3.000
FAD	'H3''	'C3''	'C2''	108.340	3.000
FAD	'O3''	'C3''	'C2''	109.470	3.000
FAD	'C3''	'O3''	'HO3''	109.470	3.000
FAD	'C3''	'C2''	'H2''	108.340	3.000
FAD	'C3''	'C2''	'O2''	109.470	3.000
FAD	'C3''	'C2''	'C1''	111.000	3.000
FAD	'H2''	'C2''	'O2''	109.470	3.000
FAD	'H2''	'C2''	'C1''	108.340	3.000
FAD	'O2''	'C2''	'C1''	109.470	3.000
FAD	'C2''	'O2''	'HO2''	109.470	3.000
FAD	'C2''	'C1''	'H1'1'	109.470	3.000
FAD	'C2''	'C1''	'H1'2'	109.470	3.000
FAD	'C2''	'C1''	N10	109.470	3.000
FAD	'H1'1'	'C1''	'H1'2'	107.900	3.000
FAD	'H1'1'	'C1''	N10	109.470	3.000
FAD	'H1'2'	'C1''	N10	109.470	3.000
FAD	'C1''	N10	C9A	120.000	3.000
FAD	'C1''	N10	C10	120.000	3.000
FAD	C9A	N10	C10	120.000	3.000
FAD	N10	C9A	C9	120.000	3.000
FAD	N10	C9A	C5X	120.000	3.000
FAD	C9	C9A	C5X	120.000	3.000
FAD	C9A	C9	H9	120.000	3.000
FAD	C9A	C9	C8	120.000	3.000
FAD	H9	C9	C8	120.000	3.000
FAD	C9	C8	C8M	120.000	3.000
FAD	C9	C8	C7	120.000	3.000
FAD	C8M	C8	C7	120.000	3.000
FAD	C8	C8M	H8M3	109.470	3.000
FAD	C8	C8M	H8M2	109.470	3.000
FAD	C8	C8M	H8M1	109.470	3.000
FAD	H8M3	C8M	H8M2	109.470	3.000
FAD	H8M3	C8M	H8M1	109.470	3.000
FAD	H8M2	C8M	H8M1	109.470	3.000
FAD	C8	C7	C7M	120.000	3.000
FAD	C8	C7	C6	120.000	3.000
FAD	C7M	C7	C6	120.000	3.000
FAD	C7	C7M	H7M3	109.470	3.000
FAD	C7	C7M	H7M2	109.470	3.000
FAD	C7	C7M	H7M1	109.470	3.000
FAD	H7M3	C7M	H7M2	109.470	3.000
FAD	H7M3	C7M	H7M1	109.470	3.000
FAD	H7M2	C7M	H7M1	109.470	3.000
FAD	C7	C6	H6	120.000	3.000
FAD	C7	C6	C5X	120.000	3.000
FAD	H6	C6	C5X	120.000	3.000
FAD	C9A	C5X	N5	120.000	3.000
FAD	C9A	C5X	C6	120.000	3.000
FAD	N5	C5X	C6	120.000	3.000
FAD	C5X	N5	C4X	120.000	3.000
FAD	N10	C10	C4X	120.000	3.000
FAD	N10	C10	N1	120.000	3.000
FAD	C4X	C10	N1	120.000	3.000
FAD	C10	C4X	C4	120.000	3.000
FAD	C10	C4X	N5	120.000	3.000
FAD	C4	C4X	N5	120.000	3.000
FAD	C4X	C4	O4	120.000	3.000
FAD	C4X	C4	N3	120.000	3.000
FAD	O4	C4	N3	120.000	3.000
FAD	C4	O4	HO4	109.470	3.000
FAD	C4	N3	C2	120.000	3.000

```
FAD     N3     C2     O2      120.000    3.000
FAD     N3     C2     N1      120.000    3.000
FAD     O2     C2     N1      120.000    3.000
FAD     C10    N1     C2      120.000    3.000
FAD     P      O3P    PA      120.500    3.000
FAD     O3P    PA     O1A     108.200    3.000
FAD     O3P    PA     O2A     108.200    3.000
FAD     O3P    PA     O5B     102.600    3.000
FAD     O1A    PA     O2A     119.900    3.000
FAD     O1A    PA     O5B     108.200    3.000
FAD     O2A    PA     O5B     108.200    3.000
FAD     PA     O5B    C5B     120.500    3.000
FAD     O5B    C5B    H5B1    109.470    3.000
FAD     O5B    C5B    H5B2    109.470    3.000
FAD     O5B    C5B    C4B     109.470    3.000
FAD     H5B1   C5B    H5B2    107.900    3.000
FAD     H5B1   C5B    C4B     109.470    3.000
FAD     H5B2   C5B    C4B     109.470    3.000
FAD     C5B    C4B    H4B     108.340    3.000
FAD     C5B    C4B    C3B     111.000    3.000
FAD     C5B    C4B    O4B     109.470    3.000
FAD     H4B    C4B    C3B     108.340    3.000
FAD     H4B    C4B    O4B     109.470    3.000
FAD     C3B    C4B    O4B     109.470    3.000
FAD     C4B    C3B    H3B     108.340    3.000
FAD     C4B    C3B    O3B     109.470    3.000
FAD     C4B    C3B    C2B     111.000    3.000
FAD     H3B    C3B    O3B     109.470    3.000
FAD     H3B    C3B    C2B     108.340    3.000
FAD     O3B    C3B    C2B     109.470    3.000
FAD     C3B    O3B    HO3B    109.470    3.000
FAD     C3B    C2B    H2B     108.340    3.000
FAD     C3B    C2B    O2B     109.470    3.000
FAD     C3B    C2B    C1B     111.000    3.000
FAD     H2B    C2B    O2B     109.470    3.000
FAD     H2B    C2B    C1B     108.340    3.000
FAD     O2B    C2B    C1B     109.470    3.000
FAD     C2B    O2B    HO2B    109.470    3.000
FAD     C2B    C1B    H1B     108.340    3.000
FAD     C2B    C1B    O4B     109.470    3.000
FAD     C2B    C1B    N9A     109.470    3.000
FAD     H1B    C1B    O4B     109.470    3.000
FAD     H1B    C1B    N9A     109.470    3.000
FAD     O4B    C1B    N9A     109.470    3.000
FAD     C1B    O4B    C4B     111.800    3.000
FAD     C1B    N9A    C4A     126.000    3.000
FAD     C1B    N9A    C8A     126.000    3.000
FAD     C4A    N9A    C8A     108.000    3.000
FAD     N9A    C4A    C5A     108.000    3.000
FAD     N9A    C4A    N3A     132.000    3.000
FAD     C5A    C4A    N3A     120.000    3.000
FAD     C4A    C5A    N7A     108.000    3.000
FAD     C4A    C5A    C6A     120.000    3.000
FAD     N7A    C5A    C6A     132.000    3.000
FAD     C5A    N7A    C8A     108.000    3.000
FAD     N7A    C8A    H8A     126.000    3.000
FAD     N7A    C8A    N9A     108.000    3.000
FAD     H8A    C8A    N9A     126.000    3.000
FAD     C4A    N3A    C2A     120.000    3.000
FAD     N3A    C2A    H2A     120.000    3.000
FAD     N3A    C2A    N1A     120.000    3.000
FAD     H2A    C2A    N1A     120.000    3.000
FAD     C2A    N1A    C6A     120.000    3.000
FAD     N1A    C6A    N6A     120.000    3.000
FAD     N1A    C6A    C5A     120.000    3.000
FAD     N6A    C6A    C5A     120.000    3.000
FAD     C6A    N6A    HN6A    120.000    3.000
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
```

```
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 FAD      var_1    O2P    P      'O5''  'C5''    175.000   20.000   1
 FAD      var_2    P      'O5''  'C5''  'C4''    175.000   20.000   1
 FAD      var_3    'O5''  'C5''  'C4''  'C3''    175.000   20.000   3
 FAD      var_4    'C5''  'C4''  'O4''  'HO4''   175.000   20.000   1
 FAD      var_5    'C5''  'C4''  'C3''  'C2''    175.000   20.000   3
 FAD      var_6    'C4''  'C3''  'O3''  'HO3''   175.000   20.000   1
 FAD      var_7    'C4''  'C3''  'C2''  'C1''    175.000   20.000   3
 FAD      var_8    'C3''  'C2''  'O2''  'HO2''   175.000   20.000   1
 FAD      var_9    'C3''  'C2''  'C1''  N10      175.000   20.000   3
 FAD      var_10   'C2''  'C1''  N10    C10      175.000   20.000   1
 FAD      CONST_1  'C1''  N10    C9A    C5X      180.000    0.000   0
 FAD      CONST_2  N10    C9A    C9     C8       180.000    0.000   0
 FAD      CONST_3  C9A    C9     C8     C7         0.000    0.000   0
 FAD      var_11   C9     C8     C8M    H8M1     175.000   20.000   1
 FAD      CONST_4  C9     C8     C7     C6         0.000    0.000   0
 FAD      var_12   C8     C7     C7M    H7M1     175.000   20.000   1
 FAD      CONST_5  C8     C7     C6     C5X        0.000    0.000   0
 FAD      CONST_6  N10    C9A    C5X    N5         0.000    0.000   0
 FAD      CONST_7  C9A    C5X    C6     C7         0.000    0.000   0
 FAD      CONST_8  C9A    C5X    N5     C4X        0.000    0.000   0
 FAD      CONST_9  'C1''  N10    C10    N1         0.000    0.000   0
 FAD      CONST_10 N10    C10    C4X    C4       180.000    0.000   0
 FAD      CONST_11 C10    C4X    N5     C5X        0.000    0.000   0
 FAD      CONST_12 C10    C4X    C4     N3         0.000    0.000   0
 FAD      var_13   C4X    C4     O4     HO4      175.000   20.000   1
 FAD      CONST_13 C4X    C4     N3     C2         0.000    0.000   0
 FAD      CONST_14 C4     N3     C2     O2       180.000    0.000   0
 FAD      CONST_15 N10    C10    N1     C2       180.000    0.000   0
 FAD      CONST_16 C10    N1     C2     N3         0.000    0.000   0
 FAD      var_14   O2P    P      O3P    PA       175.000   20.000   1
 FAD      var_15   P      O3P    PA     O5B      175.000   20.000   1
 FAD      var_16   O3P    PA     O5B    C5B      175.000   20.000   1
 FAD      var_17   PA     O5B    C5B    C4B      175.000   20.000   1
 FAD      var_18   O5B    C5B    C4B    C3B      175.000   20.000   3
 FAD      var_19   C5B    C4B    O4B    C1B     -150.000   20.000   1
 FAD      var_20   C5B    C4B    C3B    C2B      180.000   20.000   3
 FAD      var_21   C4B    C3B    O3B    HO3B     175.000   20.000   1
 FAD      var_22   C4B    C3B    C2B    C1B      -60.000   20.000   3
 FAD      var_23   C3B    C2B    O2B    HO2B     175.000   20.000   1
 FAD      var_24   C3B    C2B    C1B    N9A      150.000   20.000   3
 FAD      var_25   C2B    C1B    O4B    C4B        0.000   20.000   1
 FAD      var_26   C2B    C1B    N9A    C4A      175.000   20.000   1
 FAD      CONST_17 C1B    N9A    C8A    N7A      180.000    0.000   0
 FAD      CONST_18 C1B    N9A    C4A    N3A        0.000    0.000   0
 FAD      CONST_19 N9A    C4A    C5A    N7A        0.000    0.000   0
 FAD      CONST_20 C4A    C5A    C6A    N1A        0.000    0.000   0
 FAD      CONST_21 C4A    C5A    N7A    C8A        0.000    0.000   0
 FAD      CONST_22 C5A    N7A    C8A    N9A        0.000    0.000   0
 FAD      CONST_23 N9A    C4A    N3A    C2A      180.000    0.000   0
 FAD      CONST_24 C4A    N3A    C2A    N1A        0.000    0.000   0
 FAD      CONST_25 N3A    C2A    N1A    C6A        0.000    0.000   0
 FAD      CONST_26 C2A    N1A    C6A    N6A      180.000    0.000   0
 FAD      CONST_27 N1A    C6A    N6A    HN6A     180.000    0.000   0
loop_
_chem_comp_chir.comp_id
_chem_comp_chir.id
_chem_comp_chir.atom_id_centre
_chem_comp_chir.atom_id_1
_chem_comp_chir.atom_id_2
_chem_comp_chir.atom_id_3
_chem_comp_chir.volume_sign
 FAD      chir_01  C4B    C5B    O4B    C3B      negativ
 FAD      chir_02  C3B    C4B    O3B    C2B      negativ
 FAD      chir_03  C2B    C3B    O2B    C1B      negativ
 FAD      chir_04  C1B    O4B    C2B    N9A      positiv
 FAD      chir_05  'C2''  'C1''  'O2''  'C3''    positiv
 FAD      chir_06  'C3''  'C2''  'O3''  'C4''    positiv
 FAD      chir_07  'C4''  'C3''  'O4''  'C5''    positiv
loop_
_chem_comp_plane_atom.comp_id
_chem_comp_plane_atom.plane_id
```

```
_chem_comp_plane_atom.atom_id
_chem_comp_plane_atom.dist_esd
 FAD     plan-1    N9A      0.020
 FAD     plan-1    C1B      0.020
 FAD     plan-1    C8A      0.020
 FAD     plan-1    C4A      0.020
 FAD     plan-1    N7A      0.020
 FAD     plan-1    H8A      0.020
 FAD     plan-1    C5A      0.020
 FAD     plan-1    C6A      0.020
 FAD     plan-1    N1A      0.020
 FAD     plan-1    C2A      0.020
 FAD     plan-1    N3A      0.020
 FAD     plan-1    N6A      0.020
 FAD     plan-1    HN6A     0.020
 FAD     plan-1    H2A      0.020
 FAD     plan-2    N1       0.020
 FAD     plan-2    C2       0.020
 FAD     plan-2    C10      0.020
 FAD     plan-2    N3       0.020
 FAD     plan-2    C4       0.020
 FAD     plan-2    O2       0.020
 FAD     plan-2    O4       0.020
 FAD     plan-2    C4X      0.020
 FAD     plan-2    N5       0.020
 FAD     plan-2    N10      0.020
 FAD     plan-2    C5X      0.020
 FAD     plan-2    C6       0.020
 FAD     plan-2    C9A      0.020
 FAD     plan-2    C7       0.020
 FAD     plan-2    C8       0.020
 FAD     plan-2    C9       0.020
 FAD     plan-2    H6       0.020
 FAD     plan-2    C7M      0.020
 FAD     plan-2    C8M      0.020
 FAD     plan-2    H9       0.020
 FAD     plan-2    'C1''    0.020
#
data_comp_CSS
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.partial_charge
 CSS         OXT   O    OC       -0.500
 CSS         C     C    C         0.000
 CSS         O     O    OC       -0.500
 CSS         CA    C    CH1       0.000
 CSS         HA    H    H         0.000
 CSS         N     N    NH2       0.000
 CSS         HN2   H    H         0.000
 CSS         HN1   H    H         0.000
 CSS         CB    C    CH2       0.000
 CSS         HB1   H    H         0.000
 CSS         HB2   H    H         0.000
 CSS         SG    S    S2        0.000
 CSS         SD    S    SH1       0.000
 CSS         HD    H    H         0.000
loop_
_chem_comp_tree.comp_id
_chem_comp_tree.atom_id
_chem_comp_tree.atom_back
_chem_comp_tree.atom_forward
_chem_comp_tree.connect_type
 CSS         OXT   n/a    C     START
 CSS         C     OXT    CA     .
 CSS         O     C      .      .
 CSS         CA    C      CB     .
 CSS         HA    CA     .      .
 CSS         N     CA     HN1    .
 CSS         HN2   N      .      .
 CSS         HN1   N      .      .
```

```
CSS        CB        CA       SG       .
CSS        HB1       CB       .        .
CSS        HB2       CB       .        .
CSS        SG        CB       SD       .
CSS        SD        SG       HD       .
CSS        HD        SD       .        END
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
CSS        N         CA       single   1.450    0.020
CSS        HN1       N        single   1.010    0.020
CSS        HN2       N        single   1.010    0.020
CSS        CB        CA       single   1.524    0.020
CSS        CA        C        single   1.500    0.020
CSS        HA        CA       single   1.099    0.020
CSS        SG        CB       single   1.762    0.020
CSS        HB1       CB       single   1.092    0.020
CSS        HB2       CB       single   1.092    0.020
CSS        SD        SG       single   2.025    0.020
CSS        HD        SD       single   1.330    0.020
CSS        O         C        deloc    1.250    0.020
CSS        C         OXT      deloc    1.250    0.020
loop_
_chem_comp_angle.comp_id
_chem_comp_angle.atom_id_1
_chem_comp_angle.atom_id_2
_chem_comp_angle.atom_id_3
_chem_comp_angle.value_angle
_chem_comp_angle.value_angle_esd
CSS        OXT       C        O        123.000  3.000
CSS        OXT       C        CA       118.500  3.000
CSS        O         C        CA       118.500  3.000
CSS        C         CA       HA       108.810  3.000
CSS        C         CA       N        109.470  3.000
CSS        C         CA       CB       109.470  3.000
CSS        HA        CA       N        109.470  3.000
CSS        HA        CA       CB       108.340  3.000
CSS        N         CA       CB       109.470  3.000
CSS        CA        N        HN2      120.000  3.000
CSS        CA        N        HN1      120.000  3.000
CSS        HN2       N        HN1      120.000  3.000
CSS        CA        CB       HB1      109.470  3.000
CSS        CA        CB       HB2      109.470  3.000
CSS        CA        CB       SG       109.500  3.000
CSS        HB1       CB       HB2      107.900  3.000
CSS        HB1       CB       SG       109.500  3.000
CSS        HB2       CB       SG       109.500  3.000
CSS        CB        SG       SD       104.454  3.000
CSS        SG        SD       HD       109.500  3.000
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
CSS        var_1     OXT      C        CA       CB   175.000  20.000   3
CSS        var_2     C        CA       N        HN1  175.000  20.000   1
CSS        var_3     C        CA       CB       SG    62.578  20.000   3
CSS        var_4     CA       CB       SG       SD    58.719  20.000   1
CSS        var_5     CB       SG       SD       HD   175.000  20.000   1
loop_
_chem_comp_chir.comp_id
_chem_comp_chir.id
_chem_comp_chir.atom_id_centre
_chem_comp_chir.atom_id_1
_chem_comp_chir.atom_id_2
```

```
_chem_comp_chir.atom_id_3
_chem_comp_chir.volume_sign
 CSS      chir_01 CA    N     CB     C         negativ
loop_
_chem_comp_plane_atom.comp_id
_chem_comp_plane_atom.plane_id
_chem_comp_plane_atom.atom_id
_chem_comp_plane_atom.dist_esd
 CSS      plan-1    N         0.020
 CSS      plan-1    CA        0.020
 CSS      plan-1    HN1       0.020
 CSS      plan-1    HN2       0.020
 CSS      plan-2    C         0.020
 CSS      plan-2    CA        0.020
 CSS      plan-2    O         0.020
 CSS      plan-2    OXT       0.020
# -------------------------------------------------------
# -------------------------------------------------------
#
# --- DESCRIPTION OF LINKS ---
#
data_link_  1.61
#
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
   1.61   1 C8M    2 S        .          1.610    0.020
#
data_link_  1.62
#
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
   1.62   1 C8M    2 S        .          1.610    0.020
#
data_link_  2.04
#
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
   2.04   1 SG     2 S        .          2.036    0.020
#
data_link_  1.94
#
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
   1.94   1 SD     2 S2       .          2.036    0.020
#
data_link_  1.93
```

```
#
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
  1.93  1 SD      2 S2        .           2.036   0.020
#
data_link_  1.33
#
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
  1.33  1 C       2 N         .           1.260   0.020
#
data_link_  1.34
#
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
  1.34  1 C       2 N         .           1.260   0.020
# -------------------------------------------------------
```

Appendix table 33: Ramachandran plot[a]
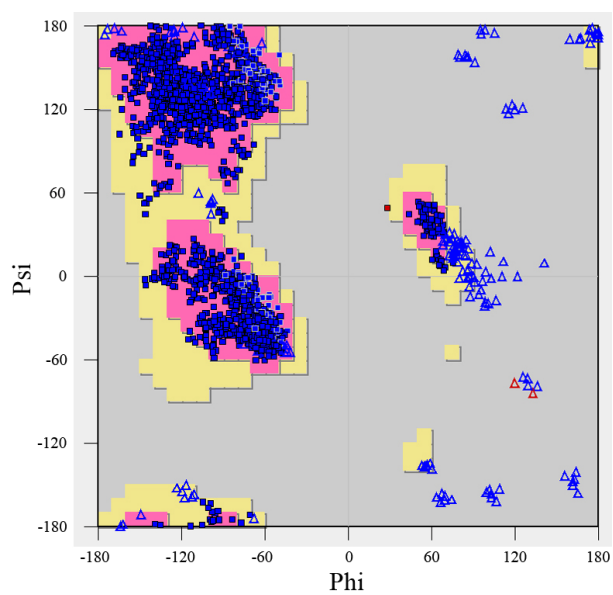
[a] for the "as-purified" SQR structure (PDB id.: 3HYV).

Appendix table 34: sponge phase crystallization screen[a]

Precipitant 1 (% w/V)		Precipitant 2 (M)		Buffer (mM)		pH	Additive (%)		notes
Jeffamine									
Jeffamine M600	20	$(NH_4)_2SO_4$	0.7	HEPES	100	8.1	1.2.3-heptanetriol	2.5	
Jeffamine ED600	12			MES	100	6.0			b
Jeffamine ED600	14			MES	100	6.0			b
Jeffamine ED600	14			Tris	100	8.0			b
Jeffamine ED600	16			MES	100	6.0			b
Jeffamine ED600	16			Tris	100	8.0			b
Jeffamine ED600	18			MES	100	6.0			b
Jeffamine ED600	18			HEPES	100	7.0			b
Jeffamine ED600	18			Tris	100	8.0			b
Jeffamine ED600	20			MES	100	6.0			b
Jeffamine ED600	20			Tris	100	8.0			b
Jeffamine ED900	12			MES	100	6.0			b
Jeffamine ED900	14			MES	100	6.0			b
Jeffamine ED900	16			MES	100	6.0			b
Jeffamine ED900	18			MES	100	6.0			b
Jeffamine ED900	20			MES	100	6.0			b
PEG									
PEG 400	40			MES	100	6.5			
PEG 400	40	$MgCl_2$	0.45						
PEG 400	42	$MgCl_2$	0.2						
PEG 400	42	$MgCl_2$	0.25						
PEG 400	42	$MgCl_2$	0.3						
PEG 400	42	$MgCl_2$	0.4						
PEG 400	45	$MgCl_2$	0.2						
PEG 400	45	$MgCl_2$	0.25						
PEG 400	45	$MgCl_2$	0.3						
PEG 400	48			HEPES	100	7.5	$CaCl_2$	200	JBM1_A5[c]
PEG 550 MME	40	$MgCl_2$	0.4						
PEG 550 MME	40	$MgCl_2$	0.8						
PEG 550 MME	50	$MgCl_2$	0.4						
MPD									
MPD	5	Ethanol	5	HEPES	100	7.5			JB8_B5
MPD	10			HEPES	100	7.5	Na-citrate	100	JB7_A1
MPD	12			Tris	100	8.5	$MgCl_2$	50	JB7_A2
MPD	15			Na-acetate	100	4.6	$CaCl_2$	20	JB7_A3

MPD	15			Na-citrate	100	5.6	$(NH_4)CH_3COO$	200	JB7_A5
MPD	15			MES	100	6.5	Mg-acetate	200	JB7_A6
MPD	15			HEPES	100	7.5	Na-citrate	200	JB7_B1
MPD	20			HEPES	100	7.5	Na-citrate	100	JB7_B2
MPD	20			Imidazole	100	8.0			JB7_B3
MPD	20	Glycerol	4				NaCl	200	JB7_B4
MPD	30			Na-citrate	100	5.6	(NH_4)-acetate	200	JB7_B6
MPD	30			HEPES	100	7.5	$(NH_4)_2SO_4$	500	JB7_C2
MPD	30			HEPES	100	7.5	Na-citrate	200	JB7_C3
MPD	30	PEG 4000	5	HEPES	100	7.5			JB7_C4
MPD	50			Tris	100	8.5	$NH_4H_2PO_4$	100	JB8_A3

[a] the screen was optimized in collaboration with Sabine Buschmann and Dr. Chitra Rajendran (Max Planck Institute of Biophysics, Frankfurt am Main, Germany).

[b] the buffer used to prepare the lipidic cubic phase was not MES 20 mM, pH 6.5, DDM 0.02 %, but MES 20 mM, pH 6.5, DDM 0.02 %, NaCl 100 mM.

[c] JBM stands for Jena Bioscience Membrane Screen, JB for Jena Bioscience Classic Screen.

Appendix table 35: command line for T-COFFEE[a]

```
t_coffee  -pdb_db pdb
          -protein_db UniRef100
          -blast LOCAL
          -mode expresso
          -in  sequences_fasta.in0
          -evaluate_mode t_coffee_slow
          -output clustalw_aln fasta_aln phylip score_html score_pdf
          -maxnseq 50
          -maxlen 2000
          -tree
          -run_name=tcfEXPR63893_7734
          -check_pdb_status
          -cache=no
          -remove_template_file=1
          -quiet=stdout
          >tcfEXPR63893_7734.tc_LOG
```

[a] the run was performed online at http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi.

Appendix table 36: template file used by T-COFFEE

```
>gi153872492__Beggiatoa _P_ 1fcdA mol:protein length:401  FCSD
>gi|255021328_Acaldus
>gi21673852___Ctepidum
>gi13542037___Tvolcanium _P_ 3h8iA mol:protein length:409  NADH oxidase
>gi15922815___Stokodaii _P_ 3h8iA mol:protein length:409  NADH oxidase
>gi16082149___Tacidophilum _P_ 3h8iA mol:protein length:409  NADH oxidase
>gi24657386___Dmelanogaster
>gi21673913___Ctepidum
>gi157092622__Amarina
>gi15899356___Ssolfataricus
>gi15607118___Aaeolicus _P_ 3hyvA mol:protein length:430  Sulfide-quinone reductase
>gi192289979_Rpalustris _P_ 1fcdA mol:protein length:401  FCSD
>gi15920835___Stokodaii
>gi8118249____Olimnetica _P_ 3hyvA mol:protein length:430  Sulfide-quinone reductase
>gi17546261___Rsolanacearum
>gi86748177___Rpalustris
>gi19112859___Spombe
>gi193211736_Cparvum
>gi21436536___Pdenitrificans _P_ 1fcdA mol:protein length:401  FCSD
>gi145218949_Pvibrioformis
>gi21672958___Ctepidum
>gi78364011___Tcrunogena
>gi38505571___SynechocystisPCC6803
>gi15605781___Aaeolicus
>gi17232004___NostocPCC7120 _P_ 3hyvA mol:protein length:430  Sulfide-quinone reductase
>gi8118252____Ahalophytica _P_ 3hyvA mol:protein length:430  Sulfide-quinone reductase
>gi19879585___Climicola _P_ 1fcdA mol:protein length:401  FCSD
>gi|21673704__Ctepidum
>gi241777285__Avinosum _P_ 1fcdA mol:protein length:401  FCSD
>gi225850055__Pmarina
>gi23015306___Mmagnetotacticum
>gi2385380____Rcapsulatus _P_ 3hyvA mol:protein length:430  Sulfide-quinone reductase
>gi78776818___Sdenitrificans
>gi15606162___Aaeolicus
>gi|193212535_Cparvum
>gi82749809___Saureus
>gi148696168__Mmusculus
>gi21464540___Tdenitrificans _P_ 3hyvA mol:protein length:430  Sulfide-quinone reductase
>gi15597541___Paeruginosa
>gi238828305__Aambivalens _P_ 3h8iA mol:protein length:409  NADH oxidase
>gi78186989___Pluteolum
>gi18313472___Paerophilum
>gi|256759353_Tintermedia
>gi62896485___Hsapiens
>gi241775880__Avinosum
>gi11498167___Afulgidus
>gi15899031___Ssolfataricus _P_ 3h8iA mol:protein length:409  NADH oxidase
>gi103492069__Aferrooxidans _P_ 3hyvA mol:protein length:430  Sulfide-quinone reductase
>gi229580793__Sislandicus _P_ 3h8iA mol:protein length:409  NADH oxidase
>gi78185961___Pluteolum
```

Appendix table 37: full sequence alignment generated by T-COFFEE

gi15607118___Aa FEDISVPLAPLLPKFNIEFI---NEKAESIDPDANTVT-------
gi21464540___Td RDDITLEAAPFLNKKNIDFI---GVGAARVHPEKNQVE-------
gi2385380____Rc RKEITVDLAPTMARKNIDFI---PVAAKKLHPAENRVE-------
gi8118249____Ol LERVQLDVGKLLKGRNIDWI---HGKVNHIDPENKTLV-------
gi103492069__Af RDDIAFPIRHYVERKGIHFI---AQEAEQIDAEAQNIT-------
gi17232004___No LESIQVSLQQRLKQKGINWI---LGRVDYLNPQNQKIS-------
gi38505571___Sy LDHMQLNVPMLAKRHGFKWV---QQKVQEIDPLQKRIY-------
gi8118252____Ah LEDVQLPLAPLLARQGINWQ---HGRVTGLDPNQKRVS-------
gi19112859___Sp VAKTRRELASLVPADKFKLH---PEFVKSLHPRENKIV-------
gi24657386___Dm LDQSHRQMADVLP-TKAKWV---KEKALSFDPDNNTVS-------
gi62896485___Hs LSSSGRPTASVIP-SGVEWI---KARVTELDPDKNCIH-------
gi148696168__Mm LSLSVRSTLSVIP-SGVQWI---QDRVAELMPDENCIR-------
gi157092622__Am LDQSAQPMNKVLP-GACDWI---KDSAVQFDPDNNTVF-------
gi82749809___Sa LKSSRKDMESVIP-EGAHWI---KQAVGSFQPENNSVI-------
gi15597541___Pa QGDTARPMAGLVP-PGVEWL---RTRVERVDPEARRLL-------
gi17546261___Rs AGRTVRQTDALIP-AGVRRI---RASATHIDTETHRVS-------
gi15920835___St PIELKRKESELLD-PRIKFLHGLKGTITKIDVANHSVI-------
gi21673852___Ct SSTLVKSRKKYIT-AGINFV---HDEITHIDDPEKKEVK------
gi11498167___Af ESEILKDVKYLLD-PLIELK---IARATKIDAEKSRVE-------
gi23015306___Mm EAELFRDQRKVLD-KRVNFF---VDPAKNIDVEKNQVT-------
gi15899356___Ss PGELVRKESELLD-PRIKFLHGLKGTISKIDVANHSVQ-------
gi18313472___Pa PDVMFKPVSKLVS-PLVKVE---REPAAKIDLSAKKIV-------
gi21672958___Ct IDDVRFKLKKVYDRWGIDYK---QAKAVSIHPEGDANISKGYVTI
gi193211736__Cp IDDVRFKLHKVYDRWGIDFK---QAKAVSIHPEGDASIDKGYVTI
gi241775880__Av ERQVTFELAPVYKKINVGFR---QARAVSIHPDGGAGHESPFVTI
gi145218949__Pv IDDVRFELKKVYDRWGIVYK---QAKALEIHPEGDASTGRGFVTI
gi78185961___Pl IDEVRFELKKVYDRWGIVFK---QAKALEIHPEGDASTGRGYVTI
gi78776818___Sd KEDVTFDLAPVYSKAGINFK---QAKALSINPEGSESGKPFVTI
gi238828305__Aa VDELKVDLSEALPEKGIQPQ---GGTVEKIDAKSSMVY-------
gi229580793__Si EDDVKVDLTKALPEKRMNFQ---QGTVEKIDASGGLVY-------
gi15899031___Ss ENDVMVDLTKALPEKRINFQ---QGTVEKIDASGGLVY-------
gi15922815___St VEQLKIDLAQALPEKGIQFQ---NGTVLKIDAKNNKVV-------
gi13542037___Tv PDDLRIDISEALSAKGIEYI---QGTVTRIDPAKNTVE-------
gi16082149___Ta AEDLRIDLAQVLPDRGIGYI---QGTVTKIDPERSLVE-------
gi21673913___Ct FRNVAFPLEKVARKHGFSLT---LDTVTAIHASRDSVT-------
gi15606162___Aa FEDVIIPLDKFARRWGVNLI---LDRVEKIDAQNKKVH-------
gi78186989___Pl YSDASIPLKKLARRHGPTLT---LDRVTAIDAESERVW-------
gi225850055__Pm LDSISIPLKDLSESHGFKVT---VDEVERISASDMKVY-------
gi78364011___Tc ADDLKIDLTNYFNRQTVEHI---EASVTDITDEGRVVH-------
gi86748177___Rp GARLKVPLGHFFEKYRIDFV---EASVTGLKDIQGRVI-------
gi241777285__Av LESIKHGYDGLR-AHGIQVV---HDSATGIDPDKKLVK-------
gi153872492__Be LDYIQFGYDSLA-KYGINVV---HSRVTGTDSVAKKVT-------
gi19879585___Cl MKDITHTYTALRTKHGVNYI---ADRVVSIDAAKGTLK-------
gi192289979__Rp LSDITHQYYALK-KRGINFI---HDWADAIEPDKVKVH-------
gi15605781___Aa LEFLIHDFLTPAAKYGYTF---NATVTDIDRDKRRVY-------
gi21436536___Pd FESLQHGYDKLS-AAGVAVV---NDTAVAVDRAARTVT-------
gi|21673704__Ct DSSMQLPLAPVLDKDETRFI---QAEVLDIDVESKKVT-------
gi|193212535_Cp DSSMQLPLAPVLDKDETRFI---QAEVLDIDVESKKVT-------
gi|255021328_Ac AKTLSMDLRSTLABDDIGFI---QAEVQALDPDAKRID-------
gi|256759353_Ti QLTMTMDILEPLHDIGARFI---QGEVTALDVDAKRVD-------

cons

gi15607118___Aa -------TQSGKK-IEYDYLVIATGPKLVFGA-EGQEE------
gi21464540___Td -------LTNGTI-VDYDFLVIATGPKLAFDEVPGLGPQG---G
gi2385380____Rc -------LENGQS-VSYDQIVIATGPELAFDEIEGFGPEG----
gi8118249____Ol -------AGE-QT-LEYDYVVVATGPELATDAIAGLGPEN---G
gi103492069__Af -------LADGNT-VHYDYLNIATGPKLAFENVPGSDPHE---G
gi17232004___No -------LGE-QS-ISYDYLIIATGAELALDAVAGLGPDG----
gi38505571___Sy -------TH--ET-LNYDYLIASGAELANDLVPGLGI-------
gi8118252____Ah -------VGEDIT-FTYDYLVITTGASLAYHLMSGLGPEE---G
gi19112859___Sp -------TQSGQE-ISYDYLVMAAGIYTDFGRIKGLTEALD--D
gi24657386___Dm -------TSGGKT-IKYDFLIIATGLQLNYGKIPGLVEALE--T
gi62896485___Hs -------TDDDEK-ISYRYLIIALGIQLDYEKIKGLPEGFA--H
gi148696168__Mm -------TDSGKE-ISYRYLIIALGIQLDYEKIKGLPEGFA--Y
gi157092622__Am -------TKHGYE-VKYDFLVCAMGLQLNYHLIKGLPEAFD--V
gi82749809___Sa -------LGDNTV-VYYDFLVVAPGLQINWSSIKGLKENIG--K
gi15597541___Pa --------LEGGDS-LKYRNLIVCPGLRLAWERIEGLEETLG--R
gi17546261___Rs --------LDDGRS-VKYSQLIVSPGLRLAWERIDGLEAALG--Q
gi15920835___St --------TADGLS-HKYDYLVITTGVEYTWDEVPGYRSA-----
gi21673852___Ct --------TK-NHT-FTYDFLVISTGCRIAPEENEGLMEAWG--K
gi11498167___Af --------LENGEV-VNYDYLVIATGPKLAFAEEMPGLVDG----
gi23015306___Mm --------TEAGRT-FKYDYLVLATGSRIMPQNIPGNTEG-----
gi15899356___Ss --------TADGVS-HSYDYLVITTGVDYTWDEIPGYRSA-----
gi18313472___Pa --------LQSGKE-LKYDYLVATCAVVKTGSYPGPGEA-----
gi21672958___Ct EYTDEEH-AGYTET-VDYDYLVNATGPKLNFEATEGLGPDK---
gi193211736__Cp EYTDPEY-AGQTEI-VDYDYLVNATGPKLNFEATEGLGPDK----
gi241775880__Av EYTDPTR-AGQSDE-IEYDYLVNATGPKLNFDATPGLGPET---G
gi145218949__Pv EYTDADN-AGTREN-VEYDYLVNATGPKLNFEATEGLGPDQ----
gi78185961___Pl EYTAGDN-SGVREN-VEYDYLVNATGPKLNFEATEGLGPDK----
gi78776818___Sd EFTSKGN-EGTREN-IEYDYLINATGPKLNFGATPGLGEGSELGE
gi238828305__Aa -YTK---PDGSMAE-EEYDYVIVGIGAHLATELVKGWDKY-----
gi229580793__Si -YTK---PDGNKTE-EDYDLIIALGAHLGTELVKGWDKY-----
gi15899031___Ss -YTK---PDGDKAE-EDYDYLIIALGAHLGTELVKGWDKY-----
gi15922815___St -YRK---PDGSETE-REYDYVIVAIGAHLATELVKGWDQY-----
gi13542037___Tv -YNK---EGAGPSK-EKYDYLVIALGSHLATELITGPEKF-----
gi16082149___Ta -YSR---EGSGGYK-EKYDFLVIALGAHLASELIAGFDKY-----
gi21673913___Ct -LE------KAGQR-SDPDFLVIALGSDKMK--HEGIEH------
gi15606162___Aa -LK------SGKVLEDPDYLVVALGQTKRK--EKGLEN------
gi78186989___Pl -LE------KGGER-SGTDAVVVALGAAKIQ--YPGMEH-----
gi225850055__Pm -LK------NSGER-SDPDYLVIALGQSKIK--HEGIEH-----
gi78364011___Tc -------TDKG-D-FNNNGLIIASGGRFI-KKLPGIE-HVI---
gi86748177___Rp -------TDRG-D-FSNDHLIVACGARFI-RKLPGIE-HAL---
gi241777285__Av -------TAGGAE-FGYDRCVVAPGIELIYDKIEGYSEEAA--A
gi153872492__Be -------TQNGNT-FTYKRLIVAPGIDFKWBAIDGYNERIA--E
gi19879585___Cl -------LAGGRV-ICYDRLVVSPGIDFKYDAIPGYSQKIA--E
gi192289979__Rp -------LRSGKT-IAYDKLIVSFGIGIKWGALPGYDQDAA--Q
gi15605781___Aa -------TEDG-Y-VEYDYLILAPGIRYNYDAWFNGDKDMA--R
gi21436536___Pd -------LAGGQV-LPYDRLVLSFGIHFKPDSVPGWSLEAA--E
gi|21673704__Ct -IQPTERPGTTTDV-LTYDYLVIALGNRLAFDKIDGFAEH-----
gi|193212535_Cp -IQPTERPGATTDV-LTYDYLVIALGNRLAFDKIDGFAEH-----
gi|255021328_Ac -FVPSERPGAAPES-MEYDYVVVAVGNRLAFDRIEGFAEH-----
gi|256759353_Ti -YVPSERPGAAPTS-TTYDYLVALGNRLAFDKIEGFAQF-----

cons

gi15607118___Aa N---ST--SI--C-TAEHALETQKK----LQE--LYAN-PGPVVI
gi21464540___Td H---TQ--SV--C-SVDHAVNAGRA----WEE--FVRN-PGPIVV
gi2385380____Rc H---TQ--SI--C-HIDHAEAAGAA----FDR--FCEN-PGPIII
gi8118249____Ol Y---TQ--SV--C-NPHHALMAKEA----WQK--FLQD-PGPIVV
gi103492069__Af P---VQ--SV--C-TVDHAERAFAE----YQA--LLRE-PGPIVI
gi17232004___No Y---TQ--SV--C-NPHHAIKAFQA----WQN--FLLA-PGPIVI
gi38505571___Sy Y---TH--SI--C-TPSHVLEARTA----WET--FCQQ-PGDLVV
gi8118252____Ah Y---TQ--SV--C-NAHHAEMARDA----WDE--FLEN-PGPLLV
gi19112859___Sp PNTPVV--TI--Y-SEKYADA----VYPWIEK-TKS---GNAIF
gi24657386___Dm PDSNVC--SI--Y-SPKYVDH----VYECLRR-TNK---GNAIF
gi62896485___Hs --PKIG--SN--Y-SVKTVEK----TWKALQD-FKE---GNALF
gi148696168__Mm --PKIG--SN--Y-SVKTVEK----TWKALQD-FKE---GNALF
gi157092622__Am -DPSLC--SN--Y-WDKTVIK----TRPAMEA-VTE---GNVIF
gi82749809___Sa --NGVC--SN--Y-SPDYVNE----TWNQISN-FKQ---GNAIF
gi15597541___Pa N--GVT--SN--Y-RYDLAPY----TWELVRG-LRG---GKALF
gi17546261___Rs H--GVT--SN--Y-RFDLAPY----TWSLVRS-LRS---GRALF
gi15920835___St ----AL--SP--Y-GIKAAEE----MREALEK-FQG---GNAIF
gi21673852___Ct N---AF--TF--Y-YKEAADQ----LRLRLKE-FDG---GKLVM
gi11498167___Af ----GH--WF--Y-NLDGAKR----LREELAR-FKK---GKVVV
gi23015306___Mm ----GH--WF--Y-DLDGAIK----LRKALNE-FQG---GKIVV
gi15899356___Ss ----GH--WF--Y-GIDDAIK----MREALEK-FQG---GTIVV
gi18313472___Pa ----WH--TL--W-TYEGAKR----LREALRR-FNG---GTIVV
gi21672958___Ct N---SL--SV--C-TYSHAAHAWEELQKSIEK--MKNGQKQRFLI
gi193211736__Cp N---SM--SV--C-TYSHAAHAWEELQKSIEK--MKAGQKQRFLI
gi241775880__Av Y---TM--SV--C-TYSHALEANEQLQKCVQE--MKAGARKTFVI
gi145218949__Pv N---SY--SV--C-TYSHAAHAWBHLQESIKK--MQAGEKQRFLI
gi78185961___Pl N---SY--SV--C-TYSHAAHAWBHLQENIKK--MQAGQKQRFLI
gi78776818___Sd F---TV--SV--C-TADHASHANEEFQKTIQK--MKKGERQKFLI
gi238828305__Aa ----GY--SV--C-EPEFATK----LREKLES-FQG---GNIAI
gi229580793__Si ----GY--SV--C-ELDYALK----LREKLKD-FKG---GTIAI
gi15899031___Ss ----GY--SV--C-ELDYALK----LREKLKD-FKG---GTIAI
gi15922815___St ----GY--SV--C-EPEYAVK----TRERLMN-WQG---GNIAI
gi13542037___Tv ----GS--SV--C-EADLALS----MWNKLNA-FKG---GNITL
gi16082149___Ta ----GS--SV--C-EADLAEN----MWKKLRD-FKG---GNITL
gi21673913___Ct ----TL--SI--C-GAPEHSIR----LKEKIDA-LIERGHKGIAF
gi15606162___Aa ----TL--SI--C-GSPEEALK----IRDRLVE-LIAKGEGKIAF
gi78186989___Pl ----AR--SI--C-CASPQNSLY----LKERVDA-IIAKGSGKIAF
gi225850055__Pm ----TL--SI--C-QSPQEALK----IRERIED-LIKIGKGKIAF
gi78364011___Tc -T-PCR------C-GIKAAEE----IKNRLDN-LSG---GTIAI
gi86748177___Rp -I-PCE------C-GIAVGEE----IGRRLDA-MKG---GTIAV
gi241777285__Av -KL--PHAW--K-AGEQTAI----LRKQLED-MADG--GTVVI
gi153872492__Be -KI--PHAW--L-AGNQTTI----LRDQLHA-MKDG--GKIII
gi19879585___Cl SKI--PHAW--Q-AGPQTVL----LHRQLQT-MKNG--GAVLI
gi192289979__Rp -IF--PHAWVPE-RGEQVLL----LRKQLEA--LPEG--GVVGF
gi15605781___Aa YAQTHYPSAF--I-PGSEHLR----LLKKVEN--FEE---GTFVL
gi21436536___Pd -IM---PHAY--K-AGPQTQL----LRRMIEA-MPQG--GVFAM
gi|21673704__Ct ----GH--TV--S-DGFYGNK----LRHYLHEDGYKG---GPIAI
gi|193212535_Cp ----GH--TV--S-DGFYGNK----LRHYLHEDGYKG---GPIAI
gi|255021328_Ac ----GH--TC--T-DFYYGNK----LRHFLEH-EYRG---GPVAI
gi|256759353_Ti ----GH--TV--T-DNTYGEA----LRQYLMH-DYKG---GPIAI

cons

gi15607118___Aa GA-------------------IPGVSCF-GPAYEFALMLNYEL
gi21464540___Td GA-------------------VQGASCY-GPAYEFANIMDTDL
gi2385380____Rc GA-------------------AQGASCF-GPAYEFTFILDTAV
gi8118249____Ol GA-------------------VPGASCF-GPAYEFALLADYVL
gi103492069__Af GA-------------------MAGASCF-GPAYEYAMIVASDL
gi17232004___No GA-------------------LPKTSCL-GPAYEFTLLADYVL
gi38505571___Sy GA-------------------APGTGCF-GPAYEVYVLLASADL
gi8118252____Ah GA-------------------VPGASCM-GPAYEFALLADYAL
gi19112859___Sp TQ-------------------PSGVLKCA-GAPQKIMWWAEDYW
gi24657386___Dm TF-------------------PNCPIKCA-GAPQKIAYISZHYF
gi62896485___Hs TF-------------------PNTPVKCA-GAPQKIMYLSEAYF
gi148696168__Mm TF-------------------PNTPVKCA-GAPQKIMYLSEAYF
gi157092622__Am TF-------------------PNSPIKCA-GAPQKIMYTEEYL
gi82749809___Sa TH-------------------PNTPIKCG-GAPMKIMYLAEDYF
gi15597541___Pa TQ-------------------PAMPIKCA-GAPQKAMYLSCDHW
gi17546261___Rs TQ-------------------PAMPIKCA-GAPQKAMYLACDHW
gi15920835___St NV-------------------AKLPHRCP-VAPLFMTLILDDYL
gi21673852___Ct NI-------------------AELPFKCP-VAPIEFVFMADWFL
gi11498167___Af SV-------------------MGIPHKCP-VAPIEVTFMLHEFF
gi23015306___Mm NV-------------------NAPHKCP-VAPLEVVFMLHDFL
gi15899356___Ss NT-------------------AKLPHKCP-VAPLEVTLILDDYL
gi18313472___Pa SV-------------------TSTPYKCP-VAPYEFLGLLNDYL
gi21672958___Ct GT-------------------GHAMATCQ-GAAFEYILNYAHEI
gi193211736__Cp GT-------------------GHAMATCQ-GAAFEYILNIAHET
gi241775880__Av GT-------------------GHGMCTCQ-GAAFEYIYNVDEVL
gi145218949__Pv GT-------------------GHAMATCQ-GAAFEYILNIAREI
gi78185961___Pl GT-------------------GHAMATCQ-GAAFEYILNIAHEI
gi78776818___Sd GT-------------------GHGLCTCQ-GAAFKYIFNIEHEL
gi238828305__Aa GSGPFYQGHNPKP--KVPENFVPNADSACE-GPVFEMSLMLHCYY
gi229580793__Si GSGSLFYQGKTPKP--KVPENYVPVADAACE-GPIFEMSLMLSGYF
gi15899031___Ss GSGSLFYQGSKPKP--KVPENYVPNADAACE-GPIFEMSLMLSGYF
gi15922815___St GSGPFYQGHNPAP--NVPKNFVPNADAACE-GPVFEMSLMLSGYF
gi13542037___Tv GSAKFIQDNKNRPT-NNPDAYAPLLESACE-GPVFEMSIMLFDYF
gi16082149___Ta GSAKFIQDTKNRPT-NTPDRYAPIADSACE-GPVFEMSIMLYDYF
gi21673913___Ct GFG-------------------GNPKDPSAVRGGPGEFFNLHHKL
gi15606162___Aa GPG-------------------GNPKDKTAVRGRPVFEVLFNVDYHL
gi78186989___Pl GPG-------------------GNPKDKSAVRGGPAFELFFNLHHKL
gi225850055__Pm GPG-------------------GNPRAKEAVRGGPVFEILFNTHNYL
gi78364011___Tc GPG-------------------GNPKEPSAMRGGPMFEPLFGIDTLL
gi86748177___Rp GFS-------------------TNPEEPGAMRGGPMFEFLFIIDTLL
gi241777285__Av AP-------------------PAAPFRCP-PGPYERASQVAYYL
gi153872492__Be AV-------------------PPKPFRCP-PGPYERVSLVAHYL
gi19879585___Cl CP-------------------PDNPFRCP-PGPYERASLVAHYL
gi192289979__Rp AI-------------------PANPFRCP-PGPYVRISLVANYL
gi15605781___Aa VV-------------------PPPHRCP-PAPYERAAMIAHVF
gi21436536___Pd VA-------------------PNPYPYRCP-PGPYERVSMVAHLL
gi|21673704__Ct GSARFHQGTKGK------LDPVPMAKAACE-GPPVEIALSLASWM
gi|193212535_Cp GSARFHQGTKGK------LDPVPMAKAACE-GPPVEMALSLASWL
gi|255021328_Ac GSARFHQGDGTKDIKNIAYDKLFFADPSAEAACE-GPPVETMLSMATWL
gi|256759353_Ti GSARFTQGDGAKGLEPYPGGSIPAAEAACE-GPPVETMLACATWL

cons

Appendix table 38: the structure of *Acidithiobacillus ferrooxidans* SQR [a]

During revision of this thesis, the anticipated structure of *Acidithiobacillus ferrooxidans* SQR overexpressed in *E. coli* (see Chapter 3.9) has become available. It was solved at 2.30, 2.30 and 2.05 Å resolution in an "as-purified", quinone-bound and C160A-mutant forms, respectively, using the *A. aeolicus* SQR structure described in this thesis as a molecular replacement model.

A. ferrooxidans SQR is characterized by the following properties:

1. it is dimeric;

2. its overall monomeric architecture is closely related to *A. aeolicus* SQR, confirming the sequence fingerprints for type I SQRs described in Chapter 3.8;

3. it is an integral monotopic membrane protein inserted about 20 Å into the lipidic bilayer;

4. it harbours a non-covalently bound FAD, whose C8M atom is positioned at a distance of 3.86 Å from Cys 128 S_γ atom. A sulfide ion is found in the electron density 4 Å apart from Cys 128 S_γ atom. The authors propose that sulfide may S-sulphydrate Cys 128. They additionally propose that the Cys128-S_γ-SH species could act as nucleophile on Cys 160 for product release;

5. Cys 160 and Cys 356 in the "as-purified" structure are connected through a branched polysulfur chain of 4 or 5 S atoms. The C160A and C356A mutants have no sulfide:quinone oxidoreduction activity. The structure of the C160A mutant shows sulfhydration of Cys 356 (Cys356-S_γ-SH). Cys 356 is also the closest cysteine near FAD atom C4A (3.2 Å between the SH group of Cys356-S_γ-SH and FAD atom C4A in the "as-purified" structure). Based on these observations, the authors propose that Cys 160 and Cys 356 form the redox active centre of *A. ferrooxidans* SQR and in particular that Cys 356 is involved in forming the covalent adduct with FAD atom C4A while Cys 160 harbours the growing polysulfur chain. However, the structure of A. ferrooxidans SQR does not contradict the mechanistic hypotheses presented in this thesis (see Chapter 3.7);

6. the product of sulfide oxidation may be released through a hydrophobic channel directly into the membrane bilayer, as suggested in this thesis (see Chapter 3.7 and Figure 3.7);

7. the position of the quinone binding site corresponds to that of *A. aeolicus* SQR as described in this thesis, but in *A. ferrooxidans* SQR one amphipatic helix may undergo a conformational change upon quinone binding (see Chapter 2.10.5 and 3.6).

[a] based on PDB entries 3KPI, 3KPG and 3KPK and on the following reference: Cherney M. M., Zhang Y., Solomonson M., Weiner J. H., James M. N. G. (2010) Crystal structure of sulfide:quinone oxidoreductase from *Acidithiobacillus ferrooxidans*: insights into sulfidotropic respiration and detoxification. *Journal of Molecular Biology* 398:292-305.

Acknowledgements.

Last but not least, I would like to thank my family for their support and belief in my work. The decision to come to Frankfurt and to pursue a scientific career is a pure consequence of their outstanding but demanding education and, although it has brought me far from them in terms of space and of life interests, I wish they can follow me and understand my perspective.

Curriculum vitae.

Personal data

Name	Marco Marcia
Date of birth	22nd November 1981
Place of birth	Cuneo, Italy

Education

since 2006, Jan 16th	Graduate student at the Max Planck Institute of Biophysics, Frankfurt am Main, Germany.
	Internal supervisor: Prof. Dr. Bernd Ludwig
	External supervisor: Prof. Dr. Hartmut Michel
2005, Dec 15th	Master degree in Pharmaceutical Chemistry and Technology, School of Pharmacy, Bologna University, Italy: highest score (110/110 *cum laude*).
	Experimental thesis: Purification and functional characterization of sulfide:quinone oxidoreductase and cytochrome *bd*:quinol oxidase from *Aquifex aeolicus*.
	Internal supervisor: Prof. Dr. Giovanni Capranico
	External supervisor: Dr. Guohong Peng
2005, Dec 15th	Licence degree at the Collegio Superiore (www.collegio.unibo.it), Università di Bologna, Italy.
2000, Jul 15th	High School degree at the Liceo Scientifico "G. Peano", Cuneo, Italy: highest score (100/100).

Invited lectures

2009, Jun 26th	Faculty of Pharmacy, Bologna University, Italy.
2009, May 11th – 14th	Joint Workshop on Current Advances of Membrane Protein Structure Determination, Jinan and Qingdao, China.
2009, May 6th	Graduate College on Sensory Photoreceptors in Natural and Artificial Systems, Regensburg University, Germany.

Selected international conferences

2009, Mar 15th – 19th	EMBO-FEMS Workshop on Microbial Sulfur Metabolism, Tomar, Portugal (poster).
2008, Jul 19th – 24th	EBEC conference, Dublin, Ireland (poster).

Teaching experience

2009, Feb 9th – 12th	X-Ray crystallography course, Practical Sessions for undergraduate students, Faculty of Biochemistry, Goethe University, Frankfurt am Main, Germany (tutor, tuition in German, responsible teacher: P.D. Dr. Ulrich Ermler).
2008, Nov 10th – 14th	IMPReS Course on X-Ray crystallography, Frankfurt am Main, Germany (tutor, responsible teacher: Prof. Dr. Hartmut Michel).

Selected courses

2007, Oct 8th – 12th	Specialised Courses on Structural Genomics (HSC6), Grenoble (poster).
2006, Oct 9th – 13th	Hercules Specialised Courses on Structural Genomics (HSC2), Grenoble (poster).
2006, Aug 27th – Sep 5th	EMBO Practical Course on protein expression, purification and crystallization (PEPC5), EMBL Hamburg Outstation (poster).