Functional and Structural Membrane Proteomics using
Aquifex aeolicus

Dissertation zur Erlangung
des Doktorgrades der Naturwissenschaften

vorgelegt beim Fachbereich 14
Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe – Universität
in Frankfurt am Main

von Ulrike Wedemeyer
aus Hannover

Frankfurt (2007)
(D30)

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

Dekan: Prof. Dr. Harald Schwalbe

1. Gutachter: Prof. Dr. Robert Tampé
2. Gutachter: Prof. Dr. Hartmut Michel

Datum der Disputation: 14. Dezember 2007

Diese Doktorarbeit wurde vom 16. Juni 2003 bis zum 30. Juni 2007 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.

Ulrike Wedemeyer
Frankfurt am Main

Publications

U. Wedemeyer, G. Peng, H. Michel and K. Hartung. (2007). Protein AQ_1862 from the hyperthermophilic bacterium *Aquifex aeolicus* is a porin and contains two simultaneous conductance pathways of different selectivity. *Biophysical J,* Volume 93, 2667-2677.

Table of Contents

Ausführliche deutschsprachige Zusammenfassung

Biologische Membranen enthalten zahlreiche Proteine, die es ihnen ermöglicht ihre jeweilige Funktion auszuführen. Membranproteine lassen sich in zwei Kategorien einteilen: polytopische Membranproteine, die in Kontakt mit beiden Seiten der Membrane sind, und monotopische Membranproteine, die nur in Kontakt mit einer Membranseite sind.

Membranproteine sind an vielen zellulären Prozessen wie dem Stoffwechsel, dem Transport von Ionen und anderen Stoffen, der Atmung, der Umsetzung von Energie und der Signaltransduktion beteiligt. Dadurch sind sie auch wichtige Angriffspunkte für Medikamente. Für ein umfassendes Verständnis ihrer Funktion sind Strukturen mit atomarer Auflösung entscheidend. Trotz ihrer Bedeutung bleiben sie unterrepräsentiert in funktionellen und strukturellen Studien. Obwohl Membranproteine 20 % bis 30 % aller vorhergesagten Proteine in den bisher sequenzierten Organismen ausmachen, stammen weniger als 1 % der bis heute bekannten Proteinstrukturen von Membranproteinen. Die meisten Proteinstrukturen wurden mittels Proteinkristallisation und Röntgenkristallographie bestimmt. Für die Proteinkristallisation ist die Qualität der Proteinprobe von besonderer Bedeutung: die Proteinprobe sollte chemisch und strukturell homogen vorliegen. Je mehr Ausgangmaterial vorliegt, desto wahrscheinlicher ist es, diese Bedingungen zu erfüllen. Die Strukturbiologie der Membraneproteine ist durch den Mangel an Proteinen eingeschränkt, die diese Voraussetzungen erfüllen.

In der Vergangenheit sind hauptsächlich solche Membranproteine funktionell und strukturell beschrieben worden, die in großer Menge in der Zellmembran vorliegen, nämlich photosynthetische Proteinkomplexe, Komplexe der Atmungskette und einige Proteine aus besonderen Quellen. Seltenere Proteine ließen sich aus nativen Zellen nur unzureichend reinigen, was ihre Untersuchung erschwerte. Die Überproduktion von Proteinen mittels verschiedener Produktionssysteme stellt einen Durchbruch für die Membranbiologie dar. Allerdings schädigt die Überproduktion von Membranproteinen häufig die Wirtszelle oder das Protein wird in nicht funktioneller Form hergestellt.

In den Genomen der bisher sequenzierten Organismen kodierten 20 % bis 50 % aller offenen Leseraster für unbekannte, so genannte „hypothetische Proteine", wobei unklar ist, ob diese tatsächlich vom Organismus hergestellt werden. Proteine mit vorhergesagten Transmembrandomänen kommen unter den „hypothetische Proteinen" besonders häufig vor. Die Analyse des Proteoms ist die einzige bisher bekannte Methode um die Existenz dieser Proteine zu zeigen. In herkömmlichen Studien zur Analyse des Proteoms werden oft gel-basierte Elektrophoresemethoden zur Trennung der Proteine vor der eigentlichen Identifizierung verwendet. Diese sind häufig nicht für Membranproteine geeignet, weshalb diese vielfach unentdeckt bleiben, beziehungsweise lassen sie sich nur schlecht in einem großen Maßstab anwenden.

Ziel dieser Arbeit war es eine Methode zur Analyse des Proteoms zu entwickeln, die auch die Identifizierung von Membranproteinen zulässt und diese Analyse mit einer funktionellen und strukturellen Untersuchung der Membranproteine in ihrem nativen Zustand zu verbinden. Dadurch sollten neue, bisher unbekannte Membranproteine aus einer nativen Quelle identifiziert und

charakterisiert werden. Hierfür wurden Membranproteine aus dem hyperthermophilen Bakterium *Aquifex* (*A.*) *aeolicus* unter Verwendung des milden, nicht denaturierenden Detergens n-Dodecyl-β-D-maltosid (DDM) aus der Membran gelöst und mittels konventioneller chromatographischer Methoden gereinigt. Die vorgereinigten Proteine konnten nach eindimensionaler Polyacrylamidgelelektrophorese durch Trypsinverdau und Massenspektrometrie identifiziert werden. Dieses Vorgehen steht im Gegensatz zu herkömmlichen Methoden, die unter denaturierenden Bedingungen stattfinden.

Im nächsten Schritt wurden Zielproteine zur weiteren Untersuchung ausgewählt. Ein wichtiges Kriterium hierfür war das Vorhandensein geeigneter Proteinmengen. Alle Zielproteine waren entweder α-helikale oder β-Faltblatt Membranproteine beziehungsweise Proteine, die mit der Membran assoziiert sind. Ein Schwerpunkt dieser Arbeit war es bisher unbekannte Proteine zu charakterisieren, daher wurden „hypothetische Proteine" bevorzugt ausgewählt. Zusätzlich wurde ein System zur heterologen Produktion der Zielproteine in dem etablierten Wirt *Escherichia* (*E.*) *coli* eingeführt. Dies ermöglichte eine Untersuchung der Zielproteine ohne durch die Menge an Proteinprobe limitiert zu sein. Drei verschiedene Expressionsvektoren mit jeweils verschiedenen Promotoren wurden verwendet: pTTQ18 mit dem *tac* Promotor, pBAD mit dem Arabinose-Promotor und pET mit dem T7-Promotor. Die Vektoren waren so verändert, dass das Protein nach Insertion als Fusionsprotein produziert wurde. Die verwendeten Fusionsproteine bestanden aus (1) einer carboxyterminalen Tabakätzvirus Proteaseschnittstelle gefolgt von einem Affinitätsanhängsel bestehend aus zehn Histidinen (deca-his-tag) oder (2) einem aminoterminalen deca-his-tag gefolgt von einer Tabakätzvirus Proteaseschnittstelle, der Proteinsequenz und einem carboxyterminalen Affinitätsanhängsel, dem strep-tag II.

Die Proteincharakterisierung umfasste die Vorhersage der Funktion mittels bioinformatischer Algorithmen, wie Sequenzvergleiche, Vergleiche der vorhergesagten tertiären Struktur mit bekannten Strukturen sowie die Suche nach konservierten Domänen und Motiven. Weiterhin wurden alle Zielprotein gereinigt und mittels biochemischer und biophysikalischer Methoden untersucht, um die Vorhersagen zu bestätigen. Das reine Protein wurde einer breit angelegten Suche nach Kristallisationsbedingungen unterzogen, um dreidimensionale Kristalle zu erhalten, die sich für die Strukturbestimmung mittels Röntgenkristallographie eignen.

1. Identifizierung von *A. aeolicus* Membranproteinen mittels Massenspektrometrie. Insgesamt wurden 75 Proteine in der Membranfraktion von *A. aeolicus* identifiziert: 16 (21 %) enthalten eine oder mehre Transmembranhelices wovon 9 als „hypothetische Proteine" annotiert sind, 24 (32 %) sind membranassoziierte Proteine, 23 (31 %) sind cytosolische und 12 (16 %) „hypothetische Proteine" ohne vorhergesagte Transmembranhelices. Membranproteine und membranassoziierte Protein machen 53 % aller identifizierten Proteine aus. Die identifizierten Proteine lassen sich in vier Gruppen einteilen: (1) Proteine, die an der Energieumsetzung und der Stoffwechselaktivität beteiligt

sind, (2) Proteine, die vorhergesagte Membrandomänen enthalten, von denen einige eine vorhergesagte Transportfunktion haben, (3) „hypothetische Proteine" und (4) alle übrigen Protein, einschließlich ribosomaler und cytosolischer Proteine.

Beinahe alle zu der ersten Gruppe gehörenden Proteine sind bereits zuvor in der Literatur beschrieben worden: die Proteinkomplexe I, III und V der Atmungskette, zwei [NiFe] Hydrogenasen, eine Dimethylsulphoxid Reduktase und eine Sulphid-Quinone Reduktase. Die hydrophoben Transmembranuntereinheiten einiger Proteinkomplexe wurden hier nicht gefunden. Alle anderen Membranproteinkomplexe, die mit der inneren Membran assoziiert sind, liegen in sehr geringer Menge vor (mit Ausnahme des Proteins AQ_1862). Nur sehr wenige Proteine sind als Transportproteine annotiert, drei als Transporter in der inneren Membran und drei in der äußeren Membran. Die Mehrheit aller identifizierten Proteine sind als „hypothetische Proteine" beziehungsweise als Proteine unbekannter Funktion annotiert, unter ihnen zwei Proteine, die möglicherweise aus der äußeren Membran stammen. Ungefähr ein Drittel aller identifizierten Proteine sind ribosomalen oder cytosolischen Ursprungs.

Unter den identifizierten Proteinen wurden drei als Zielproteine ausgewählt: (1) das Protein AQ_1862, das sehr häufig in der Membranfraktion vorkommt und bei dem es sich möglicherweise um ein Porin aus der äußern Membran handelt, (2) das Protein AQ_1558 mit zwei vorhergesagten Transmembranhelices, bei dem es sich möglicherweise um ein gattungs- oder artspezifisches Protein handelt und (3) das membranassoziierte Protein AQ_1760, das keine vorhergesagten Membrandomänen enthält.

2. Charakterisierung, Kristallisation und heterologe Produktion von AQ_1862. AQ_1862 ist als „hypothetisches Protein" annotiert. Durch Sequenzvergleiche mit anderen Proteinen konnten Ähnlichkeiten zu zahlreichen Porinen aus der äußeren Membran festgestellt werden. Diese lassen sich in zwei Familien einteilen, in die „short chain amide and urea porin (SAP)" Familie und die „*Pseudomonas* OprP porin (POP)" Familie, die am Transport von Phosphat beziehungsweise Harnstoff beteiligt sind. Vorhersagen der sekundären Struktur zeigte Ähnlichkeit zu Porinen mit 16 β-Strängen aus verschiedenen Organismen. Ein β-Faß-Vorhersageprogramm zeigte allerdings 18 β-Stränge.

AQ_1862 ist eines der häufigsten Proteine in der Membranfraktion von *A. aeolicus*. Durch eine zweistufige Reinigung bestehend aus Anionenaustausch- und Gelfiltrationschromatographie konnte das Protein gereinigt werden. Die Qualität der Probe wurde durch einen weitern Schritt bestehend aus einer Affinitätschromatographie mit „Reactive Red 120" Sepharose verbessert. Aus 120 mg totalem Membranprotein konnten 10 mg homogenes und reines AQ_1862 gereinigt werden.

Quervernetzungs- sowie analytische Gelfiltrationsexperi, zeigten eine mögliche Trimerbildung des Proteins. Um zu zeigen, dass es sich tatsächlich um ein Protein mit porinartiger Funktion handelt, wurden Lipidmembranversuche mit sogenannten schwarzen Membranen durchgeführt. Durch die

Zugabe von AQ_1862 erhöhte sich die Leitfähigkeit der Membran im Vergleich zu unbehandelten Membranen um mehrere Größenordnungen. In Versuchen mit Membranen, die viel eingefügtes Protein enthalten (makroskopische Experimente), erzeugte das Protein eine Selektivität für Anionen in Lösungen von kleinen Anionen wie Haliden oder Nitrat beziehungsweise bei pH-Werten unter pH 8. Bei pH-Werten über pH 9 oder in Lösungen großer oder organischer Anionen wie Propionat oder Gluconat erzeugte das Protein eine Selektivität für Kationen. Im Gegensatz zu den homologen Proteinen konnte keine Selektivität für Phosphat gezeigt werden. In Experimenten mit lediglich einer Leitfähigkeit erzeugenden Einheit (Molekül) in der Membran wurde eine stabile (Grund-) Leitfähigkeit von 1,4 nS in Gegenwart von 0,1 M KCl bei pH 7,5 beobachtet, die von Fluktuationen mit deiner Leitfähigkeit von circa 0,25 nS überlagert war. Da beide Ereignisse stets gleichzeitig beobachtet wurden, wird angenommen, dass beide vom selben Molekül hervorgerufen werden. Trotzdem zeigten beide Leitfähigkeiten ein sehr unterschiedliches Verhalten. Bei neutralen pH-Werten war die Grundleitfähigkeit anionenselektiv mit einer Leitfähigkeitssequenz von $Cl^- \approx Br^- \approx NO_3^- > F^- >$ Gluconat $\approx$ Acetat $\approx$ Propionat und ohne Sättigung bis zu 0,5 M KCl. In der Gegenwart großer Anionen und bei alkalischen pH-Werten wurde die Grundleitfähigkeit unselektiv und zeigte Sättigung bei niedrigen Salzkonzentrationen ($K_m \approx 20$ mM). Im Gegensatz dazu zeigte der fluktuierende Anteil der Leitfähigkeit hauptsächlich Anionenselektivität mit einer Leitfähigkeitssequenz von $K^+ \approx Rb^+ >$ $NH_4^+ > Na^+ \approx Li^+ \approx Cs^+$ und einer Sättigung bei niedrigen Salzkonzentrationen ($K_m \approx 15$ mM), die nicht vom pH-Wert beeinflusst wurde. Die Ergebnisse der makropskopischen Experimente können durch das Zusammenspiel von Grundleitfähigkeit und Fluktuationen erklärt werden. In Anbetracht der verschiedenen Eigenschaften der beiden Leitfähigkeiten kann man annehmen, dass AQ_1862 zwei verschiedene Wege für den Ionentransport besitzt und es sich bei den Leitfähigkeiten nicht um zwei verschiedene Zustände der selben Pore handelt.

Eine ausführliche, systematische Suche nach Kristallisationsbedingungen wurde durchgeführt, wobei neun verschiedene Detergenzien und eine Vielzahl von Bedingungen getestet wurden. In der Gegenwart aller Detergenzien wurden Kristalle erhalten: Kristalle, die in Gegenwart von n-Octyl-β-D-glucopyranosid erhalten wurden, beugten Röntgenstrahlen bis 3 Å. Die Qualität dieser Kristalle konnte schrittweise bis zu einer Röntgenbeugung von 1,95 Å verbessert werden. Weder mittels molekularen Ersatzes noch anomaler Dispersion von Schwermetallderivaten konnte die Struktur gelöst werden. Eine geringe Sequenzhomologie von AQ_1862 mit den verwendeten Suchmodellen verhinderte vermutlich den molekularen Ersatz, während unzureichende Bindung der Schwermetalle die Bestimmung der Struktur mittel anomaler Dispersion verhinderte.

Zwei Strategien der heterologen Produktion von AQ_1862 in *E. coli* wurde verfolgt, um die Struktur mittels anomaler Dispersion von Selenium unter Verwendung der nicht nativen Aminosäure Selenomethionin zu lösen: (1) die Produktion in funktioneller Form in der äußeren Membran von *E. coli* sowie (2) in nicht-funktioneller Form in Einschlusskörperchen. Durch Produktion in Einschlusskörperchen, Lösen mit Guanidiniumhydrochlorid und Rückfaltung durch Verdünnung des

gelösten Proteins in der Gegenwart von n-Lauryl-N,N-dimethylamin-N-oxid konnten Mengen im Milligrammbereich erhalten werden. Die Eigenschaften des zurückgefalteten Proteins wurden untersucht und mit denen des nativen Proteins verglichen. Der Anteil an sekundären Strukturelementen, die thermische Stabilität sowie der oligomere Zustand sind dem nativen Protein sehr ähnlich. Proteinkristalle konnten in einigen Bedingungen erhalten werden. Sie waren zur Bestimmung der Struktur aber noch von zu schlechter Qualität und müssen verbessert werden.

3. Charakterisierung und heterologe Produktion von AQ_1558. Das „hypothetische Protein" AQ_1558 war nur schwach homolog zu den Sequenzen beziehungsweise Faltungen bekannter Proteine: es zeigte lediglich Homologie zu einigen Bassoon-Proteinen aus Säugetieren. Möglicherweise handelt es sich um ein gattungs- oder artspezifisches Protein. Es enthält zwei Transmembrandomänen in der Nähe des Aminoterminus. In der Nähe des Carboxyterminus befinden sich zwei Sequenzmotive bestehend aus einem Cystein gefolgt von zwei beliebigen Aminosäuren und einem weiteren Cystein (CXXC).

Die Ausbeute aus der nativen Membran ist verhältnismäßig gering, nur circa 0,2 mg AQ_1558 wurden aus 120 mg totalem Membranprotein gereinigt. Um größere Mengen an Protein zu erhalten, die sich auch für strukturelle Studien eignen, wurde das Protein heterolog in *E. coli* hergestellt. Das Protein konnte erfolgreich erzeugt werden: (1) als Fusionsprotein mit einer carboxyterminalen Tabakätzvirus Proteaseschnittstelle gefolgt von einem deca-his-tag mit dem Vektor pTTQ18, sowie (2) als Fusionsprotein mit deca-his-tag gefolgt von einer Tabakätzvirus Proteaseschnittstelle, der Proteinsequenz und einem carboxyterminalen strep-tag II Affinitätsanhängsel mit dem Vektor pBAD. Pro Liter *E. coli* Kultur konnten 0,4 mg Protein erhalten werden. Um die Aufreinigung und die Kristallisation zu vereinfachen wurden die Transmembrandomänen durch genetische Modifikation entfernt. Dieses Konstrukt (AQ_1558_2) lag allerdings immer noch in der Membranfraktion vor. Die Lösung aus der Membran und die Reinigung in homogener Form war nur in Gegenwart von Detergens möglich: 1,5 mg bis 2 mg von reinem, homogenem, verkürzten Protein konnten pro Liter *E. coli* Kultur erhalten werden.

Durch Quervernetzungsexperimente und analytische Gelfiltration des vollständigen sowie des verkürzten Proteins konnte gezeigt werden, dass sie als Trimer oder Tetramer vorliegen. Da das CXXC Motiv häufig an der Bindung von Metallionen beteiligt ist, wurde eine Elementaranalyse mit Hilfe von Totalreflexions-Röntgenfluoreszenzanalyse durchgeführt, die die Anwesenheit eines Zinkatoms pro Proteinmolekül zeigte.

4. Charakterisierung und Kristallisation von AQ_1760. Das „hypothetische Protein" AQ_1760 wurde als drittes Zielprotein ausgewählt. Die Proteinsequenz enthält keine Transmembrandomänen, das Protein wurde allerdings mit der Membranfraktion zusammen aufgereinigt. Sequenzvergleiche ergaben circa 20 % Sequenzidentität zu der konservierten Proteindomäne Linocin_M18. Diese

Domäne stammt aus dem Protein Linocin aus *Brevibacterium linens*, das das Wachstum einiger Bakterien aus dem selben Lebensraum unterdrückt. Die Linocin_M18 Domäne kommt in Bakterien und Archaeen vor, allerdings wurden nur wenige Proteine, die diese Domäne enthalten, bisher in der Literatur beschrieben.

Aus 120 mg *A. aeolicus* Membranproteinen wurden circa 0,1 mg bis 0,5 mg AQ_1760 Protein erhalten. Eine heterologe Produktion in *E. coli* ergab nachweisbar AQ_1760 in der Membranfraktion. Allerdings war eine Solubilisierung nicht möglich.

So wie andere Proteine aus der Linocin_M18 Familie bildet AQ_1760 Proteinkomplexe. Durch analytische Gelfiltration wurde eine Größe zwischen 700 kDa und 20 MDa abgeschätzt. Diese konnte durch analytische Ultrazentrifugation auf circa 3,5 MDa präzisiert werden. Durch Elektronenmikroskopie konnte gezeigt werden, dass das Protein kugelförmige Partikel mit einem Durchmesser von 38 nm bildet, die entweder hohl, gefüllt beziehungsweise teilweise assemblierten vorlagen. Hohle Partikel hatten eine netzartige Hülle mit einer Dicke von 1,5 bis 2 nm, die an die Struktur eines Virus erinnert. Unter Annahme einer vollständig aufgebauten Hohlkugel aus AQ_1760 Untereinheiten, würde ein Partikel aus circa 160 bis 207 Untereinheiten bestehen und eine Masse von 5,1 bis 6,6 MDa besitzen. Kürzlich wurde die Struktur des homologen Proteins „virus-like particle" aus *Pyrococcus (P.) furiosus* mittels Röntgenkristallographie bestimmt. Es zeigt ebenfalls eine käfigartige Struktur, die an einen Virus erinnert, obwohl es vom Genom von *P. furiosus* kodiert wird.

Ähnlich dem homologen Protein bacteriocin aus *Thermotoga maritima* zeigt AQ_1760 proteolytische Aktivität mit der größten Aktivität bei pH 7. Diese Aktivität ist sehr stabil gegen Protease- und Hitzebehandlung. Der Proteinkomplex zeigte eine gelbliche Farbe. Durch Totalreflexions-Röntgenfluoreszenzanalyse konnte die Anwesenheit von zwei Eisenatomen pro Proteinmonomer gezeigt werden. AQ_1760 wurde kristallisiert und dreidimensionale Kristalle beugten Röntgenstrahlen bis circa 8 Å; die Kristalle enthielten ebenfalls Eisen.

Zusammenfassung

Membranproteine spielen eine wichtige Rolle in einer Vielzahl biologischer Prozesse, wie dem Stoffwechsel, dem Transport, der Atmung, der Umsetzung von Energie und der Signaltransduktion. Dadurch sind sie wichtige Angriffspunkte für Medikamente. Für ein umfassendes Verständnis ihrer Funktion sind Strukturen mit atomarer Auflösung von entscheidender Bedeutung. Trotz ihrer Wichtigkeit bleiben Membranproteine unterrepräsentiert in funktionellen und strukturellen Studien, lediglich 1 % aller bis heute bekannten Proteinstrukturen stammt von Membranproteinen. Auch in Studien zur Analyse des Proteoms von Zellen bleiben sie häufig unentdeckt.

Ziel dieser Arbeit war es eine Methode zur Analyse des Proteoms zu entwickeln, die die Identifizierung von Membranproteinen zulässt und diese mit einer funktionellen und strukturellen Untersuchung dieser Proteine in ihrem nativen Zustand zu verbinden. Dadurch sollten neue, bisher unbekannte Membranproteine aus einer nativen Quelle identifiziert und charakterisiert werden.

Membranproteine aus dem thermophilen Bakterium *Aquifex* (*A.*) *aeolicus* wurden durch das milde Detergens n-Dodecyl-β-D-maltosid aus der Membran gelöst und mittels konventioneller chromatographischer Methoden gereinigt. Nach eindimensionaler Polyacrylamidgelelektrophorese wurden die Proteine durch Massenspektrometrie trypsinverdauter Proben identifiziert. Zielproteine zur weitern Untersuchung wurden anhand dreier Kriterien ausgewählt: das Vorhandensein geeigneter Proteinmengen, die Assoziation mit der Membran sowie die biologische Funktion. Um auch seltenere Protein untersuchen zu können, wurde zusätzlich ein System zur heterologen Produktion der Zielproteine in dem etablierten Wirt *Escherichia* (*E.*) *coli* verwendet. Alle Zielproteine waren entweder α-helikale oder β-Faltblatt Membranproteine beziehungsweise Proteine, die mit der Membran assoziiert sind. „Hypothetische Proteine" wurden bevorzugt ausgewählt.

Insgesamt konnten 75 Proteine in der Membranfraktion identifiziert werden. Mehr als die Hälfte machen Membranproteine beziehungsweise membranassoziierte Proteine aus. Es lassen sich vier Gruppen unterscheiden, (1) Proteine, die an Energieumsetzung und Stoffwechselaktivität beteiligt sind, (2) Proteine, die vorhergesagte Membrandomänen enthalten, von denen einige eine vorhergesagte Transportfunktion haben, (3) „hypothetische Proteine" und (4) alle übrigen Protein, einschließlich ribosomaler und cytosolischer Proteine.

Zur weiteren Untersuchung wurden drei „hypothetische Proteine" ausgewählt: (i) das Protein AQ_1862, das sehr häufig in der Membranfraktion vorkommt und möglicherweise ein Porin aus der äußern Membran ist, (ii) das Protein AQ_1558 mit zwei vorhergesagten Transmembrandomänen, bei dem es sich möglicherweise um eine gattungs- oder artspezifisches Protein handelt und (iii) das membranassoziierte Protein AQ_1760, das keine vorhergesagten Membrandomänen enthält.

AQ_1862 konnte in großen Mengen aus der nativen *A. aeolicus* Membran gereingt werden. Studien an Lipidmembranen zeigten, dass es tatsächlich eine porinartige Funktion besitzt. Membranen, die das Protein enthielten zeigten eine stark erhöhte Leitfähigkeit sowie Selektivität für bestimmte Ionen. In Einzelkanalmessungen konnten zwei verschiedene Leitfähigkeiten unterschieden werden, eine

Grundleitfähigkeit, die abhängig vom pH-Wert selektiv für Anionen war und eine fluktuierende Leitfähigkeit, die unabhängig vom pH-Wert kationenselektiv war.

Es wurden Kristalle erhalten, die schrittweise bis zu einer Röntgenbeugung von besser als 2 Å verbessert werden konnten. Eine Bestimmung der Struktur war weder mittel molekularem Ersatz noch mittels anomaler Dispersion von Schwermetallderivaten möglich.

Die heterologe Produktion des Proteins konnte in Form nicht löslicher Einschlusskörperchen erreicht werden, wobei ein Protokoll zur Zurückfaltung entwickelt wurde. Das zurückgefaltete AQ_1862 war dem nativen sehr ähnlich. Es wurden Proteinkristalle erhalten, die aber noch verbessert werden müssen, um die Struktur zu bestimmen.

Das zweite Zielprotein AQ_1558 lag nur in sehr geringer Menge in der *A. aeolicus* Membran vor. Es konnte heterolog in *E. coli* produziert werden. Auch eine verkürzte Version ohne Transmembrandomänen wurde erfolgreich herstellt und gereinigt, allerdings ist auch dies mit der Membran assoziiert. Durch die geringe Sequenzhomologie mit anderen Proteinen konnten keine Rückschlüsse auf die Funktion gezogen werden. Es konnte die Gegenwart eines Zinkatoms nachgewiesen werden.

Das Protein AQ_1760 enthält die konservierte Linocin_M18 Domäne, die in Proteinen einiger Bakterien und Archaeen vorkommt, bisher aber nur wenig in der Literatur beschrieben wurde. Wie andere, die Linocin_M18 Domäne enthaltende Proteine, liegt AQ_1760 als Oligomer mit einer Größe von circa 3,5 MDa vor und zeigt proteolytische Aktivität. Mittels elektronenmikroskopischer Studien konnten kugelförmige, käfigartige Partikel mit einem Durchmesser von 38 nm und einer Hüllendicke von 1,5 nm bis 2 nm dargestellt werden, die an einen Virus erinnern. Die Gegenwart zweier Eisenatome konnte gezeigt werden, wobei deren Funktion unklar bleibt. AQ_1760 wurde kristallisiert und dreidimensional Kristalle beugten Röntgenstrahlen bis circa 8 Å.

Detailed English Summary

Every biological membrane is associated with numerous proteins that enable the membrane to conduct its specific activity. Membrane proteins are dived in two categories: polytopic membrane proteins that are in contact with both leaflets of the membrane and monotopic membrane proteins that are in contact with one leaflet.

They are involved in a variety of cellular functions, including metabolism, transport of ions and solutes, respiration, energy conversion and signal transduction. At the same time they are important drug targets. Despite their obvious importance they are underrepresented in structural and functional studies: only approximately 1 % of all structures known today are the structures of membrane proteins, compared to 20 % to 30 % membrane proteins predicted in every genome analysed so far. Most structures have been determined by X-ray crystallography. A major rule for any crystallisation attempt is a protein which is chemically and structurally homogenous. The larger the amount of available protein the higher is the chance to fulfil these conditions. One major bottleneck is the lack of availability of large amounts of pure and homogenous membrane proteins suitable for such studies.

Historically membrane proteins have been purified from the native membrane and thus the functionally and structurally best characterised membrane proteins are those of high abundance: photosynthetic membrane proteins, proteins of the respiratory chain and few proteins from specialised tissues. Less abundant proteins were not accessible to purification from native source. Using diverse expression systems was a major breakthrough in protein research. However over-production of membrane proteins is often harmful if not lethal to the host cell or membrane proteins are produced in non-functional form.

In any sequenced genome 20 % to 50 % of all identified open reading frames code for "hypothetical proteins", it remains unclear if they are transcriptionally active and proteins with predicted transmembrane domains are overrepresented among these proteins. Proteomics approaches are the only way to show their existence, however the standard proteins separation preceding protein identification is not well suited for membrane proteins and thus they often remain undetected in proteomic approaches.

The objective of this study was to link a proteomic study that is suitable for the detection of membrane proteins with a study of the proteins' function and structure. New membrane proteins from the native cell membrane shall be identified and characterised in detail. This approach is aimed to increase the knowledge on the function of membrane proteins.

Membrane proteins from the hyperthermophilic bacterium *Aquifex* (*A.*) *aeolicus* were solubilised under non-denaturing conditions using the mild detergent n-dodecyl-β-D-maltoside (DDM). Instead of the gel-based electrophoresis separation used in many standard proteomics approaches, separation was achieved by conventional chromatography using various chromatographic techniques. Proteins were identified by in-gel trypsin digestion followed by matrix assisted laser desorption/ionization time of

flight mass spectrometry (MALDI-TOF) and peptide mass fingerprinting. This procedure stands in contrast to the most commonly used approaches under denaturing conditions.

The next step was to choose proteins as targets for further detailed characterisation. An important criterion was the availability of the protein. All target proteins were either α-helical or β-barrel membrane proteins or associated with the membrane. In order not to be limited to the most abundant proteins and to introduce non-native amino acids, heterologous production of the target proteins using the well established *Escherichia* (*E.*) *coli* system was introduced. Three different vectors carrying different promotors were used: pTTQ18 with the *tac* promotor, pBAD with the arabinose promotor and pET with the T7 promotor. The multiple cloning sites were modified to produces the protein fused (1) to a tobacco etch virus protease cleavage site at the carboxy-terminus followed by a deca-histidine tag or (2) with a deca-histidine tag at the amino-terminus, followed by a tobacco etch virus protease cleavage site, as well as a carboxy-terminal strep-tag II. This work was mainly focused on the characterisation of so far unknown proteins. Therefore priority was given to proteins of unknown function i.e. "hypothetical proteins". Characterisation of the protein included prediction of the function by bioinformatic characterisation like sequence alignment, comparison of the predicted tertiary structure with known structures and search for conserved motifs and domains. Subsequently the target proteins were purified to homogeneity and the function investigated by various biochemical and biophysical assays to confirm the prediction. Furthermore target proteins were subjected to a broad screening for crystallisation conditions in order to obtain three-dimensional crystals of high quality in order to elucidate the proteins structure by X-ray crystallography.

1. **Identification of *A. aeolicus* membrane proteins by mass spectrometry.** In total 75 proteins have been identified in the membrane fraction of *A. aeolicus*: 16 (21 %) have one or more predicted transmembrane domain out of which 9 had been annotated as hypothetical proteins, 24 (32 %) are membrane associated proteins, 23 (31 %) are cytosolic and 12 (16 %) hypothetical proteins without transmembrane segments. Membrane proteins and membrane associated proteins together make up 53 % of all identified proteins. The identified proteins group in four different classes: (1) proteins involved in energy transduction and metabolism, (2) proteins of known membrane localisation, of which several are known to be involved in membrane transport, (3) "hypothetical proteins" and (4) other proteins, including ribosomal and cytosolic proteins.

Almost all proteins from the first group have been described in the literature before: the complexes I, III and V from the respiratory chain, two [NiFe] hydrogenases, a dimethylsulphoxid reductase and a sulphide-quinone reductase. For several complexes the hydrophobic transmembrane subunits were not detected. With the exception of AQ_1862, proteins associated with the inner membrane are present in very low amounts. Only very few proteins are annotated as transport proteins, three from the inner membrane and three outer membrane proteins. The majority of identified proteins are hypothetical

proteins or proteins of unknown function. Among them, two are putative outer membrane proteins. Approximately one third of all identified proteins are of cytosolic or associated with the ribosome.

Three unknown proteins were chosen as target proteins for a detailed characterisation: (i) the highly abundant putative outer membrane proteins AQ_1862, (ii) AQ_1558, a putative genus or strain specific protein with two predicted transmembrane helices and (iii) the membrane associated putative linocin homolog AQ_1760 without predicted transmembrane helices.

2.   Characterisation, crystallisation and heterologous production of AQ_1862. AQ_1862 is annotated as "hypothetical protein". Analysis by sequence alignment showed homology to several putative outer membrane proteins from two families, the short chain amide and urea porin (SAP) family and the *Pseudomonas* OprP porin (POP) family, outer membrane porins involved in the transport of amides and phosphate. Secondary structure prediction revealed homology to outer membrane proteins from various organisms with 16 transmembrane β-strands. However, a β-barrel prediction program predicted the existence of 18 transmembrane β-strands.

AQ_1862 is one of the proteins of highest abundant in the *A. aeolicus* membrane. It was purified by a two step purification procedure consisting of anion exchange chromatography followed by gel filtration chromatography. Its purity and quality was optimised by a further chromatographic step using Reactive Red 120 sepharose; approximately 10 mg of pure and homogenous protein was prepared from 120 mg of total membrane protein.

Evidence is given by cross-linking and analytical gel filtration experiments that it forms a trimer. In order to show that it is indeed a porin-like protein its function was investigated by lipid bilayer experiments. Upon introduction into a lipid bilayer, AQ_1862 was able to increase the conductance by several orders of magnitude compared to an undoped membrane. In multi-channel experiments the protein introduced anion selectivity in the presence of small anions like halides or nitrate and at pH values below 8. At pH values above 9 or in the presence of large or organic anions like propionate or gluconate membranes doped with AQ_1862 showed cation selectivity. In contrast to its homologous proteins, AQ_1862 does not show selectivity towards phosphate. In experiments with one conducting unit (molecule) a conductance of 1.4 nS was observed in 0.1 M KCl at pH 7.5. This stable (basic) conductance was superimposed by conductance fluctuations of about 0.2 nS. Because both events were always observed simultaneously it is suggested that they are caused by the same molecular entity. Nonetheless they show very different properties. The basic conductance is anion selective at neutral pH with a conductivity sequence $Cl^- \approx Br^- \approx NO_3^- > F^- >$ gluconate $\approx$ acetate $\approx$ propionate and no saturation up to 0.5 M KCl. At alkaline pH and in the presence of large the basic conductance becomes unselective and saturates at low salt concentrations ($K_m \approx 20$ mM). In contrast, the fluctuating component is mainly cation selective with a conductivity sequence $K^+ \approx Rb^+ > NH_4^+ > Na^+ \approx Li^+ \approx Cs^+$. It saturates at low salt concentrations ($K_m \approx 15$ mM) and is not affected by pH. The results of the macroscopic experiments can be explained by superposition of fluctuating and basic conductances. In

view of the diverging properties of both conductance components it seems appropriate to assume that AQ_1862 has two different conducting pathways rather than one with two different open states.

An extensive and systematic search for crystallisation condition was carried out, in which nine different detergents and a variety of different conditions were tested. Crystals were obtained in the presence of all tested detergents and in a variety of conditions; in the presence of n-octyl-β-D-glucopyranoside initial crystals diffracted X-rays to 3 Å. The quality of those crystals was improved by iterative steps to a diffraction of better than 2 Å. Neither molecular replacement nor anomalous dispersion in the presence of heavy atom derivatives allowed solving the structure. Low sequence homology of AQ_1862 to the search models impeded molecular replacement, while insufficient binding of heavy atom derivatives hampered phase determination by anomalous dispersion.

In order to obtain phase information by anomalous dispersion of selenium from the non-native amino acid seleno-methionine, heterologous production of AQ_1862 in *E. coli* was attempted. Two approaches were followed: (1) production in the functional form in the outer membrane and (2) non-functional production in insoluble inclusion bodies. By production in inclusion bodies, solubilisation by guanidinium hydrochloride and refolding by rapid dilution in the presence of n-lauryl-N,N-dimethylamine-N-oxide (LDAO) milligram amounts of AQ_1862 were achieved. The quality of the refolded protein was investigated and compared to the native protein. Its secondary structure content determined by circular dichroism spectroscopy, its thermal stability and its oligomeric state are very similar to the native protein. Crystals were obtained in several conditions that were similar to crystallisation conditions of the native protein. In order to solve the proteins´ structure the crystal quality has to be improved.

3.   Characterisation and heterologous production of AQ_1558. The "hypothetical protein" AQ_1558 shows only weak homology to the sequence or the fold of other proteins, to several bassoon proteins from mammals. However, it might be a genus- or strain specific protein. AQ_1558 contains two transmembrane domains near the amino-terminus. Near the carboxy-terminus two motifs consisting of a cysteine, two arbitrary amino acids and a second cysteine (CXXC) are present. The yield from the native membrane was rather low, approximately 0.2 mg per 120 mg of total membrane protein. In order to obtain yields suitable for structural studies the protein was produced in *E. coli*. Overproduction was achieved using the vector pTTQ18 modified to produce the protein with C-terminal deca-histidine tag and in pBAD modified to produce the protein with an N-terminal deca-histidine tag and C-terminal strep-tag II. 0.4 mg protein was achieved per litre of *E. coli* culture. In order to facilitate purification and crystallisation attempts the transmembrane helices were genetically removed, however, the construct (AQ_1558_2) was still located mainly in the membrane fraction of the cell. Separation from the membrane and purification in homogenous form was only possible in the presence of detergents, 1.5 mg to 2 mg pf pure and homogenous truncated protein was obtained per litre of *E. coli* culture.

By cross-linking and analytical gel filtration experiments the full length and the truncated version were shown to be a trimer or tetramer. Because the CXXC motif is often involved in metal ion binding, the presence of metal ions was investigated. By total reflection X-ray fluorescence (TXRF) the presence of one zinc atom per protein was shown.

4.   Characterisation and crystallisation of AQ_1760. The "hypothetical protein" AQ_1760 was chosen as third target protein. Bioinformatic prediction programs do not show a transmembrane domain, however, it was co-purified with the membrane fraction. Sequence comparison revealed approximately 20 % sequence identity to the Linocin_M18 domain. This domain originates from the protein Linocin from *Brevibacterium linens*, which inhibits growth of several bacteria from the same habitat. The Linocin_M18 domain is found in bacteria and archea, however, only few proteins sharing it have been described in the literature.

The yield from *A. aeolicus* membrane was approximately 0.1 to 0.5 mg. Heterologous produced protein was detected in the membrane fraction, however, solubilisation was not possible.

Like other members of the bacteriocin family AQ_1760 forms a homo-oligomeric complex. The size was estimated by analytical gel filtration to be between 700 kDa and 20 MDa. Analytical ultracentrifugation showed a mass of approximately 3.5 MDa. In electron micrographs spherical particles with a diameter of approximately 38 nm were detected. They appear in different forms of hollow, filled, or partially assembled particles. In the former case the shell is approximately 1.5 to 2 nm thick with a mesh- or virus-like structure. Assuming a fully assembled hollow sphere made up of AQ_1760 monomers, one particle would comprise approximately 160 to 207 subunits of AQ_1760 having a molecular mass of 5.1 to 6.6 MDa. Recently the structure of the homologous protein virus-like particle from *Pyrococcus (P.) furiosus* was solved by X-ray crystallography; it shows a cage-like structure that appears virus-like as well, although it is encoded by the genome of *P. furiosus*.

Like its homolog bacteriocin from *Thermotoga maritima*, AQ_1760 shows proteolytic activity towards a peptide with an optimal activity above pH 7. It is extremely stable towards protease digest and heat denaturation. The protein complex shows a yellowish colour and was shown by total reflection X-ray fluorescence analysis to contain two co-purified iron-ions per monomer. AQ_1760 was crystallised and three-dimensional crystals diffracted X-rays up to 8 Å and also contained iron.

Abstract

Membrane proteins are involved in a variety of cellular functions, like cell metabolism, transport of ions and solutes, respiration, energy conversion and signal transduction. Therefore they are important drug targets. Albeit their obvious importance membrane proteins remain structurally and functionally underrepresented: only approximately 1 % of all structures known today are the structures of membrane proteins. In addition, they often remain undetected in studies to characterise the proteome of the cell.

The objective of this work was to link a proteomic study that is suitable for the detection of membrane proteins with a study of the proteins' function and structure. Thereby so far unknown membrane proteins from the native cell membrane shall be identified and characterised and the knowledge on the function of membrane proteins increased.

Membrane proteins from the hyperthermophilic bacterium *Aquifex* (*A.*) *aeolicus* were solubilised under non-denaturing conditions using the mild detergent n-dodecyl-β-D-maltoside (DDM) and purified by conventional chromatography. Proteins were identified by in-gel trypsin digestion followed by matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) and peptide mass fingerprinting. Subsequently target proteins for further detailed characterisation were chosen according to three criteria: the proteins' availability, its membrane association and its biological function. In order not to limit this study to the most abundant proteins, heterologous production of the target proteins using the well established *Escherichia* (*E.*) *coli* system was introduced. All target proteins were either α-helical or β-barrel membrane proteins or associated with the cell membrane. "Hypothetical proteins" were favoured.

In total 75 proteins have been identified in the membrane fraction of *A. aeolicus*. More than half are membrane or membrane associated proteins. The identified proteins group in four different classes: (1) proteins involved in energy transduction and cell metabolism, (2) proteins of known membrane localisation, of which several are known to be involved in membrane transport, (3) "hypothetical proteins" and (4) other proteins, including ribosomal and cytosolic proteins. Three proteins were chosen as target proteins: (i) the highly abundant putative outer membrane proteins AQ_1862, (ii) AQ_1558, a putative genus or strain specific protein with two predicted transmembrane helices and (iii) the membrane associated putative linocin homolog AQ_1760 without predicted transmembrane helices.

AQ_1862 is one of the most abundant proteins in the *A. aeolicus* membrane. Lipid bilayer experiments showed that it is indeed a porin-like protein. Membranes doped with AQ_1862 showed a drastic increase in conductance as well as selectivity for certain ions. In single channel experiments, two distinct conductances were observed: a basic conductance that was selective for anions in a pH depended manner and a fluctuating conductance with a pH independent selectivity for cations.

Protein crystals were obtained that could be improved to diffract X-rays better than 2 Å. The determination of the three dimensional structure was not possible: neither molecular replacement nor

anomalous dispersion in the presence of heavy atom derivatives allowed obtaining phase information. In order to solve the phase problem by introducing the non-native amino acid seleno-methionine, heterologous production of AQ_1862 in *E. coli* was attempted. AQ_1862 was successfully produced in non-soluble inclusion bodies, solubilised and refolded into a functional form. The refolded protein resembled the native in several points. Protein crystals of refolded AQ_1862 were obtained and have to be improved in order to solve the proteins´ structure.

The second target protein, AQ_1558, was of very low abundance in the native *A. aeolicus* membrane. Overproduction of the full-length and truncated protein without transmembrane domains was achieved. However also the latter was still associated with the membrane fraction. Through its low sequence homology to other proteins no conclusions about its function can be drawn. The presence of a zinc atom was shown.

The third target protein, AQ_1760, was homologous to the Linocin_M18 domain, a domain that is found in bacteria and archaea. However, only few proteins sharing the Linocin_M18 domain have been described in the literature. Like its homologous proteins, AQ_1760 forms a homo-oligomeric complex with a mass of approximately 3.5 MDa and shows proteolytic activity. In electron micrographs spherical particles with a diameter of approximately 38 nm were detected. The shell is approximately 1.5 to 2 nm thick with a mesh- or virus-like structure. The presence of two iron atom was shown, however their function remains unclear. AQ_1760 was crystallised and three-dimensional crystals diffracted X-rays up to 8 Å.

Abbreviations

AMC	7-amido-4-methylcoumarin
APS	ammonium peroxodisulphatate
BAC	benzyldimethyl-n-hexadecylammonium chloride
BCA	bicinchoninic acid
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BLAST	Basic Local Alignment Search Algorithm
BLM	black lipid membrane
BN	Blue Native
BSA	bovine serum albumine
$C_{14}M$	n-tetradecyl-β-D-maltoside
C_8E_4	octyltetraoxyethylene
C_8E_n	octylpolyoxyethylene
C_8HESO	octyl-2-hydroxy ethylsulfoxide
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CD	circular dichroism
CHCA	α-cyano-4-hydroxy-cinamicacid
CL	cardiolipin
CMC	critical micelle concentration
COG	cluster of orthologous groups
CTAB	cetyltrimethylammonium bromide
Da	Dalton
DDM	n-dodecyl-β-D-maltoside
DM	n-decyl-β-D-maltoside
DMSO	dimethyl-sulfoxide
DNA	deoxyribonucleic acid
EBI	European Bioinformatics Institute
EDTA	ethylene diamine tetra-acetic acid
EM	electron microscopy
FASTA	FAST-All
Fos-12	FOS-CHOLINE-12
Fos-16	FOS-CHOLINE-16
GHK	Goldman-Hodgkin-Katz
GRAVY	Grand Average of Hydrophobicity
HEPES	4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid
(deca-) his-tag	affinity tag (ten histidines)
(hexa-) his-tag	affinity tag (six histidines)

HPLC	High-performance liquid chromatography
IEF	isoelectric focussing
IMAC	immobilized metal-ion affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranosid
K_m	Michaelis-Menten constant
LB	lysogeny broth
LC	liquid chromatography
LDAO	n-lauryl-N,N-dimethylamine-N-oxide
M	molecular weight
MALDI-TOF	matrix assisted laser desorption/ionization time of flight
MES	2-(N-morpholine)-ethane sulphonic acid
MPD	2-methly-2,4-pentandienol
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NCBI	National Center for Biotechnology Information
NBT	nitro blue tetrazolium
NG	n-nonyl-β-D-glucopyranoside
Ni-NTA	Nickel-nitriloacetat
NMG	N-methylglucosamine
NMR	nuclear magnetic resonance
OD	optical density
OG	n-octyl-β-D-glucopyranoside
OmpC	general porin from *Escherichia coli*
OmpF	general porin from *Escherichia coli*
OprP	Phosphoporin from *Pseudomonas aeruginosa*
ORF	open reading frames
OTG	1-S-n-octyl-β-D-thioglucopyranoside
Oxa1	Cytochrome-Oxidase-Biogenesis
PAGE	polyacrylamide gel electrophoresis
PE	phosphatidylethanolamine
PEG	polyethylene glycol
pdb	protein data bank
PhoE	Phosphoporin from *Escherichia coli*
pI	isoelectric point
POP	*Pseudomonas* OprP porin
PSI-BLAST	Position Specific Iterated BLAST
RND	resistance-nodulation-cell division

S.D.	standard deviation
SAP	short chain amide and urea porin
SDS	sodium dodecyl sulfat
Se-Met	seleno methionine
SMART	Simple Modular Architecture Research Tool
TB	terrific broth
TEMED	N,N,N,N-tetramethyl-p-ethylendiamid
TEV	tobacco etch virus
TFA	trifluoric acid
TMH	transmembrane helix
Tris	tris-hydroxymethyl-9-aminomethane
TROSY	transverse relaxation-optimised spectroscopy
TRP	tetratricopeptide
TXRF	total reflection X-ray fluorescence
VDAC	voltage-dependent anion channel
v/v	volume per volume
vlp	virus-like particle
w/v	weight per volume
w/w	weight per weight
zcp	zero current potential

Å	Angstrom
ε	Extinction coefficient
°C	degree Celsius

1 Introduction

1.1 Membranes

Biological membranes are the boundaries of living cells shielding them from environmental influences and separating compartments for special purposes in eukaryotic cells. They are no static barriers but enable transport of solutes. To function efficiently they must selectively pass ions and molecules from one side to the other. Biological membranes are complex systems consisting essentially of lipids and proteins. Lipids are amphiphilic compounds with a hydrophilic head group attached to a hydrophobic hydrocarbon tail. Amphiphilic molecules tend to form ordered aggregates. The energetically most favourable form for many lipids in an aqueous surrounding is a two-dimensional bilayer where the polar head groups face the aqueous compartment and the apolar chains are sandwiched between them. Water soluble compounds within the cell and surrounding it are not soluble within the lipid bilayer and pass the membrane slowly or not at all. Therefore a hydrophobic lipid bilayer is the best barrier between aqueous compartments (Adam *et al.*, 1995; Singer *et al.*, 1972).

Gram-negative bacteria, and according to the endosymbiontic theory also mitochondria and plastids, are surrounded by two lipid bilayers. The inner membrane is often a phospholipid bilayer that separates the cell; the outer membrane comprises lipopolysaccharides in the outer leaflet, maintains the shape, and shields the cell.

1.1.1 Membrane Proteins

Every biological membrane is associated with numerous proteins that enable the membrane to conduct its specific activity. Proteins can be associated with the membrane in several ways. They are dived in two categories: polytopic and monotopic membrane proteins. Polytopic membrane proteins are in contact with both leaflets of the membrane while monotopic membrane proteins are only in contact with one. Usually, when referring to membrane proteins the focus is on integral membrane proteins which are embedded into the lipid membrane with at least one part of the molecule. Most of them contain hydrophobic helical segments spanning the membrane one to more than 20 times. A minority forms a hydrophobic β-barrel spanning the outer membrane of Gram-negative bacteria, mitochondria, and plastids and even in a group of Gram-positive bacteria. Monotopic membrane proteins lack an integral membrane domain. They are associated with the membrane by one or more amphipathic helices embedded into the membrane but not spanning it, by hydrophobic loops, by interaction with integral membrane proteins or polar lipid head groups or by covalent association with a hydrocarbon chain.

1.1.1.1 Inner Membrane Proteins

In every genome sequence investigated so far 20 % to 30 % of all open reading frames code for proteins with predicted membrane transmembrane segments (Wallin *et al.*, 1998). Proteins residing in

the bacterial inner membrane fulfil important functions including metabolism, solute transport, respiration, energy conversion, and signal transduction. The generation of ATP by oxidative phosphorylation and photosynthesis occur in the inner membrane.

Among the most prominent bacterial inner membrane proteins are the complexes of the respiratory chain, four membrane associated complexes that couple oxidation of NADH and $FADH_2$ with the reduction of oxygen and establish a proton gradient across the membrane that drives amongst others the ATP production. The electron transport system of aerobic respiratory consists of the complex I or NADH:ubiquinone oxidoreductase, complex II or succinate:quinone reductase, complex III or ubiquinole: cytochrome *c* oxidoreductase and complex IV or cytochrome c oxidase. ATP is generated from ADP by the ATP synthase. Depending on the organism and the respiratory conditions, different kinds of complexes e.g. different terminal oxidases or different subunit compositions are present.

Proteins involved in transport differ according to the mechanism and the substrate. In general, transporter proteins are classified in three groups: ion channels and primary and secondary active transporters (Saier, 2000). Ion channels allow diffusion of solutes along a concentration gradient. Transporters can transport their substrates against a concentration gradient: primary transporters obtain the energy from the use of ATP while secondary transporters make use of a secondary energy source like ion gradients established by ATP-dependent ion pumps or respiration. Most families of channel proteins possess three or fewer TMHs per polypeptide chain, while almost all families of secondary carriers include members that possess eight or more TMHs.

1.1.1.2 Outer Membrane Proteins

β-barrel membrane proteins are found in the outer membrane of Gram-negative bacteria, mitochondria, and plastids. Approximately 2 % to 3 % of the genome of Gram-negative bacteria encode β-barrels forming proteins (Wimley, 2003). A variety of functions is provided by those proteins: the most prominent is permeation through the outer membrane by pore forming proteins called porins; others are adhesins, proteases, lipases, or pore-forming toxins.

The TransportDB (Saier, 2000) describes several different classes of outer membrane proteins according to substrates specificity based on biochemical analysis or phylogenetic data. However low sequence conservation and inconsistent functional characterisation makes it difficult to compare and classify porins from different species. Therefore classification is most commonly performed based on substrate specificity and the mode of permeation according to Nikaido (Nikaido, 2003): Three classes of porins are generally distinguished: general non specific porins, substrate specific porins and active transporters. The longest known and best studied porins are the "classical porins" OmpF, OmpC and PhoE from *Escherichia* (*E.*) *coli*, belonging to the non specific class. They form water-filled pores of 16 stranded β-barrel trimers (Cowan *et al.*, 1992; Schirmer, 1998; Weiss *et al.*, 1991) which facilitate the diffusion of small polar substrates (less than 600 Da) across the outer membrane, but do not have a

binding site for their substrates (Benz *et al.*, 1985; Rosenbusch, 1974). They are only slightly selective for cations (OmpF and OmpC) or anions (PhoE).

The substrate specific porins such as the nucleoside transporter Tsx (Benz *et al.*, 1988; Hantke, 1976) or the maltose channel LamB (Szmelcman *et al.*, 1975) have a saturable binding site for their substrate and much lower single channel conductances. Most specific porins form 18 stranded β-barrels and facilitate the diffusion of carbohydrates, nucleotides, or lipids with mM to μM binding constants. OprO and OprP from *Pseudomonas* (*P.*) *aeruginosa* are the only porins specific for small ions, pyrophosphate and phosphate, respectively. OprP is a trimer of a 16 stranded β-barrel (Moraes *et al.*, 2007).

The active transporters like FhuA bind a substrate and are able to transport it against a concentration gradient (Ferguson *et al.*, 1998; Locher *et al.*, 1998). The energy required for the transport is supplied by the inner membrane protein TonB.

Activity of porins is most commonly assayed by determination of ion or solute fluxes across the membrane. Four different techniques are most commonly used today: determination of the flux of ions or hydrophilic solutes across the outer membrane of intact cells; reconstitution of the protein into liposomes for liposome swelling assays, reconstitution into planar lipid bilayers, few studies have used the patch clamp technique on liposomes. Using planar lipid bilayers or the patch clamp technique has the advantage that the action of single channels can be measured. Since the planar lipid bilayer technique, which is also called black lipid membrane (BLM) technique, was established in 1963 by Müller and co-workers (Müller *et al.*, 1963) it is frequently used to investigate current flows across a lipid bilayer: it is the most frequently used method to investigate outer membrane porins (Nikaido, 2003), the results can be compared with those from other proteins and it is comparably easy to set up.
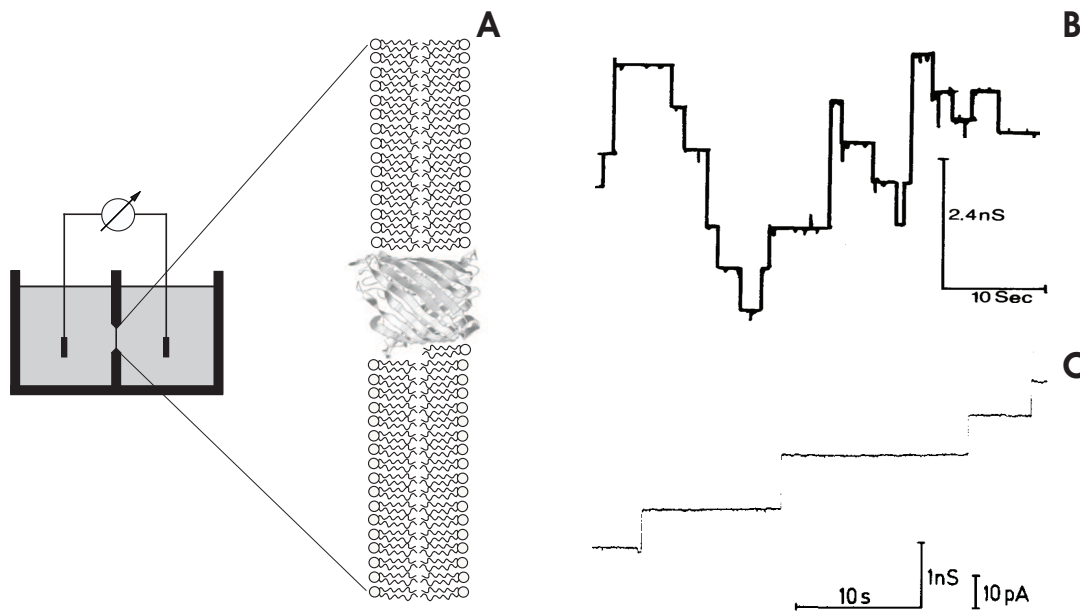
Figure 1.1-1: A: Schematic view of a black lipid membrane. B and C: Single channel recording of OmpF from *E. coli*. B: (Phale *et al.*, 1997), C: (Benz *et al.*, 1978)

The planar lipid bilayer in a BLM setup is formed from a lipid dissolved in a non-volatile solvent (e.g. n-decane) which is painted across the aperture separating the two half cells of the Teflon cuvette. Upon optical inspection the lipid-solvent mixture shows interference colours immediately after painting. Through diffusion the organic solvent thins out and collects at the rim of the aperture leaving behind a lipid bilayer with a thickness of 5 to 9 nm, which appears black; this optical effect has given the technique its name.

This system offers accessibility of both compartments and thereby allows investigation of transport processes across the membrane. Integration of membrane proteins is easy to realise, most outer membrane proteins are added to one aqueous compartment in their detergent solubilised form and integration appears spontaneous.

In addition to the determination of channel conductance permeability ratios for cations over anions (P_a/P_c) can be determined. According to the model of Goldman-Hodgkin-Katz (GHK) the selectivity of the macroscopic conductance is obtained from the potential over the membrane in the presence of an ion gradient (Goldman, 1944; Hodgkin *et al.*, 1949):

$$zcp = \frac{R*T}{F} \ln\left(\frac{P_c c^{trans} + P_a c^{cis}}{P_c c^{cis} + P_a c^{trans}} \right)$$

$$\frac{P_a}{P_c} = \frac{c^{trans} * \exp\left(\dfrac{zcp*F}{R*T} \right) - c^{cis}}{c^{trans} - c^{cis} * \exp\left(\dfrac{zcp*F}{R*T} \right)}$$

Formula 1.1-1: Goldman-Hodgkin-Katz (GHK) equation (Goldman, 1944; Hodgkin *et al.*, 1949).

where zcp is the zero current potential (reversal potential), c^{cis} and c^{trans} are the salt concentrations on the concentrated and the diluted side of the membrane, respectively. P_a and P_c are the permeability coefficients of the anion and the cation, respectively and R, T, and F have their usual meanings.

The functional properties of the general outer membrane porins have been investigated since decades mainly by planar bilayer lipids. Still functional data is often difficult to compare because of diverging purification methods, buffer conditions or the definition of "single-channel" conductance (monomer vs. trimer). Nikaido suggested referring to an OmpF conductance of 0.7 nS for a monomer in 1 M KCl. It is still a matter of debate if the porin pore is able to close (Figure 1.1-1 B) or if it remains a static open channel after insertion (Figure 1.1-1 C).

1.1.1.2.1   Phosphate-selective Porins

The porin PhoE from *E. coli* is still called the phosphate selective porin because it is induced under phosphate starvation. However, functional characterisation showed that it is only slightly anion selective and does not have a phosphate binding site. Today it is considered as general porin.

In contrast to this the porins OprO and OprP, which are induced under phosphate starvation in *P. aeruginosa*, are highly anion-selective. OprP has a single channel conductance of 160 pS in 100 mM chloride solutions (Benz *et al.*, 1987) and a high affinity binding site for phosphate with a K_d of approximately 0.1 µM (Benz *et al.*, 1993). Its three-dimensional structure give evidence for its phosphate selectivity: a "ladder" of nine arginine residues span the barrel from the exterior to the constriction zone where a phosphate ion is bound (Moraes *et al.*, 2007). Towards the periplasmic side twelve lysines create a positively charged surface that pulls the phosphate ions into the cell.

Several proteins have been classified as "phosphate-selective porin O and P" and grouped into the outer membrane protein family *Pseudomonas* OprP porin family (POP), and are predicted to be anion specific with a binding site for phosphate, although none have been studied so far.

1.1.2 Membrane Protein Production and Purification

A major bottleneck in understanding membrane proteins is still the lack of high amounts of pure and homogenous membrane protein suitable for functional and structural studies for many membrane protein families.

Only very few membrane protein exist in the native membrane in abundance high enough to meet the needs of such studies. Historically photosynthetic membrane proteins were first characterised functionally and their structure determined by X-ray crystallography (Deisenhofer *et al.*, 1985; Essen *et al.*, 1998; Ferreira *et al.*, 2004; Jordan *et al.*, 2001). Proteins of the respiratory chain and many bacterial outer membrane porins occur in relatively high amounts and many are functionally and structurally well characterised. Depending on the kind of cell or organism other proteins are highly abundant: the aquaporin water channel from bovine lens (Gonen *et al.*, 2004) or red blood cells

(Murata *et al.*, 2000), the acetylcholine receptor from torpedo muscle (Miyazawa *et al.*, 2003) or rhodopsin from the bovine rod outer segment (Palczewski *et al.*, 2000). A major drawback in using native membrane protein is the restricted number of proteins produced natively in an amount high enough to be purified in high yields and purity. In addition, native protein complexes can have an inhomogeneous composition of subunits which is a possible source for inhomogeneity of the protein sample. In many cases it is not possible to introduce non-native or labelled amino acids that are helpful for NMR or X-ray crystallography based studies, often impeding structure determination.

Less abundant proteins like many channels and transporters are not accessible to purification from native sources. Therefore over-production is needed for such proteins to attain high amounts suitable for such studies. Using diverse expression systems was a major breakthrough in protein research. By this means hundreds of structures of soluble proteins could be solved, the number of protein data bank (pdb) entries is increasing exponentially. Also membrane proteins are included in these approaches: The structures of several membrane proteins including primary and secondary transporters and ion channels could be elucidated and broadened the understanding of the cell membrane.

However, over-production of both homologous and heterologous membrane proteins is often harmful to the membrane and hence to the growth of the host cell, if not lethal (Miroux *et al.*, 1996). Membrane protein production includes not only transcription and translation but also insertion and proper folding of the protein within the membrane. Until today no ideal membrane protein production system exists. In general heterologous over-production has been most successful to obtain functional protein (Bannwarth *et al.*, 2003) in cases where the target protein and the expression host is closely related and for homologous production. The relation of the produced protein and the host can be critical for proper membrane insertion and folding. Especially outer membrane porins failed to be produced functionally if the target protein and the expression host are not closely related. However only few organisms are applicable for plasmid-based expression systems and genetic modification is often difficult and time consuming. Therefore most frequently the *E. coli* expression system is used to produce proteins. *E. coli* is simple to handle, inexpensive to grow and in addition a variety of different expression vectors with diverse promotors are available. It has also been proven to successfully produce many prokaryotic membrane proteins (Surade *et al.*, 2006; Wang *et al.*, 2003) in a functional, membrane inserted form suitable in quality and quantity for crystallisation attempts. However many outer membrane porins failed to be produced in a functional form. Surprisingly the production of many porins in non-functional inclusion bodies with subsequent solubilisation under denaturing condition and refolding in the presence of detergent into the native and functional form was successful.

Integral membrane proteins are more complicated to handle than their soluble counterparts. For their purification the native membrane has to be disrupted and replaced by detergent molecules as membrane proteins are not soluble in aqueous buffers. Detergents are amphiphilic compound like lipids, but in contrast to lipids, do not form a bilayer under many conditions. At low concentrations

they form a monolayer at the water-air interface. Above a certain concentration, the critical micelle concentration (CMC), they rearrange to form micelles: the hydrophobic tails of the detergent molecules are oriented toward the inside, while the polar head group faces the aqueous surrounding. The membrane spanning part of the protein is buried by the detergent micelle in a belt-like fashion, and the micelle shields it from the contact with water. Detergents have to be present throughout all purification steps, and during functional and structural studies. Still membrane proteins tend to form aggregates hampering the efficiency of purification and reducing the yield. The choice of detergent, ionic strength and pH is therefore essential for effective, non-denaturing solubilisation and needs to be optimised for each membrane protein.

Many standard chromatographic purification techniques are frequently applied to membrane proteins, i.e. gel filtration, ion exchange or affinity chromatography including dye, ligand, or metal affinity chromatography or density gradient centrifugation and native isoelectric focusing. However, the presence of detergents can restrict implementation of certain purification techniques or lower the resolution.

Purification of membrane proteins usually involves several separation steps; however, the time for isolation should be kept as short as possible to avoid proteolytic degradation and with the lowest number of purification steps involved to prevent delipidation and potential loss of activity and structural integrity. Once a membrane proteins is available in milligram amounts the next logical step is to investigate its three-dimensional structure.

1.2 Structure Determination

Despite their importance in many cellular functions just very few membrane proteins structures have been investigated in detail. Today only approximately 120 unique structures of membrane proteins have been determined, many of them outer membrane proteins (see below). This is a very small number compared to more than 15000 non-redundant structures of soluble proteins. There are three major tools available to obtain structural information of a protein: nuclear magnetic resonance (NMR) spectroscopy, electron microscopy (EM), and X-ray crystallography. Most of these structures have been determined by X-ray crystallography.

NMR can be used to study the three-dimensional structure of a protein in solution, in contrast to electron crystallography and X-ray crystallography which need crystalline samples. Therefore it offers the possibility to obtain a structure under conditions more similar to the physiological condition of the protein of interest. NMR is frequently used in structural genomics approaches, especially if no crystals suitable for X-ray crystallography can be obtained. NMR provides information on the atomic level but it is limited by the overall size of the protein. The limit on the molecular weight is at about 30 kDa, therefore investigations are often limited to small proteins, single domains or to synthetic peptides derived from a loop or a transmembrane helix of a membrane protein (Dmitriev *et al.*, 1999; Jung *et al.*, 1995; Wüthrich, 1998). Technical progress such as the improvement of sensitivity by new high

field magnets and the methodically advances like solid state NMR or TROSY (transverse relaxation-optimised spectroscopy) made it possible to exceed NMR studies to structures with a size of more than 100 kDa, such as membrane proteins in micelles or lipid vesicles (DeAngelis *et al.*, 2006; Wüthrich, 1998). However only few three-dimensional structures of membrane proteins could be solved by NMR spectroscopy including the backbone assignments, and three-dimensional structures determined are the α-helical glycophorin A (MacKenzie *et al.*, 1997) from human erythrocytes, mistic membrane-integrating protein from *Bacillus subtilis* (Roosild *et al.*, 2005) and the ATPase modulator phospholamban from human sarcoplasmic reticulum (Oxenoid *et al.*, 2005) and the β-barrel *E. coli* outer membrane proteins OmpA (Arora *et al.*, 2001), OmpX (Fernandez *et al.*, 2004) and PagP (Hwang *et al.*, 2002).

Electron microscopy is applicable to biological specimen with dimensions spanning several orders of magnitude from only few kilo Dalton to several mega Daltons. Complexes that are larger than approximately 100 kDa can be visualised by electron microscopy from negatively stain samples. This is a fast and simple method that requires only micrograms of starting material and also applicable to determine the homogeneity of a sample. However, the macromolecular complex has to be treated with heavy atoms in order to enhance the contrast, a treatment which does not reflect native conditions. Observation of molecules in an aqueous environment can be achieved by cryo-EM. The molecules are immobilised in vitreous ice which offers the possibility of trapping functional states. In both cases two or three-dimensional reconstruction is achieved on randomly oriented single particles. Single particle analysis can give structural information, and in combination with X-ray crystallography or NMR it can provide information e.g. on the assembly of macromolecular complexes.

For complexes that are much smaller than 100 kDa two-dimensional crystals, crystalline arrays of the membrane protein together with lipid molecules, are required which are studied by electron crystallography. The protein structure is obtained in a membrane embedded form instead of a detergent micelle which might better reflect the native membrane. In addition many membrane proteins form two-dimensional crystals much more readily than three-dimensional crystals (Renault *et al.*, 2006). Electron crystallography revealed several structures of membrane proteins at medium resolution (5 Å to 10 Å) that allow the assignment of secondary structure motifs such as transmembrane α-helices. However, only very few membrane proteins yielded crystals diffracting to atomic resolution. The water channel Aqp0 provides an exceptionally high resolution of 1.9 Å (Gonen *et al.*, 2005), a resolution which is very close to 1.4 Å, the highest resolution of a membrane protein provided by X-ray crystallography for the ammonia transporter AmtB (Khademi *et al.*, 2004).

1.2.1 X-ray Crystallography

A major prerequisite for any crystallisation attempt is a protein which is chemically and structurally homogenous. The larger the amount of available protein the higher is the chance to fulfil these conditions. A large amount of protein usually means a favourable relation of protein to impurity and

the utilisation of only the purest fraction during purification. Another aspect is that rigid and stable proteins have a much higher propensity to crystallise. Flexibility is impeding crystals growth in many cases.

In general, the crystallisation techniques for membrane proteins are the same as for soluble protein: in most cases the traditional vapour diffusion technique is used. Since established more than two decades ago, the most successful technique is to crystallise a membrane protein from an aqueous detergent solution. However, the presence of detergent adds an additional level of complexity. Similar to soluble proteins membrane protein crystals are mainly stabilised by polar contacts between hydrophilic parts. The amphilic nature of membrane proteins comprises a limited hydrophilic surface capable of mediating crystal contacts. The hydrophobic part that is usually buried by the membrane does not contribute much to the order of the crystal. Therefore the size of the detergent micelle and thus the choice of detergent is a crucial factor in the crystallisation of membrane proteins. A detergent micelle too large in size will impede crystal contacts while a too small micelle might not shield the hydrophobic part and thus lead to protein denaturation. The size of the detergent micelle depends on the type of detergent, the ionic strength, and pH and also on the protein itself. It is possible to alter the size of the micelle by adding small amphilic molecules like heptane 1,2,3 triol.

Proteins with large hydrophilic domains are more likely to crystallise, in some approaches additional hydrophilic domains have been introduced by co-crystallisation with antibody (F_v) fragments (Ostermeier *et al.*, 1995). However, outer membrane proteins formed crystals of high quality although most of the protein is buried by the membrane. The major reason might be the extreme stability of the protein in addition to strong polar crystal contacts.

1.3 Proteomics

Since the first entire genome of a free-living organism, *Haemophilus influenzae*, was sequenced in 1995 (Fleischmann *et al.*, 1995), the genome of several species especially bacteria have been sequenced and many more are to come. Among others two fields of study benefit most from the accessibility of genomic data: structural genomics which aims to determine the structure of the proteins and proteomics, directed towards the identification of the proteins.

However, in any newly sequenced genome, for between 20 % and 50 % of the open reading frames (ORF) it is not known whether they are transcriptionally active. Their potential gene products are known as hypothetical proteins. Two classes of hypothetical proteins can be distinguished: hypothetical proteins which have homologous proteins in other organisms and hypothetical proteins which do not. The first group is usually referred to as conserved hypothetical proteins while the second group consists of species- or genus-specific genes as well as wrongly annotated ORF, which are not transcriptionally active. The function as well as the cellular localisation for these hypothetical proteins is not established and remains unclear. Proteins with predicted transmembrane helices are overrepresented among the unknown proteins (Schwartz *et al.*, 2001).

Proteomic studies are the only way to prove the existence of the hypothetical protein and simultaneously necessitate the renaming from hypothetical to existent proteins. To discover the function of these unknown proteins will be a very important step towards understanding of the cellular function (Galperin, 2001; Lubec *et al.*, 2005; Yakunin *et al.*, 2004). The function of uncharacterized proteins can be deduced using sequence-level similarity searches of known homologous proteins, protein domains, or families (Lubec *et al.*, 2005). Tentative functions can subsequently be experimentally confirmed (Murzin *et al.*, 1999). Determination of the structure is also a valid approach to discover a protein's function (Vitkup *et al.*, 2001; Zarembinski *et al.*, 1998).

By definition the proteome of a cell is the collectivity of all proteins present in the cell under defined conditions. Due to the complexity of the mixture usually subfractions like the membrane or mitochondrial proteome after multi-dimensional separation are studied.

In general a gel-based separation has prevailed, typically by isoelectric focussing (IEF) followed by SDS PAGE. IEF separates polyions like proteins in a pH-gradient according to their net charge (Svensson, 1961; Svensson, 1962). At the pH value where the net charge is zero, the isoelectric point, the mobility in the electric field becomes zero as well, they finally come to rest. SDS PAGE is a well established technique developed by Laemmli to separate proteins according to their apparent molecular weight in a denatured, SDS bound form (Laemmli, 1970). The combination of both methods has proven to be successful in many proteomics studies. Membrane proteins however are frequently undetected in such studies; in general no integral membrane protein is detected. In contrast to a separation by SDS PAGE, which is well suited even for very hydrophobic membrane proteins, these proteins tend to precipitate during IEF at their isoelectric point and are then not capable of being transferred to the second PAGE dimension (Santoni *et al.*, 2000). The loss of membrane proteins is avoided by adding detergents to the isoelectric focussing buffer in order to increase the solubility of the membrane proteins, especially zwitterionic like CHAPS or Zwittergent detergents have been successfully applied (Aivaliotis *et al.*, 2006). Other approaches avoid the IEF and replace it by a different PAGE system e.g. double SDS PAGE with different acryl amide concentrations or addition of urea in one dimension (Rais *et al.*, 2004) or by replacing SDS with a different detergent, e.g. benzyldimethyl-n-hexadecylammonium chloride (BAC) (Macfarlane, 1983) or cetyltrimethylammonium bromide (CTAB) (Helling *et al.*, 2006) have been successfully used in several studies. However, in all of these approaches the protein is not required in its native state and is separated using mainly denaturing conditions. Thereby information on protein complexes is lost. Blue Native (BN) PAGE in the first dimension offers a solution for the analysis of protein complexes. Membrane proteins bound to the dye Coomassie blue G-250 are separated according to their size in the acryl amide gradient gel. The method was developed for the separation of membrane proteins and membrane proteins often retain their activity (Schägger *et al.*, 1991). However, all gel-based approaches were optimised for small scale preparations of microgram amounts of membrane protein.

10

Up scaling to a preparative scale satisfying the need of functional or structural studies, is often difficult if not impossible.

In order to avoid denaturation at early purification stage and limitation to small scale preparations, liquid chromatography under native conditions is used to pre-fractionate the protein mixture prior to one-dimensional SDS PAGE. It allows the choice of a wide variety of detergents, ionic strength, and additives during solubilisation and purification, and chromatographic techniques under non-denaturating conditions prior to the final SDS PAGE. Although not very frequently used, it has already proved successful in the enrichment and identification of soluble and membrane proteins, even very low abundant ones (Everberg *et al.*, 2006; Fountoulakis *et al.*, 1999; Schlüsener *et al.*, 2005; Szponarski *et al.*, 2004). Schlüsener and co-workers showed that the number of integral membrane proteins could be increased compared to gel-based separation without losing information on size, occurrence and protein-protein interaction and protein complexes (Schlüsener *et al.*, 2005).

Proteins are most frequently identified by in-gel trypsin digestion and matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) followed by peptide mass fingerprinting. Protein bands are excised from SDS PAGE gel, treated with the protease trypsin and the peptides analysed by mass spectrometry. The technique was optimised to the need membrane proteins (Shevchenko *et al.*, 1996). Proteins are identified by matching the constituent peptide masses to the theoretical peptide masses generated from genome databases. Several peptides have to be detected in order to gain sequence coverage that is high enough to confidently identify the protein. However, the identification of membrane proteins comprises several problems. In the hydrophobic regions of proteins only few trypsin cleavage sites are present, hydrophobic peptides are hardly soluble and difficult to extract and handle. Therefore hydrophobic parts remain generally undetected and the sequence coverage of membrane proteins is low, especially it the respective protein does not contain extended hydrophilic domains. Thus hydrophobic membrane proteins are often not detected by approaches based on peptide mass fingerprinting following MALDI-TOF MS. The probability of identifying membrane proteins is increased by tandem mass spectrometry (MS/MS). It involves fragmentation of the respective peptide and thereby identification of the peptide sequences. The identification of even very hydrophobic membrane proteins is feasible because only one or few peptides of a protein needs to be present in order to identify the protein.

Currently most success in identification of membrane proteins is achieved by so called shot gun approaches: high resolution liquid chromatography of peptides is followed by tandem mass spectrometry (LC-MS/MS) identification of the proteins (Li *et al.*, 2004; Zhang *et al.*, 2007; Zhang *et al.*, 2004). However a major drawback of this method is the loss of information on the size and amount of the protein and on protein-protein interactions.

So far the membrane proteome of only one thermophilic bacterium was determined, *Geobacillus thermoleovorans*, using a shot gun method (Graham *et al.*, 2006).

1.4 Bioinformatic Characterisation

Sequence data obtained from genome sequencing projects is available on several web based databases. Most genome databases already provide information on genes and proteins including bioinformatic data like automated annotation of the function, homologous sequences and domains, or related literature. To achieve further information about a gene or protein on the basis of its sequence a variety of methods are available today.

Among the most commonly used tools to achieve information on a protein is the sequence alignment. It offers the possibility to deduce the function of a protein, to classify it in a protein family or to predict its evolutionary origin by comparison to proteins of known function.

Two programs are most commonly used today: FASTA (FAST-All) (Pearson *et al.*, 1988) and BLAST (Basic Local Alignment Search Algorithm) (Altschul *et al.*, 1990). Both programs compare the protein on basis of local alignment (Smith *et al.*, 1981), an algorithm optimised for dissimilar sequences or similar regions within a larger context. The similarity is computed on basis of a substitution matrix that contains scores for all possible substitutions of an amino acid with another (Henikoff *et al.*, 1992). In most cases the matrix BLOSUM62 is used, a matrix which is calculated from observed substitution of residues in proteins with a sequence identity of 62 %. In order to speed up the search time both programs are heuristic, i.e. they restrict the search space and search only against selected sequences.

The statistics of scoring are usually given in the so called E values, a measure for the probability that two sequences are not similar. However, for FASTA and BLAST it is only an approximation; the accuracy of the score is neglected in order to accelerate the search. FASTA estimates the scores on basis of real database searches while BLAST estimates them before hand on basis of comparison of random sequences. Therefore, the similarity between two sequences should always be evaluated by means of sequence identity and homology as well. In general two sequences can only be considered as significantly homologous if at least 25 % of the residues are identical.

One major problem of alignments are the so called regions of low complexity, regions with highly biased composition. Here, alignments have little meaning and are therefore excluded from the sequence analysis (Wootton *et al.*, 1993). Up to one fourth of all residues are located in such regions.

Many functional important protein similarities are detectable only through comparison of structural information rather than through sequence comparison (Brenner *et al.*, 1998). If there is no three-dimensional structure available, pattern of conservation which are derived from the alignment of related sequences aid in the recognition of distant similarities. Algorithms like the Position Specific Iterated BLAST (PSI-BLAST) are able to uncover relationships between proteins, which have not been discovered by sequence alignment and were previously only detected from the three-dimensional structure (Altschul *et al.*, 1997; Huynen *et al.*, 1998; Mushegian *et al.*, 1997) by taking the sequence conservation at each position into account.

Fold recognition follows the line of structural comparison. It attempts to recognise similarity between the proteins´ three-dimensional structure that are not accompanied by significant similarity of the

sequence level. The sequences are fitted to known structures and best matches are recognised. The success rate depends on several factors including the algorithm and the correct prediction of the fold. However, fold recognition will not be able to recognise a "new" fold.

Most proteins are made up of functional domains, which are repeated within the protein or within different protein from different species and might thereby contain attributes of several ancestral genes. A structural domain is a part of the protein that forms a stable structure and is able to fold independently. These structural domains are often associated with a biochemical function. Another approach is to refer to the homology towards functional protein domains and motifs rather than whole proteins in order to deduce the proteins function.

A variety of databases are available online to detect protein domains, among them several composite databases like InterProScan from the European Bioinformatics Institute (EBI). Integrated into this server are several algorithms to detect domains, TMH, or signal peptides. Another important program is SMART (Simple Modular Architecture Research Tool), that identifies domains and also includes signal peptides (Letunic *et al.*, 2006).

The prediction of transmembrane helix is achieved by bioinformatic algorithms. The simplest bioinformatic algorithm is to search for a segment of 15 to 35 hydrophobic amino acids (von Heijne, 1999). A measure for hydrophobicity is the GRAVY-index (Grand Average of Hydrophobicity) based on empirical values for each amino acid of a protein (Kyte *et al.*, 1982), a positive value stands for hydrophobic amino acids, a negative for hydrophilic. More elaborate algorithms make use of the charge distribution as well ("positive inside rule") (von Heijne, 1986). There are several web based algorithms available, one of the most specific and most commonly used algorithms of today is the prediction program TMHMM (Krogh *et al.*, 2001).

β-barrel proteins spanning the outer membrane are made up by 4 to 22 strands in 1 to 3 protein subunits. In most cases these strands are interconnected by short turn sequences facing the periplasmic side and longer loop sequences facing the external space. The comparison and identification of outer membrane proteins by means of the sequence alignments is extremely difficult because porins lack extended hydrophobic stretches (Cowan *et al.*, 1994). Relative simple algorithms rely on the detection of a 9 to 10 amino acid long stretch with an alternation of hydrophobic and polar amino acid exposed to the membrane and the aqueous phase, respectively, on β-turn exposed to the aqueous phase or on hydrophobicity profiles (Paul *et al.*, 1985; Vogel *et al.*, 1986). The sequence conservation is usually low especially in the external loops between two transmembrane β-strands because these regions undergo rapid mutation due to environmental influences (Nikaido, 2003) thus hampering identification by sequence alignments. The sequence conservation is only high within one group of bacteria (e.g. within the α or the γ proteobacteria the sequence conservation is high but α compared to γ it is low). Therefore one way to compare porin sequences is first to identify the transmembrane β-strands and compare the sequences on basis of the strands rather than the loops. One among the few web-based

prediction methods is the program PRED-TMBB (Bagos *et al.*, 2004), based on a Hidden Markov Model.

The bioinformatic prediction of monotopic membrane proteins is biased. Protein with covalently attached lipid anchors can be recognised by sequence motifs. However no bioinformatic algorithm is available to predict proteins containing amphipathic helices.

Signal peptides, short hydrophobic peptides located at the N-terminus that direct post-translational transport of the protein, are often wrongly annotated as transmembrane domains (Krogh *et al.*, 2001). Several algorithms are available to identify signal peptides (SignalP, SMART), some of them also provide information on the probability if and where the peptide is cleaved.

1.5   *Aquifex aeolicus*

In 1992 Huber, Stetter and co-workers described a new genus of thermophilic eubacteria which they called *Aquifex*, the water-maker, according to its ability to make water from hydrogen (Huber *et al.*, 1992). Among the best studied representatives is *Aquifex* (*A.*) *aeolicus*. It was first obtained at the Aeolic Islands (Italy).

It is a rod shaped non-sporefoming, Gram-negative bacterium. The optimal temperature for growth is 85 °C; the minimum is 65 °C and the maximum is 95 °C. As an obligate autotroph, it grows in an atmosphere of hydrogen, carbon dioxide and oxygen (79.5 : 19.5 : 1.0) in a medium containing exclusively inorganic compounds and does not grow on a number of organic compounds. Energy is mainly obtained from the oxidation of hydrogen. The organism is strictly aerobic, up to now it was not possible to culture it with an alternative electron acceptor than oxygen. The *Aquificaceae* are among the deepest branching family within the bacterial domain. Based on analysis of the 16S ribosomal RNA sequence, it is closest to the last common ancestor of bacteria and archea, although on the basis of individual proteins the placement relative to other groups varies (Griffiths *et al.*, 2004).

Its genome (Deckert *et al.*, 1998) was the second of a hyperthermophile to be sequenced and the first of a hyperthermophile eubacterium. The availability of the genome sequence made the organism applicable for genomic and proteomic studies. The genome is relatively small, 1.5 Mb, which is only one third the size of the *E. coli* genome, and codes for 1522 proteins. The genome is densely packed, with many genes overlapping slightly. Most genes are expressed in polycistronic operons but unlike other organisms, where genes belonging to one pathway are grouped in one operon, these genes are often dispersed throughout the genome or found in novel operons in *A. aeolicus*. Even genes coding for subunits of the same enzyme are often separated. This is especially true for genes coding for the biosynthesis of amino acids. Additionally there often occur two genes coding for one protein or subunit isoforms. In contrast, genes coding for proteins of the respiratory chain, hydrogenase subunits, are often placed in related operons. To date, genetic manipulation in the organism by a genetic system is not possible.

1.5.1 Inner Membrane Proteins

A. aeolicus is a hydrogen oxidising bacterium, and its genome contains three [NiFe] hydrogenases: one soluble and two which are associated with the membrane through an integral b-type heme containing protein. All three have been described in the literature to be present in the cell (Brugna-Guiral *et al.*, 2003).

All complexes of the respiratory chain are present according to database annotation. The NADH:ubiquinone oxidoreductase (complex I of the respiratory chain) with 24 putative subunits, many of which are different isoforms of a single subunit, was described by Peng and co-workers in 2003. A fumarate reductase (complex II of the respiratory chain) with one flavoprotein subunit and two different iron-sulphur protein subunits is also present in the genome; the presence of the complex has not been shown yet. A ubiquinol:cytochrome *c* oxidoreductase (*bc*$_1$ complex or complex III of the respiratory chain) was described by Schütz and co-workers (Schütz *et al.*, 2003); in addition, two genes encoding for cytochrome *c* molecules are present and both proteins have been described by Baymann and co-workers (Baymann *et al.*, 2001). A cytochrome *c* oxidase (complex IV of the respiratory chain) with two different genes encoding for subunits I and subunit II, respectively, and an alternative oxidase with a cytochrome *bd* ubiquinole oxidase are also present in the genome. Other metabolic pathways in *A. aeolicus* have been described in the literature, including sulphide oxidation by the membrane associated sulphide:quinone oxidoreductase (Nübel *et al.*, 2000) and reduction of sulphur by the dimethyl-sulphoxid-reductase complex (DMSO reductase) (Guiral *et al.*, 2005). Numerous oxidoreductases are present. Several membrane proteins are involved in biogenesis of membrane proteins and protein complexes, e.g. in flagellar or cytochrome *c* biosynthesis.

The TransportDB (http://www.membranetransport.org, (Ren *et al.*, 2004)) annotated 64 protein from *A. aeolicus* as transport proteins: 25 primary transporters, namely 19 are subunits of ABC-transporters, 3 are subunits of the F_1F_O-ATP synthase and 3 are P-type ATPases, 7 ion channels out of four families, 32 secondary transporters out of 18 families and one unclassified transporter. 55 % of all proteins with predicted membrane spanning domains are hypothetical proteins although several of those are annotated for example as transport proteins.

1.5.2 Outer Membrane Proteins

According to NCBI annotation *A. aeolicus* has six outer membrane porins: OstA, an liposaccharide export protein; OtnA, a polysaccharide biosynthesis protein; Omp, an outer membrane insertion protein; OprC, a TonB dependent receptor; Pal, an OmpA-OmpF-like protein and GspD, a general secretion pathway protein D protein. The TransportDB (Ren *et al.*, 2004) predicts several additional outer membrane proteins; however most are still annotated as "hypothetical proteins". No biochemical study has been carried out on the outer membrane of *A. aeolicus* so far to prove these predictions.

1.6 Proteins from Thermophilic Organisms

Proteins from thermophilic organisms have been of interest for decades. They are known to be stable and active under extreme conditions although they have the same overall fold and make use of the same molecular mechanism for their function as proteins from mesophilic organisms (Russell *et al.*, 1997; Vieille *et al.*, 1995). Much interest has been directed in understanding the thermal stability of thermophilic proteins (Perutz, 1978; Perutz *et al.*, 1975; Razvi *et al.*, 2006). It became evident that the stability cannot be assigned to a single factor, but is made up of a number of factors (Jaenicke *et al.*, 1998; Perutz *et al.*, 1975). Proteins from thermophilic organisms have been reported to comprise networks of ion pairs especially at the protein surface (Maes *et al.*, 1999; Pace, 2000; Perl *et al.*, 2000). Thermophilic proteins are often shorter than their mesophilic counterparts, especially in exposed loop regions (Thompson *et al.*, 1999). An increase of protein compactness i.e. the buried surface (Pace, 1992; Szilágyi *et al.*, 2000) is also observed. The most prominent feature of thermophilic proteins compared to proteins from mesophilic origin is the proportion of charged versus polar residues (Cambillau *et al.*, 2000; Suhre *et al.*, 2003).

Proteins from thermophilic organisms have a higher thermal stability by an increase of the melting temperature, T_m, i.e. the midpoint of the transition between folded and unfolded state, compared to their mesophilic counterparts (Razvi *et al.*, 2006). Thereby thermophilic proteins are more stable at ambient temperatures than mesophilic proteins. Thermal stability is an important factor in protein crystallisation. Therefore, thermophilic organisms are utilised as source organisms by numerous groups aiming to solve three-dimensional protein structures. *Thermus thermophilus* is often used as thermophilic source organism with a growth temperature optimum around 55 °C. However, the observation of hyperthermophilic organisms including *A. aeolicus* with a growth temperature optimum of more than 80 °C offered a source for proteins of even higher thermostability. Also in the case of membrane proteins crystallisation attempts were successful for those derived from hyperthermophilic organisms (Jiang *et al.*, 2003; Lunin *et al.*, 2006; Yamashita *et al.*, 2005). *A. aeolicus*, together with other hyperthermophilic organisms, have become source organisms in many structural genomics consortia. Currently, 58 non-redundant structures from *A. aeolicus* proteins are present in the protein data base (http://www.rcsb.org/pdb/) many of those deposited by structural genomics approaches and including the structure of one membrane proteins, a bacterial homologue of the Na^+/Cl^--dependent neurotransmitter transporters (Yamashita *et al.*, 2005),

1.7 Bacteriocins

In general bacteriocin is the generic term for a broad class of peptides which are released from mainly Gram-positive bacteria and display an anti-microbial spectrum against other strains than the host. It was suggested that mechanism of growth inhibition involves interaction with the membrane and disruption of the membrane potential (La Rocca *et al.*, 1999; Sanchez-Barrena *et al.*, 2003). However, Valdes-Stauber and co-workers discovered a protein from *Brevibacterium linens* they called Linocin

M18 which has the ability to inhibit growth of several Gram-positive bacteria from the same habitat but which is not a peptide but rather a multimeric protein complex.

Genome sequences of several bacteria and few archaea revealed the existence of homologous proteins sharing similarity to the Linocin_M18 domain. Only a few other members of this family have been described: bacteriocin in *Thermotoga* (*T.*) *maritima* (Hicks *et al.*, 1998), virus-like particle, vlp, *Pyrococcus* (*P.*) *furiosus* (Akita *et al.*, 2007; Namba *et al.*, 2005) and CFP29 from *Mycobacterium* (*M.*) *tuberculosis* (Rosenkrands *et al.*, 1998), conducting a variety of functions. However, inhibition of bacterial growth has not been observed for any of these proteins. The functionally best characterised protein is the bacteriocin from *T. maritima*. In the study by Hicks and co-workers in 1998 a proteolytic activity was determined but it remains unclear if and how this is related to a putative antimicrobial effect (Hicks *et al.*, 1998). A common feature of all bacteriocin proteins of this type is the appearance as a complex of high molecular weight, at least 800 kDa, which appear spherical with dimensions of 19 nm to 30 nm or tube shaped with a length of 60 nm to 80 nm by electron microscopy. The structurally best characterised is vlp from *P. furiosus*. Its crystal structure (Akita *et al.*, 2007) reveals spherical particles with a diameter of 30 nm, similar to many phages and viruses. Akita and co-worker suggest that the gene is a remnant of a bacteriophage in the genome of *P. furiosus* and thus the proteins´ function is packing. However these authors do not discuss the sequence similarity to the Linocin_M18 domain. It can only be speculated how the different functions match each other and how the function is correlated to the oligomeric state and the appearance.

The genome of *A. aeolicus* contains one Linocin_M18 domain containing protein, AQ_1760, which is annotated as "hypothetical protein", although the sequence conservation is rather high. In the study of Hick and co-workers in 1998 it was already suggested, that this protein is a bacteriocin homolog (Hicks *et al.*, 1998).

1.8 *Aim of this Work*

The aim of this work is to purify membrane proteins from *A. aeolicus* in a native form, identify them, to choose target proteins of unknown structure and function and to characterise them functionally and structurally. Membrane proteins and protein complexes shall be crystallised and their structure determined by X-ray crystallography. Thereby the understanding about the function and mechanism of membrane proteins and the cell membrane shall be increased. This study should not be limited to proteins of high abundance, that have already been well characterised, but should also include proteins of lower abundance, protein complexes and proteins of unknown function.

Target proteins are selected according to three criteria: their availability, membrane association i.e. they should be integral membrane protein or proteins associated with the membrane, and their biological function, proteins of unknown function shall be favoured. In order to study the proteins functionally and structurally the accessibility of suitable amounts is obligatory. Depending on the method or assay the protein should be available in microgram to milligram range. Proteins which are

present in suitable amounts are purified to homogeneity; proteins which are only present in small amounts are heterologously produced and purified.

All target proteins shall be investigated using bioinformatics in order to obtain indications on the function. The suggested function shall be tested by biochemical assays. The proteins shall be applied to crystallisation attempts in order to obtain three-dimensional crystals suitable for structure determination by X-ray crystallography.

The hypertheromophilic eubacterium *Aquifex aeolicus* was chosen as source organism because it membrane proteins are expected to be rigid and stable, in addition its genomic sequence is known and it contains only a limited number of target proteins that shall facilitate protein purification and identification.

2 Results

2.1 Identification

2.1.1 Membrane Proteins from *Aquifex aeolicus*

The genome of *A. aeolicus* (Genbank No: AE000657) was sequenced by Deckert and co-workers in 1998 (Deckert *et al.*, 1998). Along with the availability of the cell material, this information fulfils the most important requirement for a proteomics project. The *A. aeolicus* genome contains 1522 predicted open reading frames and is therefore small compared to that of *E. coli* (4237). The number of proteins containing one or more transmembrane domains is accordingly low, 315 compared to 1035 in *E. coli*. Thus α-helical membrane proteins therefore make up approximately 21 % of all predicted proteins. The value is towards the lower end of what has been observed for bacteria so far (20 % to 30 %, Wallin *et al.*, 1998). Almost one third (104) of all predicted membrane proteins contain only one TMH as seen in a plot of the number of predicted membrane proteins versus the number of predicted TMH (Figure 2.1-1). However, the high number of proteins with one predicted TMH is most likely caused by difficulties of the standard database programs to distinguish between an N-terminal transmembrane helix and a signal peptide, a short peptide chain of approximately 6-60 basic and hydrophobic amino acids, that directs the transmembrane transport of the protein (Nielsen *et al.*, 1997). Fewer proteins have two or three transmembrane domains (20 and 26, respectively). From four to 20 TMH the number of proteins decays almost steadily. Figure 2.1-1 shows a preference for proteins with 4 and 8 TMH which is in contrast to the preference for 6 and 12 TMH usually observed in unicellular organisms (Wallin *et al.*, 1998). Approximately half of the predicted transmembrane proteins are hypothetical proteins (155).
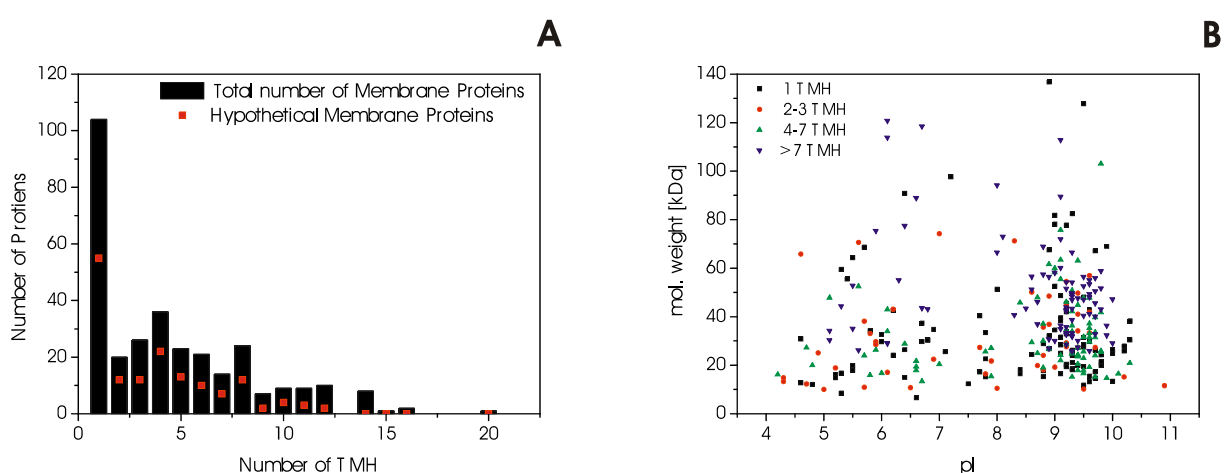

Figure 2.1-1 A: Distribution of the number of transmembrane helices for the total number of predicted membrane proteins and the hypothetical proteins from *A. aeolicus*. B: Distribution of theoretical pI and molecular weight the predicted membrane proteins.

23 proteins are predicted to be outer membrane proteins, 20 are annotated in the TransportDB (Ren *et al.*, 2004), and three additional ones are predicted in the PEDANT database to be outer membrane

proteins. It should be noted that several of the predicted outer membrane proteins are also classified as protein with one TMH because of their N-terminal signal peptide.

The distribution of the isoelectric point (pI) of the proteins was determined in order to find an optimal pH range for purification and to avoid precipitation of proteins caused by their low solubility at the isoelectric point. The distribution is bimodal with one population having a pI around 4.5 and the other around 9 while only very few proteins have a theoretical pI at neutral pH.

2.1.2 Purification and Identification Strategy

Membrane preparation was carried out according to a standard protocol of differential centrifugation to isolate cell membranes. Membrane proteins (120 mg per purification) were extracted from the membrane using n-dodecyl-β-D-maltoside (DDM) as detergent and separated by anion exchange chromatography at pH 7.4 (Peng *et al.*, 2006; Peng *et al.*, 2003). According to the distribution of pI most membrane proteins shall be soluble. The gradient of sodium chloride (NaCl) was designed in segments which allow an optimal resolution already in this first separation step.

The solubilised membrane proteins elute from the anion exchange resin in four major peaks (Figure 2.1-2). The first peak contains the unbound material (fractions 1-14, elution volume 0 to 20 ml). The second peak contains proteins eluting between 50 mM and 150 mM NaCl with a shoulder containing proteins which are eluted when the gradient is kept at 150 mM. Proteins from the third peak elute between 150 mM and 300 mM NaCl (fractions 40 to 60, elution volume of 80 ml to 115 ml) followed by proteins in two small peaks (fractions 61 to 63 and 64 to 65, elution volume 115 to 130 ml). The highest absorbing peak elutes between 300 mM and 1 M NaCl (fractions 68 to 72, elution volume 130 to 140 ml).
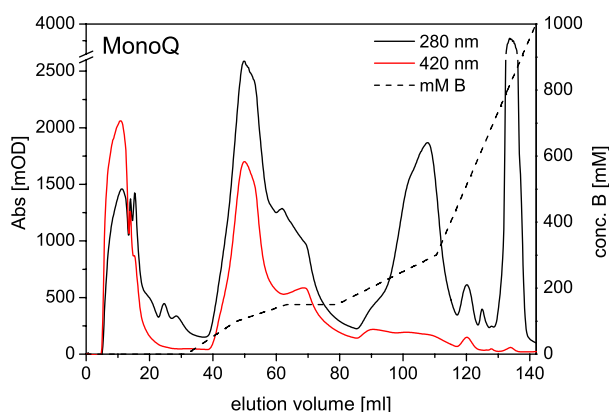

Figure 2.1-2: MonoQ chromatogram of the DDM solubilised membrane proteins from *A. aeolicus*. The MonoQ column was equilibrated with buffer A (20 mM Tris-Cl pH 7.4, 0.05 % (w/v) NaN$_3$ and 0.05 % (w/v) DDM) and proteins eluted with a segmented gradient of buffer B (buffer A and 1 M NaCl).

This work concentrates on identification of proteins from the unbound material as well as proteins and protein complexes which elute at salt concentration higher than 250 mM NaCl. The purification and characterisation of the NADH:ubiquinone oxidoreductase (complex I of the respiratory chain) and the F_1F_O ATP synthase from this preparation has already been described (Peng *et al.*, 2006; Peng *et al.*, 2003); they are part of the group of proteins eluting between 150 mM and 300 mM NaCl.

An aliquot from each fraction was analysed by SDS PAGE gels. In order to assure optimal separation, each fraction was treated at different temperatures in SDS sample buffer. For a majority of fractions, SDS PAGE was sufficient to resolve most of the proteins (Figure 2.1-3 and Figure 2.1-6). In cases where protein bands overlapped in SDS PAGE gels an additional gel filtration step fractionated the anion exchange pool further, resulting in better separation on SDS PAGE gels. This minimised the risk of not detecting protein bands with low intensity by interference with other protein bands. The most abundant protein bands were then analysed by in-gel trypsin digestion followed by mass spectrometry and peptide mass fingerprinting of the obtained peptides. Sample preparation and analysis by mass spectrometry was adopted from Karas and co-workers (Rais *et al.*, 2004).

2.1.3 Membrane Protein Purification and Identification

Figure 2.1-3 shows the zoomed MonoQ chromatogram in the section of the unbound material. In these fractions proteins with an alkaline pI of higher than 7.4 are expected, which make up the majority of membrane proteins (Figure 2.1-1).
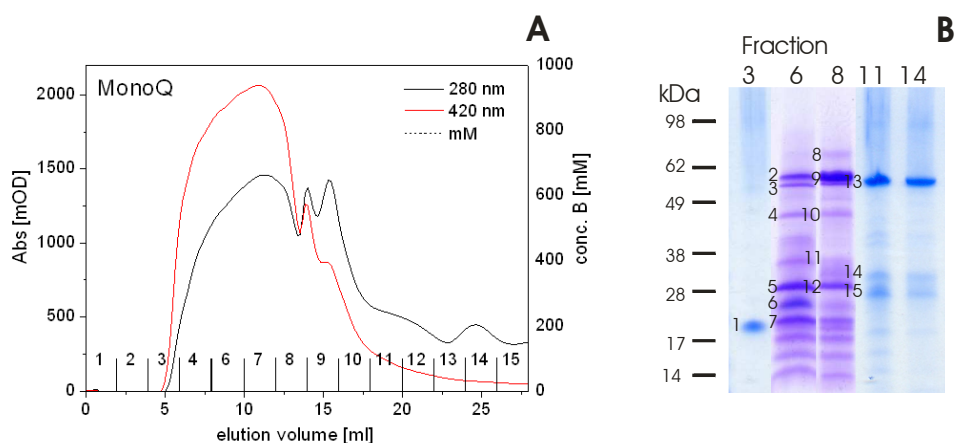

Figure 2.1-3: Cutaway MonoQ chromatogram showing fractions 1 to 15 and SDS PAGE gel (Gel A) of the MonoQ fractions 3, 6, 8, 11 and 14. The small number on the gel refer to the identity of the protein in Table 2.1-1 to Table 2.1-4.

However analysis of the proteins in these fraction revealed that approximately half of the proteins have a predicted an acidic pI of less than 7.4. Three very hydrophobic membrane proteins (GRAVY-score > 0) were identified: the protein export membrane protein SecD (AQ_973; Figure 2.1-3, band 4) containing six TMH, cytochrome *b* (AQ_044; Figure 2.1-3, band 5) containing eight TMH and the Rieske-I iron sulphur protein (AQ_045; Figure 2.1-3, band 6) containing one TMH. A YidC homolog

(AQ_175; Figure 2.1-3, band 10) contains five TMH and has a GRAVY-score of -0.03. In addition, two membrane associated proteins and two protein complexes were identified, cytochrome *c* (AQ_042; Figure 2.1-3, band 5) and cytochrome c_{552} (AQ_792; Figure 2.1-3, band 7) and [NiFe] hydrogenase 1 and 2 with large and small subunit (Figure 2.1-3, band 2 and 3),

In order to identify more proteins with alkaline pI that might be of low concentration and covered by the observed protein bands, the species in the flow through fractions were separated further by cation exchange chromatography. MonoS resin equilibrated at pH 6 yielded the best results among the different column material and pH values tested.
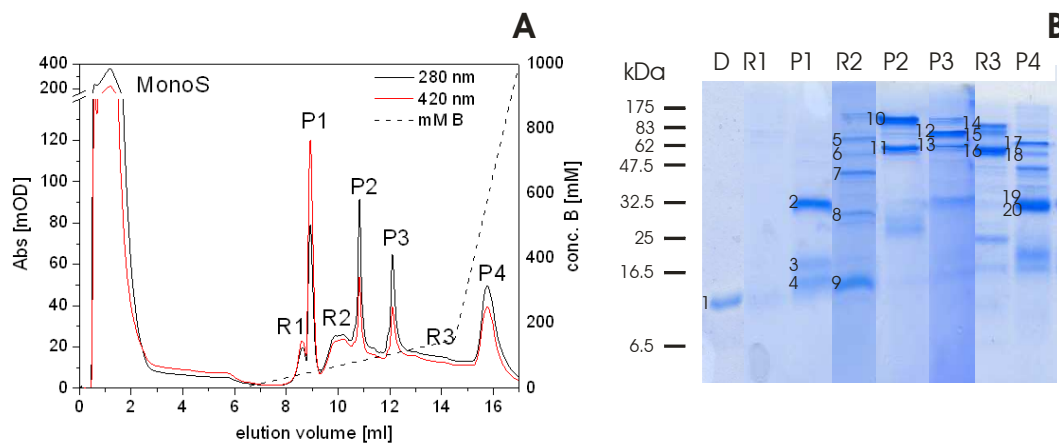

Figure 2.1-4: Chromatogram and Coomassie stained SDS PAGE gel (Gel B) of MonoS purification. The fractions 3 to 14 from the standard MonoQ chromatography were concentrated and applied to a MonoS column equilibrated at pH 6.

The majority of material did not bind to the resin, and even at pH 4.5, only a small fraction of the total protein bound to the column (data not shown). Three additional membrane proteins were identified: the putative proteins AQ_1923 (Figure 2.1-3, band 20) and AQ_2189 (Figure 2.1-3, band 2), both containing one TMH and a NADH dehydrogenase I chain nueM (AQ_135; Figure 2.1-3, 20). In the SDS PAGE gel of the unbound material of the MonoQ resin they are most likely covered by the band of cytochrome *b* (AQ_044; Figure 2.1-3, band 5 and 12). From the MonoS resin cytochrome *b* eluted mainly in peak P1 together with AQ_2189, while AQ_1923 and NADH dehydrogenase eluted in peak P4.

In the following figures (Figure 2.1-5 to Figure 2.1-10) the gel filtration or ion exchange chromatograms and respective SDS PAGE gels of fractions 40 to 67 are shown.
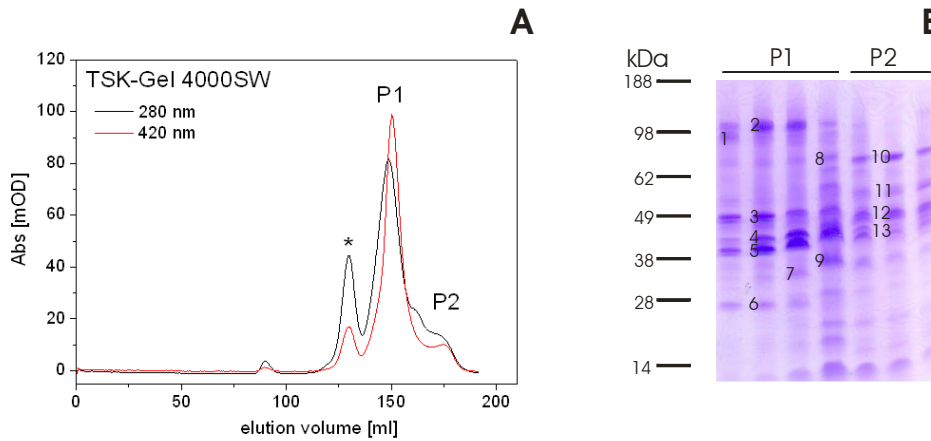
Figure 2.1-5: TSK-GEL 4000SW chromatogram of the MonoQ fractions 40 to 50 and Coomassie stained SDS PAGE gel (Gel C). The gel was performed on the major (P1) and the small (P2) peak with an elution volume of approximately 150 ml and 175 ml, respectively. The peak with a elution volume of 130 ml (*) contains mainly the NADH:ubiquinone oxidoreductase (complex I of the respiratory chain) as described by (Peng *et al.*, 2003).

The MonoQ fractions 40 to 50 contain the NADH:ubiquinone oxidoreductase (complex I of the respiratory chain), which is separated from the remaining proteins by gel filtration chromatography (Figure 2.1-5). The latter (peaks P1 and P2 in Figure 2.1-5) were analysed. Four membrane proteins were identified: cytochrome oxidase *d* subunit I (AQ_1357; Figure 2.1-5, band 9) with nine predicted TMH, outer membrane protein C (AQ_529, Figure 2.1-5, band 8) and the hypothetical proteins AQ_1558 (Figure 2.1-5, band 5) and AQ_1862 (Figure 2.1-5, band 4) with two and one predicted TMH, respectively. Bioinformatic analysis revealed that AQ_1862 and AQ_1259 (Figure 2.1-5, band 3) are porin-like proteins (see below).
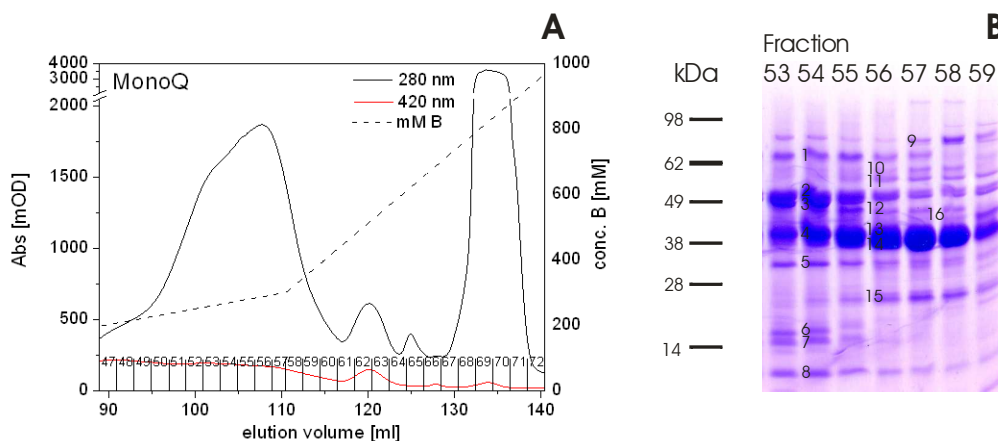
Figure 2.1-6: Cutaway MonoQ chromatogram showing fractions 47 to 72 and Coomassie stained SDS PAGE gel (Gel D) of the MonoQ fractions 53 to 59.

The SDS PAGE gel of fractions 53 to 59 (Figure 2.1-6) shows many intense protein bands including subunits of the F_1F_O ATP synthase. No additional membrane proteins were identified here. Purification of the F_1F_O ATP synthase complex and further separation of the remaining proteins was achieved by gel filtration chromatography (Figure 2.1-7). The F_1F_O ATP synthase appears in a peak

with an elution volume of approximately 130 ml (indicated by *) using a TSK-GEL 4000SW gel filtration column (Figure 2.1-7) as described by Peng and co-workers in 2006. The SDS PAGE gel in Figure 2.1-7 shows fractions from the peak indicated by P with an elution volume of approximately 160 ml. Three additional proteins were identified: the "hypothetical proteins" AQ_1754 (Figure 2.1-7, band 5) AQ_1163, AQ_1192 (Figure 2.1-7, band 9) and (Figure 2.1-7, band 10), none of these contains predicted TMH.
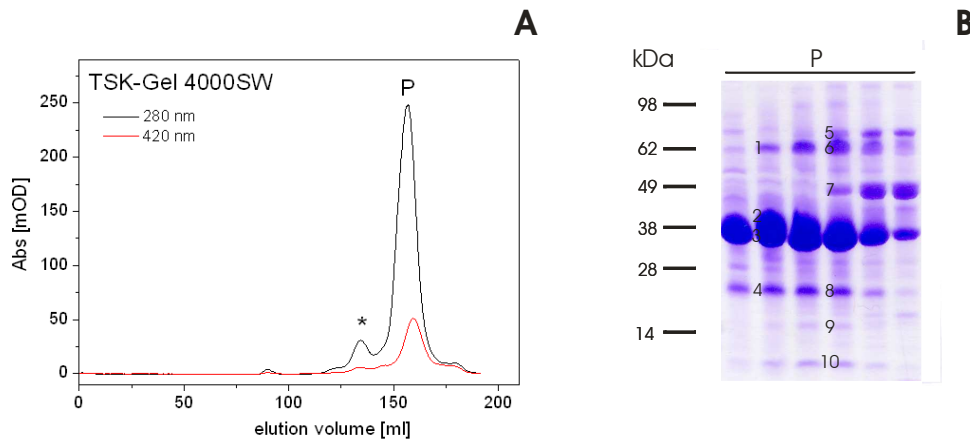
Figure 2.1-7: TSK-GEL 4000SW chromatogram of the MonoQ fractions 52 to 59 and Coomassie stained SDS PAGE gel (Gel E). The gel was prepared from the major peak with an elution volume of approximately 160 ml (P). The peak with an elution volume of 130 ml (*) contains mainly the F_1F_0 ATP synthase as described by Peng and co-workers in 2006.

Although MonoQ fractions 63 to 67 showed a rather low UV absorbance compared to the other peaks, SDS PAGE analysis revealed many protein bands (data not shown), therefore further separation was achieved by gel filtration chromatography.
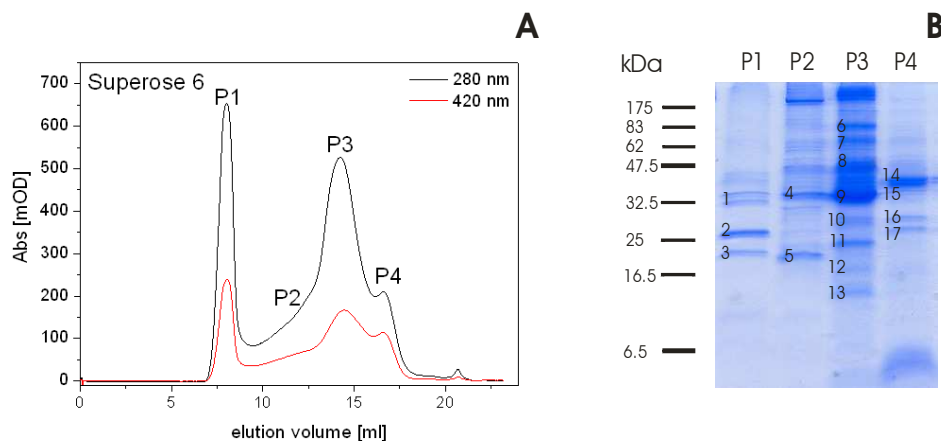

Figure 2.1-8: Superose 6 chromatogram and Coomassie stained SDS PAGE (Gel F) of the MonoQ fractions 61 to 63.

Three additional membrane proteins were identified, all are hypothetical proteins: AQ_389 (Figure 2.1-8, band 13, one TMH), AQ_911 (Figure 2.1-8, band 3, two TMH), AQ_1436 (Figure 2.1-8, band

11, one TMH) and AQ_1439 (Figure 2.1-8, band 7, two TMH). AQ_389 and AQ_911 are very hydrophobic with a GRAVY-score above zero.
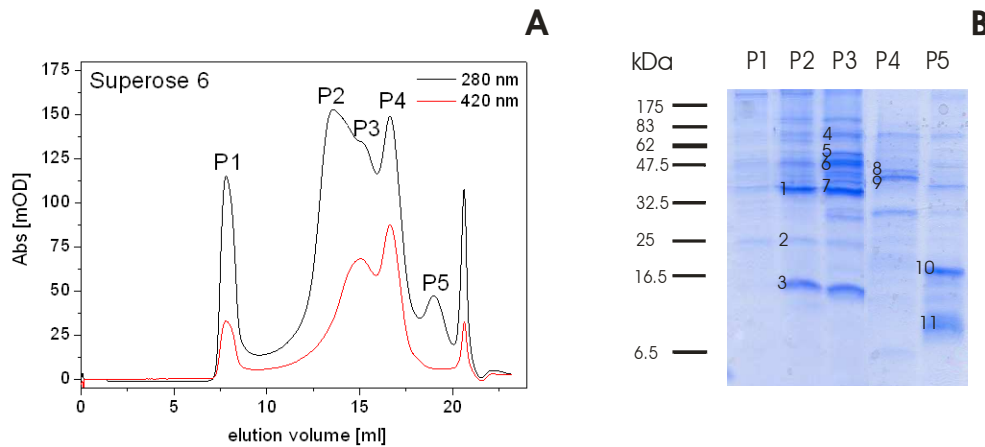

Figure 2.1-9: Superose 6 chromatogram and Coomassie stained SDS PAGE gel (Gel G) of the MonoQ fractions 64 and 65.
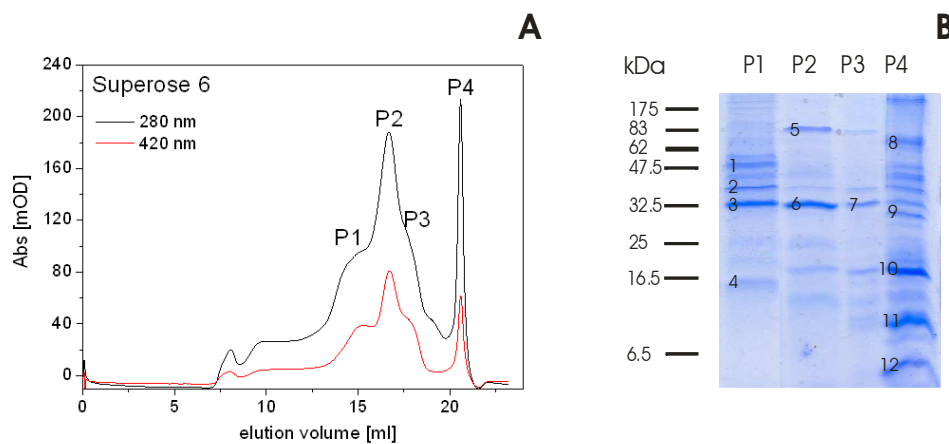
Figure 2.1-10: Superose 6 chromatogram and Coomassie stained SDS PAGE gel (Gel H) of the MonoQ fractions 66 and 67.

In fractions 65 to 67 no additional membrane proteins were identified.

The UV absorption of fractions 68 to 72 was very high; it is the most intense peak. However, the amount of protein detected by bicinchoninic acid (BCA) protein assay was low, only approximately 5 % of what would be expected from the OD_{280nm}. UV spectra revealed a maximum at 260 nm indicating the presence of nucleic acids. Gel filtration analysis of the respective fractions revealed that the material is entirely in the void volume of the column indicative for very large complexes (>700 kDa). The proteins identified from these fractions are exclusively of ribosomal origin (data not shown), suggesting that these fractions contained ribosomes and no substantial amount of any other protein.

The culture medium of *A. aeolicus* was analysed for the presence of secreted proteins. The supernatant from a culture was concentrated approximately hundredfold using an Amicon chamber with a 10 kDa cut-off membrane. However none of the identified proteins is expected to by natively secreted from the cell. Rather they are of periplasmic or outer membrane origin and are detected in the supernatant most likely due to cell lyses or cell disruption during cell harvest.

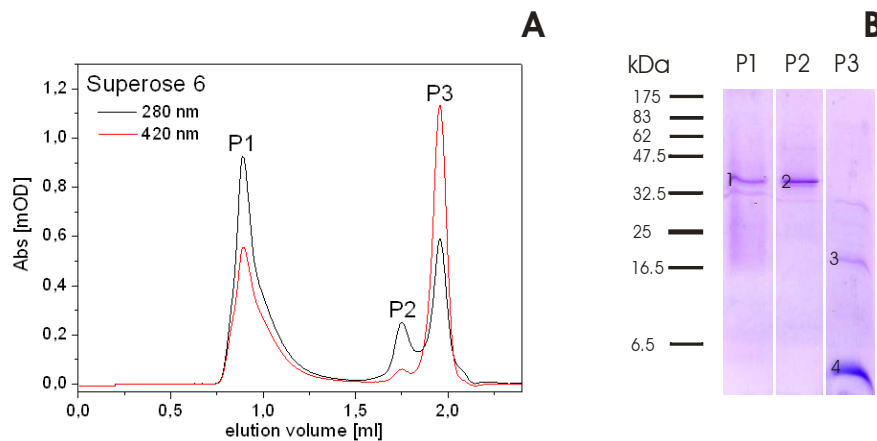**Figure 2.1-11: Superose 6 chromatogram and Coomassie stained SDS PAGE gel (Gel S) of the concentrated *A. aeolicus* culture supernatant.** The three lanes on SDS PAGE gel are the three major peaks in the chromatogram with the elution volume of 0.9 ml (P1), 1.75 ml (P2), and 1.9 ml (P3), respectively.

The initial analysis of the identified proteins from the membrane fraction of *A. aeolicus* revealed the presence of only one annotated outer membrane protein (Outer Membrane Protein C, AQ_529, Figure 2.1-5, band 8) and of two putative porins (AQ_1259 and AQ_1862). In order to investigate the protein content of the outer membrane, separation of outer and inner membranes by means of sucrose gradient centrifugation was attempted, however no separation of the two membranes could be achieved. Because outer membrane proteins are usually associated with the peptidoglycan layer as well, separation of the peptidoglycan layer was attempted. This separation was tried according to the method established by Lakey and co-worker in 1985 (Lakey *et al.*, 1985). From the crude total membrane lipids and inner membrane proteins are removed by solubilisation using SDS, the association of protein with the peptidoglycan is not broken by this treatment and these proteins remain in the pellet fraction. The association to the peptidoglycan layer is subsequently broken by solubilisation in the presence of SDS, β-mercaptoethanol, and high ionic strength. The separated proteins were first solubilised using Triton X-100 followed by SDS. Triton X-100 solubilised only one protein, AQ_1862. Among the remaining proteins are two annotated outer membrane proteins, the outer membrane protein (AQ_1300, Figure 2.1-12, band 2) and the outer membrane protein C (AQ_529, Figure 2.1-12, band 2) and the peptidoglycan associated lipoprotein (AQ_2147 Figure 2.1-12, No 9). The other proteins are ribosomal proteins and the flavocytochrome C sulphide dehydrogenase. Three annotated and one putative outer membrane proteins were identified.
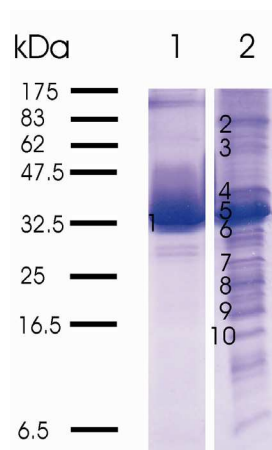
Figure 2.1-12: Coomassie stained SDS PAGE gel of peptidoglycan associated proteins (Gel P). In lane one is the protein fraction soluble in Triton X-100; in lane 2 the pellet alone in SDS.

The following tables give a detailed overview of all identified proteins. In the first column of each table the predicted location (PSORT) of the respective protein is given, in column two and three the protein and the gene name, respectively, columns four five and six shows the theoretical molecular mass and pI (M/pI; ProtParam), the GRAVY-score (ProtParam) and the number of predicted TMH (TMHMM), column number seven indicate the gel (letter) and the respective band (digit) from which the protein was identified.

The identified proteins and subunits of protein complexes group in four different classes: (1) proteins involved in energy transduction and metabolism, (2) proteins of known membrane/extra cellular localisation, of which several are known to be involved in membrane transport, (3) "hypothetical proteins" and (4) other proteins, including ribosomal and cytosolic proteins.

Table 2.1-1: Proteins involved in energy transduction and metabolism

Location	Protein	AQ ...	M/pI	Gravy	TMH	Location
Membrane	cytochrome c	Aq_042	27645.3/9.2	-0.154	2	A5, A6, B3
Membrane	cytochrome b	Aq_044	46928.1/9.30	0.585	8	A5, A12, B2, B8
Membrane	Rieske-I iron sulfur protein	Aq_045	19405.0/9.17	0.097	1	A6, B4
Unknown	NADH dehydrogenase (ubiquinone)	Aq_135	36340.7/9.15	-0.199	0	B20
Membrane	flavocytochrome C sulphide dehydrogenase	Aq_232	48902.0/9.18	-0.276	0	C12, F4, F9, G1, G7, S2, P4
Unknown	Rieske iron-sulfur protein soxF	Aq_234	26875 /9.09	-0.224	0	S3
Cytoplasmic	NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain G	Aq_437	72776.7/5.83	-0.329	0	D10
Unknown	hydrogenase 1small chain	Aq_660	38618.7/6.98	-0.092	0	A14, A15
Periplasmic	hydrogenase 1 large chain	Aq_662	71965.8/6.63	-0.368	0	A2, A3, A8, A13, B5, B12, B14, B15, B16
Cytoplasmic	ATP synthase F1 alpha subunit	Aq_679	55544/5.20	-0.140	0	C11, D2, E7
Unknown	cytochrome c_{552}	Aq_792	18485.5/8.98	-0.836	0	A1, A7, B1
Periplasmic	hydrogenase 2 large chain	Aq_960	64738.5/6.87	-0.422	0	A2, A3, A9, B6,B10, B11, B13, E1, E6
Unknown	hydrogenase (2) small subunit	Aq_965	38187.2/7.91	-0.104	0	A9, B10
Periplasmic	DMSO reductase chain A	Aq_1234	112678/8.36	-0.383	0	C1
Membrane	cytochrome oxidase d subunit I	Aq_1357	52822/5.55	0.504	9	C9
Cytoplasmic	NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain nuoD2	Aq_1314	67892.8/5.95	-0.387	0	D10
Periplasmic	cytochrome c_{552}	Aq_1550	10950/9.15	0.119	0	S4
Membrane	ATP synthase F0 subunit b	Aq_1586	16709/5.34	-0.442	1	D8
Unknown	ATP synthase F0 subunit b	Aq_1587	21241.6/8.79	-0.478	0	D7
Cytoplasmic	ATP synthase F1 delta chain	Aq_1588	20723/9.66	-0.133	0	D6

Cytoplasmic	ATP synthase F1 beta subunit	Aq_2038	53288/5.22	-0.124	0	C11, D3, E7
Cytoplasmic	H+-transporting two-sector ATPase (ATP synthase F1 gamma subunit)	Aq_2041	33531/6.89	-0.378	0	D5
Membrane	sulphide-quinone reductase	Aq_2186	47449.0/6.53	-0.032	0	C12

Table 2.1-2: Proteins involved in membrane transport and proteins of known membrane/extracellular localisation

Location	Protein	AQ_...	M/pI	Gravy	TMH	Location
Membrane	hypothetical protein (YidC homolog)	Aq_175	58225.0/9.04	-0.029	5	A4, A10, B18
Cytoplasmic	cell division protein FtsA	Aq_523	45731.5/6.31	-0.193	0	D12, F15, G8
Outer Mem.	outer membrane protein c	Aq_529	77502/8.70	-0.503	0	C8, P3
Membrane	Erythrocyte Band 7 homolog	Aq_911	28576.8/5.91	0.192	2	F3
Membrane	protein export membrane protein SecD	Aq_973	55459.7/9.10	0.277	6	A4, A10, B7, B18
Unknown	phosphate transport ATP binding protein	Aq_1055	29232/8.56	-0.246	0	C9
Outer Mem.	outer membrane protein	Aq_1300	89999/ 9.07	-0.425	0	P2
Extracellular	flagellin	Aq_1998	55177.7/4.55	-0.089	0	G4
Unknown	peptidoglycan associated lipoprotein	Aq_2147	22856/ 7.63	-0.360	0	P9

Table 2.1-3: Hypothetical proteins

Location	Protein	AQ_...	M/pI	Gravy	TMH	Location
Cytoplasmic	conserved hypothetical protein	Aq_124a	8747.5/ 9.4	0.025	0	H12
Cytoplasmic	hypothetical protein aq_142	Aq_142	16321.1/5.69	-0.079	0	C6, D15, E4, E8
Unknown	hypothetical protein	Aq_271	46506.5/4.99	-0.202	0	G6
Cytoplasmic	hypothetical protein	Aq_324	30383.2/4.94	-0.368	0	F16
Cytoplasmic	hypothetical protein	Aq_331	34599.7/4.99	-0.599	0	F1
Membrane	hypothetical protein	Aq_389	19894.0/5.48	0.341	1	F13
Unknown	conserved hypothetical protein	Aq_814	32603/6.01	-0.058	1	C7
Cytoplasmic	hypothetical protein	Aq_863	118127/5.85	-0.429	0	C2
Unknown	hypothetical protein	Aq_1163	14894.6/5.06	-0.735	0	E10
Cytoplasmic	hypothetical protein	Aq_1192	21288.9/6.60	-0.237	0	E9
Unknown	hypothetical protein	Aq_1259	49390/5.12	-0.314	0	C3
Unknown	hypothetical protein	Aq_1292	20555.5/9.9	-0.235	0	G10, H10
Membrane	hypothetical protein	Aq_1436	30746.0/4.63	-0.221	1	F11
Membrane	hypothetical protein	Aq_1439	65819.0/4.63	-0.130	2	F7
Membrane	hypothetical protein	Aq_1558	43050.5/6.22	-0.119	2	A11, B19, C5, C13
Cytoplasmic	conserved hypothetical protein	Aq_1754	40960/4.79	-0.312	0	D9, E5
Unknown	hypothetical protein	Aq_1760	31846.9/5.29	-0.052	0	F2
Membrane	hypothetical protein	Aq_1862	42671/6.25	-0.185	1	C4, D4, D13, D14, E2, E3, F10, G7, H2, S1, P1, P5
Membrane	hypothetical protein	Aq_1923	34545/9.04	-0.486	1	A12, B20
Membrane	putative protein	Aq_2189	30389.5/6.84	0.022	1	B2

Table 2.1-4: Other proteins

Location	Protein	AQ_...	M/pI	Gravy	TMH	Location
Cytoplasmic	translation elongation factor EF-Tu	Aq_005	44743.5/5.15	-0.202	0	D16, F14
Unknown	ribosomal protein L03	Aq_009	27017/9.89	-0.429	0	P7
Cytoplasmic	ribosomal protein L02	Aq_013	34593/10.42	-0.711	0	P6
Cytoplasmic	30S ribosomal protein S19	Aq_015	25770/ 9.92	-0.779	0	P8
Cytoplasmic	single stranded DNA-binding protein	Aq_064	17132.4/4.85	-0.738	0	G2, G3
Cytoplasmic	glutamine synthetase	Aq_111	53294.7/4.96	-0.341	0	G5
Cytoplasmic	polyribonucleotide nucleotidyltransferase	Aq_221	86435/5.36	-0.247	0	D9, E5
Cytoplasmic	ADP-ribosylglycohydrolase	Aq_534	37049.0/ 4.6	-0.157	0	H7
Cytoplasmic	dihydrolipoamide dehydrogenase	Aq_736	51564.3/5,61	-0.222	0	F6
Unknown	NH(3)-dependent NAD+ synthetase	Aq_959	65272.1/6.25	-0.339	0	F6
Cytoplasmic	phosphoribosyl formylglycinamidine synthase I	Aq_1105	25819.7/5.05	-0.229	0	F5
Cytoplasmic	succinyl-CoA ligase beta subunit	Aq_1306	48349.1/5.11	-0.209	0	F8
Cytoplasmic	ATP-dependent Clp protease proteolytic subunit	Aq_1339	22451.1/5.36	-0.101	0	F12
Cytoplasmic	Ribosomal Protein S01	Aq_1485	60788.3/9.24	-0.472	0	F7
Unknown	superoxide dismutase (Fe/Mn family)	Aq_1499	24425.9/5.68	-0.472	0	F5, F13
Cytoplasmic	pyruvate carboxylase C-terminal domain	Aq_1520	73566/5.47	-0.423	0	C10, D1, E1, E6

Cytoplasmic	ribosomal protein L30	Aq_1644	27017.4/9.89	-0.654	0	G11
Unknown	ribosomal protein S07	Aq_1832	18624/9.87	-0.656	0	P10
Membrane	phosphoribosylformylglycinamidine synthase II	Aq_1836	83213.8/ 5.0	-0.213	0	F6
Unknown	succinyl-CoA ligase alpha subunit	Aq_1888	39828.0/5.98	-0.032	0	F9
Cytoplasmic	D-3-phosphoglycerate dehydrogenase	Aq_1905	59448.1/5.33	-0.065	0	D11
Cytoplasmic	translation elongation factor EF-Tu	Aq_1928	44744.2/5.15	-0.197	0	D16, F14
Cytoplasmic	ribosomal protein S02	Aq_2007	31447.5/6.6	-0.410	0	F17, G9

In total, 75 proteins have been identified, 16 (21 %) have one or more predicted transmembrane helix out of which 9 (12 %) are hypothetical proteins, 24 (32 %) are membrane associated i.e. subunits of membrane protein complexes, that do not contain a transmembrane helix, 23 (31 %) are cytosolic and 12 (16 %) hypothetical proteins without transmembrane domains. Membrane proteins and membrane associated proteins together make up 53 % of all identified proteins.

The majority of proteins identified in the membrane fraction of *A. aeolicus* were complexes involved in bioenergetic processes (Table 2.1-1). This is not surprising since on the one hand those complexes are in general the most abundant membrane proteins and on the other hand several complexes have been described in the literature to be present in the membrane fraction.

Among theses are the two [NiFe] hydrogenases 1 and 2, each with a large and small subunit (AQ_960 and AQ_965, AQ_662 and AQ_660). They are predicted to be located in the periplasm and presumably catalyse electron transfer from molecular hydrogen to quinol (Brugna-Guiral *et al.*, 2003). However, the respective cytochrome *b* membrane anchor subunits were not observed. The identified cytochrome bc_1 complex (Schütz *et al.*, 2003) with subunits cytochrome *c* (AQ_042), cytochrome *b* (AQ_044) and the Rieske-I iron sulphur protein (AQ_045) transfers electrons from quinol to cytochrome *c*. Both cytochrome c_{552} versions annotated in the genome were detected in the proteome. Cytochrome c_{552} (AQ_792), present in the membrane fraction, and cytochrome c_{552} (AQ_1550), found in the culture medium, transfer electrons from the cytochrome bc_1 complex to cytochrome *c* oxidase. A subunit of the alternative *bd*-type oxidase was identified: cytochrome oxidase *d* subunit I (AQ_1357). Frequently occurring in many fractions was the flavocytochrome *c* sulphide dehydrogenase (Aq_232), which presumably reduces sulphide and transfers the electrons to cytochrome *c*. The sulphide-quinone oxidoreductase (AQ_2186) oxidises sulphide and transfers the electrons to quinone (Nübel *et al.*, 2000).

The second group of proteins contains those known to be associated with the membrane and often involved in membrane transport. All of them are of very low abundance: they can be visualised by SDS PAGE analysis but the amounts were too low to attempt purification. SecD (AQ_973) is a subunit of the SecDFyajC pre-protein translocase complex and is classified by the TransportDB as resistance-nodulation-cell division (RND) superfamily protein. YidC (AQ_175) is grouped with the cytochrome-oxidase-biogenesis (Oxa1) family proteins and has been described to form a complex with SecDFyajC. Indeed, in most of the cases both proteins were found in the same fraction. FtsA (AQ_523) (Lai *et al.*, 2004) is a cytosolic protein, but it is known to interact with membrane-bound cell division proteins and plays a role in cell division (Jensen *et al.*, 2005). However, none of the latter

were identified here. The erythrocyte band 7 homolog (AQ_911) is highly conserved among species but its function is unknown. The phosphate transport ATP binding protein (AQ_1055) is the soluble ABC domain of a putative phosphate transporting ABC-transporter according to the TransportDB. However, none of the associated domains (the transmembrane domains AQ_2018 and AQ_2019 and the phosphate-binding periplasmic protein AQ_2016) were identified in this approach. Flagellin (AQ_1998) is often present in membrane proteomes (Bunai *et al.*, 2003; Zhang *et al.*, 2007).

The third group contains hypothetical proteins with and without predicted transmembrane domains. Seven hypothetical proteins were predicted by PSORT to be inner membrane proteins with one or two transmembrane helices. The other proteins are of unknown or cytoplasmic localisation. Among these proteins are very high abundant ones. Protein AQ_1862 is one of the most abundant proteins, approximately 10 mg can be obtained from 120 mg total membrane protein; AQ_1558 and AQ_1760 are less abundant but can be purified in 100 µg amounts per purification. "Hypothetical proteins" in general can be assigned to two groups: conserved hypothetical proteins with homologues in other organisms and species- or genus-specific proteins without homologues (Galperin *et al.*, 2004). The following proteins are conserved among species, the putative function is given in brackets: AQ_124 (function unknown), AQ_331 (putative bacterioferritin), AQ_389 (function unknown), AQ_814 (putative periplasmic serin protease), AQ_863 (glucocorticoid receptor signature), AQ_1192 (putative CO dehydrogenase subunit), AQ_1259 (putative porin), AQ_1292 (function unknown), AQ_1754 (function unknown), AQ_1760 (putative bacteriocin), AQ_1862 (putative porin), AQ_1923 (function unknown) and AQ_2189 (function unknown). In the group of species- or genus-specific proteins are the proteins AQ_142, AQ_271, AQ_1436, AQ_1439, and AQ_1558.

The fourth group contains other proteins mainly of ribosomal and cytosolic origin. These proteins are present in low amounts; however, they comprise the largest group of proteins. Many of them have already been described in other membrane proteomics approaches to be present in the membrane fraction. Those include the translation elongation factor EF-Tu (AQ_005 and AQ_1928) (Zhang *et al.*, 2007), the ATP-dependent Clp protease proteolytic subunit (AQ_1339) (Bunai *et al.*, 2003) and the polyribonucleotide nucleotidyltransferase (Zhang *et al.*, 2007). Some of those are referred to ambiguously as contaminants or membrane associated proteins, including ribosomal proteins and different dehydrogenases (Everberg *et al.*, 2006).

2.1.4 Target Selection

The first and most important criterion of target selection was the availability of protein. However, the number of proteins accessible in high amounts from the native membrane is limited. The most abundant proteins identified in this study were mainly proteins involved in respiration and energy transduction and the hypothetical protein AQ_1862. The latter was chosen as first target protein. The next most abundant group of proteins comprises many "hypothetical proteins" including AQ_1558 and

AQ_1760: these proteins are present at concentrations of approximately 1 % of the highest abundant proteins. Both were chosen as target proteins.

The second criterion was that the target proteins should be associated with the membrane; in this work all targets were identified in the membrane fraction. AQ_1862 shows evidence that it is an outer membrane porin. Prediction programs identified two TMH in the sequence of AQ_1558. Bioinformatic analysis indicated that AQ_1760 does not contain transmembrane domains but homologous proteins have been shown to be associated with the membrane or secreted.

The third criterion was to characterise new membrane proteins. Proteins of unknown function or whose function is not well understood were preferentially selected. AQ_1558, AQ_1760, and AQ_1862 are all hypothetical proteins without annotated function. AQ_1558 has no highly homologous proteins in other organisms; hence it might be a species- or genus-specific protein. Bioinformatic analysis gave no indication for its function. AQ_1760 and AQ_1862 are conserved hypothetical proteins with homologues in other organisms; however, their function has not been investigated so far.

2.2 AQ_1862

2.2.1 Similarity

The protein AQ_1862 (Swiss-Prot # O67714) has a molecular weight of 42.7 kDa (385 amino acids). Sequence analysis revealed the presence of an N-terminal signal peptide with a high probability (0.97) to be cleaved between amino acid 23 and 24 (SignalP).

Initially, sequence similarity revealed significant homology (with an E-value of less than 3×10^{-8}) to FmdC from *Geobacter sulfurreducens* and to several hypothetical proteins (from *Thiomicrospira crunogena*, *Geobacter* (*G.*) *metallireducens,* and *Desulfuromonas* (*D.*) *acetoxidans*). FmdC was first described as a putative porin encoded by a gene within the formamidase (FmdA) operon of *Methylophilus* (*M.*) *methylotrophus*. It is induced by formamide, acetamide and urea (Mills *et al.*, 1997) but so far no experimental evidence for example by means of electrophysiological analysis, has been reported to indicate that FmdC is indeed a porin. FmdC and the FmdC homologous proteins were nevertheless grouped in the short chain amide and urea porin (SAP) family (Transport Classification Database). Several of these initial "hypothetical proteins", including those from *G. metallireducens* and *D. acetoxidans*, were subsequently annotated by the database as "phosphate-selective porin O and P". The phosphate-selective porin O and P family (Pfam # PF07396) or *Pseudomonas* OprP porin (POP) family (Transport Classification Database) comprises proteins homologous to OprO and OprP. AQ_1862 has 23 % identity (and 38 % homology) for 267 aligned residues to the phosphate selective porin O and P from *G. metallireducens* (Swiss-Prot # Q39Y99). The sequence similarity to the *P. aeruginosa* porins is low: 16 % identity (and 28 % homology) for 374 aligned residues to OprO, no homology was detected towards OprP. However, tertiary structure prediction programs show a high similarity to the fold of OprP. A PSI-BLAST search supported the results of the BLAST search, and

proteins from the phosphate-selective porin O and P family and putative FmdC-like outer membrane porins were detected with E values of less than 10^{-30}. A PSI-BLAST search revealed homology to three more *A. aeolicus* proteins which has not been detected by sequence alignment before. All of them annotated as hypothetical proteins: AQ_693 with 19 % identity (and 33 % homology) for 349 aligned residues (E = 6 x 10^{-47}), AQ_1259 with 9 % identity (and 26 % homology) for 351 aligned residues (E = 4 x 10^{-35}) and AQ_2168 with 10 % identity (and 24 % homology) for 321 aligned residues (E = 3 x 10^{-04}).

A secondary structure prediction (β-barrel recognition by PRED-TMBB) was carried out indicating an 18 stranded β-barrel with a reasonable distribution of transmembrane segments, loops and turns: short turns facing the periplasm and long loops facing the outside; loop 6 is exceptionally extended. This kind of distribution is common throughout outer membrane porins: in most porin structures one loop is more extended and folds into the barrel. In addition to similarity to the fold of OprP, tertiary structure predictions (3D-PSSM and HH-PRED) revealed similarity to the osmoporin OmpK36 from *Klebsiella* (*K.*) *pneumoniae*, to the Phosphoporin (PhoE) from *E. coli* and to other porins from the general diffusion pathway. All of these proteins form trimers of 16-stranded β-barrels. The homology to 18-stranded β-barrel porins is lower. Figure 2.2-1 shows the locations of predicted transmembrane segments according to the secondary (light grey) and tertiary (dark grey) structure predictions.

```
AQ_1862      1    ---nkkglllagavfslgvlagsana-----------vtiri---------deeqvlkfgargiidaiikddrgivdgne
FmdC [Gu]    1    minnkktgilsatlllgltlntgtafagpr--------itfgp---------edqgalqidykgqlqntvr-dtgsgennd
FmdC [Mm]    1    mmmlkkcalhtavanacglgaqlanag----------atisf---------gedkyisagf-gfitsynsvedaaangsd
OprO/P [Gm]  1    minnrkistisgvalgtalllagtafagpr--------itfgp---------edqgalqidykgqfqntvrdngsgadgns
OprO [Pa]    1    n--irkhsl--gfvasalalavsaqafagtvttdgadiviktkgglevattdkefsfklggr--lqadysrfdgfytkng

AQ_1862      58   dksdivfthr-------dariyfkyklnkiisygfqas-----------lfapgkag-dievfd--sfinlafa------
FmdC [Gu]    63   d-tttnfnfr-------rnrlafngkygenlslyvqteftedqnvttlgvastnqgt-efqlld--avnrfkln------
FmdC [Mm]    61   rsnd--ftld-------sarlylsgsfnkyikgnlnte-----------ksgggsansnfevid--anvqfqlt------
OprO/P [Gm]  64   tttn--fnfr-------rnrlafngkygdnvslyvqte-----------ftedpnvg-tlgvsd--qsantefqlldavv
OprO [Pa]    75   ntadaayfrrafielggtaykdwkyqinfdlshntgss-------------------dngyfdeasvtytgfn------

AQ_1862      111  -----pefniiagsykvpferhsglqsywt--------ylfptwl-----gnsa-----iagykh---------aftnpv
FmdC [Gu]    126  -----dsfhvnagkfkynlsr----enlea--------cenpltl-----drsl-----fi--rt-------aytt--
FmdC [Mm]    113  -----pevaiwagrflspsdrannagpyys--------nrgggywa----nias-----rygwnggvigrddgvavvtsf
OprO/P [Gm]  121  rfklndafrvnvgkfkygfsr----enlea--------cenpltl-----drsl-----fi---------------
OprO [Pa]    129  -----p-vnlkfgrfdpdfglekatsskwvtapernaayeladwinthqdgngaqvnstladnnay---------lsagvs

AQ_1862      159  andh--------kpfrsgsrsagltawgniadgnfkyyvgvfd-----asdegatsnaktaysirvqftptnlgykaekg
FmdC [Gu]    166  -----------------trdtgvavwgnlfndifqyrvdanregrkavsgdtapssnfr--ysfrghvt--lldpendyg
FmdC [Mm]    170  nedh--------vavsfgafe--------gdnifr-fsgvga-----qseangtkd-klnyagrvqid----fwdaepg
OprO/P [Gm]  160  ---r--------apfvs-trdngvgiwgnvledkiqyradvnegrk--agdedgnghkspdssfrytfrghvslldpesd
OprO [Pa]    194  akdadds dgdsvkqfnfrgvfapnheagnv------lhvgv---------nyayrdlddtafdsrirprlgnrgiatsgg

AQ_1862      226  yvlkntyvgkkdvltiglsytsqp---fgega-----k------dqdswgvdlnrwe---------------qkf---gti
FmdC [Gu]    224  y--kgtylgkkkvltiggayqfepeiaygdtvtrtdkk------dyqawtidgffeypl-------------egi---gtv
FmdC [Mm]    222  yygtgnyfgakdilaigiagrnrkk---dgvvs-----tipgvvgdyksysvdfll e-----------kkdvgpgtf
OprO/P [Gm]  226  ygykgtynrgkkqvltfgaayqfepdvvygnpaaksdkk------dyqawtvdgffeypv-----------eav---gtf
OprO [Pa]    259  ----ndagdngnratfg-gvsnsp---ags-y-----k------ddsvwglegawanrgpfsaqaeylarklkad---dna

AQ_1862      274  vp-------diqi--gyvkhtn-wagtdg--d--------drtgwlvq-g---qllydqkv--flgkpalaary--vqtef
FmdC [Gu]    281  tasaayedvdldd-aykganp-dagtiglng--------qkngwyvk-g---gynlpnlplqffg------ry--ekwrf
FmdC [Mm]    279  sa-------eaay-yyydtddvflgeqg--k--------aysagl-------gylfndpv-gwgkinpivry----qkf
OprO/P [Gm]  285  tlsaayadydldk-syntain--altep--dkdaiglngekngwynkag---ynlpnlplq-ffg------ry--ekwsf
OprO [Pa]    316  yk-------dikakgyyaqla-ytlt-g--e-------srq-ykle-gakfdsvkpenk--eigawevfyrydnikved

AQ_1862      327  d--------rpg-----------s--dpklksfgvainyyikgptnrialwvenvkldn---------vnlteeedsytdv
FmdC [Gu]    339  a--------elnni---------s--dqkvdwygggganyyfrgqnlkltfeaahtafdkdgt------vngvtskdfttfv
FmdC [Mm]    328  d--------adgltt aasvntrt--sadtkrfeiganyvi-apynsv-itatygktdt--------p--aassdnffrl
OprO/P [Gm]  348  a--------nln----------niydqdvtwygggvnyyiwgqnlkltaeisrtdfdkegvasgiqgtnlktedfttfvt
OprO [Pa]    373  dnvvadtatrev----------g--dtkakahnlgvnwyvndavkisaayvk-aktdk---------itnnngdddgdgf

AQ_1862      378  g-lsvfynf
FmdC [Gu]    395  tqlqlif--
FmdC [Mm]    385  a-lqnqf--
OprO/P [Gm]  410  q-lqllf--
OprO [Pa]    431  v-trlqyvf
```

Figure 2.2-1 Sequence alignments with AQ_1862 from *A. aeolicus* (Swiss-Prot # O67714) using the program Clone Manager 7 with the algorithm BLOSUM62. FmdC [Gu] is the putative porin FmdC from *Geobacter uraniumreducens* (Swiss-Prot # Q2DQD8), FmdC [Mm] is putative porin FmdC from *M. methylotrophus* (Swiss-Prot # O05123), OprO/P [Gm] is the phosphate-selective porin O and P from *Geobacter metallireducens* Q39Y99 and OprO [Pa] the Pyrophosphate-specific outer membrane porin from *P. aeruginosa* (Swiss-Prot # P32977). In the upper row the model for the transmembrane arrangement of AQ_1862 is shown. Light gray: 18-stranded model obtained using the TMBB-PRED algorithm (Bagos *et al.*, 2004). Dark gray: 16 stranded model using OmpK36 from *K. pneumoniae* as template with the HH-PRED algorithm (Söding *et al.*, 2005).

Sequence and fold similarity provide strong evidence, that the protein AQ_1862 is a porin.

2.2.2 Purification

2.2.2.1 Native

AQ_1862 is probably the most abundant protein in the solubilised membrane fraction; it comprises the most prominent band in SDS PAGE analysis in fractions 50 to 60 (Figure 2.1-6 B). When proteins from those fractions were applied to crystallisation attempts, crystals were readily obtained. They had a cuvette-like shape and diffracted X-rays to a maximal resolution of 4 Å. In order to improve those crystals, the purity and homogeneity of the sample had to be improved (see chapter 2.2.4.1.1).

Further purification was achieved by gel filtration chromatography. The TSK-GEL 4000SW column separated two peaks (Figure 2.1-6 A): the first (indicated by *) contained mainly the F_1F_O ATP synthase (Peng *et al.*, 2006) and the second (indicated by 1) AQ_1862. In order to obtain the protein in a pure and homogenous form suitable for crystallisation attempts and functional analysis, four different purification methods were tried: hydroxyapatite, different anion exchange resins, different hydrophobic interaction resins, and different reactive dye agarose.

The mechanism of hydroxyapatite (a form of calcium apatite, $Ca_5(PO_4)_3(OH)$) chromatography is a mixed mode of ion exchange and affinity; proteins are usually eluted by a buffer with increasing phosphate concentrations. The protein sample containing AQ_1862 was loaded onto a hydroxyapatite column equilibrated with 10 mM Na-phosphate buffered at pH 7.4; almost all protein bound to the material and very little protein was in the flow through (Figure 2.2-2). A step gradient of phosphate buffer at different concentrations eluted the proteins. Most impurities were removed by washing with 300 mM Na-phosphate, and protein AQ_1862 eluted at 500 mM (Figure 2.2-2). A fraction of the protein could only be eluted using 500 mM Na-phosphate containing 2 M NaCl. Both fractions contain the protein in a very pure form. Since phosphate buffer and high ionic strength can impede crystallisation attempts, the buffer was exchanged to the standard gel filtration buffer (20 mM Tris-Cl pH 7.4, 150 mM NaCl) by concentrating in Centripreps (10 kDa cut-off) and diluting in the desired buffer. Concentrating the protein from a buffer containing 500 mM Na-phosphate with addition of 2 M NaCl was difficult because of salt precipitation. In addition, residual phosphate buffer caused massive appearance of salt crystals in the crystallisation trials, especially when calcium was present. To ensure a complete buffer exchange an additional desalting (PD-10 Desalting Column) step was necessary. The yield after the hydroxyapatite chromatography was only 50 %. Because of the loss of protein and the difficulties in handling the protein in high ionic strength buffers this purification method was abandoned.
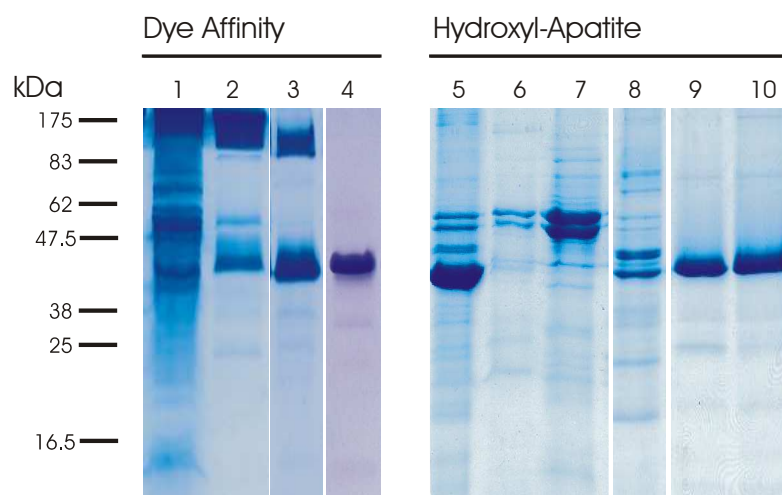
Figure 2.2-2: Coomassie stained SDS PAGE gel of AQ_1862 purified using reactive red sepharose and hydroxyl apatite. 1: DDM solubilised membrane, 2: MonoQ fractions 50 to 60, 3: peak after gel filtration, 4: elution from reactive red, 5 MonoQ fractions 50 to 60, 6: flow through, 7 to 9: elution with 100 mM, 300 mM, 500 mM Na-phosphate, 10 elution with 500 mM Na-phosphate, plus 2 M NaCl.

AQ_1862 containing fractions were applied to weak and strong anion exchange resins including MonoQ, MiniQ, Q-Sepharose, and DEAE-Sepharose at different pH values. MiniQ chromatography at pH 7.5 with a more shallow sodium chloride gradient yielded purest AQ_1862. The protein eluted in two peaks, one very pure, and one containing several impurities. The capacity of MiniQ is limited; therefore this approach was only used in special cases (see below).

Hydrophobic interaction chromatography was ineffective in purifying AQ_1862, because the protein did not elute from the resin under all investigated conditions.

Reactive dyes are textile dyes which have been found to bind proteins, especially those with affinity to various nucleotides (Kopperschläger *et al.*, 1968). Therefore the dye Reactive Red 120 was initially used to separate F_1F_O ATP synthase from its impurities (see (Peng *et al.*, 2006). The F_1F_O ATP synthase was located in the non-bound fraction, despite its putative nucleotide-binding site. Among the proteins binding to the resin is AQ_1862, it elutes with 1.5 M NaCl in almost pure form.

AQ_1862 containing fractions were applied to different dye affinity materials (Reactive Brown 10, Reactive Yellow 3, Reactive Green 5, Cibacron Blue 3GA, Reactive Yellow 86, Reactive Green 19, Reactive Blue 72, and Reactive Red 120). The protein AQ_1862 binds to all different materials and elutes in almost pure form. Almost no difference was observed between the dye conjugates, AQ_1862 eluted in pure form in the presence of 1.5 M NaCl. The only exception is given by Cibacron blue 3GA, from which the protein can only be eluted with 2.5 M NaCl. The yield after the dye affinity chromatography was more than 80 %. Dye affinity materials offer a very fast and simple way to obtain pure and homogeneous protein (Figure 2.2-2).

A variety of detergents were tested for purification of AQ_1862, including n-decyl-β-D-maltoside (DM), FOS-CHOLINE-12 (Fos-12), n-nonyl-β-D-glucopyranoside (NG), n-octyl-β-D-glucopyranoside (OG), 1-S-n-octyl-β-D-thioglucopyranoside (OTG), n-lauryl-N,N-dimethylamine-N-

oxide (LDAO), octyl-2-hydroxy ethylsulfoxide (C_8HESO), octyltetraoxyethylene (C_8E_4) and octylpolyoxyethylene (C_8E_n). The detergent was exchanged during dye affinity chromatography. The protein was loaded on the DDM equilibrated column, washed with 10 to 15 column volumes of new detergent buffer, and eluted in the new detergent. Only in exceptional cases like in the presence of C_8HESO, C_8E_5, and C_8E_n, which are not soluble in the buffer containing 1.5 M NaCl, the detergent exchange was carried out on the MiniQ column.

This procedure yielded approximately 10 to 12 mg of protein per purification.

2.2.2.2 Heterologous Production

Native AQ_1862 was purified in milligram amounts from the membrane of *A. aeolicus*. However, having access to the genetic system has several advantages. Non-native amino acids like seleno-methionine (Se-Met) for structure determination by X-ray crystallography can be introduced. After determination of the structure, variants for functional analysis can be obtained by site directed mutagenesis. Therefore it was aimed to establish heterologous production of AQ_1862 in *E. coli*.

Heterologous production of outer membrane proteins can be achieved by two different approaches: (1) by functional production of the protein in the outer membrane and solubilisation using detergent or (2) by non-functional production into insoluble inclusion bodies, solubilisation using urea or guanidinium hydrochloride and refolding in the presence of detergent. In the first approach, protein production in *E. coli* and isolation from the outer membrane fraction was attempted. The gene coding for AQ_1862 was cloned together with its native signal sequence into the expression vectors pTTQ18 and pBAD in the modified version A and C (Surade *et al.*, 2006) using the primer A1862.For and A1862.Rev (Table 4.1-5). In short, proteins produced in the A version contain a TEV protease cleavage site at the C-terminus followed by a deca-histidine tag to facilitate purification; in the C version proteins contain a deca-histidine tag (his-tag) at the N-terminus, followed by a TEV protease cleavage site, and a C-terminal strep-tag II to allow a second affinity purification step. The pTTQ18 constructs were tested for production in *E. coli* NM554 and BL21(DE3) cells, the pBAD constructs in *E. coli* TOP10 and in BL21(DE3).

In almost all tested plasmid/host combinations either no production or premature lysis of the host cell was observed. Only the pBAD-A construct in *E. coli* Top10 cells resulted in protein production as shown by Western blot analysis and cell lysis was moderate when the induced cells were kept at 20 °C (data not shown). The protein was located in the membrane fraction and was solubilised using DDM. After solubilisation an additional band at approximately 30 kDa appeared, while the signal of the full length protein was weaker. Most likely the protein is degraded during solubilisation. Several solubilisation conditions and additives like protease inhibitors were tested. No conditions could be found under which the protein was stable.

Since purification of AQ_1862 from the membrane fraction was not possible, the second approach of production was attempted: non-soluble inclusion bodies. To direct the protein into inclusion bodies,

the DNA coding for the signal peptide (amino acid 1 to 23) was removed and the protein encoding gene expressed using the plasmid pET-24a(+) (primer AQ1862pET.For and AQ1862pET.Rev, Table 4.1-5). The methionine start codon was placed in front of the first amino acid (valine 24 which was identified as N-terminal amino acid by N-terminal sequencing) of the mature protein. The stop codon at the C-terminus was replaced by a short linker peptide followed by a hexa-histidine tag. Protein production and inclusion body preparation was done as described in Materials and Methods. Inclusion bodies were solubilised using 6 M guanidinum hydrochloride or 8 M urea. The monodispersity of the sample was investigated by gel filtration chromatography. Urea in the pH range from pH 5 to pH 9 solubilised only minor amounts of protein. 6 M guanidinum hydrochloride solubilised the protein quantitatively in mainly non-aggregated form. For refolding trials, the rapid dilution technique as well as dialysis were applied. The solubilised protein was diluted 1:1, 1:5 or 1:10 in a buffer containing a high concentration of detergent (> 100 x CMC) or dialysed against the same buffer containing 3 x CMC. Different types of detergents (DDM, OG, LDAO, C_8E_n, C_8HESO, and Triton X-100), different salt concentrations, additives, and different pH values were tried. Refolding was followed by the elution volume in gel filtration and by circular dichroism (CD) spectroscopy, a method to determine the secondary structure content of a molecule. The measurement allows the fraction of a protein having α-helical, β-sheet or random coil conformation to be estimated. Since outer membrane porins have almost exclusively in β-sheet character, an increase in β-sheet content of the protein can be followed by the increasing CD signal. However none of the tested conditions yielded stable refolded protein with appropriate elution volume in gel filtration and a CD signal similar to the native protein.

For a porin from *Rhodopseudomonas blastica* (Kreusch *et al.*, 1994; Schmid *et al.*, 1996) it was shown that it is crucial to keep the exact number of amino acids in order to yield stably refolded protein. In this porin the NH_2- at the N-terminus from one subunit forms a structurally important salt bridge with the C-terminal COOH from phenylalanine of another subunit. Therefore, in a second approach the gene was constructed in a way to retain the exact number of amino acids. The signal peptide was removed and valine 24 was replaced by the methionine for starting gene translation. The native stop codon was kept; phenylalanine 365 is the C-terminal amino acid (primer AQ1862pET_MTIR.For and AQ1862pET_stop.Rev, Table 4.1-5).

Protein production and inclusion body preparation was performed as described in Materials and Methods. Again, 6 M guanidinum hydrochloride solubilised the protein quantitatively and in mainly non-aggregated form. Using 8 M urea or guanidinum hydrochloride at concentrations below 6 M yielded aggregated, incompletely solubilised protein (Figure 2.2-3). Exchanging the solvent to 8 M urea after the solubilisation with guanidinum hydrochloride to facilitate the subsequent anion exchange chromatography was not successful, the protein precipitated.
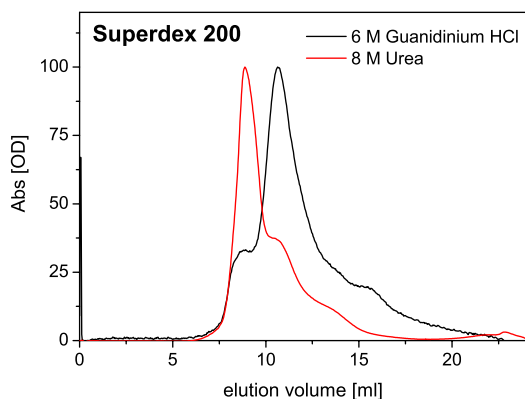
Figure 2.2-3: Gel filtration pattern of an inclusion body protein sample dissolved in 8 M urea or 6 M guanidinium hydrochloride. 8 M urea or 6 M guanidinium hydrochloride were used as corresponding chromatography buffer on a Superdex 200 HR 10/30 column. Using urea the majority of the protein is in the void volume of the column (8 ml) while with guanidinium hydrochloride the protein is mainly non-aggregated form and not aggregated.

Refolding was induced and followed as described above. Again different types and concentrations of detergents were tested. LDAO was the only detergent able to refold the protein into a stable and monodisperse form. Best results were obtained by diluting the solubilised inclusion bodies drop by drop 1:10 in 5 % (w/v) LDAO at neutral pH. Gel filtration could not be used to separate the folded from the misfolded protein because the elution volumes of the two forms are very similar. Therefore folded protein was purified by anion exchange chromatography (Figure 2.2-4 A). Protein from the bound fraction was concentrated and applied to gel filtration chromatography using a Superose 6 column (Figure 2.2-4 B). The protein eluted in a major peak with an elution volume (14 ml) identical to the native protein. A small shoulder with higher elution volume (16.5 ml) contains traces of protein (see band 5 in Figure 2.2-4 D) but has a CD signal which does not suggest a β-sheet fold. Both native and refolded protein show only β-sheet signals with a minimum at 218.5 nm (Figure 2.2-4 C). The CD signal per residue of the refolded protein was identical in some preparations, while other preparation showed up to 20 % less CD signal than the signal of the native protein, meaning that a fraction of protein is still unfolded or the protein it self is only partially folded compared to the native protein. That the latter is not the case will be seen in the coming chapters, however, the CD signal could not be increased by further chromatographic steps. The anion exchange chromatography step was used to exchange the detergent from LDAO to the desired detergent.
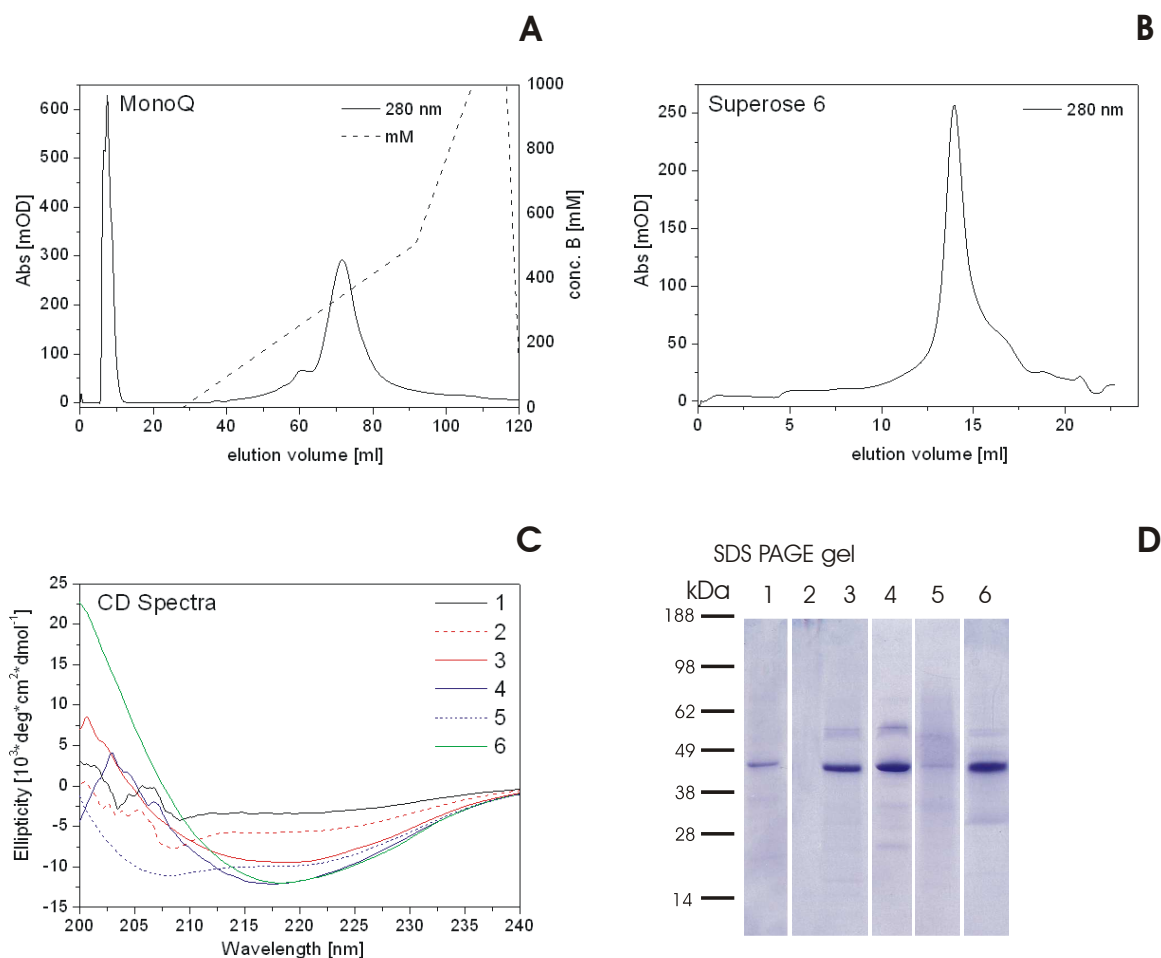
Figure 2.2-4: A: Anion exchange chromatogram (MonoQ HR10/10) of the refolded protein AQ_1862 and B: gel filtration pattern (Superose 6 HR 10/30) of the protein AQ_1862 after elution from the anion exchange. C and D: CD spectra and Coomassie stained SDS PAGE gel of the respective purification steps. In each lane 2 µg of protein is loaded. 1: sample from inclusion body solubilised in 6 M guanidinium hydrochloride, 2 and 3 is the flow through and elution from the MonoQ (A), respectively, 4 and 5 is the main peak and the shoulder from the Superose 6 (B), respectively, and 6 is a sample of the native protein.

2.2.3 Functional Characterisation

2.2.3.1 Biochemical Characterisation

2.2.3.1.1 The N-terminal Signal Peptide

Like most outer membrane proteins, AQ_1862 uses a signal peptide to guide the protein into the outer membrane. The peptide is cleaved during membrane incorporation. To show that the peptide is indeed not present in the native protein and to determine whether cleavage site after amino acid 23 has occurred as depicted above, mass spectrometric analysis was performed. In the mass spectrum of the trypsin digested protein, data for the N-terminal domain is missing. However, hydrophobic amino acid stretches are usually not detectable in mass spectrometric investigations. A peak at 1442.8 m/z appears in the mass spectrum of the trypsin digested protein and most likely belongs to the peptide VTIRIDEEQVLK. This peptide can only be present if the protein sequence begins with amino acid 24. N-terminal sequencing confirmed that the N-terminal signal peptide is indeed cleaved and the protein sequence starts with VTIRI (data not shown).
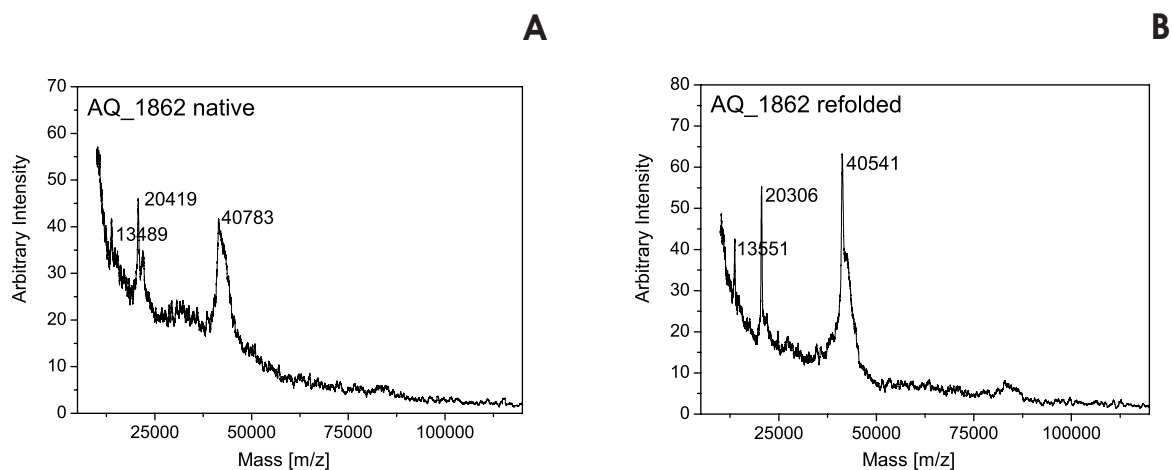
Figure 2.2-5: Mass spectra of the native and refolded AQ_1862. The native protein has an apparent mass of 40.7 kDa, the refolded protein of 40.5 kDa, the peaks at 20.4 kDa, 13.4 kDa, 20.3 kDa and 13.5 kDa are the double and triple charged protein, respectively.

In the mass spectrum of the full length native protein (Figure 2.2-5 A) the major peak appears at 40.7 kDa, which best fits with the mass of the amino-terminally truncated protein (40.526 kDa). The peaks at 20.4 kDa and 13.4 kDa correspond to the doubly and triply charged protein, respectively. From Figure 2.2-4 it is already obvious that the refolded protein resembles the native one in the elution volume in gel filtration chromatography i.e. the oligomeric state and in the size in SDS PAGE analysis. In the mass spectrum of the full length protein (Figure 2.2-5 B) the size of 40.5 kDa for the refolded protein was confirmed.

2.2.3.1.2   Electrophoretic Behaviour

Trimeric porins often appear at the size of the trimer in SDS PAGE analysis; all porin proteins with similar tertiary structure detected by fold recognition (see chapter 2.2.1) are trimeric porins. Figure 2.2-6 shows an SDS PAGE gel of the purified sample incubated with SDS sample buffer with 2.5 % SDS or 0.2 % SDS without addition of β-mercaptoethanol at different temperatures. AQ_1862 has an apparent mass of approximately 40 kDa, the size of the protein monomer. No indication of oligomer formation is detected even if the SDS concentration is decreased to 0.2 %. At temperatures below 80 °C the protein runs as a double band of approximately 38 kDa and 42 kDa. Both bands were analysed by mass spectrometry and contain exclusively AQ_1862. After incubation at 100 °C the upper band disappears and the protein appears as a single band. Hence the upper band is most likely the non-denatured form of the protein and the lower the denatured form. The protein either exists as a monomeric porin in nature or it forms oligomers which are very sensitive to SDS treatment.

The stability of the protein against limited proteolysis was investigated. Conclusions can be drawn whether the entire protein forms one barrel or whether is it is a two domain protein like OmpA. AQ_1862 was found to be stable towards trypsin treatment. Even after 24 h no degradation product was detected and the intensity of the band was unchanged.
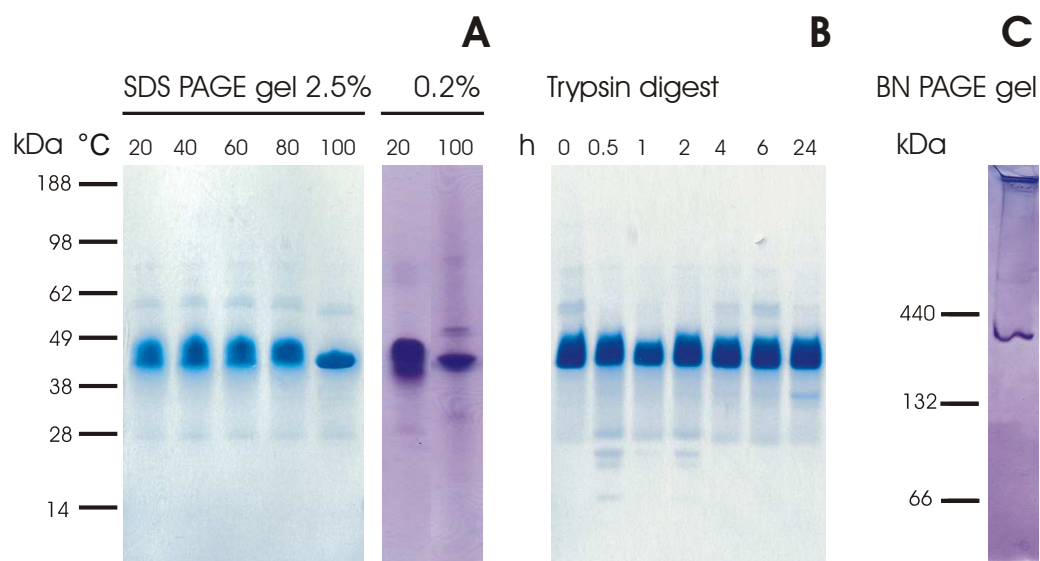
Figure 2.2-6: A and B: Coomassie stained SDS PAGE gel and C: BN PAGE gel of the protein AQ_1862.
A: The purified protein was incubated at different temperatures and different SDS concentrations without addition of β-mercaptoethanol. B: Limited proteolysis of the protein AQ_1862 by trypsin (protein incubated with trypsin (20 to 1, w/w) in 50 mM Tris-Cl pH 8.0, 1 mM $CaCl_2$, 0.03 % (w/v) DDM).

In BN PAGE analysis the protein has an apparent molecular mass of more than 132 kDa. This data suggest that the native form of the protein is not a monomer. In its native form the protein might not bind Coomassie in the same way as BSA or α-helical membrane protein do, so it is difficult to judge what the native size of the protein is.

2.2.3.1.3   Size Estimation by Gel Filtration Chromatography

In order to investigate the oligomeric state of AQ_1862 further, analytical gel filtration chromatography using a Superose 12 column calibrated with soluble standard proteins was performed. Figure 2.2-7 shows a gel filtration chromatogram of the native and the refolded protein. Both have a similar elution volume: 1.23 ml and 1.19 ml, respectively. A single peak in both samples suggests monodispersity of the samples. However, the refolded protein shows a small peak at 0.9 ml, suggesting protein aggregation. In some preparations a shoulder towards higher elution volume which contains unfolded protein (see Figure 2.2-7) is also observed.
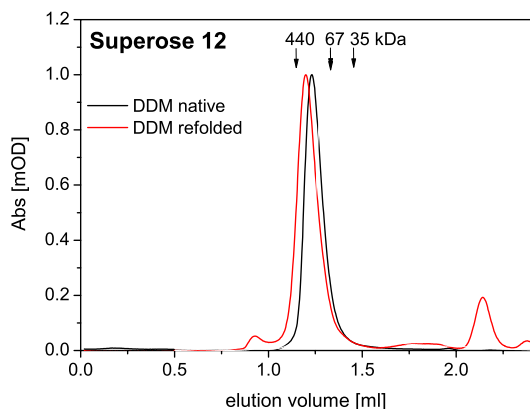
Figure 2.2-7: Superose 12 gel filtration chromatogram of a native and a refolded sample of AQ_1862. The elution volumes of the soluble standard proteins are indicated with arrows and the molecular weights of the respective proteins.

Calibration of the column with the indicated soluble standard proteins (440 kDa: ferritin, 67 kDa: BSA, 35 kDa: β-Lactoglobulin) was applied to determine the apparent molecular weight of the protein. The elution volumes correspond to molecular weights of approximately 190 kDa and 250 kDa, respectively. The molecular weight of the protein monomer as calculated from the amino acid sequence and determined from mass spectrometry (Figure 2.2-5) is 40.5 kDa. Assuming a micelle containing 50 to 100 detergent monomers (le Maire *et al.*, 2000), the extra molecular weight would make up 25 to 50 kDa per monomer. The data agree with a trimeric form of the protein (3 x 40.5 kDa = 121.5 kDa for the protein) and a detergent micelle of 68 kDa and 131 kDa for the native and the refolded protein, respectively.

2.2.3.1.4   Cross-linking Experiments

Data obtained from gel filtration chromatography and SDS PAGE analyses are contradictory, the former suggest an oligomer and the latter a monomer. Therefore cross-linking experiments with a bifunctional reagent were applied. Glutaraldehyde was chosen because it was already successfully for several other outer membrane porins. The protein was exposed to glutaraldeyde and samples were taken at various time points and analysed by SDS PAGE (Figure 2.2-8). During exposure to glutaraldehyde at least two additional bands appear. While the signal belonging to the monomeric form decreased in intensity, the signal indicated by II increased but became weaker after 6 hours. The band indicated by III appeared after two hours, gaining intensity, and broadening with time (IIIa). The three bands can be assigned as monomeric form (I in Figure 2.2-8 A), two (II) and three (III) polypeptide chains cross-linked, respectively. The band indicated by IIIa is most likely an aggregate of oligomers which appears due to an excess of glutaraldehyde or to the prolonged treatment.

The cross-linking pattern of the refolded protein is similar (Figure 2.2-8 B). However, the subpopulation of cross-linked protein was lower (band III) and no broadening of the band of the putative trimer was detected. More material is aggregated and was not able to enter the gel (indicated

by *). The apparent molecular weight of the putative oligomeric forms of the refolded protein are shifted compared to the native protein for both dimer and trimer.

Obviously both native and refolded protein forms a trimer as shown by gel filtration chromatography and cross-linking of the protein with glutaraldehyde. The results from SDS PAGE indicated that the trimer is very sensitive to SDS treatment, a feature which is rather atypical for outer membrane porins.
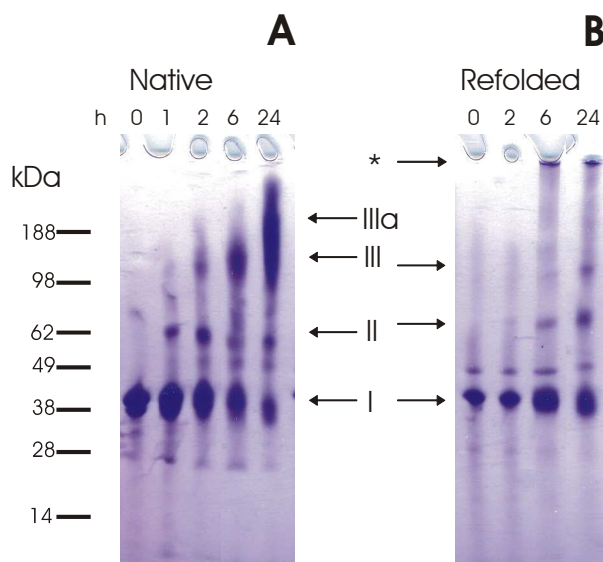

Figure 2.2-8: Coomassie stained SDS PAGE gel of the cross-linking experiment. 0.5 mg/ml of the native and the refolded protein (in DDM) were mixed with 0.04 % (v/v) glutaraldehyde; after the indicated time points samples were boiled in SDS sample buffer for 3 min. The monomeric form of the protein is indicated with I, the putative dimeric and trimeric form with II and III, respectively. After 24 hours an additional form of the native protein appears, IIIa, which most likely is due to aggregation. The refolded protein shows aggregated protein which was not able to enter the gel, indicated by *.

2.2.3.1.5 Circular Dichroism Spectroscopy

Circular Dichroism spectroscopy, which was already applied to follow the refolding of the heterologously produced protein, was also used to investigate the protein's stability. The sensitivity towards SDS, heat treatment and the influence of different detergents was analysed.

The CD spectrum of the native protein showed a minimum at 218.5 nm, a value close to 217 nm and 218 nm reported for other porins like OmpF (Eisele *et al.*, 1990) and PorB (Dahan *et al.*, 1996), respectively, suggesting a majority of β-sheet structure. The minimum of the refolded protein is slightly shifted to 218 nm. Upon heating to 94 °C the signal shifts 1 nm to 219.7 nm for both native and refolded protein but the signal intensity is approximately 10 % lower. The loss of signal results from irreversible precipitation of protein. In the presence of 1 % (w/v) SDS a loss of the CD signal is apparent for the native protein already at 20 °C. Under the same conditions a broadening with an additional minimum at 210 nm indicative for structural changes is observed for the refolded protein. Upon heating the broadening becomes obvious for both refolded and native protein.
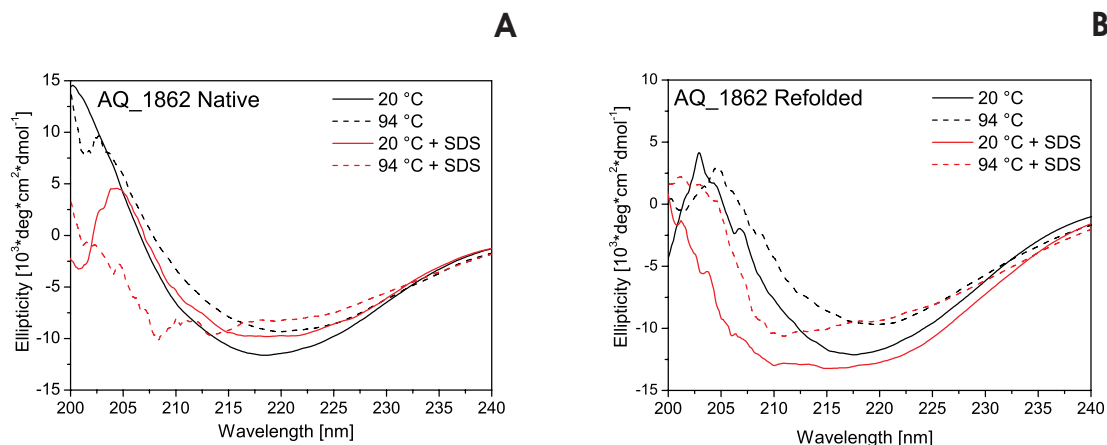
Figure 2.2-9: CD spectra of the native and the refolded protein AQ_1862. Native and refolded protein show a minimum at 218.5 nm and 218 nm, respectively, indicative for a majority of β-sheet structure. Upon heating or in the presence of 1 % (w/v) SDS the overall structure is kept but the signal intensity decreases. Heating in the presence of SDS unfolds the protein, the spectra are broadened.

The CD signal of the protein is not influenced by different detergents, as it is almost identical in the presence of OG or DDM (data not shown). The difference in lipid content as depicted in chapter 2.2.3.1.6 has obviously no influence on the fold of the protein. The lower signal detected in the spectrum of the refolded protein results therefore most likely from a lower amount of β-sheet in the refolded protein preparation, suggesting that it is not completely folded.

The stability of both native and refolded protein was followed by heat treatment of the two samples.

The variation of $\left[\dfrac{\theta_{217nm}^{T}}{\theta_{217nm}^{20°C}}\right]$ is indicative for changes in the β-sheet structure from 20 °C. Both samples

show gradual structural changes above 60 °C but no sudden transition. Proteins from thermophilic organisms such as *A. aeolicus* are usually more resistant to heat denaturation, and AQ_1862 behaviour is expected.
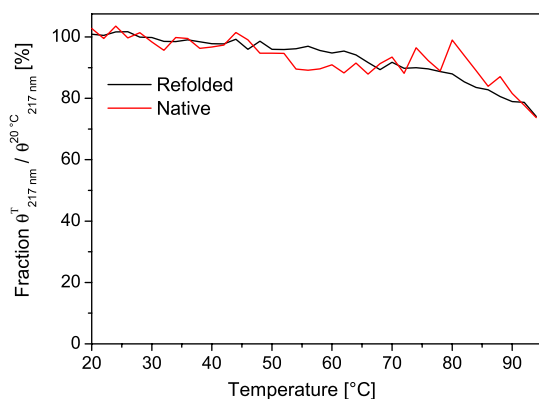

Figure 2.2-10: Heat stability of the native and the refolded protein AQ_1862. The fraction of folded protein is determined by the CD signal at the minimum (217 nm) at the respective temperature divided by the signal at

20 °C $\left[\dfrac{\theta_{217nm}^{T}}{\theta_{217nm}^{20°C}}\right]$.

44

2.2.3.1.6 Lipid Content

Lipids co-purified with AQ_1862 were analysed by HPLC after chloroform-methanol extraction of the purified sample. Different protein preparations and the *A. aeolicus* membrane were compared. In the chromatogram of native *A. aeolicus* membrane, two major peaks are present. Comparison of the elution volumes to the standard mixture revealed similar elution volumes as phophatidylinositol (PI) and maltoside detergents. However, in the literature it was shown, that *A. aeolicus* contains mainly fatty acids, glycerol esters and mono and di substituted glycerol ethers (Jahnke *et al.*, 2001). It can be suggested that the peaks reveal the two derivatives of glycerol esters and ethers.

The strongest peak observed in the sample of the protein purified in the presence of DDM has an elution volume of 13.5 ml, corresponding to maltoside detergents. Two other prominent peaks resemble the elution volume of the two peaks present in the native membrane. Both peaks are absent if the detergent was exchanged against OG. Two major peaks correspond to maltoside detergent (13.5 ml) and to glucoside detergent (8 ml). The data suggest (1) that detergent exchange by washing with ten column volumes of buffer containing OG decreases the amount of bound native lipids under the detection limit and (2) that a substantial amount of DDM remains in the preparation. The refolded protein shows only one major peak with an elution volume of 13.5 ml indicative for the maltoside detergent. The additional small peak with an elution volume of 6.5 ml might originate from remnants of LDAO bound to the protein from the refolding buffer.
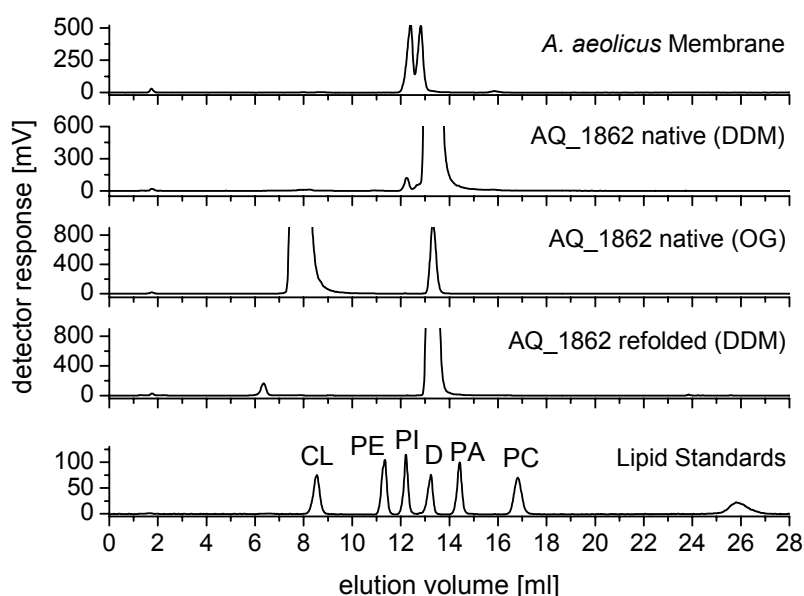

Figure 2.2-11: HPLC chromatogram of different AQ_1862 samples after extraction to determination of the lipid content. Lipid standard: cardiolipin (CL), phosphatidylethanolamine (PE), phophatidylinositol (PI), detergent (D), phosphatidylacide (PA) and phosphatidylcholine (PC).

2.2.3.2   Electrophysiological Characterisation

2.2.3.2.1   Multi-channel Experiments

In order to obtain further evidence that AQ_1862 is indeed a porin like structure, lipid bilayer experiments were performed: this method is often used to show the pore forming ability of a porin protein either by following the increase of conductance in multi-channel experiments or by investigating the conductance of a single molecule inserted into the lipid bilayer. In a first approach multi-channel experiments were performed: a Teflon cuvette with two compartments that are connected by a hole of approximately 1 mm diameter (see Figure 1.1-1) was filled with buffer and a membrane was formed from diphytanoyl phosphatidylcholine. AQ_1862 purified from the native membrane of *A. aeolicus* purified in the presence of DDM was, because purification in DDM leaves most native lipids bound and the resulting protein resembles most the native form found in the outer membrane of *A. aeolicus* most. Small amounts of DDM solubilised AQ_1862 were added to the aqueous compartment. No change in conductance could be monitored, even when the protein concentration was raised to 50 µg/ml. Only when the membrane was formed after addition of AQ_1862 the conductance of such membranes was much higher than in control experiments without the addition of protein. Membranes formed in the presence of protein (0.5 – 2.0 µg/ml) had conductances of up to ≈ 0.2 µS/mm^2 but only of 0.15 nS/mm^2 in the absence of protein and addition of detergent alone did not increase the conductance of the membrane. Once a membrane was formed the conductance was very stable and further incorporation of protein was a very rare event. The addition of large amounts of the protein (up to 50 µg/ml) after the membrane was formed resulted in no increase of conductance but rather in destabilisation and destruction of the membrane. Figure 2.2-12 shows current voltage relations obtained for three different membranes doped with AQ_1862 in 100 mM KCl at pH 7.5 containing different numbers of molecules and for a control membrane without protein addition. The slope conductance is constant up to 60 mV, at higher voltages the slope increases slightly and becomes superlinear. No voltage dependent closing of the channel was detected. The protein is able to increase the conductance of a membrane, but it is not able to insert into the membrane independently: an increase in conductance can only be monitored when the protein inserted into the membrane during formation, at least under the given conditions.
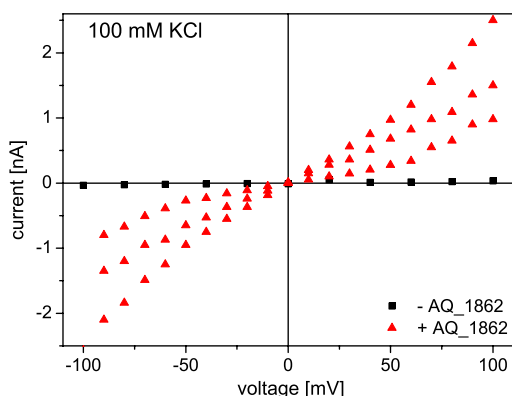
Figure 2.2-12: Current-voltage relation of a membrane doped with the protein AQ_1862 (▲) and a control membrane (■) without added protein. 0.1 M KCl on both sides, pH 7.5.

2.2.3.2.2   Selectivity Measurements in Multi-channel Experiments

The selectivity of the conductance induced by AQ_1862 was studied by zero current membrane potential (zcp) measurements with concentration gradients of various electrolytes across the membrane. After the formation of a membrane doped with AQ_1862 the salt concentration on the *cis*-side (connected to the amplifier) of the membrane was increased by adding small amounts of a concentrated stock solution (see Material and Methods). The initial concentration was 10 mM on both sides. The zero current potential was measured either with a current amplifier applying a voltage to the *trans*-side which drives current to zero or with a voltage amplifier. In both cases the results were identical. A negative zcp indicates anion selectivity of the membrane.

From the observed zcp values the permeability ratios P_a/P_c were calculated according to Formula 1.1-1. Values are given in Table 2.2-1 for a variety of potassium salts with a 1:10 gradient (10 mM vs. 100 mM). The zcp is approximately -40 mV in the presence of K-halides, indicating an anion selectivity of the membrane. Using larger anions like K-gluconate or K-propionate the zcp changes to +43 mV, indicating cation selectivity. Obviously the selectivity of the membrane switches from anion selectivity to cation selectivity when the size of the anion increases. The P_a/P_c ratio decreases from 10 to 0.1 in the order fluoride, bromide, chloride, formate, nitrate, acetate, HEPES, gluconate, and propionate. Also, sulphate and phosphate have a low permeability as indicated by the zcp (Table 2.2-2). Table 2.1-1 shows the zcp obtained with different chloride salts. The membrane is anion selective for all tested chloride salts with the highest selectivity for chloride observed with the relatively large cations Tris$^+$ and Rb$^+$ while it is lowest with Na$^+$ and Li$^+$, suggesting that the permeability for cations is size dependent, too.

Table 2.2-1: Zero current potential (zcp) and P_a/P_c ratios calculated from the corresponding zcp's for different potassium salts. Concentration gradients were always 1:10 (10 mM vs. 100 mM).

Electrolyte	zcp [mV]		P_a/P_c
KF	-43	±0.7	11
KBr	-41	±0.3	10
KCl	-39	±3.5	8
K-formate	-34	±2.0	6
KNO$_3$	-23	±1.1	3
K-acetate	+9	±1.9	0.7
K-HEPES	+9	±1.6	0.6
K-gluconate	+26	±1.0	0.3
K$_2$SO$_4$	+35	±1.3	-
K-phosphate	+42	±2.3	-
K-propionate	+43	±1.0	0.1

Table 2.2-2: Zero-current potentials and calculated permeability ratios obtained with different Cl$^-$ salts. The concentration gradient was 1:10 (10 mM vs. 100 mM).

Electrolyte	zcp [mV]		P_a/P_c
LiCl	-33	±2.1	5
NaCl	-33	±1.5	5
KCl	-39	±3.5	8
RbCl	-42	±3.4	10
CsCl	-39	±1.5	8
NH$_4$Cl	-39	±4.2	8
CaCl$_2$	-37	±3.5	-
Tris-Cl	-42	±1.9	10

In Figure 2.2-13 A the zcp is given as a function of the ratio c^{cis}/c^{trans} across the membrane exemplarily for KCl, LiCl, K-acetate and K-gluconate. Lines are drawn according to the GHK equation (Formula 1.1-1) with the fixed permeability (P_a/P_c) ratios given in Table 2.2-1 and Table 2.2-2, $c^{trans} = 100$ mM and c^{cis} varied. For chloride solutions the concentration ratio dependence of zcp can be described by the GHK equation with a fixed permeability ratio. This does not hold for K-acetate and K-gluconate where large deviations from the GHK equation are observed indicating that in the presence of large anions the permeability ratio is dependent c^{cis}/c^{trans}. Obviously the simple model of GHK cannot describe the permeability ratio in the presence of acetate and gluconate.
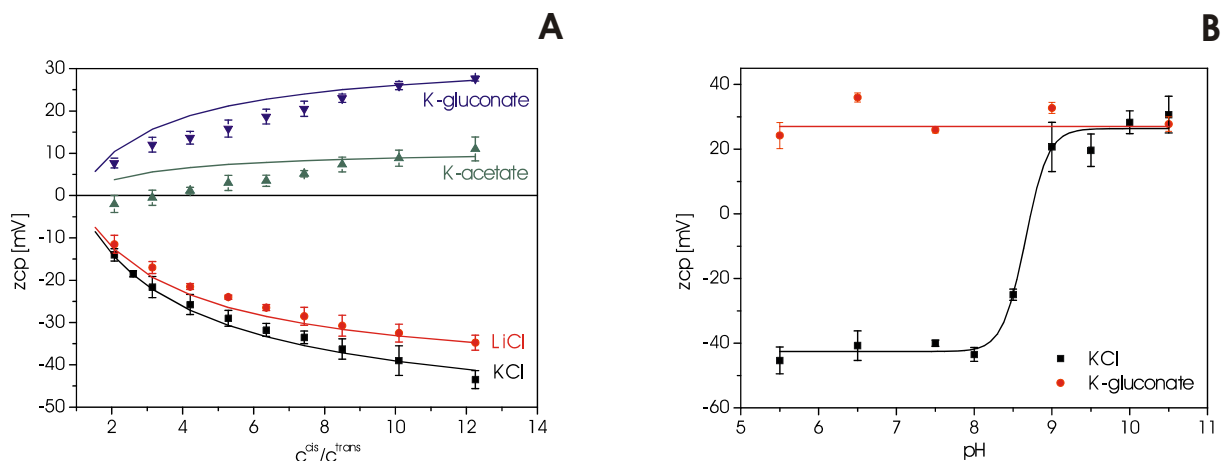

Figure 2.2-13 Zero current potential (zcp) as a function of the electrolyte concentration gradient (c^{cis}/c^{trans}) across the membrane and as a function of pH. Each point represents the average of three membranes. A: Exemplary data obtained with KCl (■), LiCl (●), K-acetate (▲), and K-gluconate (▼) are shown. Lines are drawn according to the Goldman-Hodgkin-Katz (GHK, Formula 1.1-1) equation using the P_a/P_c values at c^{cis}/c^{trans} =10:1 from Table 2.2-1 and Table 2.2-2. For Cl$^-$ solutions the concentration dependence of zcp can be described by the GHK equation with a fixed permeability ratio. This does not hold for acetate and gluconate where large deviations from the GHK equation are observed indicating that in the presence of large anions the permeability ratio is concentration dependent. B: Exemplary data obtained with KCl (■) and K-gluconate (●). For K-gluconate the zcp is positive between pH 5.5 and 10.5 and pH independent (+35 mV). In the presence of KCl the zcp is negative (-40 mV) below pH 8. Above pH 8 the zcp becomes more positive i.e. the anion selectivity decreases and above pH 9 the zcp approaches +35 mV, i.e. it is cation selective as in the presence of gluconate.

To investigate whether a titrable charged residue determines the selectivity of the presumed pore, the pH dependence of the zcp was studied. Exemplarily the pH dependence of zcp for KCl and K-gluconate was determined with the concentration gradient at 1:10 (10 mM vs. 100 mM) in both cases. Initially the pH was changed by adding small amounts of HCl or KOH after the salt gradient was established. Due to difficulties in measuring the pH in the cuvette and destabilisation of the membrane upon HCl or KOH addition, each data point in Figure 2.2-13 B was determined in three independent experiments. In the presence of KCl the pore is anion selective (P_{Cl}/P_K = 1:8) from pH 5.5 to pH 8.5, at higher pH the zcp changes its sign, and above pH 9.0 the membrane becomes cation selective (P_{Cl}/P_K = 3:1). For K-gluconate the zcp is positive and pH independent between pH 5.5 and 10.5 i.e. the pore shows a pH independent cation selectivity ($P_{gluconate}/P_K$ = 3:1).

These data suggest a selectivity filter, which lets the small anions like Cl^- pass approximately eight times better than K^+ so that the membrane is anion selective. Larger or organic anions cannot pass, in this case the selectivity of the membrane dominated by the cation, the membrane is cation selective. The selectivity filter for small anions is pH dependent, above pH 8 it does not let anions pass anymore, and the selectivity is again dominated by the cation.

2.2.3.2.3   Single Channel Measurements

To obtain more detailed information on the ion selectivity and gating properties of AQ_1862, experiments with only one channel incorporated into the membrane were attempted. To increase the probability of incorporating only one channel the area of the lipid membrane was reduced to 1 % of the area used for multi-channel experiments, and the final concentration of the protein added was reduced to 10 %. Incorporation was again achieved by forming the membrane in the presence of protein; still it was a rare event. Typically a stable conductance of about 1.4 nS in 100 mM KCl was observed when AQ_1862 inserted into the membrane (more than 200 times the conductance of an undoped membrane). It lasted for several hours without any interruption. Channel-like fluctuations of 0.25 nS were usually observed superimposed on this conductance (Figure 2.2-14).
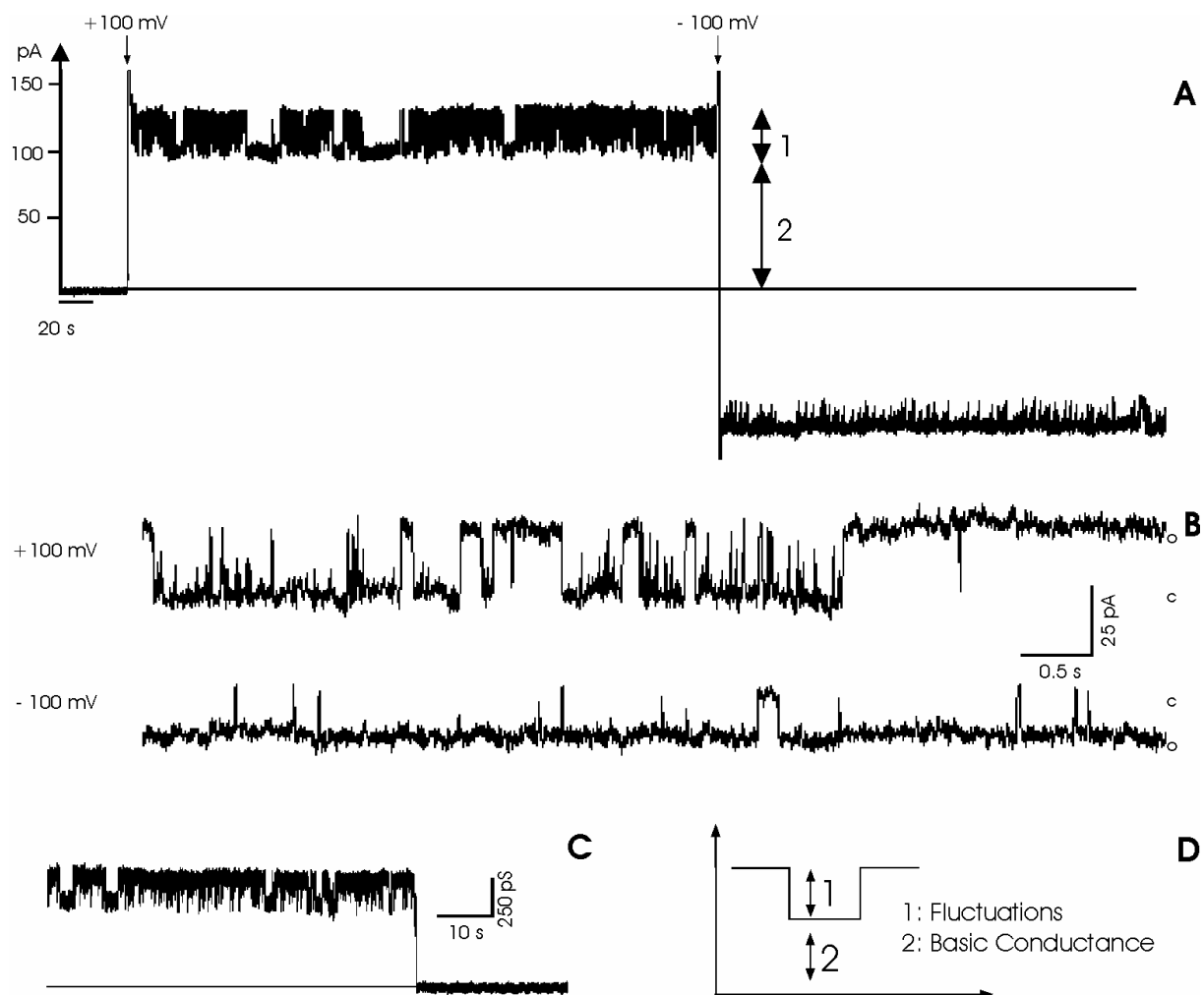
Figure 2.2-14: Currents observed in a typical BLM experiment with a single molecule of AQ_1862 inserted into a diphytanoyl phosphatidylcholine bilayer (0.1 M KCl on both sides, pH 7.5, 25 °C). Switching the voltage from 0 to ± 100 mV induces a large current of 100 pA, here indicated by 2. Superimposed on the large current step there are rapid outward fluctuations of about 25 pA here indicated by 1. B: On the expanded scale the fluctuations can be seen as channel opening and closing events. Note that the amplitude of fluctuations is smaller at negative voltage. C: Simultaneous break-down of the basic and fluctuating conductance. (0.1 M KCl both sides, pH 7.5, +40 mV). D: For definition of the fluctuating and the basic conductance: When a voltage is applied the current reaches a basic level (2) upon which a fluctuating component (1) resides.

Both the stable and fluctuating conductance were observed simultaneously and eventually disappeared simultaneously (Figure 2.2-14 C). We coined the terms basic conductance and fluctuation (Figure 2.2-14 D for definition) instead of open state and substate because the latter terms imply that both states are attributed to the same conductance pathway which may not necessarily be the case for AQ_1862 (see below and discussion). Figure 2.2-14 A shows a typical experiment: switching the voltage from 0 to +100 mV induces a large current of 100 pA (indicated with 2). Superimposed on the large current step there are rapid outward fluctuations of about 25 pA (indicated by 1, see expanded scale Figure 2.2-14 B). The conductance of the fluctuating component shows a voltage dependent asymmetry. In approximately 75 % of the cases the conductance is larger at positive than at negative voltages. Furthermore, the probability for the fluctuating part to be open is larger at the higher conductances. The basic conductance does not show any voltage dependence.

The basic conductance is 1.4 nS ±0.3 nS (mean ± S.D.) in 0.1 M KCl at pH 7.5 while the conductance of the fluctuations is about 0.24 nS ±0.014 nS (Table 2.2-3). At higher chloride concentrations (> 0.3 M to 0.5 M) the fluctuations cannot be distinguished from the noise anymore (see below). Therefore in the following chapter all experiments were carried out at a concentration of 100 mM because the fluctuating part can still be monitored and the differences in the basic conductance are more pronounced than at 10 mM.

Although both events are strongly correlated and therefore most likely related to the same molecular entity they have different properties with respect to ion selectivity, concentration and pH dependence, as will be shown below.

2.2.3.2.4   Size and Selectivity of the Basic and Fluctuating Conductance for Cations and Anions

To investigate the selectivity of the basic and the fluctuating conductances, single channel experiments with AQ_1862 were carried out in a variety of electrolytes.

Figure 2.2-15 A and Figure 2.2-16 A show recordings of the fluctuations in the presence of different potassium and chloride salt solutions (100 mM, pH 7.5, 100 mV). Figure 2.2-15 B and C show current-voltage plots in 100 mM salt solution at pH 7.5 of the fluctuating and the basic conductance respectively.
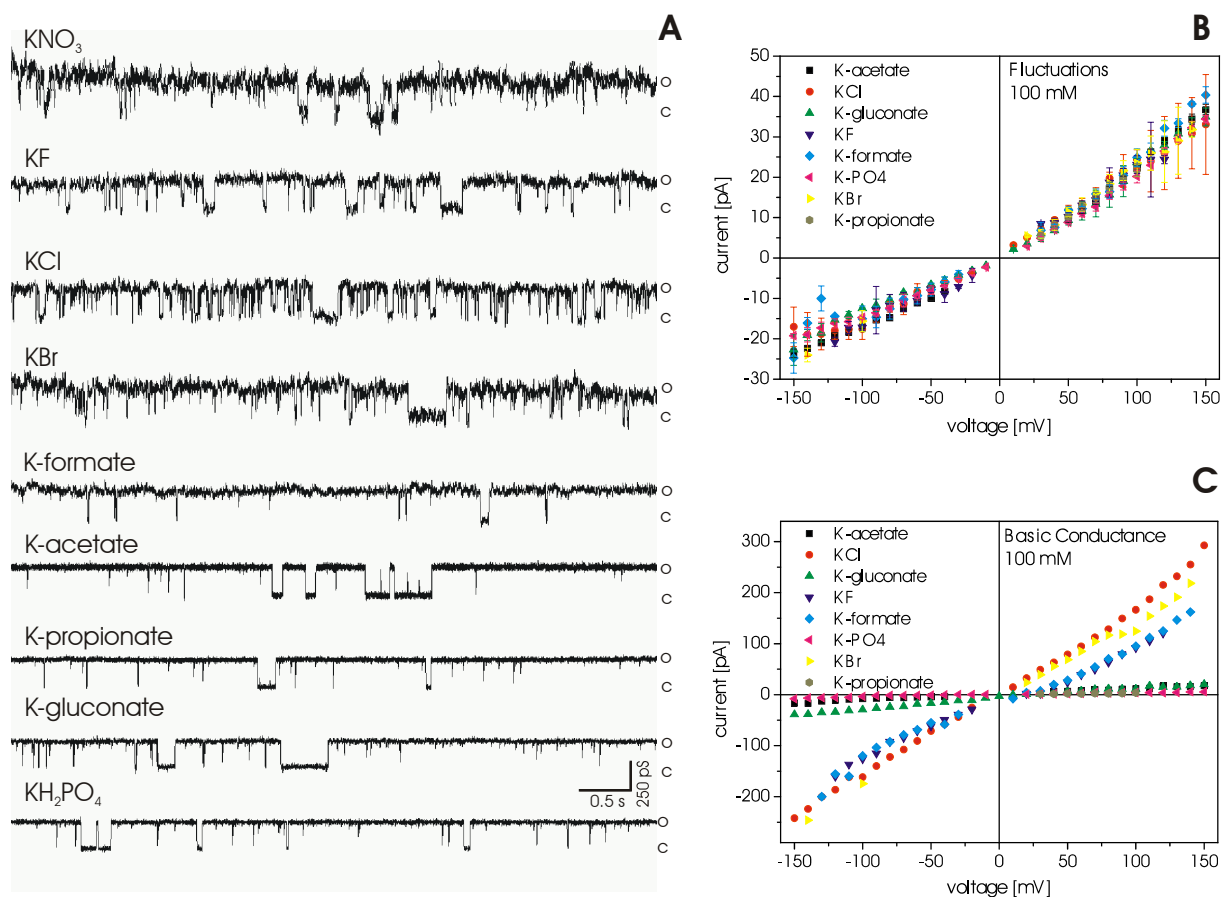
Figure 2.2-15: Current fluctuations, current-voltage relation of the fluctuating part (B) and of the basic conductance (C) recorded in different potassium salt solutions (100 mM on both sides, 100 mV). A: The amplitude of the channel-like fluctuations is independent of the anion but the traces are much noisier in the presence of small anions (see text for details). Open and closed states are marked with o and c, respectively. B: The fluctuating part is independent of the anion. C: The basic conductance using high in the relatively small anions chloride, fluoride or formate and low using the relatively large or organic anions acetate, gluconate, or propionate.

The amplitude of the fluctuations is identical in all potassium salt solutions (Figure 2.2-15 A and B) independent of the anion. This result suggests that the fluctuations are cation selective. On the other hand the recordings obtained in halides or nitrate appear to be more noisy. This observation is most likely based on the large basic conductance and the related increased thermal noise in the presence of high concentrations of small anions. As Table 2.2-3 and Figure 2.2-15 C show, the basic conductance is much higher in the presence of these anions than in the presence of phosphate or organic anions. The frequency of the fluctuations can vary from experiment to experiment. In addition it might be influenced by the kind of anion, but this question has not been addressed in detail.

Table 2.2-3: Size of the basic and fluctuating conductances for different electrolytes. 0.1 M salt solutions on both sides, pH 7.5, 100 mV, 25 °C.

Electrolyte	Fluctuation [pS]	S. D.	Basic Conductance [pS]	S. D.
KF	241	17.6	660	300
KCl	241	13.8	1380	300
KCl pH 10.5	274	14.4	50	20
KBr	233	25.9	1370	320
KNO$_3$	247	14.1	1250	250
KNO$_3$ pH 10.5	213	22.8	300	100
K-phosphate	224	4.7	75	60
K-formate	255	25.0	720	550
K-acetate	245	22.6	300	170
K-propionate	230	10.8	80	20
K-gluconate	241	13.2	190	160
K-gluconate pH 10.5	256	23.5	200	100
NH$_4$Cl	111	8.7	990	260
NH$_4$-acetate	116	8.7	210	220
NaCl	60	4.1	1440	320
Na-acetate	58	5.2	200	80
Na-gluconate	62	4.7	190	100
LiCl	41	3.3	960	300
Li-acetate	40	5.5	200	90
RbCl	235	28.6	1090	210
Rb-acetate	203	1.5	190	90
CsCl	29	5.7	750	210
Cs-acetate	32	7.1	220	100
NMG-Cl	-	-	200	100
NMG-acetate	-	-	20	10

Table 2.2-3 shows the size of the fluctuating conductance and of the basic conductance at 100 mV and pH 7.5 for different electrolytes (100 mM on both sides). For the tested salts of K^+, Na^+ or Cs^+ the size of the fluctuating conductance is independent of the anion present (240 pS, 60 pS and 30 pS respectively). The rank order of the size of the fluctuating conductance for cations is $K^+ \approx Rb^+ > NH_4^+ > Na^+ \approx Li^+ \approx Cs^+$. Using divalent or very large ions like Ca^{2+} or N-methylglucosamine (NMG), no fluctuations are observed. The fluctuating part is independent of the anion used. For the basic conductance, the sequence for anions is $Cl^- \approx Br^- \approx NO_3^- > F^- >$ gluconate $\approx$ acetate $\approx$ phosphate $\approx$ propionate (100 mM, 100 mV, pH 7.5), decreasing from 1300 pS in Cl^- to 80 pS in propionate. The basic conductance is not affected by the kind of cation. The variability of the basic conductance is larger than that of the fluctuating component indicated by the large standard deviation. It is also evident that although the basic conductance is mainly determined by the anion, cations are also important as can be shown by the conductances in the presence of KCl, LiCl, CsCl, and NMGCl which are 1.4 nS, 0.96 nS, 0.75 nS and 0.22 nS respectively.

Figure 2.2-16 shows single channel recordings and the current-voltage relations of different cations as acetate salts of the fluctuating and the basic conductance, respectively.
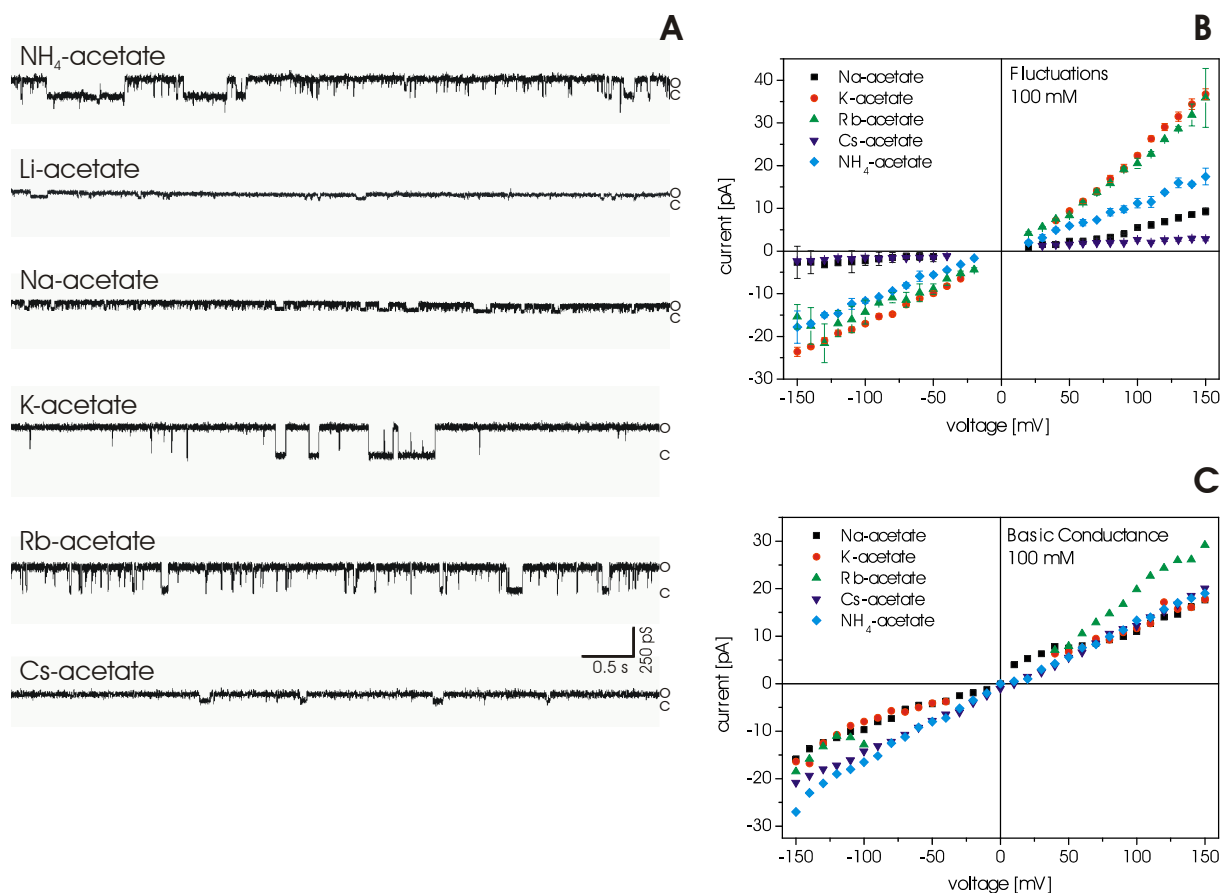
Figure 2.2-16: **Current fluctuations, current-voltage relation of the fluctuating part (B) and of the basic conductance (C) recorded in different acetate salt solutions (100 mM on both sides, 100 mV).** A: The amplitude of the channel-like fluctuations is independent of the anion but depends on the cation the conductance (high in K^+ and Rb^+, intermediate in ammonium and low in Na^+ and Cs^+). Open and closed states are marked with o and c, respectively. B and C: The fluctuating part is dependent on the cation, while the basic conductance is independent of the cation.

The current-voltage relationship was linear for the basic conductance. The fluctuating part has a linear current-voltage relation in the range of ± 60 mV; at more positive voltages it becomes superlinear, at more negative voltages sublinear.

The differential selectivity of the fluctuating and the basic conductance was investigated further by determination of the zcp in a single channel experiment (10 mM vs. 100 mM gradients using different electrolytes). The reversal potential is determined by the zero point of the abscissa. Figure 2.2-17 shows exemplarily the current-voltage relation for KCl. Figure 2.2-17 A illustrates the disappearance of the fluctuations at +40 mV indicating cation-selectivity, while the basic conductance reverses its sign at negative potentials indicating anion-selectivity. In Figure 2.2-17 B current-voltage relations of the basic conductance and of the fluctuating component show a reversal potential of +40 mV for the fluctuating part and of -45 mV for the basic conductance.
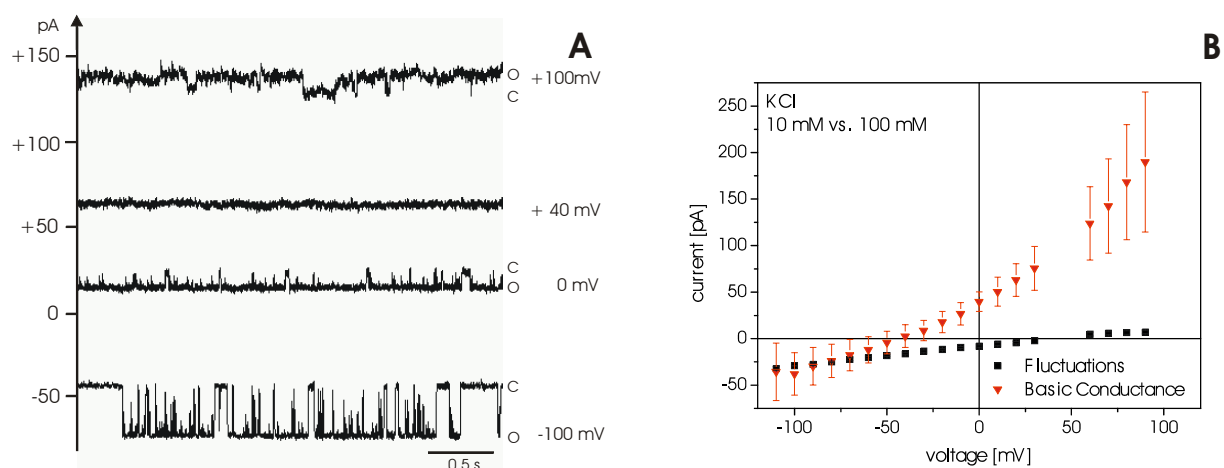
Figure 2.2-17: Current-voltage relation the presence of an ion gradient (10 mM vs. 100 mM KCl). A: Current traces at different applied voltages (right hand side). o and c mark the open and closed state of the fluctuating conductance, respectively. Note that the amplitude of the fluctuations disappears at + 40 mV. Current attributed to the basic conductance may be estimated from the current level marked as closed state (c). This current component reverses sign at approximately -45 mV: Current-voltage relation of the fluctuating (■) and the basic conductance (▼). The zero current potentials are +40 mV and -45 mV, respectively.

A variety of electrolytes were analysed in the same way (Figure 2.2-18). The zcp of the fluctuations is in the range of +40 mV for all tested cations (K^+, Na^+, Cs^+, NH_4^+, Rb^+, Li^+, as chloride or acetate salt; Figure 2.2-18 A) independent of the type of anion (acetate, NO_3^-, Cl^-, gluconate, propionate and formate; Figure 2.2-18 C). The zcp of the basic conductance is -45 mV for the relatively small anions chloride, nitrate and formate, while it is close to zero for acetate, gluconate and propionate (Figure 2.2-18 D). It is not changed by the type of cation (Figure 2.2-18 B).
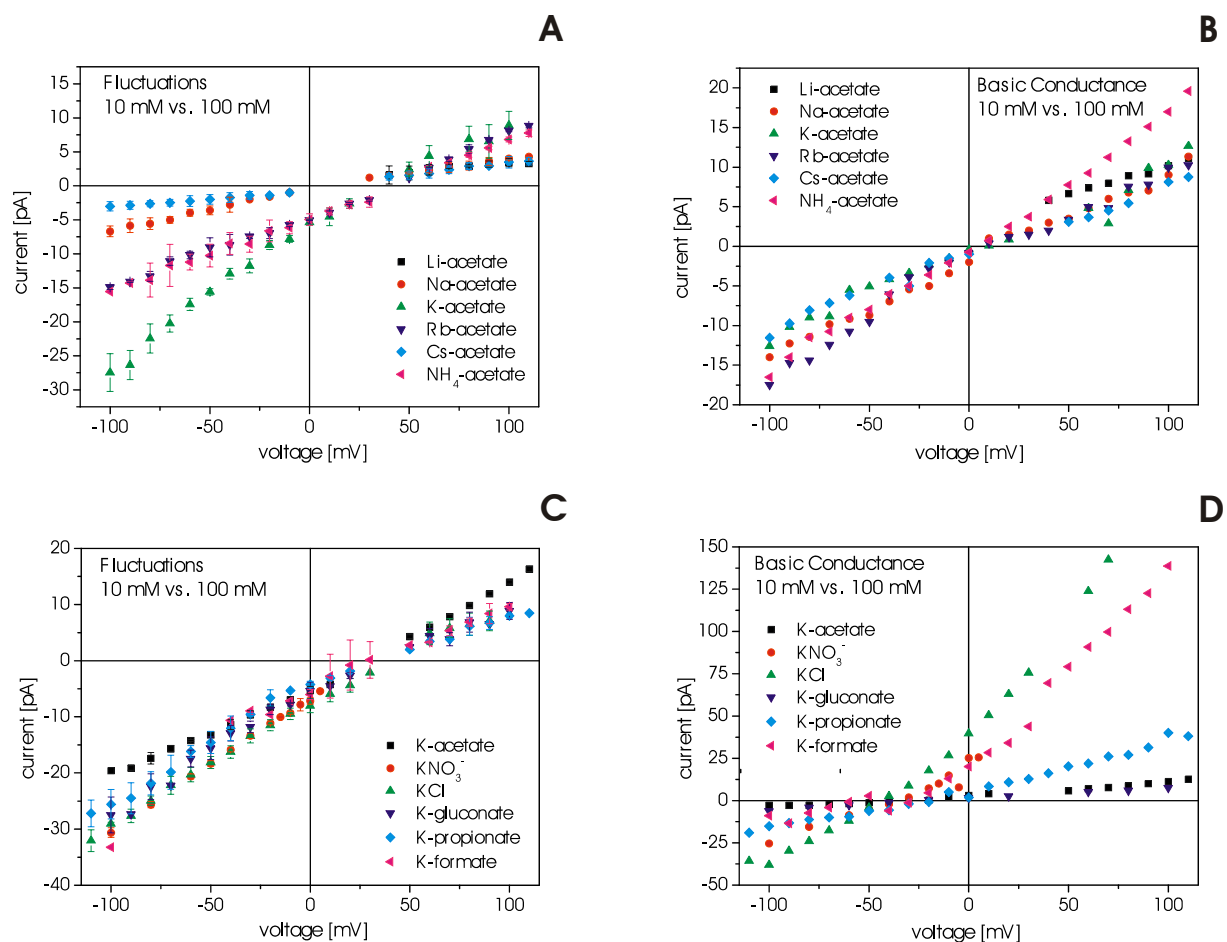
Figure 2.2-18: Current-voltage relation the presence of an ion gradient (10 mM vs. 100 mM KCl) of the fluctuating (A and C) and the basic conductance (B and D). The zcp of the fluctuating part are +40 mV independent of the anion. The zcp of the basic conductance are 0 mV for all acetate salts independent of the cation and 0 mV for acetate, gluconate and propionate and -45 mV in the presence of chloride, nitrate and formate.

In conclusion it should be noted that anions which induce anion-selectivity in macroscopic experiments generate high basic conductance and anion-selectivity of this component in single channel experiments. Anions which generate cation selectivity in multi-channel experiments induce low basic conductance and non-selectivity of the basic conductance in single channel experiments.

Since the zcp of KCl was highly affected by pH in the multi-channel experiments, single channel experiments were carried out at different pH values in solutions with a gradient of 10 mM to 100 mM. KCl was chosen as electrolyte because it showed a large effect in the multi channel experiments. The current-voltage relation of the fluctuating part at pH between 6.5 and 10.5 is depicted in Figure 2.2-19 A. Neither the slope (conductance) nor the zcp shows strong pH dependence. On the other hand the curves measured for the basic component (Figure 2.2-19 B) show a remarkable shift of the reversal potential from anion selectivity at pH 6.5 (zcp = -40 mV) to non-selectivity at pH 10.5 (zcp = 0 mV).
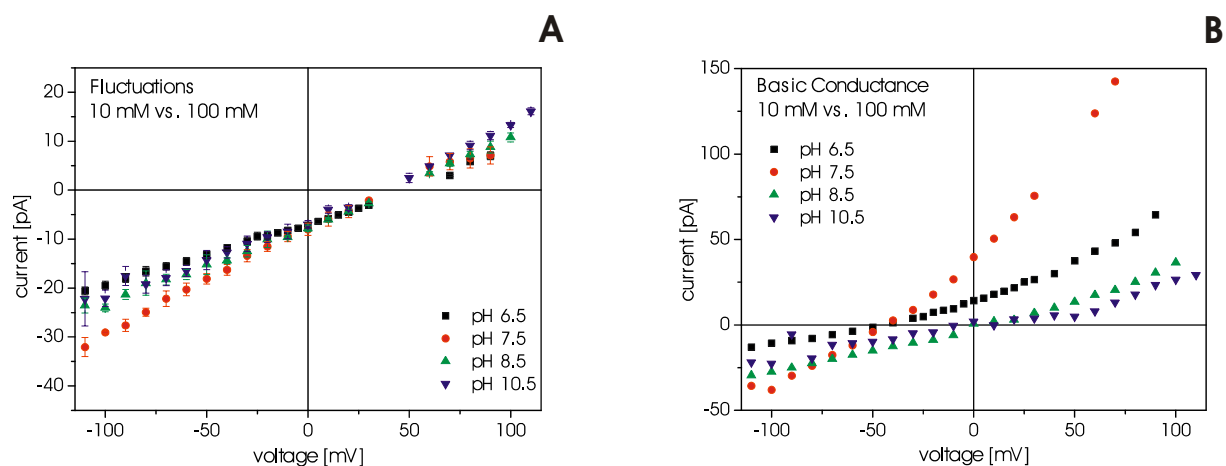
Figure 2.2-19: pH dependence of current-voltage relations of the fluctuating (A) and the basic (B) part of the conductance in the presence of a KCl concentration gradient (10 mM vs. 100 mM). The zcp of the fluctuating component is pH independent between pH 6.5 to 10.5 while the zcp of the basic conductance is shifted from -45 mV at pH 6.5 to 0 mV at pH 10.5. There is also a shift of the slope conductance which increases from pH 6.5 to 7.5 and then decreases again.

2.2.3.2.5   Concentration and pH Dependence

The concentration dependence of the conductance was determined by forming a membrane in 10 mM salt solution and then increasing the salt concentration by adding small amounts from a concentrated stock solution to both sides. Figure 2.2-20 A shows the conductance of the fluctuations for KCl at pH 7.5 and pH 10.5 and for K-acetate at pH 7.5 at ± 100 mV. As mentioned above the conductance is voltage dependent even in symmetric solutions. The conductance for both positive and negative voltages saturates at about 100 mM. A fit with the Michaelis-Menten equation yields a K_m of about 9 mM for KCl.

$$\Lambda = \frac{\Lambda_{max} * c}{K_m + c}$$

Formula 2.2-1: Michaelis-Menten formula modified for the conductance Λ and the concentration c.

For all cations tested (Na^+, K^+, Rb^+ and NH_4^+ in combination with Cl^- or acetate Figure 2.2-20 A and C) the K_m values of the fluctuating component are between 5 and 20 mM and maximal conductances vary between 60 and 250 pS, depending on the cation.

In contrast, the concentration dependence of the basic conductance depends on the type of anion. Since the basic conductances do not show any voltage dependent asymmetry only the results for +100 mV are shown. Figure 2.2-20 B shows that the basic conductance at pH 7.5 increases linearly up to at least 500 mM KCl, the largest concentration measured. On the other hand at pH 10.5, in K-acetate and KCl this conductance component saturates at rather low concentrations ($\approx$ 100 mM).
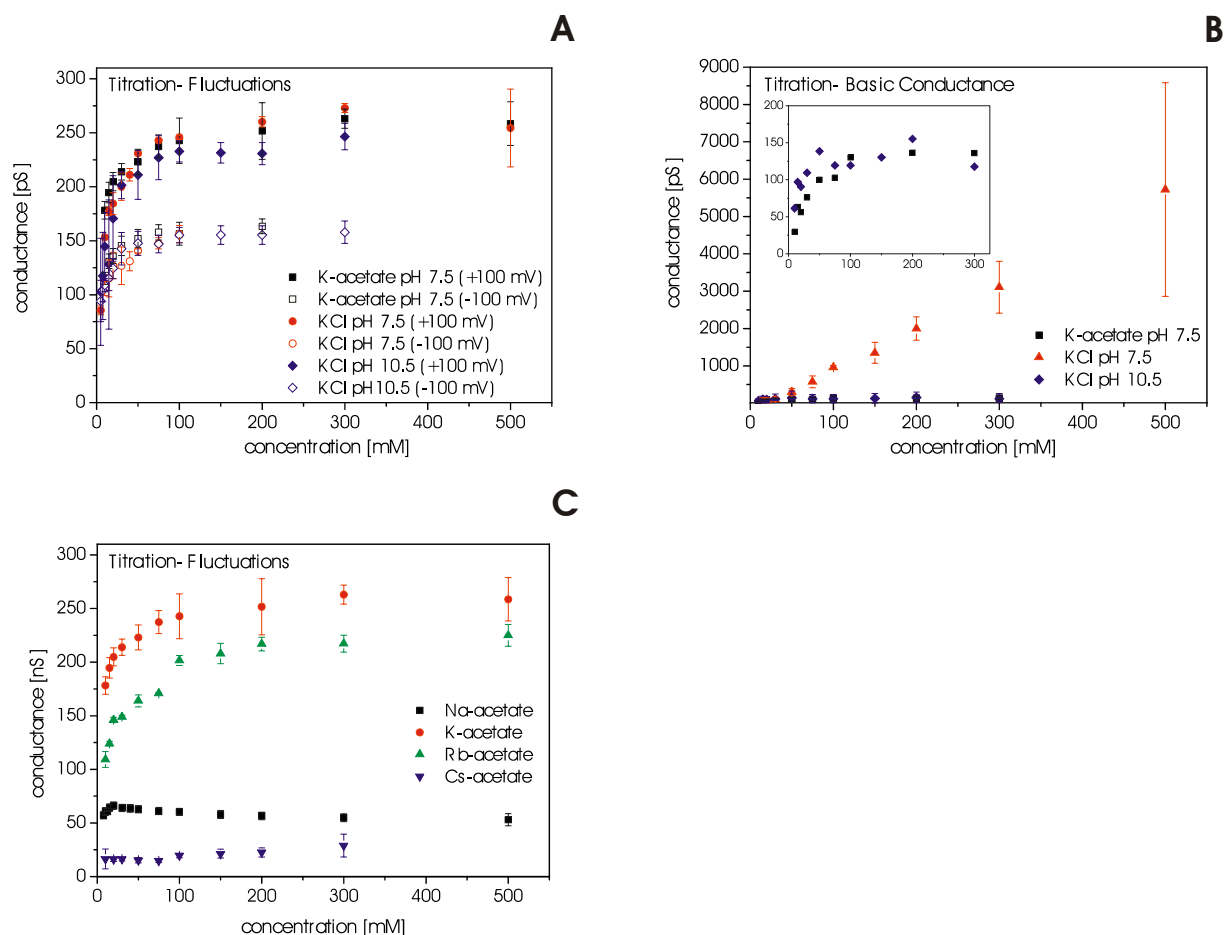
Figure 2.2-20: Concentration dependence of the fluctuating (A) and basic conductance (B) in the presence of K-acetate and KCl and different cations (C). The fluctuating component saturates at low salt concentration ($K_m \approx 15$ mM) irrespective of the anion. At negative voltages the conductance is smaller by approximately 1/3. Since the basic conductance does not show any voltage dependent asymmetry only the result for +100 mV are shown. The concentration dependence of the basic conductance differs for Cl^- and acetate. In the presence of Cl^- the conductance increases approximately linearly with the concentration whereas in the presence of acetate it does not significantly change with increasing concentration. In both cases the conductance is voltage independent. C: Concentration dependence of the fluctuating conductance in the presence of Na^+, K^+, Rb^+ and Cs^+.

The difference in the concentration dependence of both conductance components in KCl explains why we never observed fluctuations at high KCl concentrations. The relatively small fluctuations which saturate at about 20 mM are not visible at high concentrations anymore because of the large noise related to the basic conductance. All anions which induce a high basic conductance (Table 2.2-3) show a linear increase of this conductance with rising concentrations up to at least 500 mM, while all anions with low basic conductance saturate at low concentrations.

2.2.3.2.6 Kinetic Properties of AQ_1862

Both the basic conductance and the superimposed fluctuations are very stable and can be observed for several hours. Occasionally a complete closing of the channel was observed (Figure 2.2-14 C).

A proper evaluation of the open times of the fluctuating component is not possible because of the limited time resolution of the recording system (<300 Hz) and the number of brief openings. However,

a rough estimate of the closed time indicates that the open probability is pH dependent. At high pH the fluctuating part is almost always open with only short flickerings to the closed state. At pH 8.5 the closing events become more often and longer and at pH 6.5 the channel transition between open and closed state are most frequently.

The open probability is about 80 % at pH 7.5 and increases to 99 % at pH 10.5.

2.2.3.2.7   Refolded AQ_1862

In order to determine if the refolded protein is indeed similar to the native protein it was applied to lipid bilayer experiments. It was used in the presence of DDM like the native protein. Upon adding the protein to the aqueous compartment it was also not able to incorporate into the membrane from the aqueous compartment. If the membrane was formed in the presence of the protein it is readily inserted in the membrane as judged by an increase of conductance. In the case of the refolded protein incorporation was a more frequent event.

2.2.4   Crystallisation

2.2.4.1   *Native Protein*

2.2.4.1.1   Crystals in Presence of DDM

AQ_1862 crystals obtained from protein purified by the dye-affinity chromatography in the presence of DDM were grown at pH 4.6 using different polyethylene glycol (PEG) 4000 and Zn-acetate as precipitant. The crystals were large (up to 0.7 x 0.3 x 0.3 mm) with a cuvette-like shape and were observed in hanging and in sitting drop vapour diffusion approaches [Peng, unpublished data]. The crystals diffracted X-rays to a maximal resolution of 4 Å, however the crystal quality i.e. mosaicity had to be improved and the phase problem had to be solved.

In order to improve the crystal quality different approaches were followed: (i) improve the protein quality and homogeneity by the respective purification attempts (ii) screening of reservoir solution using different precipitant and salt concentrations (iii) control of crystal growths (iv) and addition of additives

The protein was purified by different strategies (chapter 2.2.2.1): dye affinity, hydroxyapatite, and anion exchange, of protein fraction obtained from anion exchange and gel filtration chromatography. No difference between the samples in crystal quality, size, or quantity was determined. Therefore protein from the most effective and fastest method, purification by dye affinity, was used further on.

The reservoir solution was varied using PEGs, salts, or alcohols at varying concentrations as precipitants. Crystals were obtained at Zn-acetate concentrations between 0.08 M and 0.12 M and PEG 4000 concentrations between 7 % and 10 %. In the presence of ammonium-sulphate, NaCl as well as 2-methyl-2,4-pentanediol needles could be obtained, but they could not be improved to three-dimensional crystals. Independent on the protein concentration crystals could only be obtained

between pH 4.4 and 5.0. Dependent on the PEG concentration crystals were obtained with protein concentrations of 4 to 7 mg/ml which diffracted X-rays up to 5 Å. The space group was $P2_12_12_1$.

The crystals grew very fast; first crystals are already observed after 1 h. Since the speed of crystal growth might negatively interfere with the crystal quality, different approaches were undertaken to regulate the crystal growth: ultracentrifugation, different crystallisation techniques, and modification of the drop size and the protein content. The protein sample was applied to ultracentrifugation before and after mixing with the crystallisation buffer in order to remove microcrystals or other seeds of premature crystal growth. In the first case the crystal quality was not improved in the latter no crystals could be obtained. Crystallisation by vapour diffusion in sitting drop devices yielded larger crystals but with comparable crystal quality. Batch crystallisation is an often used alternative to vapour diffusion technique and can sometimes improve the crystal quality. The standard vapour diffusion conditions were used for batch crystallisation. The drops were sealed from the environment by paraffin oil, which allows little to no diffusion of water, silicone oil, which does allow diffusion of water or Al´s oil, which is a mixture of both. Crystals could be obtained in a range from 10 % to 13 % of PEG 4000 but the quality was comparable to the original conditions. Drop size and protein content did not influence the speed of crystal growth. Different additives were tested to improve the crystal quality. Especially benzamidine, hexanetriol, heptantriol, and OG improved the quality; crystals which diffracted X-rays up to 4 Å were obtained.

X-ray data collection is easiest and fastest when crystals are frozen and measured in a nitrogen stream at 100 K. Different cryo-protectants including sucrose, glucose, PEG 400 and glycerol at different concentration have been tested. Best results were obtained when the crystals were soaked for a short time ($\approx$ 30 s) in the crystallisation buffer containing 15 % to 20 % (v/v) glycerol and 20 % (v/v) PEG 4000.

To solve the phase problem during structure determination, heavy atom derivatives were co-crystallised with the protein. At the same time the presence of zinc in the crystallisation buffer was utilised: MAD experiments were performed at the zinc edge. However the anomalous signal was too weak for structure determination. A heavy atom screen was performed; heavy atom compounds were supplemented as additives in a concentration of 1 to 10 mM from a stock solution of 100 mM to the crystallisation drop. AQ_1862 purified in the presence of DDM crystallised in the presence of several heavy atoms (K_2PtCl_6(IV), KCl_4Au(III), Hg acetate, Pb acetate and K_2OsO_4). The addition of 1 mM of K_2PtCl_6 (IV) or K_2OsO_4 yielded crystals diffracting to approximately 6 Å, while the crystal quality in the presence of all other derivatives was low. However, fluorescence scans did not show a clear signal at the platinum edge, maybe due to the presence of zinc.

Since crystal growth and quality could not be controlled using the initial conditions. AQ_1862 was applied to a wide grid screen of crystallisation conditions. Crystals were obtained in several other conditions. Two conditions were optimised to obtain three-dimensional crystals suitable for X-ray data collection. Crystals grown in the presence of PEG 400 and Li-sulphate using protein concentrations

from 5 to 15 mg/ml diffracted up to 10 Å. Crystals were obtained from a condition very similar to the initial condition but at neutral pH with PEG 6000 and a mixture of Zn-acetate and $ZnCl_2$ which diffracted up to 5 Å. Again MAD experiments at the zinc edge were performed, but the anomalous signal was weak as well. Co-crystallisation experiments with heavy atom derivatives were not attempted since zinc is present in the condition as well and the same problems are expected.

2.2.4.1.2 Different Detergents

Up to now most structures of membrane proteins have been obtained by crystallisation of the membrane protein from aqueous solution in the presence of detergents. The choice of detergent is very crucial for the success of the experiment. Most outer membrane proteins have been crystallised using C_8E_4, OG, or LDAO. Ten different detergents were chosen at concentrations 1 to 4 x CMC: DDM, DM, Fos-12, NG, OG, OTG, LDAO, C_8E_4, C_8E_n, C_8HESO. Detergent was exchanged using Reactive Red 120 after MonoQ anion exchange chromatography and gel filtration chromatography. The protein concentration was adjusted to 5 mg/ml and 10 mg/ml in a buffer containing 150 mM NaCl and 20 mM Tris-Cl buffered at pH 7.4. AQ_1862 was applied to broad sparse matrix screens of commercially available crystallisation conditions (standard crystallisation screening) using the vapour diffusion technique in sitting drop experiments. Crystals or crystalline precipitate could be obtained in all tested detergents in a variety of conditions in standard crystallisation screening, but only few could be reproduced and improved to three-dimensional crystals suitable for X-ray data collection.

Table 2.2-4: The crystallisation conditions of the protein AQ_1862 in the presence of different detergents.

| Detergent	Precipitant	Buffer	Salt	Shape	Space group	Resolution
DDM	7 - 10 % PEG 4000	Na-acetate pH 4.6	0.08 - 0.12 M Zn-acetate	Cuvette/ rectangle	$P2_12_12_1$	4-5 Å
	5 - 12 % PEG 6000	Tris-Cl pH 7.5	0.15 M Zn acetate, 0.05 M $ZnCl_2$	Hexagon	$P2_12_12_1$	4-5 Å
	20 - 28 % PEG 400	Tris-Cl pH 8.5	0.2 M Li-sulphate	Hexagon		10 Å
DM	25 -28 % PEG 400	Na-HEPES pH 7.5	0.1 - 0.15 M NaCl	Needle		10 Å
	28 % PEG 400	Tris-Cl pH 8.5	0.18 - 0.22 M Li-sulphate	Needle		10 Å
NG	16 - 24 % PEG 400	Na-HEPES pH 7.5	0.08 - 0.1 M $MgCl_2$	Cubic	R3	2.5 Å
OG/ OTG	25 - 30 % PEG 400	Na-HEPES pH 7.5	0.18 - 0.2 M $CaCl_2$	Cubic/ tetrahedral	R3	2 Å
	28 - 30 % PEG 400	Na-HEPES pH 7.5	0.18 - 0.22 M $MgCl_2$	Cubic	R3	6 Å
	30 - 32 % PEG 400	Na-HEPES pH 7.5	0.18 - 0.22 M $CaCl_2$	Cubic	R3	6 Å
LDAO	25 - 30 % PEG 400	Na-acetate pH 4.6	0.08 - 0.12 M $CaCl_2$	Cubic		10 Å
C_8E_4 / C_8E_n	20 /25 % PEG 4000	Tris-Cl pH 8.5	0.18 - 0.22 M Li-sulphate	Plates		10 Å
C_8HESO	25 - 30 % PEG 4000	-	-	Cubic		10 Å
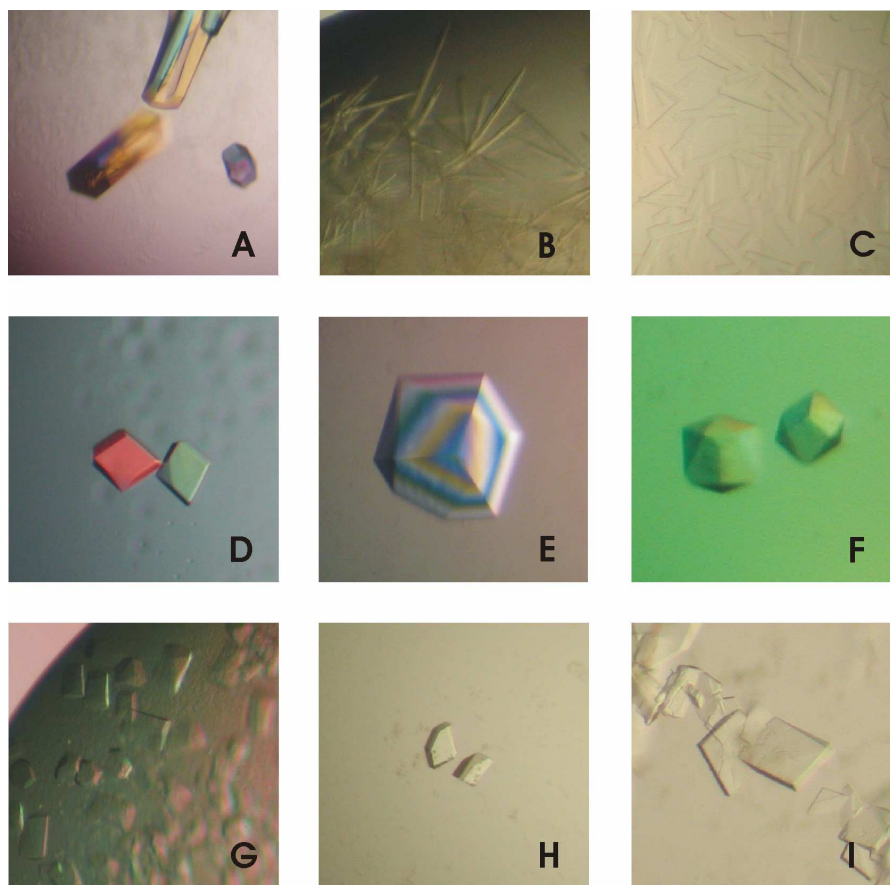
Figure 2.2-21: Crystals of protein in the presence of different detergents. A: DDM, B: DM, C: Fos-12, D: NG, E: OG, F: OTG, G: LDAO, H: C_8E_5, I: C_8E_n

Crystals obtained in the presence of C_8E_4 and C_8E_n (20 to 25 % PEG 4000, 0.18 to 0.22 M Li-sulphate at pH 8.5), C_8HESO (25 to 30 % PEG 4000) and LDAO (25 to 30 % PEG 400, 0.08 to 0.12 M $CaCl_2$ at pH 4.6) lacked reproducibility and did not diffract X-ray. The detergents NG, OG, and OTG are chemically very similar and crystals appeared in similar conditions: 25 to 30 % PEG 400, 0.1 to 0.25 M of $CaCl_2$ or $MgCl_2$ at pH 4.6 and pH 7.5 with protein concentration of 5 to 10 mg/ml (Figure 2.2-21 D to F). Crystals obtained from the initial screen grown in the presence of OG or OTG and $CaCl_2$ or NG in combination with $MgCl_2$ at neutral pH diffract X-ray to approximately 3 Å (see Figure 2.2-22), crystals grown in the presence of OG or OTG and $MgCl_2$ diffracted only to approximately 6 Å.

2.2.4.1.3   Crystals in the Presence of Alkylglycosides

Crystals obtained in the presence of OG, NG, or OTG and $CaCl_2$ showed the highest resolution. Crystals appeared in two different macroscopic forms from the same crystallisation condition, one cubical and one tetrahedral (see Figure 2.2-22). However the crystal quality was low. Diffraction patterns showed a high mosaicity, smeared, and split spots (Figure 2.2-22). Therefore they were not useful for X-ray structure determination. In order to improve the crystal quality different approaches were followed: (i) screening of reservoir solution using different precipitant and salt concentrations (ii)

variation of pH, temperature, protein concentration (iii) addition of additives (iv) and different crystallisation techniques.
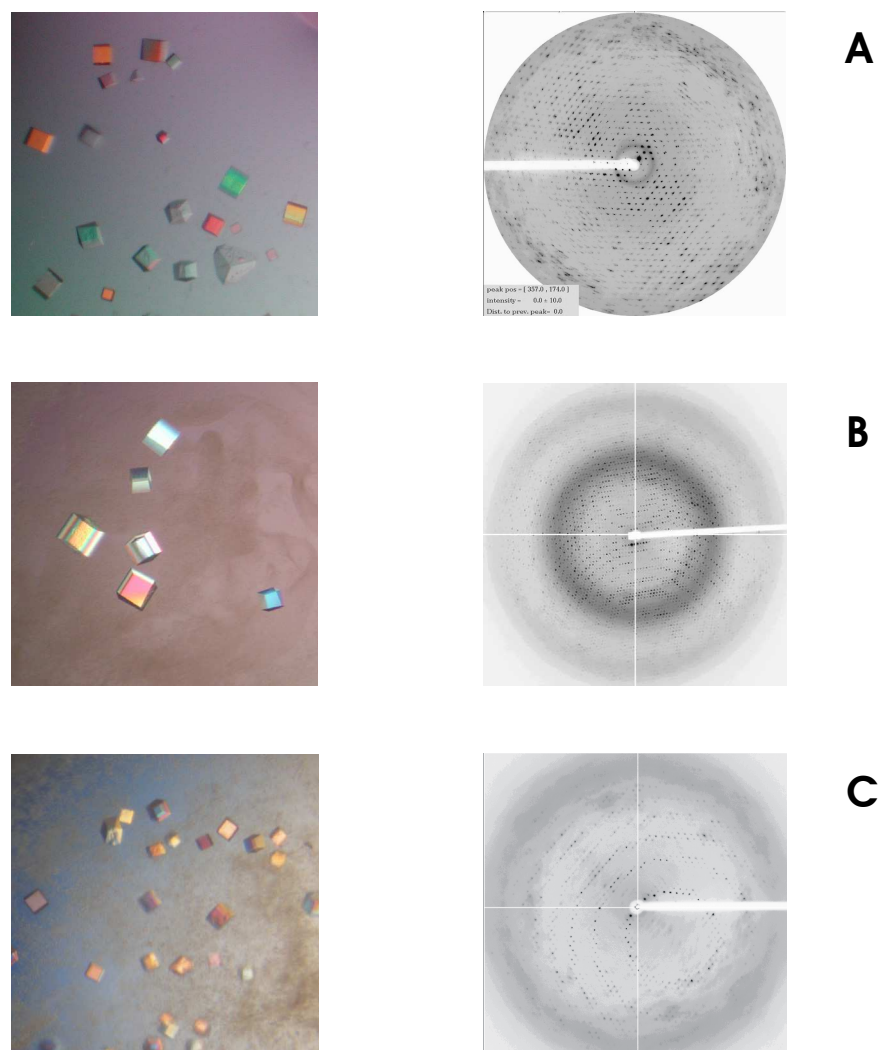

Figure 2.2-22: Crystals and corresponding diffraction pattern (0.5° rotation) in the presence of OG. A: Crystals from the initial sparse matrix screen. Diffraction pattern of a crystal soaked in crystallisation buffer supplemented with 10 % glycerol for cryo-protection; X-ray diffraction resolution of ≈ 3 Å, but high mosaicity. B and C: Improved crystallisation conditions (B: pH 7.5 and C: pH 4.6), no additional supplementation of cryo-protectants; X-ray diffraction resolution of ≈ 2 Å.

The pH of the reservoir buffer was varied within a large range, between pH 4.0 and pH 10.5. Crystals were obtained at all pH values; diffraction reached 3 Å or better from Na-acetate (pH 4.0 to pH 5.0), Na-HEPES (pH 7.0 to pH 8.0) and Tris-Cl (pH 8.5 to pH 9.5) buffered condition. Crystals appeared using $CaCl_2$ concentrations between 0.05 M and 0.25 M, however, diffraction was optimal between 0.18 M and 0.22 M. Crystallisation was carried out at different temperatures. Below 10 °C phase separation appeared which abolished crystals growth, above 25 °C crystals reached larger dimensions, however with lower quality, therefore all further experiments were carried out at 18 °C. The protein concentration was kept between 6 mg/ml and 8 mg/ml. For each batch of protein the optimal PEG, salt

and protein concentration for each pH values had to be determined by a grid screen (20 % to 30 % PEG 400, 0.18 M to 0.22 M CaCl$_2$).

A variety of additives were tested, including salts, small amphiphilic molecules, and detergents. Only the addition of hexanetriol and heptanetriol as additive improved the crystal quality.

In order to permit X-ray data collection in a nitrogen stream different cryo-protectants were tested (glycerol, PEG 400, sucrose, glucose). However, soaking the crystal in any kind of buffer drastically reduced the crystal quality and the diffraction. In addition, crystallisation buffer containing 25 % PEG 400 was a sufficient cryo-protectant however crystals frozen within the standard crystallisation buffer show water rings in some cases.

All screens were performed using the hanging drop vapour diffusion technique. Sitting drops generally yielded larger crystals thus the optimised conditions were carried out using this method. Crystallisation in batch under oil (see above) did not improve the crystal quality and crystals remained rather small.

A typical crystal grown in the presence of OG appears within 2 to 3 days and reaches a size of 0.4 x 0.4 x 0.4 mm within one week (Figure 2.2-22). An X-ray resolution of approximately 2 Å was obtained. Statistics for data collection are summarised in Table 2.2-5. Some crystals appeared in a larger unit cell with a longer c-axis of 512 Å.

Table 2.2-5: Data Collection Statistics. $R_{sym}=\Sigma(I-<I>)/ \Sigma I$, where I is the intensity for a given reflection and <I> is the average intensity for multiple measurements of this reflection Highest resolution shell is shown in parentheses

Space group	R3
Cell dimensions	a=b=108.8
	c=267.7
	$\alpha=\beta=90°$
	$\gamma=120°$
Resolution range	30 - 1.95 Å
Number of reflections	
total	177676 (12902)
unique	61994 (4436)
Completeness (%)	98.9 (95.2)
I/σI	10.0 (5.86)
R$_{sym}$ (%)	6.9 (18.8)

Two approaches for structure determination were followed: molecular replacement using a structurally related protein as search model and phase determination by anomalous dispersion experiments in the presence of a heavy atom compound. All proteins that showed similarity of the tertiary structure using the fold recognition programs were used as search model for molecular replacement using the program EPMR (Kissinger *et al.*, 1999). None of the model revealed a reasonable solution: the symmetry related molecules penetrate each other in the xy plane; in the z plane the molecules are far apart that no crystal contact is possible.

Determination of the phases by heavy atom derivatives was attempted by soaking and co-crystallisation. The crystals were soaked with buffer containing heavy metal compounds (Pt, U, Pb, Hg, Os, Lu, Sm, Au, Ag, W, Gd, Yb, Dy, Pr, Eu, Zn, Ir, La) in different concentrations: 0.5 mM to 50 mM depending on the solubility and the sensitivity of the crystal. Crystals in the small unit cell were preferred for anomalous dispersion experiments because this crystal forms contains presumable fewer molecules in the asymmetric unit.

In addition to soaking experiments co-crystallisation experiments in the presence of heavy metal compounds were attempted. The protein was crystallised under standard condition (pH 7.5 and pH 4.6) with addition of 0.2 mM to 50 mM of the respective heavy metal compound. Some derivatives completely impeded crystal growth or the shape of the crystal was changed to needle or drop-like form. In these cases X-ray data collection was not possible. In general, most heavy atom derivatives improved crystal quality. The diffraction was improved to 2 Å and mosaicity was lowered. At neutral pH six different Pt derivatives, two U, two W, and two Au were tested, all of them induce crystals in the long cell dimension. Dy, Pr, Gd, Yb, and Pb at neutral pH and Pt, Au, Hg and Pb at pH 4.6 induced only crystals in the short unit cell. However neither heavy atom co-crystallised nor soaked crystals showed an anomalous signal suitable for structure determination.

Table 2.2-6: Heavy atom derivatives that have been used phase determination. Both soaking (soak) and co-crystallisation (co) have been tried. Column "Cry" indicated when crystals appeared and "data" if a data set was collected.

Heavy atom	Derivative	soak	co	Cry	data
Ag	Ag acetate				
	Ag(NO3)				
Lu	LuCl$_3$	+	+	+	
W	Na$_2$WO$_4$	+	+	+	
	(NH$_4$)$_2$WS$_4$	+	+	+	+
Ir	(NH$_4$)$_3$IrCl$_6$	+			
	K$_3$Cl$_6$Ir (III)	+	+	-	
	K$_3$Br$_6$Ir (III)		+	-	
	IrCl$_3$	+			
Pt	(NH$_4$)$_2$PtCl$_4$ (II)	+	+	+	
	K$_2$Pt(CN)$_4$ (II)	+	+	+	
	K$_2$Pt(CN)$_6$ (IV)		+	+	
	K$_2$PtCl$_6$ (IV)		+	+	
	K$_2$Pt(NO$_2$)$_4$ (II)	+	+	+	
	[Pt$_2$I$_2$(H$_2$NCH$_2$CH$_2$NH$_2$)$_2$](NO$_3$)$_2$ (II)		+	+	
	Chloro-terpyridinePt(II)PbCl$_2$		+	-	
	K$_2$PtCl$_4$	+	+	+	+
Au	KCl$_4$Au(III)		+	+	
	KAu(CN)$_2$(I)	+	+	+	
Hg	Na(ethylmercuricthiosalicylat)	+			+
Pb	PbCl$_2$	+	+	+	
	Pb acetate		+	+	
Lanthanoides	PrCl$_3$		+	+	
	SmCl$_3$		+	+	+

	Sm acetate		+	+	
	EuCl$_3$		+	+	+
	GdCl$_3$		+	+	+
	DyCl$_3$		+	+	
	YbCl$_3$		+	+	
U	UO$_2$(acetate)$_2$	+			
	(NH$_4$)$_2$U$_2$O$_7$	+	+		
	UO$_2$(NO$_3$)$_2$	+	+	-	+
	UCl$_3$		+	+	

In order to obtain crystals in an alternative crystal form i.e. a different space group, conditions that showed microcystals in the initial standard crystallisation screen were subjected to a broad screen to optimise the crystal quality. However so far no conditions could be discovered suitable for X-ray data collection.

2.2.4.2 Refolded Protein

Since phasing of the AQ_1862 crystals derived from the native source failed, crystallisation and subsequent phasing using heterologously produced protein containing seleno-methionine was attempted. Heterologous production, refolding and purification was successful (see chapter 2.2.2.2), however a batch to batch variation of the CD signal was observed (chapter 2.2.3.1.5) suggesting incomplete folding.

Initially the refolded protein was applied to crystallisation attempts in LDAO, the same detergent in which it was refolded. Needle like crystals were observed in various conditions after 4 to 8 weeks using protein concentrations between 5 and 15 mg/ml. The crystals were very small and fragile. At the same time they were difficult to reproduce. Therefore, the protein was subjected to a wider crystallisation screen. Different detergents and more conditions were tried. Using DDM as detergent, crystals were obtained in conditions similar to the native DDM condition. With 15 to 25 % PEG 400, 0.1 M CaCl$_2$ or Zn-acetate at pH 4.6 crystals were obtained with dimensions up to 0.5 x 0.1 x 0.1 mm. Crystals grown in the presence of CaCl$_2$ diffracted X-rays to 5 Å but nonisotropically (Figure 2.2-23). Crystals obtained in the presence of Zn-acetate diffracted to 10 Å. Interestingly crystals appeared even if the CD signal was approximately 10 % lower than the signal of the native protein, with comparable quality to those obtained from protein with native CD signal.
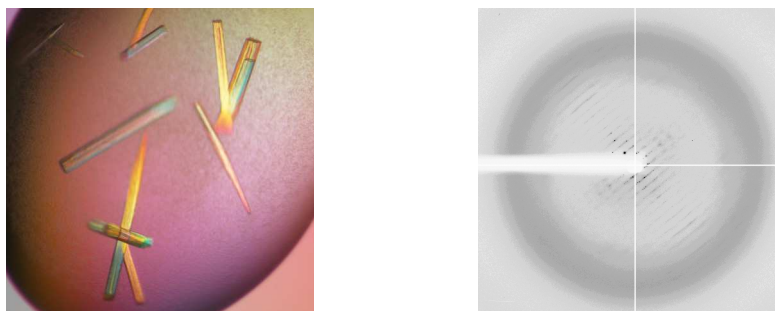
Figure 2.2-23: Crystals of refolded AQ_1862 and corresponding diffraction pattern (0.5° rotation) in the presence of DDM. Crystals grown in the presence of CaCl$_2$. Diffraction pattern of a crystal soaked in crystallisation buffer supplemented with 10 % glycerol for cryo-protection; X-ray diffraction resolution of $\approx$ 5 Å, but anisotropic.

In order to improve the diffraction the quality of the refolded protein has to be improved and a wider screen of detergents and additives has to be applied.

2.3 AQ_1558

2.3.1 Similarity

AQ_1558 (Swiss-Prot # O67508) has a molecular weight of 43.05 kDa (365 amino acids). It comprises two predicted transmembrane helices near the N-terminus (from amino acid 7 to 24 and 39 to 61). SMART and SignalP (Schultz *et al.*, 1998) but not Interpro (Zdobnov *et al.*, 2001) identified the first TMH as a signal peptide, however the probability of the peptide to be cleaved was rather ambiguous (SignalP; p = 0.5). The N- and C-terminus are predicted to be cytosolic.

Conserved domain search show a coiled coil region between amino acid 138 and 202 and a region of low complexity between amino acid 292 and 304. From amino acid 135 to 187 Interpro Scan predicted a tetratricopeptide (TRP)-like domain (IPR011990). From amino acid 120 to 180 the sequence is similar to the COG2956 domain (Predicted N-acetylglucosaminyl transferase, E-value 0.001) a domain whose function is not well studied.

Sequence alignment revealed similarities between AQ_1558 and proteins annotated in the database but with low likelihood and the result varied drastically among the programs used. With the BLAST program (Altschul *et al.*, 1997) the result is dependent on the setting of the "compositional adjustments". This adjustment compensates for the different amino acid composition of the compared sequence to yield more accurate E-values. The standard NCBI BLAST setting "composition-based statistics" revealed only very low similarity to a band 7 protein from *Anabaena variabilis* (E = 0.46) and to several response regulator proteins e.g. from *Pseudomonas syringae* (E = 0.49): 49 % homologous and 28 % identical for 170 aligned residues and 57 % homologous and 28 % identical for 66 aligned residues. Using the NCBI BLAST program in the "not adjusted" mode or the ExPASy BLAST program in the general mode, evidence for similarity of the central domain ($\approx$ 80 to 210) can be detected. It is 29 % identical and 54 % similar for 103 aligned residues to a "hypothetical protein" from *Plasmodium falciparum* (Swiss-Prot # Q6LF44, E = 0.023) and to several bassoon proteins from

different mammals including *Rattus norvegicus* (Swiss-Prot # O88778) with 31% identical and 51 % similar for 113 aligned residues (E = 0.023) ; *Mus musculus* (Swiss-Prot # O88737) and *Homo sapiens* (Swiss-Prot # Q9UPA5)). In the aligned region the bassoon protein contains a coiled coil motif and a region of low complexity which might be irrelevant for the protein's function. The C-terminal domain is similar to DNA repair proteins (e.g. 62 % homologous and 48 % identical for 26 aligned residues to the DNA repair protein RadA from *Synechococcus* sp. (Q2JXD1_SYNJA)).

PSI-BLAST and PHI-BLAST did not reveal similarity to more proteins. AQ_1558 possesses two CXXC motifs near the C-terminus, characteristic for a variety of functions including metal binding. This motif is also conserved in the DNA repair protein RadA.

327-W E **C** E E **C** G K E Y N K Y T S V **C** R N **C** L G V-349

Figure 2.3-1: Sequence of AQ_1558, amino acids 327 to 349. Two CXXC motifs are highlighted.

Tertiary structure prediction shows only very low homology (E >1) to mammalian farnesyl- and geranyltransferases as well as to the vesicular transport protein Sec 17 from *Saccharomyces cerevisiae*. AQ_1558 shows no clear homologys to proteins in other bacteria and might be an *A. aeolicus* specific protein.

The protein was chosen as a target protein because its function is entirely unknown, and the low similarity of the sequence and the fold to other proteins can only give rise to speculations. In the membrane it is present in the several hundred microgram range per preparation.

2.3.2 Purification

2.3.2.1 *Native*

The protein is present in the fractions 40 to 50 of the MonoQ elution (Figure 2.1-4); however it is not highly abundant. It is among the most intense band in the SDS PAGE gel after gel filtration chromatography. Further purification was achieved by an additional anion exchange chromatography using MonoQ resin at pH 6.0. The yield was very low, only approximately 0.2 mg per purification, too little for broad screening of crystallisation conditions. In addition a large batch to batch variation of the amount was detected. Therefore heterologous production in *E. coli* was undertaken.

2.3.2.2 *Heterologous Production*

The gene encoding for the protein AQ_1558 was amplified from genomic *A. aeolicus* DNA by PCR using the oligonucleotide primers AQ1558.1.For and AQ1558.Rev (Table 4.1-5). It was cloned in the vectors pTTQ18 and pBAD in the modified version A and C were utilised (Surade *et al.*, 2006).

Western blot analysis of four clones for each construct monitored protein production in small scale. A band of approximately 40 kDa in whole cell lysate using all the different constructs indicated protein

production. The highest production levels were obtained with pTTQ18-A and pBAD-C. In pBAD-C both N-terminal deca-histidine tag and C-terminal strep-tag II can be detected, suggesting that the hydrophobic domain near the N-terminus is indeed a TMH and not a cleaved signal peptide.

A small scale membrane preparation from cells grown in a 50 ml culture showed the membrane localisation of the Western blot signal (Figure 2.3-2 A). The signal in the insoluble pellet fraction from centrifugation at 10000 g can result from non-quantitative cell disruption or inclusion body formation.
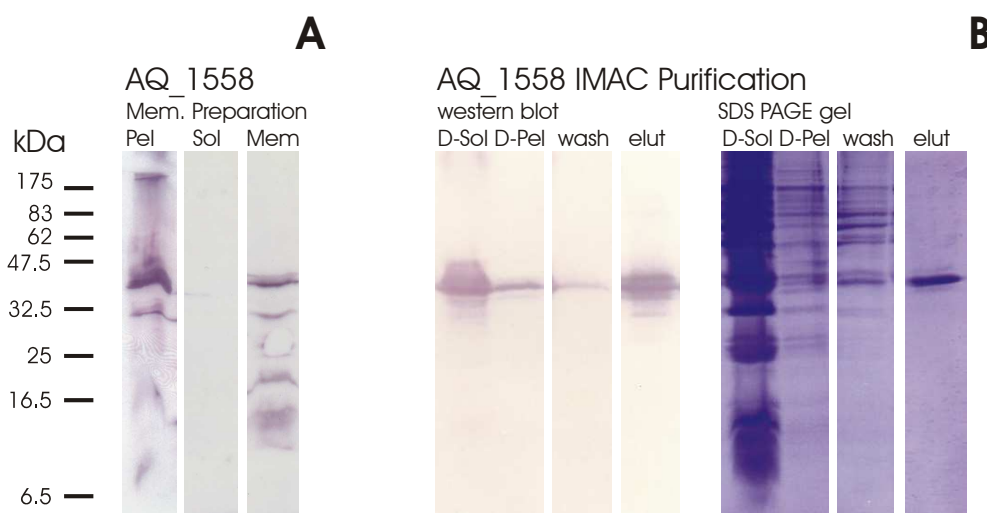
Figure 2.3-2: Western blots and Coomassie stained SDS PAGE gel of a membrane preparation and a purification procedure. A: Western blot (Anti-His) of a membrane preparation from *E. coli* BL21(DE3) containing pTTQ18-A-1558. Pel: pellet from 10000 g centrifugation, Sol and Mem: soluble and pellet fraction from 100000 g centrifugation step, respectively. B: Western blot and Coomassie stained SDS PAGE gel from IMAC purification of AQ_1558. D-Sol and D-Pel: soluble and pellet fraction of DDM solubilised membrane fraction after 100000 g centrifugation; wash and elut.: washing with 200 mM imidazole and elution with a gradient from 200 mM to 500 mM imidazole fractions from HisTrap[TM].

For preparative protein production pTTQ18-A and pBAD-C were tested. Both showed comparable protein production levels. The highest yield was obtained when production was induced in mid-log phase ($OD_{600nm} \approx 0.5$) with 1 mM IPTG (for pTTQ18) or with 0.02 % (w/v) arabinose (for pBAD). The cells were harvested by centrifugation after 4 hours at 37 °C or after overnight growth at 20 °C. The yield of cells per litre culture increased from approximately 4 g wet weight per litre in lysogeny broth (LB) to 6.0 g wet weight per litre in terrific broth (TB). The production level judged by Western blot signal remained similar.

Initial purification was carried out with the protein product from pTTQ18-A. A variety of detergents were tested for their ability to solubilise AQ_1558 from the membrane (DM, DDM, NG, OG LDAO and FOS 12), all were able to do so. The amount of solubilised protein was only slightly affected by pH and salt concentration and best results were obtained, using DDM, pH 8.5 and 300 mM NaCl. Since the protein contains a deca-his-tag at its N-terminus, purification was achieved by immobilized metal-ion affinity chromatography (IMAC) using a HisTrap[TM] column equilibrated with 50 mM Tris-Cl pH 8.5, 300 mM NaCl and 0.03 % DDM. The protein shows a single band in SDS PAGE and Western blot. Washing with two steps of imidazole buffer (150 mM and 200 mM imidazole) removed

most impurities and the protein eluted on a gradient from 300 mM to 500 mM imidazole (Figure 2.3-2 B).

Since no activity assay is known for AQ_1558 to show its active state the elution volume in gel filtration chromatography compared to the native protein was taken as measure for its native state. The protein eluted in two peaks, one with the same elution volume as the native protein (Figure 2.3-3, "native" and "initial"), the other with smaller elution volume, most likely containing an aggregated form of the protein (Figure 2.3-3, "initial"). Once the correctly folded protein was separated from the aggregated protein, almost no additional aggregation was evident, even after storage for one week at room temperature (Figure 2.3-3, "one week"). Overnight storage or dialysis of the IMAC eluted protein increased the proportion of aggregated protein. Hence, gel filtration chromatography was carried out subsequent to the IMAC step. The tailing edge of the IMAC elution peak contained most of the aggregated protein and was discarded. Intensive screening of production, solubilisation, and purification conditions decreased but did not completely remove the aggregated protein peak. The protein was most stable at pH values above pH 7.5; it is precipitating at pH values below pH 6.5. The addition of β-mercaptoethanol to the solubilisation buffer had a positive effect on the aggregation while different ionic strength or glycerol did not improve the stability. The temperature after induction of the cells had the greatest influence on the aggregation behaviour: keeping the cells at 25 to 30 °C drastically decreased the amount of aggregated protein. However the yield was very low, only 0.1 to 0.2 mg protein per litre of *E. coli* culture were obtained after gel filtration chromatography.

Affinity tags, especially histidine tags, are very helpful in protein purification; however, due to their charge density and flexibility they can complicate or even abolish protein crystallisation. Therefore the expression vectors contain a TEV cleavage site between his-tag and the protein sequence in order to remove the his-tag. TEV digestion of the protein almost quantitatively yielded a pure and non-aggregated protein preparation (Figure 2.3-3, "TEV cut").

Protein production using the pTTQ18-A construct showed large batch to batch variation especially at low production temperatures; the reason for this observation is unclear. The pBAD-C construct on the other hand showed a stable production level. Solubilisation and purification were similar to protein produced in the pTTQ18-A vector, but protein from the pBAD-C vector was not as aggregation prone (Figure 2.3-3, "pBAD-C"). TEV digestion of the N-terminal deca-histidine tag remained unsuccessful in this construct. Possibly the N-terminus of the protein is not accessible for the TEV protease while the C-terminus is freely accessible. Despite this complication, the pBAD-C-construct was used for all further studies. Approximately 0.4 mg of non-aggregated pure protein was obtained per litre of culture.

The choice of detergent is crucial to keep a membrane protein in its native state and to obtain well ordered three-dimensional crystals suitable for structure determination by X-ray crystallography. Hence, variety of detergents were screened in protein solubilisation and purification including DM, n-tetradecyl-β-D-maltoside ($C_{14}M$), Fos-12, Fos-16, OG, $C_{10}E_5$ and LDAO. In the presence of Fos-12 the amount of AQ_1558 solubilised from the membrane was higher than with DDM but most of the

protein was aggregated. Purification in the presence of DM, $C_{14}M$, and LDAO was similar to the purification in DDM. However, in the presence of LDAO the protein aggregated completely within one week. Exchanging the detergent to OG, $C_{10}E_5$ or Fos-16 substantially reduced the amount of the protein binding to the HisTrap^{TM} column, and the resulting preparations showed precipitation or aggregation. Only in the presence of the maltoside detergents DM, DDM and $C_{14}M$ the proteins was stable and did not aggregate.
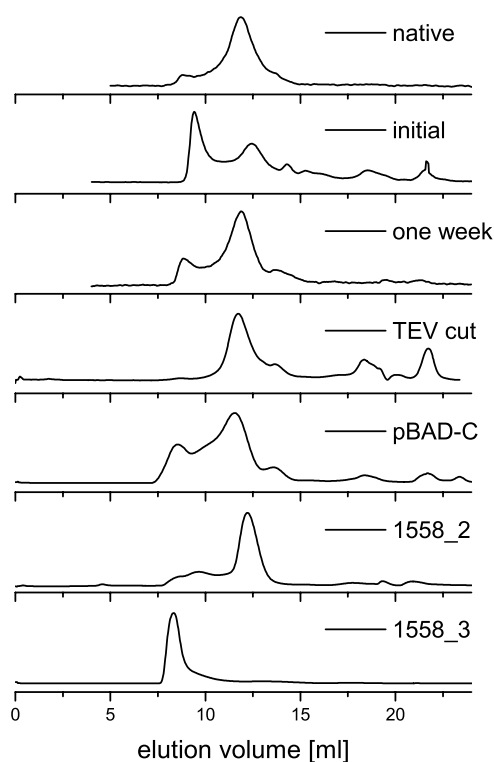
Figure 2.3-3: Gel filtration chromatograms using a Superose 12 column. Native protein, initial purification using the A construct, protein from non-aggregated peak from the initial gel filtration kept for 7 days at room temperature, protein from non-aggregated peak from the initial gel filtration TEV digestion and cleaved protein purified by Ni-NTA, purification using the pBAD-C construct, 1558_2 in the pBAD-C construct, 1558_3 in the C construct. Buffer: 20 mM Tris-Cl, 150 mM NaCl, 0.03 % DDM.

The protein's two TMHs close two the N-terminus suggest that the protein consist of two domains: a membrane anchor and a soluble domain. For anchored membrane proteins crystallisation is most promising when the membrane anchor is removed by a protease or genetically. Therefore, N-terminally truncated versions were constructed, where gene translation starts behind the last predicated transmembrane helix of the full-length construct. The first construct, AQ_1558_1, starts at position 67, five amino acids behind the second predicted transmembrane helix, the constructs AQ_1558_ 2, AQ_1558_3 and AQ_1558_4 start at positions 85, 101 and 115, respectively. The oligonucleotide primer AQ1558.67.For, AQ1558.85.For, AQ1558.101.For and AQ1558.115.For were used in combination with AQ1558.Rev (Table 4.1-5). According to the program TMHMM, the C-terminus is located in the cytosol; therefore cytosolic location of the product was probed. All constructs were

produced in high amounts judged by Western blot analysis: AQ_1558_2 and AQ_1558_3 produced the highest yield.
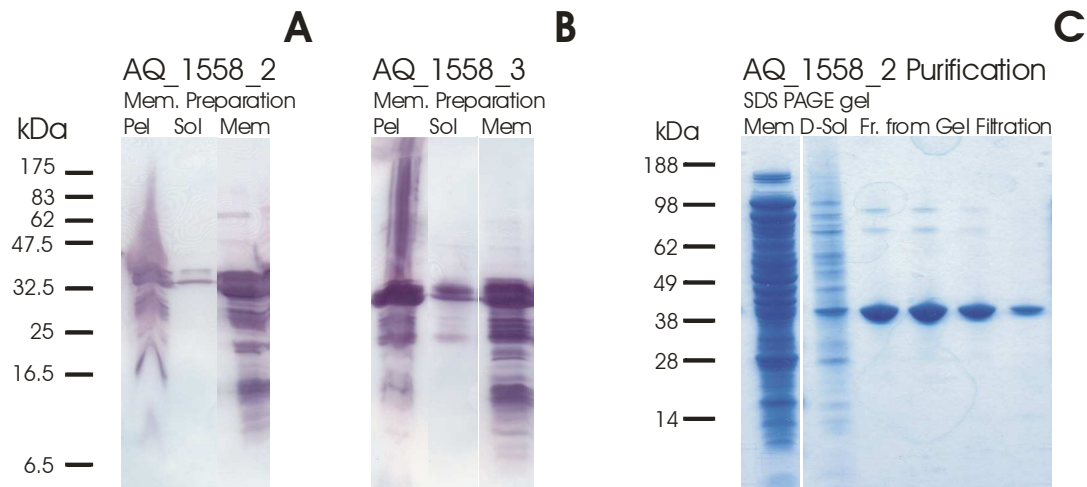

Figure 2.3-4: Western blots and Coomassie stained SDS PAGE gel of a membrane preparation and a purification procedure. A and B: Western blot of a membrane preparation from *E. coli* BL21(DE3) containing pBAD-C-1558_2 and pBAD-C-1558_3, respectively. Pel: pellet from 10000 g centrifugation, Sol and Mem: soluble and pellet fraction from 100000 g centrifugation step, respectively. B: Western blot and Coomassie stained SDS PAGE gel from IMAC purification of AQ_1558. D-Sol and D-Pel: soluble and pellet fraction of DDM solubilised membrane fraction after 100000 g centrifugation; wash and elut.: washing with 100 mM imidazole and elution with a gradient from 100 mM to 500 mM imidazole fractions from HisTrap^{TM}.

Since production and purification of the full length protein was most successful using the pBAD-C construct, the truncated versions AQ_1558_2 and AQ_1558_3 were also produced in this vector. Production was carried out similar to the full length construct the cells were induced by 0.02 % (w/v) arabinose in the mid-log phase and grown at 20 °C over night.

Both AQ_1558_2 and AQ_1558_3 were distributed in all cellular fractions: membrane, cytosol, and insoluble fraction; however, the strongest signal was detected in the membrane fraction (Figure 2.3-4 A and B), although no predicted TMH is present in these constructs anymore.

Purification of the fraction present in the soluble cytosolic fraction was attempted by IMAC using a HisTrap^{TM} column, no binding was achieved. As the majority of the protein was present in the membrane fraction, extraction and purification from this fraction was attempted. Neither protein constructs could not be removed from the membrane by washing with buffer containing a low amount of detergent (0.03 % DDM) but could be by solubilisation using high amounts of detergent (2 % (w/v) DDM). Both were successfully purified by means of IMAC in the presence of 0.03 % DDM under the conditions optimised for the full length construct. In gel filtration chromatography using a Superose 12 column, AQ_1558_3 eluted in the void volume of the column (Figure 2.3-3, "1558_3"). It was not possible to establish purification conditions for 1558_3 that result in non-aggregated protein. AQ_1558_2 however showed a peak at similar elution volume to that of the native, full length protein (Figure 2.3-3, "1558_2"), though it was not stable to aggregation without the addition of detergent above CMC concentrations. Per litre of culture 1.5 to 2 mg of AQ_1558_2 were obtained. TEV

digestion removed the C-terminal deca-histidine tag of the protein almost quantitatively in the standard TEV buffer at 37°C over night.

Heterologous production and optimisation of purification conditions yielded stable and homogenous protein AQ_1558 in full length and truncated form.

2.3.3 Functional Characterisation

2.3.3.1 Electrophoretic Behaviour

The native protein has a predicted mass of 43 kDa, truncated version (AQ_1558_2) of 32.9 kDa. The produced proteins in the C-version contain additional 37 amino acids from the N- and C-terminal tags, resulting in a molecular weight of 47.4 kDa for the full-length protein and 37.3 kDa for the truncated version.

In a mass spectrum of AQ_1558 and AQ_1558_2, the predicted molecular masses were confirmed (Figure 2.3-5), however, the observed masses are 0.6 kDa and 0.4 kDa larger than predicted, respectively. This shift might be due to a partial formylation of the protein during preparation of the sample for mass analysis.
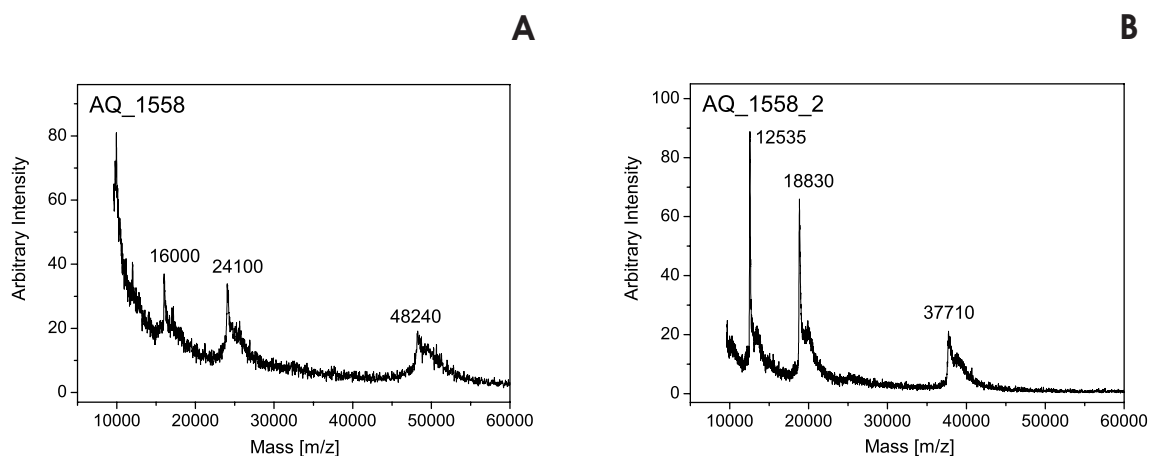
Figure 2.3-5: Mass spectra of AQ_1558 and the truncated version AQ_1558_2. AQ_1558 has an apparent mass of 48.2 kDa, the truncated protein of 37.7 kDa, the peaks at 24.1 kDa, 16.0 kDa, 18.8 kDa and 12.5 kDa are the doubly and triply charged protein, respectively.

In SDS PAGE analysis the full-length and the truncated versions of protein AQ_1558 have an apparent molecular mass of approximately 45 kDa and 38 kDa: the size of the respective protein monomer. A band at higher molecular mass, indicative of a protein dimer, was detected in both cases (Figure 2.3-6) and in some cases even a putative trimer is observed. It was shown by Western blot analysis and mass spectrometry that the band consists exclusively of the protein AQ_1558 (data not shown). The bands of the oligomers appear even if the sample is incubated at 90 °C in SDS sample buffer. SDS concentration, addition of β-mercaptoethanol, or incubation temperature did not alter the electrophoretic behaviour. In BN PAGE the protein runs in multiple bands, with the lowest mass at approximately 130 kDa (indicated by I). The bands at approximately 250 kDa (indicated by II) and 400 kDa (indicated by III) are putative dimers and trimers of the first band. A substantial amount of

protein was not able to enter the gel and remained in the pocket (indicated by *). The running behaviour in SDS and BN PAGE gels suggest that the protein exists as an oligomer.
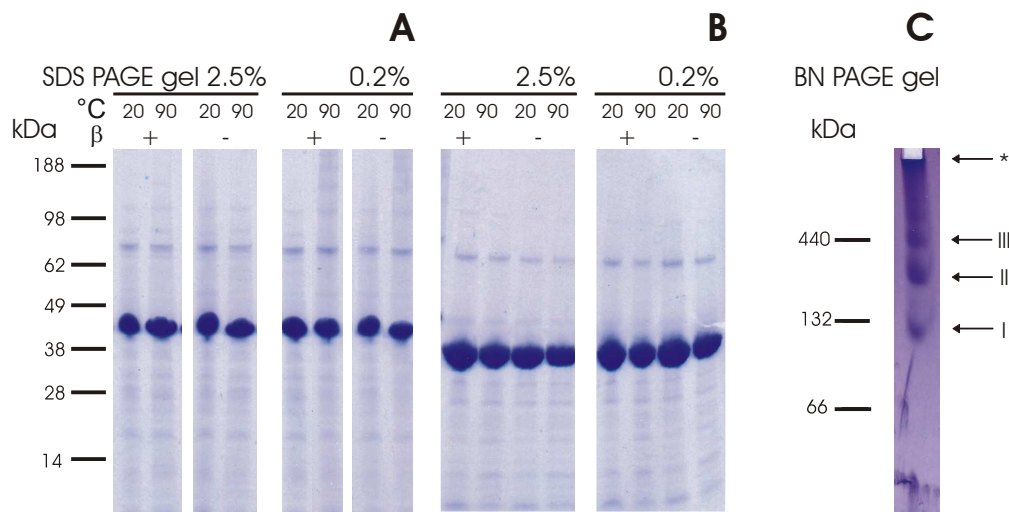

Figure 2.3-6: Coomassie stained SDS PAGE and BN PAGE gels of the purified AQ_1558 protein (left hand side full-length protein, right hand side truncated protein) incubated at different temperatures and different SDS concentrations with or without addition of β-mercaptoethanol (indicated by β).

2.3.3.2 *Size Estimation by Gel Filtration Chromatography*

Analytical gel filtration chromatography was used to investigate the oligomeric state of the two proteins further. Figure 2.3-7 shows that native and produced versions (Figure 2.3-7, "native" and "1558_C") of AQ_1558 have an identical elution volume in gel filtration chromatography, 1.19 ml. The truncated version has an elution volume of 1.22 ml (Figure 2.3-7, "1558_2C"). Calibration of the column with the indicated soluble protein standards was used to determine the oligomeric state of the protein. The elution volumes correspond to molecular weights of approximately 270 kDa, for native and produced protein and 205 kDa for the truncated version.
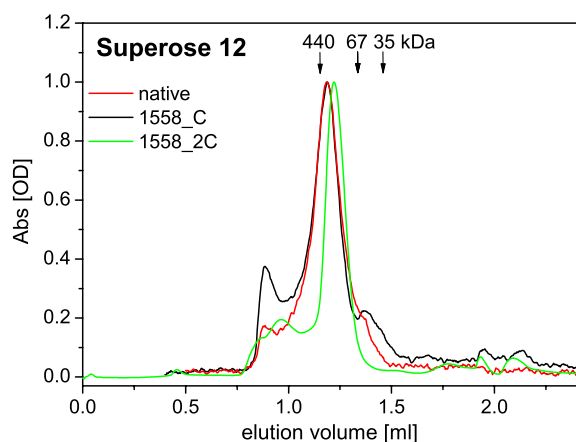

Figure 2.3-7: Superose 12 gel filtration chromatogram. The elution volumes of the soluble standard proteins are indicated with an arrow and the molecular weight of the respective protein.

The data would agree with a trimeric or tetrameric form of the full-length protein with the addition of 130 kDa or 80 kDa of detergent, respectively (3 x 47.4 kDa = 142.2 kDa protein + 127.8 kDa = 3 x 42.6 kDa detergent or 4 x 47.4 kDa = 189.6 kDa protein + 80.4 kDa = 4 x 20.1 kDa detergent). The truncated version does not contain the predicted transmembrane domains anymore and thus it is questionable whether it is covered by a detergent micelle. However, the observed molecular mass would agree with a trimeric or terameric form of the protein as well with the addition of 90 kDa or 56 kDa of detergent, respectively (3 x 37.2 kDa = 111.6 kDa protein + 93.4 kDa = 3 x 31.1 kDa detergent or 4 x 37.2 kDa = 148.8 kDa protein + 56.2 kDa = 4 x 14.1 kDa detergent) or without addition of detergent, a pentameric form.

2.3.3.3 Cross-linking Experiments

In order to further investigate the oligomeric state of the protein, cross-linking experiments using glutaraldehyde were performed. The bands at 80 kDa and 120 kDa, indicative for dimer and trimer are present in the cross-linked sample and in the control (-), however in the glutaraldehyde-treated sample the bands are shifted, perhaps due to internal cross-linking of the protein (Figure 2.3-8). An additional band (indicated by IV) at higher molecular weight appeared suggesting a tetramer, however a substantial amount of cross-linked protein is not able to enter the gel (indicated by *), most likely because of aggregation. The band of the monomeric protein has nearly disappeared suggesting an almost quantitative cross-linking reaction. Similar results have been obtained for the truncated protein.
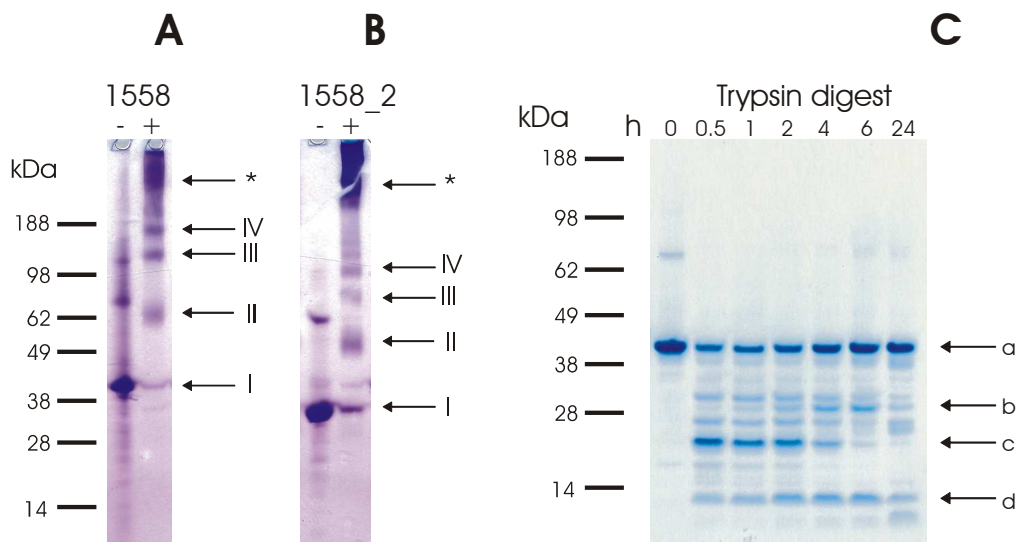

Figure 2.3-8: Coomassie stained SDS PAGE gels showing the results from cross-linking and limited proteolysis experiments. A and B: 0.5 mg/ml of the full-length (A) and the truncated (B) protein (in DDM) are mixed with 0.04 % glutaraldehyde. After six hours the samples are boiled in SDS sample buffer for 3 min. The monomeric form of the protein is indicated with I, the putative dimeric, trimeric and tetrameric form with II, III and IV respectively. Aggregated protein which was not or not able to enter the gel is indicated by *. C: AQ_1558 incubated with trypsin (20 to 1, w/w) in 50 mM Tris-Cl pH 8.0, 1 mM CaCl2, 0.03 % (w/v) DDM. Samples were taken after the indicated time and incubated at 90 °C in SDS sample buffer in order to inactivate the protease. The full-length protein is indicated by a, the major digestion products by c to d.

2.3.3.4   Limited Proteolysis

The stability of the protein was investigated by limited proteolysis using trypsin. By this means flexible regions can be removed or single domains detected and subsequently produced. This procedure can yield rigid and stable fragments or domains and facilitate protein crystallisation. After the indicated time of digestion, samples were taken and boiled in SDS sample buffer in order to deactivate the protease. Already after 30 minutes several digestion products in different concentrations occur. However, the band of the full-length protein remains almost unchanged in size and intensity, no shift in molecular weight compared to the control becomes obvious. Three major products indicated by b to c are observed, all of them were identified as AQ_1558 by mass spectrometry. However it was not possible to determine the cleavage sites of the degradation products by mass spectrometry. The protein contains 54 possible trypsin digestion sites leaving a large variety of possible constructs. None of the degradation products is still present at high concentration after 24 hours of digestion, only the full-length protein shows an intense signal. Obviously only a subpopulation of the protein is degraded while the remaining protein is stable. The band of the dimeric form is weaker after digestion, suggesting that mainly the oligomeric form was cleaved. The full-length monomeric form of AQ_1558 appears to be the most stable form of the protein.

2.3.3.5   Lipid Content

The co-purified lipids from the different preparations were identified by chloroform-methanol extraction and subsequent HPLC analysis and compared to *E. coli* membranes and standard lipids (Figure 2.3-9). The strongest peak observed in the HPLC chromatogram of the full-length construct (Figure 2.3-9, "AQ_1558") has an elution volume of 14 ml and corresponds to maltoside detergents. All other peaks resemble the lipid content of the *E. coli* membrane. The second prominent peak has an elution volume of 11.5 ml suggesting that it consist of phosphatidylethanolamine (PE). The double peak at 8.5 ml most likely consists of cardiolipin (CL) and phosphatidylglycerol (PG). Comparison to the lipid standards revealed the presence of 6.6 molecules of PE, 0.8 molecules of CL and 2.6 molecules of PG per molecule of full length protein.

The lipid content of the truncated version AQ_1558_2 resembles the lipid content of the full-length construct qualitatively (Figure 2.3-9, "AQ_1558_2"). The strongest peak is indicated as maltoside detergent, and PE, CL and PG are present. However, the amount of bound detergent and lipids is much lower, perhaps due to the missing hydrophobic transmembrane part. Only 0.1 molecules of CL, 0.2 molecules of PG and 0.5 molecules of PE were detected per molecule.
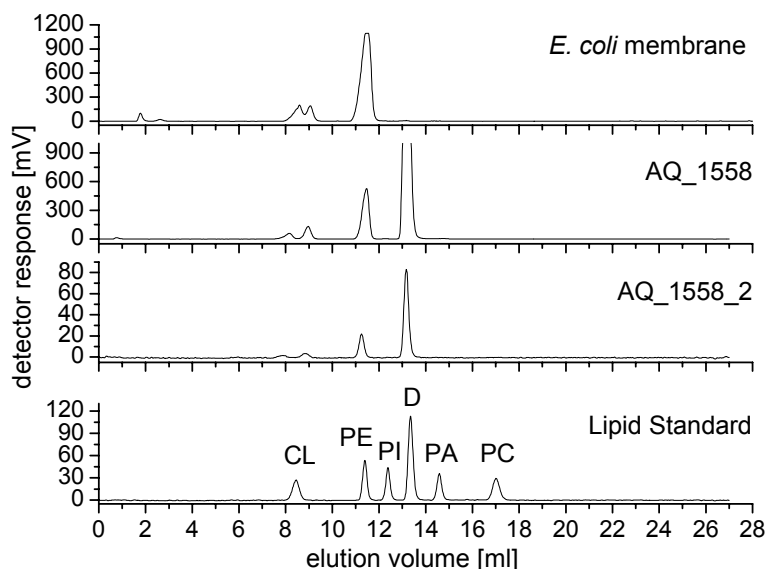
Figure 2.3-9: HPLC chromatogram of *E. coli* membrane, AQ_1558 and AQ_1558_2 after extraction to determination of the lipid content.

2.3.3.6   Cofactor Identification

The protein has two prominent CXXC motifs near its C-terminus. This motif is present in a variety of proteins and can mediate metal binding.

On order to investigate if AQ_1558 indeed contains metal, total reflection X-ray fluorescence spectrometry (TXRF) measurements of the full-length and the truncated version were conducted. As control a protein sample was purified in the presence of 5 mM EDTA and then desalted by concentrating and dilution the sample, in order to minimise the risk of metal contamination. One protein monomer of the full-length protein comprises ten methionines and four cysteines: together fourteen sulphur atoms. The truncated version has seven methionines and 4 cysteines: together ten sulphur atoms. The amount of sulphur detected in the protein sample is in good agreement with the amount expected from the protein concentration, 13.6, 11.3, and 11.6, respectively. Per monomer, approximately one zinc atom was detected, 1.3 and 1.2 in the full-length and the truncated version, respectively and 1.0 if the protein was purified in the presence of EDTA. Since the zinc is also present in the truncated version it is likely that it is bound by the two CXXC motifs.

In addition, 12.4 phosphate atoms per monomer of full-length protein were detected. This is in approximate agreement with the phosphorous content determined by the lipid content of the full-length protein, 10.8 (see chapter 2.3.3.5). The amount of phosphorous in the truncated version was smaller, only 2 per monomer, approximately twice as much as expected from the lipid content per molecule.
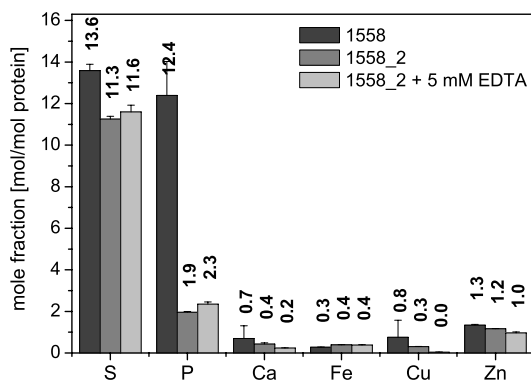
Figure 2.3-10: TXRF measurement of AQ_1558, AQ_1558_2, and AQ_1558_2 purified in the presence of 5 mM EDTA. (20 µM in 20 mM Tris-acetate pH 7.5, 0.03 % DDM)

2.4 AQ_1760

2.4.1 Similarity

AQ_1760 (Swiss-Prot # O67639) has a molecular weight of 31.8 kDa (281 amino acids). According to the program TMHMM it does not contain transmembrane helices and the location is periplasmic. A conserved domain search discovered the Linocin_M18 domain. This domain derives from Linocin_M18 from *Brevibacterium linens* (Swiss-Prot # Q45296), a protein which is frequently detected in the culture supernatant. Sequence alignment revealed similarity to proteins from bacteria and archaea with a sequence identity of up to 25 %. All of these also share the Linocin_M18 domain, although several are categorised as "hypothetical proteins". The sequence is 20 % identical and 39 % similar to Linocin_M18 from *Brevibacterium linens* 23 %, identical and 46 % similar to bacteriocin from *T. maritima* (Swiss-Prot # Q9WZP2) and 25 % identical and 48 % similar to the C-terminal 200 amino acids of the vlp from *P. furious* (Swiss-Prot # Q401P1). The N-terminal part of the vlp from *P. furiosus* shows only weak homology to AQ_1760. PSI-BLAST supported these hits but also revealed homology to the 29 kDa antigen from *M. tuberculosis*.

Tertiary structure prediction revealed similarity to the fold of the vlp from *P. furious* (probability of 100 %) and to the major capsid protein from Bacteriophage HK97 (HHPRED: (probability ≈ 96 %; and Phyre: probability ≈ 75 %). Rather low similarity (probability ≈ 30 %) was detected by HH-PRED to the fold of several cytochrome *c* proteins.

Although it is still annotated as "hypothetical protein" in the databases, sequence similarity provides strong evidence that it belongs to the bacteriocin-Linocin_M18 family.

```
AQ_1760 [Aa]    1    -----nef l qr dqapl t aeewdqi dkt a----------ye-----------vf ks t v-vcr kf npvvgpf----------
Bacter. [Tm]    1    nvn--nef l kr s f apl t ekqwqei dnr a----------r e-----------i f kt ql -ygr kf vdvdgpy----------
Virus-l [Pf]    1    nl s i npt l i nr dk-pyt keel nei l r l ai i ael dai nl yeqnar ys edenvr ki l l dvar eekahvgef nal l l nl dpeq
Linocin [Bl]    1    -----nnnl yr el api pgpawdei eeed-----------r r -----------t f kr ni - agr r i vdvdgpt-----------
CFP29 [Mt]      1    -----nmnl yr dl apvf eaawdei el ed-----------ar -----------t f kr hi - agr r vvdvs dpg-----------
```

```
AQ_1760 [Aa]    44   -----g--- aghqv-------------------vs y------dvl ygvepgvcevk-------pgqeykvc-----epvr
Bacter. [Tm]    47   -----gweyaahpl -----------------gev------evl --------------------s denev--------vk
Virus-l [Pf]    80   vt el kg--- gf eevkel t gi eahi ndnkkees nveyf ekl r s al l dgvnkgr s l l khl pvt r i egqs f r vdi i kf edgvr
Linocin [Bl]    44   -----g--- f et s a-----------------vt t ------ghi r dvqs et s gl q-------vkqr i -----------
CFP29 [Mt]      44   -----g-----------------------------------pvt aavs -------t gr l i dvk-----apt -
```

```
AQ_1760 [Aa]    79   t ger khv---- pvpt l ykdf vi s wr dl ehwr qf nl pvdt t gvaaaas s l avaedkl i l f gnqengi egf l t akg----t l
Bacter. [Tm]    70   wgl r ks l ---- pl i el r at f t l dl wel dnl er gkpnvdl s s l eet vr kvaef edevi f r gceks gvkgl l s f e-------
Virus-l [Pf]    157  vvkqeyk---- pi pl kkkf yvgi r el ------ndgt ydvs i at kagel l vkdees l v-------i r ei l s t eg----i k
Linocin [Bl]    72   --vqeyi ---- el r t --- pf t vt r qai ddvar gs gds dwqpvkdaat t i anaedr ai l hgl daagl ggi vpgs s ----na
CFP29 [Mt]      63   ngvi ahl r as kpl vr l r vpf t l s r nei ddver gs kds dwepvkeaakkl af vedr t i f egys aas i egi r s as s npal t l
```

```
AQ_1760 [Aa]    151  r eel s dwekvgnaf qdvv- kgi s r l vekgf yt nyyl i vnpkr yf l l nr i hdnt gl l el eq- i kkvvk--e- vyqt pi i pe
Bacter. [Tm]    139  er ki ecgs t pkdl l eai v- r al s i f s kdgi egpyt l vi nt dr w--i nf l keeaghypl ekr veecl r ggk- i i t t pr i - e
Virus-l [Pf]    216  knkl s s wdnpeeal ndl n- nal qe- as nas agpf gl i i npkr yakl l ki yeks gknl vev- l kei f r --ggi i vt l ni ni
Linocin [Bl]    139  avai pd-- avedf adava- qal s vl r t vgvdgpys l l l s s aeyt kvs es t dh- gypi r eh- l s r ql gage- i i wapal --
CFP29 [Mt]      143  pedpr e------- i pdvi s qal s el r l agvdgpys vl l s advyt kvs et s dh- gypi r eh- l nr l vdg- d- i i wapai --
```

```
AQ_1760 [Aa]    226  di vl l vs as panf dl ai al dvnvaf vet s nnnht f r vnenvvpr i --kr peai l i f s s -----
Bacter. [Tm]    214  d- al vvs er ggdf kl i l gqdl s i gyedr ekdavr l f i i t et f t s r l s t r r p-------------
Virus-l [Pf]    291  nkvi i f ant pavl dvvvgqdvt l qel gpegddvaf l vs eai gi r i --knpeai vvl e------
Linocin [Bl]    211  egal l vs t r ggdyel hl gqdl s i gy-----ys hds et vel yl qet --f gf l al t des s vpl s l
CFP29 [Mt]      210  dgaf vl t t r ggdf dl ql gt dvai gyas hdt dt vr l yl qet l t f l c--yt aeas val s h-----
```

Figure 2.4-1: Sequence alignments with AQ_1760 from *A. aeolicus* (Swiss-Prot # O67639) using the program Clone Manager 7 with the algorithm BLOSUM62. Bacter. [Tm] is the Bacteriocin from *T. maritima* (Swiss-Prot # Q9WZP2), Virus-l [Pf] is the Virus-like particle from *Pyrococcus furious* (Swiss-Prot # Q401P1), Linocin [Bl] is the Linocin_M18 from *Brevibacterium linens* (Swiss-Prot # Q45296) and CFP29 [Mt] is the 29 kDa antigen from *Mycobacterium tuberculosis* (Swiss-Prot # O07181).

The protein was chosen as a target because its biological function is poorly understood. The functions of its few homologs have been studied with contradicting results, and their biological roles remain unclear. Although it does not contain transmembrane helices homologous proteins have been shown to be associated with the membrane and AQ_1760 is purified from the membrane fraction. It is present in the several hundred microgram range per preparation.

2.4.2   Purification

2.4.2.1   *Native*

The gel filtration profile of the MonoQ fractions 61 to 63 showed a major peak at 8 ml elution volume, which is in the void volume of the column (see P1 in Figure 2.1-8). The major component of this peak is the protein AQ_1760 (Gel F, band 2 in Figure 2.1-8). To further purify the protein different purification strategies were applied: different anion exchange resins and different reactive dye agarose. Most successful was an additional MonoQ step at pH 8.5 with a shallow sodium chloride gradient. This procedure removed most of the additional bands. In some purification an additional band of 35 kDa was observed, belonging to the "hypothetical protein" AQ_331 [Swiss-Prot # O66666], a putative bacterioferritin. The archaeal homologs (e.g. vlp from *P. furiosus* (Namba *et al.*, 2005)) have

an additional ferritin or bacterioferritin-like domain N-terminal to their bacteriocin domain. Proteins homologous to AQ_1760 have been purified from the culture medium, in *A. aeolicus* no traces of AQ_1760 have been found in culture medium or the cytosol; it is exclusively present in the membrane fraction.

As already depicted in chapter 2.4.1, AQ_1760 does not contain any transmembrane segments. The protein remained soluble when the detergent was removed in anion exchange chromatography. The effects of removing the detergent are described in chapter 2.4.3 and 2.4.4. The purification procedure yielded approximately 0.1 to 0.5 mg of protein per purification.

2.4.2.2   *Heterologous Production*

AQ_1760 could be purified from the membrane fraction of *A. aeolicus*, but with low yield. Therefore it was attempted to produce the protein in *E. coli.* Cloning and production was analogous to the procedure described for protein AQ_1558.

The gene encoding for protein AQ_1760 was amplified from genomic *A. aeolicus* DNA by PCR using the oligonucleotide primers AQ1760.For and AQ1760.Rev (Table 4.1-5) and cloned in the pTTQ18 and pBAD vectors. A band at approximately 30 kDa in whole cell lysates indicated protein production of AQ_1760 from both the pTTQ18-C and the pBAD vectors. Western blot analysis showed both the N-terminal histidine tag and C-terminal strep-tag II using the pTTQ18-C construct suggesting that N- and C-terminus are present. Membrane localisation was shown for protein produced from the pTTQ18-C construct but not for the pBAD constructs. The latter was exclusively found in the insoluble fraction of 10000 g centrifugation and hence was most likely produced as inclusion bodies. None of the constructs was present in the soluble fraction.

Although a large variety of detergents, temperatures, pH-values, and salt concentrations were tested, only minor amounts of the protein could be solubilised. The protein did not bind to the Ni-chelating material. Possibly the signal in the membrane fraction was caused by contamination with inclusion bodies. Purification of the heterologously produced protein was not possible.

2.4.3   Functional Characterisation

2.4.3.1   *Properties of the Complex*

The protein complex is resolved as a single clear band of approximately 30 kDa on an SDS PAGE gel if the protein is heated with 2 (w/v) % SDS at more than 80 °C (Figure 2.4-2A). A "smear" above 50 kDa might be due to partial unfolding of the protein complex but no defined oligomers are detected. A substantial amount of protein failed to enter the gel. The amount of material did not decrease when the sample was heated or β-mercaptoethanol or protease treated. The band of 30 kDa was identified by mass spectrometry to be AQ_1760. Mass spectrometry of the protein revealed a molecular weight of 32.5 kDa (Figure 2.4-2 C), approximately 0.68 kDa more than the theoretical weight. The protein

complex failed to enter the acryl amide gel in BN PAGE as well. Apparently the acryl amide concentration of the stacking gel, 4 %, might form a mesh with too small pores for the complex.

To test the protease stability, the complex was treated with trypsin. Almost no degradation product can be detected even after 48 h (Figure 2.4-2 B).
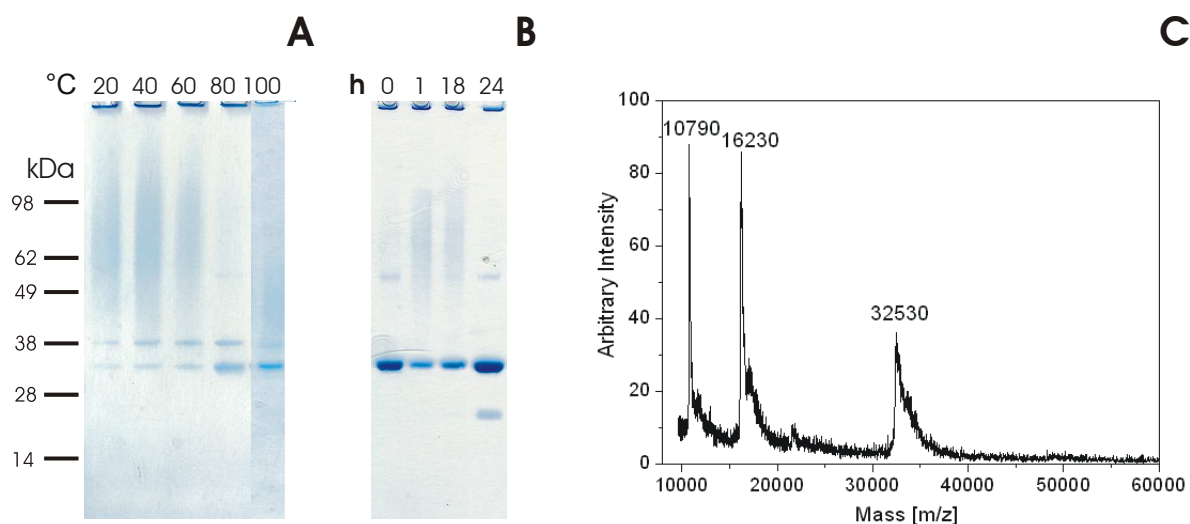

Figure 2.4-2: Coomassie stained SDS PAGE gel and mass spectrum of AQ_1760. A: the protein was treated with SDS-sample buffer at different temperatures. Only above 60 °C the complex disassembles and the 30 kDa protein becomes visible. B: the protein was treated with trypsin for indicated times at 37 °C and then boiled in SDS sample buffer. Even after 48 hours the complex remains stable and almost no degradation product is visible. B: the protein has a molecular weight of 32.5 kDa, the peaks at 16.23 kDa and 10.79 kDa are the doubly and triply charged protein, respectively.

2.4.3.2 Determination of the Size

As already described in chapter 2.1.2, the protein complex elutes near the void volume before any of the marker proteins on a Superose 6 gel filtration column, indicating that the protein complex in its native form is very large, at least 700 kDa, the highest mass that can be resolved on a Superose 6 column. To estimate the size of the complex more exactly and to exclude that the complex is non-specifically aggregated protein, different methods including an additional gel filtration chromatography step using Sephacryl-500 HR, analytical ultracentrifugation and electron microscopy were applied.

The gel filtration resin Sephacryl-500 HR is used to separate dextrans with sizes between 40 kDa and 20 MDa. In a few cases it was used for the separation of complex protein mixtures, including DNA and RNA-protein complexes. AQ_1760 appears in a symmetric peak suggesting a homogeneous sample. However, the column was not calibrated since no protein calibration standards of matching size are available, so the size can only be estimated to be between 700 kDa and 20 MDa.
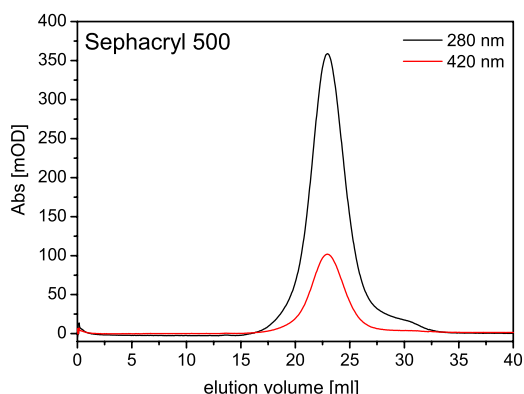
Figure 2.4-3: Sephacryl-500 HR gel filtration chromatogram. The column volume was 40 ml. The protein elutes in symmetric peak suggesting a homogenous sample.

Since no detailed information about the size of the complex could be determined from gel filtration chromatography, the oligomeric state was studied by sedimentation velocity experiments based on a continuous distribution level for S-values between 10 and 400 S. The protein was examined in the presence of 0.01 % DDM and without addition of detergent. Using DDM the c(S) distribution could be fitted to the experimental data exhibiting the presence of a peak at approximately 70 S with a shoulder at 100 S. These data suggest that at least two species of complexes are present. The heterogeneity and the uncertainty of the amount of bound detergent complicate the determination of the effective molar mass; however, assuming the partial specific volume of the protein to be 0.74 ml/g, a typical value for proteins, the effective molar mass of the 70 S complex can be estimated to be approximately 3.5 MDa. Without addition of detergent the c(S) distribution is fitting the experimental data showing several species of complexes with the major species at a size of approximately 70 S, 90 S, 120 S and 130 S. These data suggest that the protein forms a complex of approximately 70 S (3.5 MDa), and without addition of detergent the complex forms oligomers.
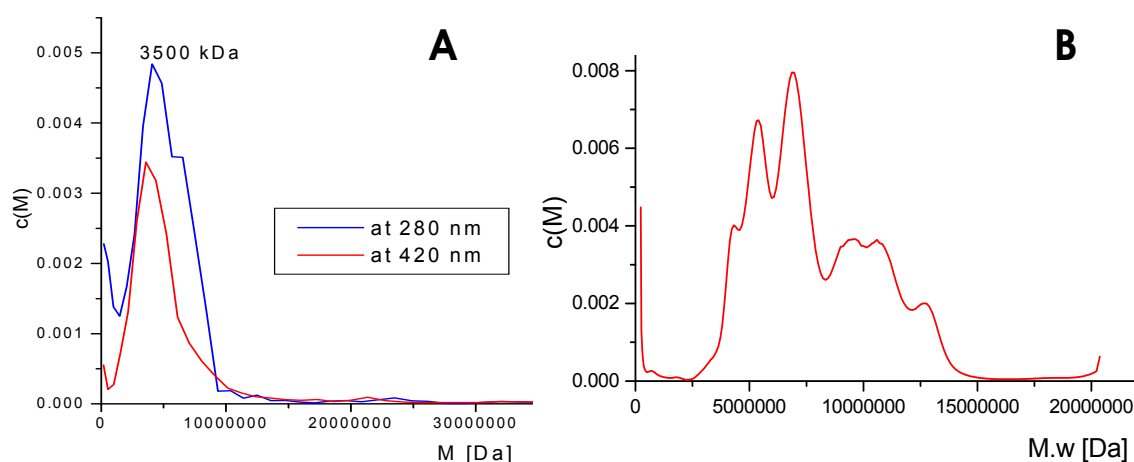

Figure 2.4-4: Sedimentation velocity analysis of AQ_1760 in the presence (A) or absence (B) of 0.03 % DDM. Best fit distribution calculated using the Lamm equation based on the model of continuous size distribution.

82

2.4.3.3 Electron Microscopy

By gel filtration and analytical ultracentrifugation only limited information on the oligomerisation state of the complex could be obtained. However, it could be suggested that it is both of appropriate size and homogeneity to obtain information from electron microscopy of the detergent supplemented sample. Therefore, AQ_1760 was analyzed by electron microscopy from negatively stained preparations. By this means the size in terms of nm dimensions and the homogeneity can be analysed at the same time. Four different staining solutions were tried: 1 % uranyl acetate, 2 % phosphotungstic acid, 2 % methylamine tungstate, and methylamine vanadate. The best results could be obtained with preparations in methylamine tungstate. Most projections are circular with an approximate diameter of 38 nm suggesting that the structure of the particle is spherical (Figure 2.4-5). The images show substantial particle variation. Some particles appear as solid spheres, while other particles exhibit a dark, stain filled centre. In the latter case the negative stained outer shell is approximately 1.5 nm to 2 nm thick and the stain filled inner core accounts for 35 nm. Some particles show distortions along one side, or seem to resemble broken vesicles. Obviously such structures can accommodate many 30 kDa subunits.

In several micrographs spheres touch each other suggesting multimers of particles. Along with the spherical particles amorphous material is detected.
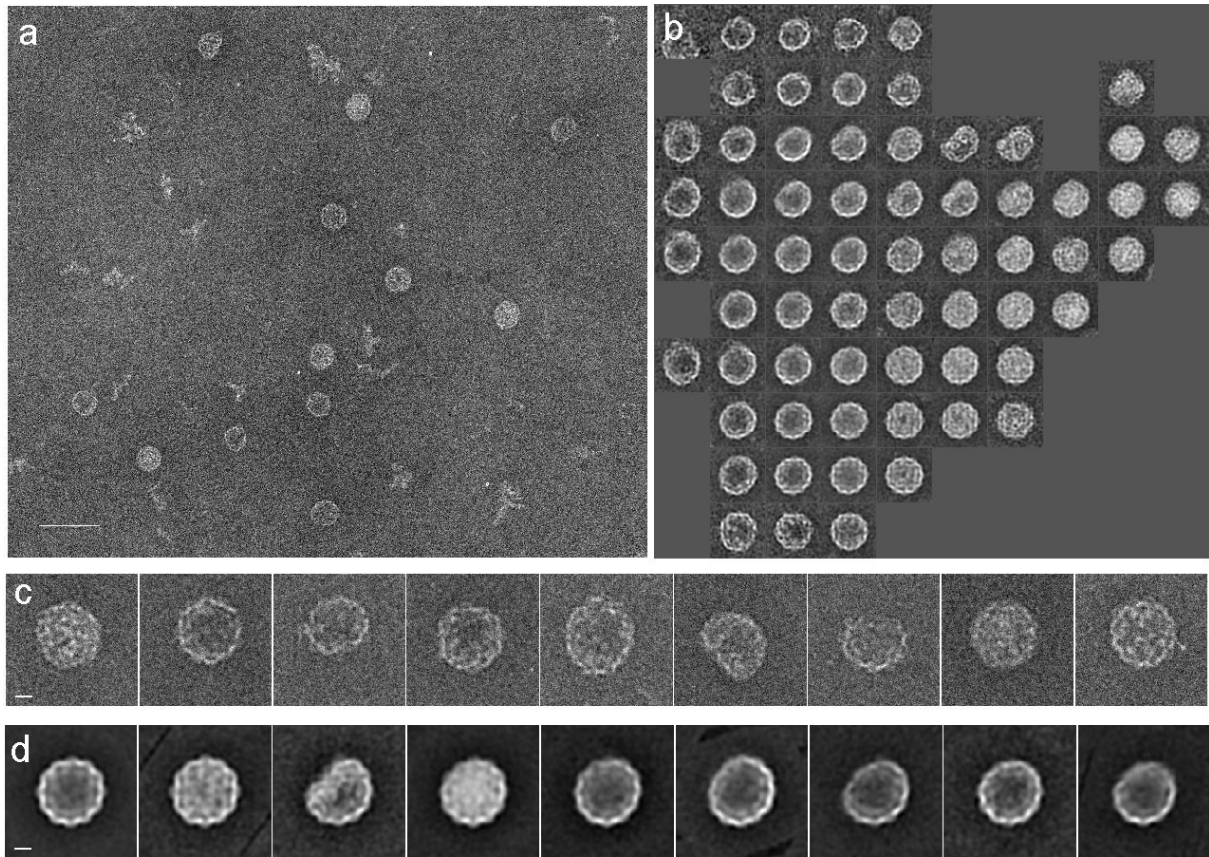
Figure 2.4-5: Electron micrograph of AQ_1760, stained with 2 % methylamine tungstate. Scale bar 100 nm. b) The first two factors of a multivariate analysis of the image set, displaying the largest variability in the data. Towards the left particles with a stain filled centre can be observed, towards the right solid particles are visible. c) Handpicked single images of AQ_1760 provide an overview of the particle shapes found in the sample. Scale bar 10 nm. d) Nine class averages as a result of the multivariate analysis of the image set followed by classification. The numbers of particles in each class average (from left to right) are: 73, 26, 10, 75, 97, 43, 23, 17, and 51. Scale bar 10 nm.

Already from the raw images, structure on the surface of the particles can be identified. White protrusions which are sometimes elongated, sometimes roundish are silhouetted against darker parts.

The images displayed in Figure 2.4-5 C show a hand-selected subset of images of individual particles as examples of the typical variations found in the sample. 415 particles were selected from 17 micrographs and were subjected to translational/rotational alignment and classification. Correspondence analysis (Figure 2.4-5 B) and the classification of the data set into nine classes (Figure 2.4-5 D) give a more quantitative display of the variations present in the sample. The class-averages are dominated by the outer shape and therefore lack some of the details that can be recognized in some of the individual images (Figure 2.4-5 C).

The appearance of the particles resembles those of viruses and phages. AQ_1760 was shown to be homologous to the "virus-like particle" purified from *P. furiosus* (Namba *et al.*, 2005), which was given its name due to its virus-like appearance. Also the predicted fold of AQ_1760 was similar to vlp and to gp5 from Bacteriophage HK97 [Swiss-Prot # P49861], which is the major capsid-forming protein. The appearance of the AQ_1760 protein complex resembles the early stages of phage maturation, the Prohead I, where the phage is a spherical particle of 47 nm diameter.

2.4.3.4   Proteolytic Activity

One of the closest related proteins to AQ_1760 on the basis of sequence comparison is bacteriocin from *T. maritima.* It was shown to exhibit proteolytic activity against peptides and galantine in zymograms (Hicks *et al.*, 1998). Similar experiments were performed with AQ_1760 in order to investigate whether its function is related.

Standard zymograms were performed to test the activity towards galantine. The protein complex was incubated at the indicated temperatures in SDS sample buffer before loading on the zymogram. After electrophoresis the zymogram was renatured and incubated at 80 °C overnight to allow cleavage of the galantine by AQ_1760. A clear band at around 30 kDa indicates the digestion of galantine by protein AQ_1760 (Figure 2.4-6 A). Additional bands at higher molecular weight might be due to oligomeric forms of the protein and to partial unfolding in SDS. When a lower developing temperature (30 °C or 60 °C) was chosen the bands were weaker and difficult to detect (data not shown).
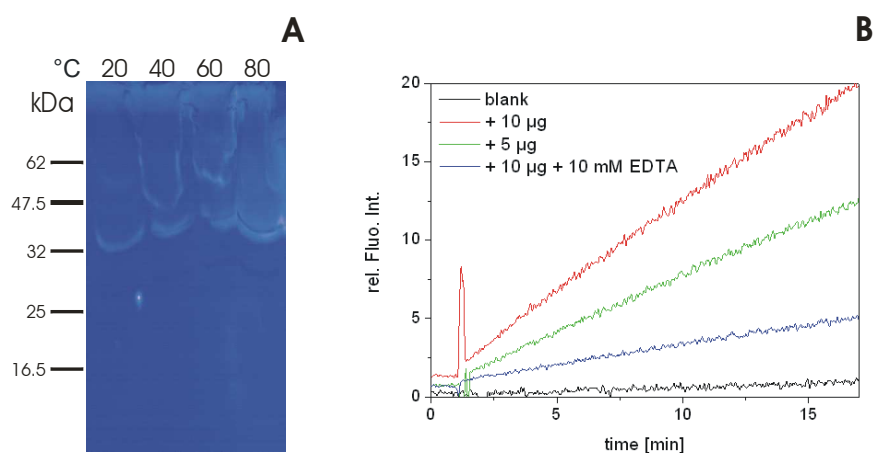

Figure 2.4-6: Coomassie stained zymogram containing 0.1 % galantine and proteolytic assay. A: Protein AQ_1760 was incubated in SDS sample buffer at the indicated temperatures for 3 min. B: The fluorescence increase of AMC from the peptide VKM-AMC is measured. Fluorescence is excited at 383 nm and detected at 455 nm.

Further characterisation of the proteolytic activity was accomplished by determination of the activity towards a variety of 7-amido-4-methylcoumarin (AMC)-linked peptides, the most pronounced feature of bacteriocin from *T. maritima*. The activity of AQ_1760 against three different peptides (Z-Val-Lys-Met-AMC, Suc-Ala-Ala-Phe-AMC and Z-Phe-Val-Arg-AMC) was tested by an assay adopted from Zimmermann and co-workers (Zimmerman *et al.*, 1978). The activity is monitored by the fluorescent signal of AMC which is released from the carboxy-terminus of the N-terminally blocked peptide (Figure 2.4-6 B).

Unlike the bacteriocin from *T. maritima* which shows a very broad specificity towards different peptides, no activity towards AAF-AMC or FVR-AMC was associated with AQ_1760. In contrast, the activity for Val-Lys-Met (VKM)-AMC at 30 °C and pH 7.5 was 8.0 nmol*min^{-1}*mg^{-1} which is in the same range determined by Hicks and co-workers (Hicks *et al.*, 1998) where an activity of 2.3 nmol*min^{-1}*mg^{-1} was determined. However, the proteolytic assay is not specific for AQ_1760 but

for proteases in general and thus it is not the only active species in the membrane fraction of *A. aeolicus*.
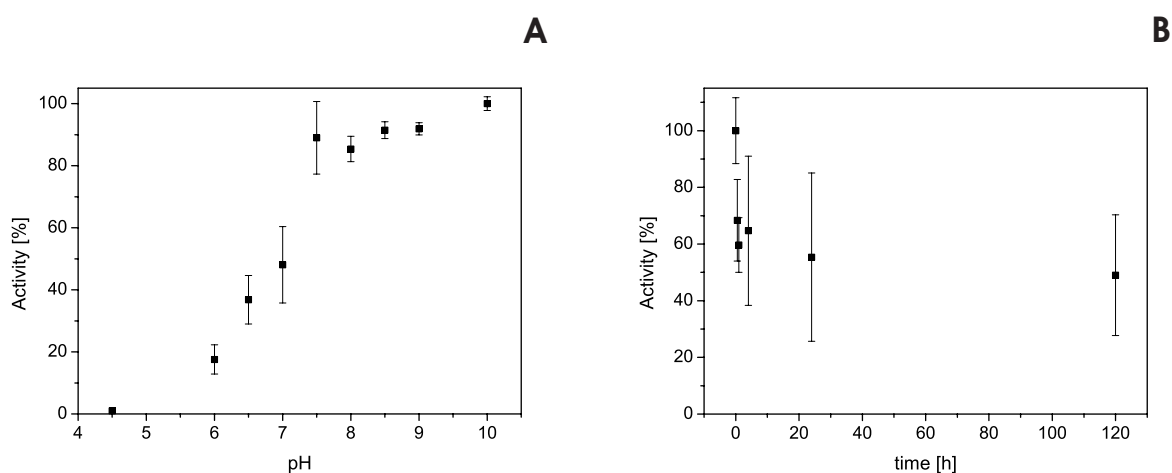

Figure 2.4-7: pH dependence and dependence of treatment at 85 °C of AQ_1760 in proteolytic assay.

The activity increases with pH: below pH 6.0 it is almost inactive, and above pH 7.5 it is 100 % active. The activity could be reduced to 25 % by addition of 10 mM EDTA. The dithionite reduced form is practically inactive. There is no dependence on the salt concentration (NaCl from 0 to 2 M). The stability of the protein was investigated by incubating the sample at 85 °C. It loses approximately 30 % activity within the first 30 minutes, but the remaining activity is very stable, and even after 5 days 50 % activity is retained.

2.4.3.5   *Identification of Cofactors*

The purified protein as well as the crystals (chapter 2.4.4) showed a yellowish colour suggesting the presence of a cofactor. The presence of metals was investigated by X-ray absorption spectra of protein crystals. An absorption signal near the iron edge (K edge, 7.11 keV) was detected, indicating the presence of iron. The iron content was quantified by TXRF measurements (Figure 2.4-9) to 1.8 irons per 8.7 sulphur (9 sulphur per protein monomer).
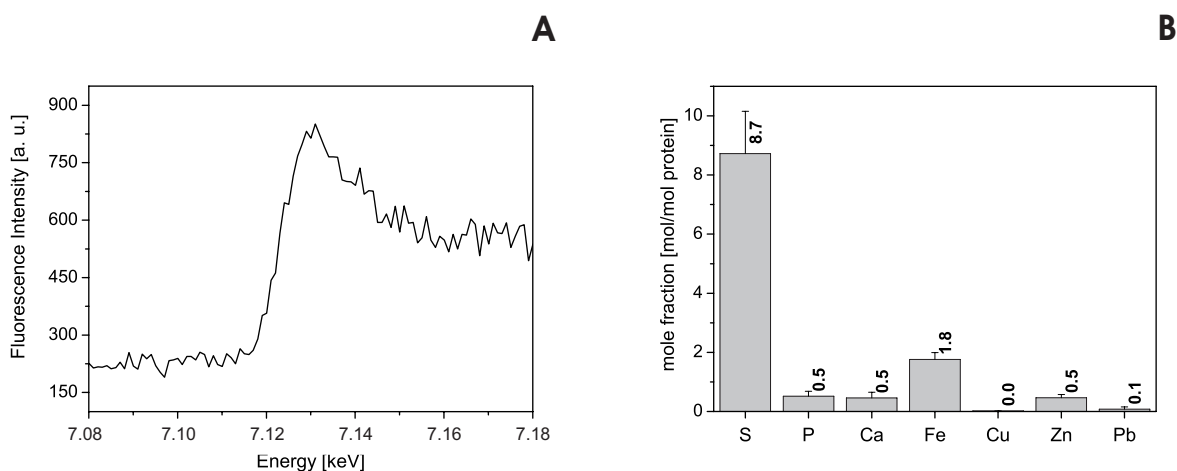
Figure 2.4-8: Energy scan of an AQ_1760 crystal and TXRF analysis of the protein in solution (20 μM in 20 mM Tris-acetate pH 7.5)

Iron is often present in the form of a heme and bound heme or flavins can be detected by spectroscopy or SDS PAGE based methods.

UV/VIS spectra (Figure 2.4-9) showed a prominent peak at 420 nm which is well known to be associated with heme bound protein (Soret-band). The characterisation of heme is most commonly done by SDS PAGE based heme staining of the covalently bound heme after electrophoresis or in REDOX spectra. No staining of the protein was detected in SDS PAGE analysis indicating absence of a covalently linked heme. Since the protein complex was not able to enter acryl amide gels in its native form it was not possible to determine the heme content in the native protein complex by BN PAGE. In air oxidised minus dithionite reduced spectra of the protein in solution, as well as in the crystal, a peak at 559 nm appeared, suggesting a *b*-type heme.
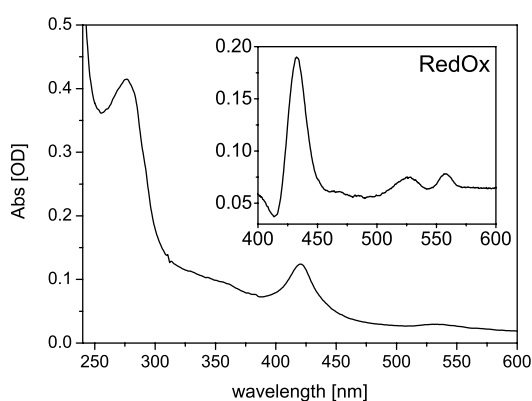

Figure 2.4-9: UV-VIS spectrum of an AQ_1760 crystal. In the inset an air-oxidised minus dithionite reduced spectrum is shown.

Quantification of the heme *b* content by comparison of the absorption at 280 nm and 415 nm and from absorption difference spectra at 565 nm-minus-575 nm revealed a substoichiometric amount of heme, only 0.2 and 0.02 heme per protein monomer, respectively. The values for ε(420 nm) and ε(565 nm

minus 575 nm) were taken from the literature, ε(280 nm) from the program ProtPram. Obviously the heme is only associated with the complex but not a stoichiometric part of it. The amount of phosphor per mole of protein (Figure 2.4-9) does not suggest stoichiometric binding of phosphorus lipids or DNA.

No fluorescence was associated with the band of AQ_1760 in an SDS PAGE gel illuminated with UV-light indicating that no flavin cofactor is covalently bound to the protein (data not shown). The characterisation of flavin can also be accomplished by UV/VIS spectrophotomerty and elementary analysis by TXRF measurements. An absorption maximum of FAD at 350 nm was not detectable and TXRF measurements of the protein sample showed only approximately 0.5 phosphorus per protein monomer (Figure 2.4-9) instead of two if one FAD group was present.

In conclusion, the protein contains two iron molecules per monomer but only substoichiometric amounts of heme and no flavin. However, the mode of binding remains unclear.

2.4.4 Crystallisation

Initially the protein was subjected to crystallisation attempts directly after gel filtration chromatography and in the presence of DDM. Three-dimensional crystals were obtained in hanging drop vapour diffusion technique, however crystals were often of leaf-shaped appearance (Figure 2.4-10 E), or spherulites appeared. In the presence of 20 % MPD, 4 % glycerol and 0.2 M NaCl (Figure 2.4-10 A) or citrate in a non buffered system, in 12 % MPD, 50 mM MgCl2 at pH 8.5 (Figure 2.4-10 B) as well as using 12 % PEG 4000 at pH 8.5 (Figure 2.4-10 C) crystals appeared at protein concentrations of 5 to 8 mg/ml after 3 to 4 weeks.
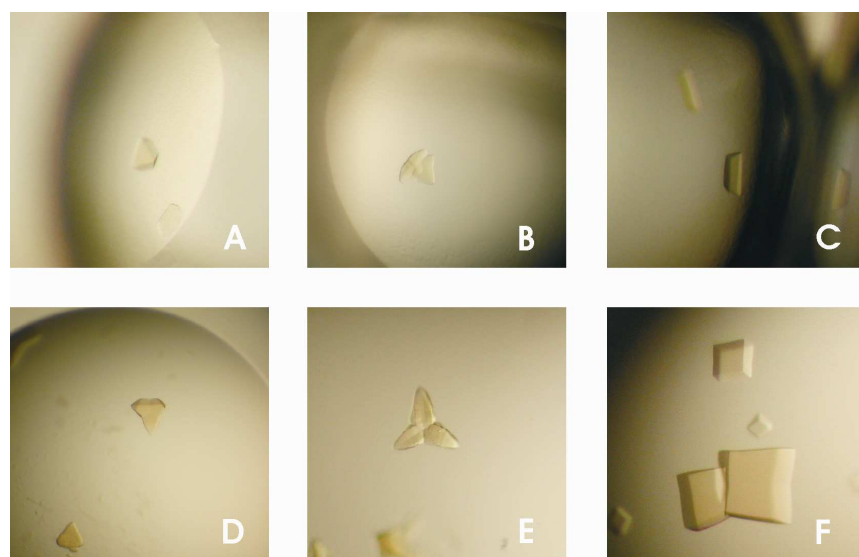

Figure 2.4-10: Protein crystals from AQ_1760. A to C initial crystals of the protein AQ_1760 derived from screening, D to F improved crystallisation conditions (D is improved A, E improved B and F improved C)

Dissolved crystals were analysed on an SDS PAGE gel and showed a double band (Figure 2.4-11 A). Both band were analysed by mass spectrometry and consisted of the protein AQ_1760, in both cases the N- and the C-terminal peptide was present. The crystals did not diffract X-rays.
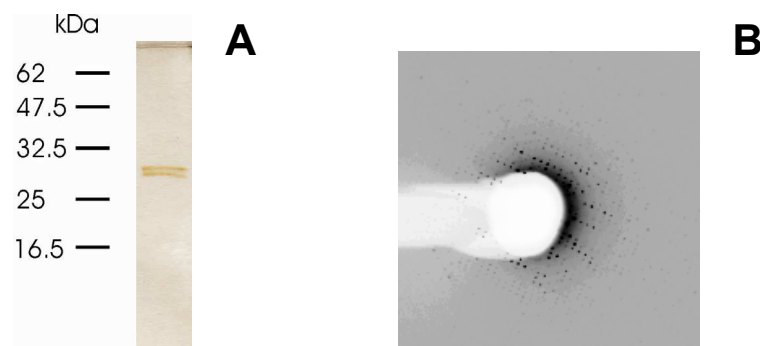

Figure 2.4-11: SDS PAGE gel of a dissolved crystal and diffraction pattern (0.5 ° rotation) of an AQ_1760 crystal.

Crystals grown in the presence PEG 4000 could be improved by removing the detergent by an additional anion exchange step. Using 8 to 12 % PEG 4000, cubic crystals with sharp edges and dimensions up to 0.3 x 0.3 x 0.3 mm appeared after 2 to 5 months. For X-ray data collection crystals were soaked in cryo buffer (10 % Glycerol, 30 % PEG 4000 at pH 8.5). Only when the PEG 4000 concentration was increased to more than 20 %, the crystals diffract X-rays, in the best cases to 8 Å (Figure 2.4-11 B). The crystals were very sensitive to radiation damage, and after the collection of 4 to 5 images the crystal did not diffract X-rays anymore. The quality of the crystals could not be improved. The major reason for this was the limited amount of protein, which prohibits large scale or grid screening. A second reason was the slow growth of the crystals obtained.

3 Discussion

3.1 Proteomics

3.1.1 The Method

This works demonstrates the first membrane proteomic approach using the hyperthermophilic eubacterium *A. aeolicus*. At the same time it is the first approach to investigate the feasibility combining membrane proteomics with non-denaturing purification in order to obtain intact integral membrane proteins and membrane protein complexes including peripheral subunits and other interaction partners for functional and structural characterisation.

The method of membrane preparation, solubilisation, and protein purification applied here was shown to be well suited for proteomics studies and at the same time for functional and structural studies of membrane proteins.

Membranes were prepared on basis of differential ultracentrifugation according to standard methods for membrane protein purification. Conventional methods of membrane preparation in membrane proteomics projects typically include density gradient centrifugation, treatment with urea or buffer at high ionic strength in order to remove remnants of cytosolic proteins, had to be modified in order to accommodate a preparative scale and to keep proteins in a native state. Treatment with mild buffer reduced the amount of cytosolic protein without denaturing the membrane proteins. Membrane associated proteins were retained in the membrane fraction by this means, either naturally attached or associated artificially during purification. Thereby contamination with non-membrane attached proteins and ribosomes, which, indeed, were detectable (Table 2.1-4) was allowed.

Membranes were not treated with any washing other than 50 mM Tris-Cl pH 8.0. Hence, cytosolic protein might be present and hamper the detection of membrane proteins by covering bands in SDS PAGE gels analysis and thereby complicating mass spectrometry analysis. However, a compromise between purity and yield had to be found. In most cases contamination with non-membrane proteins did not impede protein purification and a variety of different membrane proteins were identified. This mild treatment was shown to remove only a small fraction of the contaminating non-membrane proteins (Schlüsener *et al.*, 2005) in *Corynebacterium* (*C.*) *glutamicum*. The fact that a fraction of 31 % of the identified proteins were cytosolic proteins shows that the same is true in our approach. However, the risk of denaturing membrane protein was low and the chance of obtaining a preparation containing native membrane proteins maximised.

The choice of detergent is crucial for the success of any approach dealing with membrane proteins including proteomics studies. A compromise between (1) a detergent solubilising as many membrane proteins as possible (2) that is suitable with for the method of separation and (3) which is mild enough to retain membrane proteins and complexes in their native form has to be found. Here emphasis was placed on the mild treatment rather than on completeness of the membrane proteome. In order to keep most of the proteins in their native form, no harsh treatment was applied. The relatively mild detergent DDM was chosen for solubilisation and purification, as it had already been shown to successfully

solubilise many native and overproduced membrane proteins in proteomic and crystallisation approaches. The membrane protein extraction efficiency using DDM at the given detergent to protein ration was demonstrated to be 70 % to 80 % in other membrane proteomics approaches (Everberg *et al.*, 2006; Schlüsener *et al.*, 2005) compared to 95 % for SDS and approximately 80 % to 90 % for Aminosulfobetaine-14 and Zwittergent 3-10. In contrast to SDS, DDM is compatible with all applied chromatographic steps and with protein crystallisation. However, it is obviously not able to extract all proteins from the membrane. Very hydrophobic proteins might need a harsher treatment.

The method of choice for protein separation in many membrane proteomic approaches is multi-dimensional gel electrophoresis. IEF, which is commonly used for qualitative separation of proteins in proteomic approaches, was avoided here. It comprises several problems if applied to membrane proteins including low solubility at the pI and incompatibility with salt content or the detergent with the electrophoresis system. In addition, it is difficult in preparative handling of native proteins, especially for the use of non-coloured proteins. Depending on the system, protein fractions cannot be divided by their UV signal but have to be chosen randomly. Few approaches made use of a BAC or a BN PAGE gels to separate proteins in the first dimension. Those approaches were successful in identifying numerous membrane proteins and especially for BN PAGE gels additional information on complexes sizes is obtained. However only microgram amounts can be recovered from those gels, too little for many functional studies and three-dimensional crystallisation. Our approach of native purification followed by SDS PAGE analysis as the last separation step avoided those problems. Protein separation was achieved according to the pI by anion and cation exchange chromatography. Using a buffer system with a pH of 7.4 is suitable for most membrane proteins, because only few membrane proteins have a neutral pI (see Figure 2.1-1 B and below). This minimised the risk of protein precipitation. In addition, chromatography facilitates upscaling of protein purification in order to obtain enough material for structural and functional studies. Prefractionation of protein samples by liquid chromatography followed by SDS PAGE gels has already proven successful in some cases of soluble (Butt *et al.*, 2001; Fountoulakis *et al.*, 1999) and of membrane proteins (Schlüsener *et al.*, 2005; Szponarski *et al.*, 2004), however it is still an unusual approach. The chromatography-based purification technique shall concentrate low abundance protein and enhance protein identification; however multi-step purification on the other hand might not be applicable for those proteins, as they might get lost during purification and remain undetected. Chromatographic protein separation and purification can complement electrophoresis based proteomics studies.

Generally the identification of membrane proteins by approaches based on peptide mass fingerprinting following MALDI-TOF MS is often hampered the hydrophobic nature of the membrane regions and the absence of large hydrophilic domains. When only few peptide fragments are accessible for mass spectrometry, sequence coverage will be low and the results obtained from e.g. MASCOT are ambiguous. Unlike conventional proteomic approaches, here the proteins were identified after the sample passed several purification steps and only one or few proteins were present in an SDS PAGE

gel band. By this means the risk of missing the identification of a protein was minimised. On the other hand it is likely that information on less abundant membrane proteins that were not resolved as a single band on SDS PAGE gels was lost. However, the project did not have the objective to identify all membrane proteins, but only the most abundant ones. To overcome problems inherent in MALDI-TOF mass spectrometry and peptide mass fingerprinting, tandem mass spectrometry approaches can be conducted. Here a single peptide is fragmented and analysed; from the size of the fragments its sequence is deduced. Thereby a single peptide is sufficient to identify a protein.

3.1.2 The Organism

This is the first study on the proteome of *A. aeolicus*; neither the soluble nor the membrane proteome has been studied so far. It is the second study of the membrane proteome of thermophilic organism and the first of a hyperthermophile, the membrane proteome of *Geobacillus thermoleovorans* was determined using a shot gun approach (Graham *et al.*, 2006). *A. aeolicus* has a comparably low number of genes (Deckert *et al.*, 1998) encoding for putative target proteins. Thus the number of identified proteins was lower: 16 proteins with transmembrane segments were identified in comparison to 50 for a similar approach using *C. glutamicum* (Schlüsener *et al.*, 2005). However, compared to the total number of transmembrane proteins, 6.5 % (out of 772) of the predicted membrane proteins in *C. glutamicum* were identified and 5.1 % (out of 315) in *A. aeolicus*. It should be taken into account, that in the final assessment only approximately two thirds of proteins existent in the elution profile of *A. aeolicus* were examined in this work.

Proteins from thermophilic organism are believed to be rigid and stable. Proteins identified here were indeed stable toward heat denaturation i.e. AQ_1862 is almost resistant towards heat denaturation. In addition purification, functional analysis and crystallisation at 18 °C was possible.

However, *A. aeolicus* is not easy to grow: it needs an atmosphere of mainly hydrogen and carbon dioxide at high temperature and thus is expensive. Hence the growth of the cells is the limiting step in the process of obtaining protein. In addition, it cannot be genetically modified so far; implementation of a genetic system will be challenging considering the limitations in cell culture. Without the possibility of genetic modification it will not be possible to introduce artificial amino acids (e.g. seleno-methionine) or mutations in the native protein, hampering structure determination and functional characterisation of target proteins. In addition, protein purification can be facilitated or the amount of protein increased by introducing affinity tags or placing the coding sequence behind a strong promoter.

3.1.3 The Proteins

The distribution of predicted pI of the ORF containing at least one predicted TMH is bimodal as described already for cytosolic as well as membrane proteins from several species of bacteria, archea

and euchariota (Schwartz *et al.*, 2001; Wu *et al.*, 2006). It was assumed that this feature is due to the neutral pH within the cell and the low solubility of proteins at their isoelectric point.

In total, 75 proteins were identified, 16 membrane proteins and 24 membrane associated proteins, that can be grouped into four classes: (1) respiration and energy transduction, (2) transport, (3) "hypothetical proteins" and (4) other, mainly cytosolic proteins. Only few are present in high amounts. The proteins present in high amounts are mainly those from the first group. Out of these all have been described previously in the literature and some have already been studied in detail. However, most studies were focused on only one or a few membrane protein complexes. This is the first study to give a more general overview on *A. aeolicus* although only parts of the membrane proteome were analysed. The presence to two kinds of [NiFe] hydrogenases in the membrane was confirmed, both have been studied by Brugna and co-workers before (Brugna-Guiral *et al.*, 2003); however, the respective cytochrome *b* subunits could not be identified here. A third hydrogenase without predicted membrane anchor was not detected in the membrane fraction. Guiral and co-workers suggested the existence of two super-complexes: (1) hydrogenase 2 and DMSO reductase and (2) hydrogenase 1 and bc_1 complex (Brugna-Guiral *et al.*, 2003; Guiral *et al.*, 2005). The DMSO reductase A chain (AQ_1234) was present in our study, but not in the same fraction as the hydrogenase and much less abundant. The DMSO reductase B and C subunits were not detected here. A super-complex of hydrogenase 1 and bc_1 complex might also be present in our initial preparation (Figure 2.1-2, lane 6), however, by cation exchange chromatography (Figure 2.1-4) both could be separated (P1 and P2), suggesting co-purification, rather than complex formation.

Two versions of cytochrome c_{552} were identified, one in the membrane fraction (AQ_792) and one in the culture supernatant (AQ_1550). AQ_1550 obviously does not interact strongly with the membrane and reached the supernatant most likely due to cell disruption before centrifugation, while AQ_792 does interact with the membrane and hence is present in the membrane fraction, although no transmembrane domain is present. Both have already been described to be membrane bound and soluble, respectively (Baymann *et al.*, 2001), although referred to as c_{555} in this publication. The sulphide-quinone oxidoreductase (AQ_2186) has also already been described as being present in the membrane fraction of *A. aeolicus* (Nübel *et al.*, 2000).

The proteome reflects the metabolism of *A. aeolicus*, which primarily makes use of hydrogen (through the two [NiFe] hydrogenases) and sulphide (through sulphide-quinone oxidoreductase). Compared to components of these two systems, only small amounts of NADH:ubiquinone oxidoreductase (complex I of the respiratory chain) is present and no complex II was detected, suggesting that degradation of sugars and amino acids by the citric acid cycle plays only a minor role in energy transduction. *A. aeolicus* was reported to exclusively utilise oxygen as the final electron acceptor. Cytochrome *c* oxidase is present in the membrane and was identified (Peng, unpublished data). The amount and proportion of the different complexes varies depending on culture condition.

The only protein which was present in similar amounts to the proteins described above, i.e. approximately 10 % of all membrane proteins (the accurate amount of the other proteins of high abundance was not determined), was the "hypothetical protein" AQ_1862. So far it has not been described in the literature and its sequence similarity to annotated proteins was not strong enough to propose a function. However, it can be considered a conserved protein because several proteins in other organisms show at least a weak sequence similarity.

Other proteins associated with the inner membrane identified in this approach were present in concentrations of approximately 1 % of the highest abundant proteins, many even less. It should be noted that only few identified membrane proteins identified here are annotated as transport proteins of the inner membrane by the TransporterDB. These are: AQ_1055, which is a subunit of a putative phosphate transporting ABC-transporter from the ABC superfamily; the F-type ATPase was already described by Peng and co-workers in 2006. Secondary transporters include AQ_175, which is member of the cytochrome oxidase biogenesis (Oxa1) family and SecD (AQ_973) of the resistance-nodulation-cell division (RND) superfamily. It is tempting to suggest that the low number of transport proteins is due to the fact that as an autotrophic bacterium *A. aeolicus* does not depend on uptake of nutrients or metabolites like sugar, fatty acids, or amino acids.

A protein which is frequently identified in membrane proteomics projects is the erythrocyte band 7 homolog (AQ_911), a protein which is highly conserved among species: the human and the *A. aeolicus* protein share 70 % homology and 46 % identity for 230 aligned residues. It was named according to SDS PAGE analysis of erythrocyte membrane proteins as the seventh largest protein; however its function is still unclear. Patients lacking this protein or with a defective copy suffer from a distorted sodium and potassium balance in erythrocytes (Gallagher *et al.*, 1995). A homologous protein in *Caenorhabditis elegans*, MEC2 is involved in sodium channel regulation (Goodman *et al.*, 2002).

By annotation only few proteins identified here were predicted (TransporterDB) as outer membrane proteins: outer membrane protein (AQ_1300) as a member of the outer membrane protein insertion porin (OmpIP) family, the outer membrane protein C (AQ_529) as member of outer membrane receptor (OMR) family that is an outer membrane receptor TonB homolog, and a peptidoglycan associated lipoprotein (AQ_2147) as member of the OmpA-OmpF porin (OOP) family involved in cell envelope biogenesis. Most of them do not seem to be involved in ion transport across the outer membrane. The only putative porin which seems to be of relevance for ion or solute transport through *A. aeolicus* outer membrane is AQ_1862 and possibly its homolog AQ_1259. The comparative abundance of AQ_1862 to AQ_1259 can be estimated to be 100 to 1 but it is still possible that AQ_1259 plays an important role in ion or solute transport as well but this was not investigated here. In order to investigate more outer membrane porins by functional proteomics, *A. aeolicus* has to be cultured under different conditions, for example under phosphate limitation, to induce production of additional porins. AQ_1862 was also one of the major components in the culture supernatant.

Interestingly, the porin protein OprP from *P. aeruginosa* was identified as a widely distributed dissolved component in the sea water (Kimata *et al.*, 2004; Tanoue *et al.*, 1995).

Out of the 29 identified proteins annotated as "hypothetical proteins", 9 contain one or more predicted transmembrane segments: 55 % of all membrane proteins are "hypothetical proteins" within the membrane proteins. However, 6 out of the 9 have only one transmembrane segment, which might also be a signal peptide. Usually membrane proteins are overrepresented in the group of hypothetical proteins. 12 hypothetical proteins without TMHs and of either unknown or cytoplasmic location were detected; however, they still might be membrane associated proteins.

Only very few cytosolic proteins were detected in amounts suitable for further analysis including glutamine synthase and elongation factor Tu, they were however excluded from the investigation. These proteins might be present in the membrane fraction due to cross-contamination because of their high abundance in the cytosol. Other proteins, like ribosomal or Clp protease subunits might be present in the membrane fraction due to their high molecular weight: they are pelleted during membrane preparation.

Ribosomes are frequently found in membrane proteomic approaches and regarded as ambiguous, either as contaminant (Klein *et al.*, 2005; Schlüsener *et al.*, 2005) or as membrane associated proteins (Everberg *et al.*, 2006). They were also observed when the membrane was treated with high pH buffers, which is considered to be an efficient method for removal of peripheral and loosely attached membrane proteins. The presence of ribosomes in membrane preparations is a matter of debate. Everberg and co-workers suggested that the interaction is presumed due to the mild treatment of the membrane. On the other hand Klein and co-workers suggested non-specific binding of ribosomal components to the membrane by electrostatic and/or hydrophobic interactions. Several ribosomal proteins and most likely also RNA were detected in the MonoQ fractions 68 to 71. The ribosomal proteins remained after sucrose gradient centrifugation, so it can be assumed that the content of ribosomes is no co-sedimentation. In our approach the contamination of most membrane protein containing fractions with ribosomal proteins was avoided since the ribosomal proteins eluted at high ionic strength and only in the fractions preceding the ribosomal fractions was some contamination observed.

3.1.4 The Targets

The major aim of this project was to increase knowledge on membrane proteins through functional and structural characterisation. In general the identity of analysed membrane proteins from native origin reflects the abundance of the respective proteins, the higher the abundance, or availability the better a proteins is studied. Accordingly, the best functional and structurally studied are those involved in respiration, energy transduction, or metabolism except for few exceptions of membrane proteins purified from specialised tissue, like aquaporins from the eye lens or erythrocyte or rhodopsin from retina. The same holds true for our study on *A. aeolicus*. The membrane proteins that were identified

here as highly abundant and as possible targets for functional and structural characterisation are those involved in energy transduction and respiration; however these are already well known and have already been studied in detail either from *A. aeolicus* or from different sources.

The only protein of unknown function and high abundance was the "hypothetical protein" AQ_1862. It was chosen as target protein for characterisation because its high abundance facilitates purification and taking the abundance into account indicates it may be of great biological importance to the organism. From 120 mg membrane 10 to 12 mg of pure and homogenous protein were obtained. Bioinformatic characterisation suggest that it is a putative outer membrane porin. The knowledge on the outer membrane of *A. aeolicus* was still very limited and so far most studies have been carried out on outer membrane proteins from proteobacteria and much less is known on the outer membrane of other bacteria especially those existing in extreme environments. By choosing AQ_1862 as target protein the identity and the function of the major outer membrane protein from *A. aeolicus* was elucidated; its structure determination is underway.

Two other hypothetical proteins were chosen as targets for further studies: AQ_1760 and AQ_1558. They were much less abundant than AQ_1862- from 120 mg of membrane protein only 0.3 to 0.5 mg of pure protein was obtained. This amount is towards the lower limit of purification feasibility and too little for a broad functional and structural screening. Therefore heterologous production of the target proteins was attempted. AQ_1760 can be considered to be a conserved hypothetical protein since it shares similarity to functionally characterised proteins but the sequence identity was not high enough for annotation. It is a putative bacteriocin involved in antimicrobial effects. Almost nothing is known on antimicrobial effects in the thermophilic environment. Biochemical and structural similarities to homologous proteins were discovered. AQ_1558 does not share much similarity to any other protein so it might be a species- or genus specific protein. Proteins without homologous sequences are a great challenge in the investigation of the proteins function. The biological function of both remained unclear.

3.2   AQ_1862

3.2.1   Similarity

The protein AQ_1862 shows features typical of a porin originating from the bacterial outer membrane: the sequence contains an N-terminal signal peptide which is cleaved in the mature protein and a phenylalanine at the C-terminus, characteristic of localisation in the outer-membrane (Struyve *et al.*, 1991). It is one of the most abundant proteins of the DDM solubilised *A. aeolicus* membrane. Whether the protein indeed originated from the outer membrane of *A. aeolicus* could not shown definitely. Separation of the inner and outer membranes by classical means remained unsuccessful. But the facts mentioned above, together with a strong interaction with the peptidoglycan layer gives rise to the assumption that the protein is indeed associated with the outer membrane.

96

Sequence comparisons show that the protein has similarities to two families of outer membrane porins: the short chain amide and urea porin (SAP) family and the *Pseudomonas* OprP porin (POP) family. Only recently the database annotation for several AQ_1862 homologous proteins was changed from "hypothetical protein" to "phosphate-selective porin O and P", although their sequence homology to the *P. aeruginosa* porins OprO and OprP is rather low. The sequence conservation of AQ_1862 is obviously too low for bioinformatics annotation by means of standard sequence comparison done by databases: it is still annotated as "hypothetical" protein. The similarity of AQ_1862 to putative porins from *Geobacter* and *Thermotoga* on the other hand is high compared to porins from proteobacteria. However, none of these similar proteins were studied biochemically or electrophysiologically: their annotation is based on sequence comparison and the physiological function is unclear.

By tertiary structure prediction, however, the similarity to the POP family was confirmed. AQ_1862 showed very high similarity to the fold of OprP from *P. aeruginosa*, and HH-PRED gave a probability score of 100 %. Homology was also detected to OmpK36 from *K. pneumoniae* and PhoE from *E. coli* and other 16 stranded "classical porins", that are present as trimers.

One of the few freely available web-based prediction programs for outer membrane β-barrels is the program PRED-TMBB (based on a hidden Markov model). For AQ_1862 it predicted an 18 stranded β-barrel with reasonable distribution of loops and turns (Figure 3.2-1): elongated loops facing the outside with an exceptionally long loop which might fold into the barrel as a constriction and short turns towards the periplasm. This is a typical feature in most porin structures.
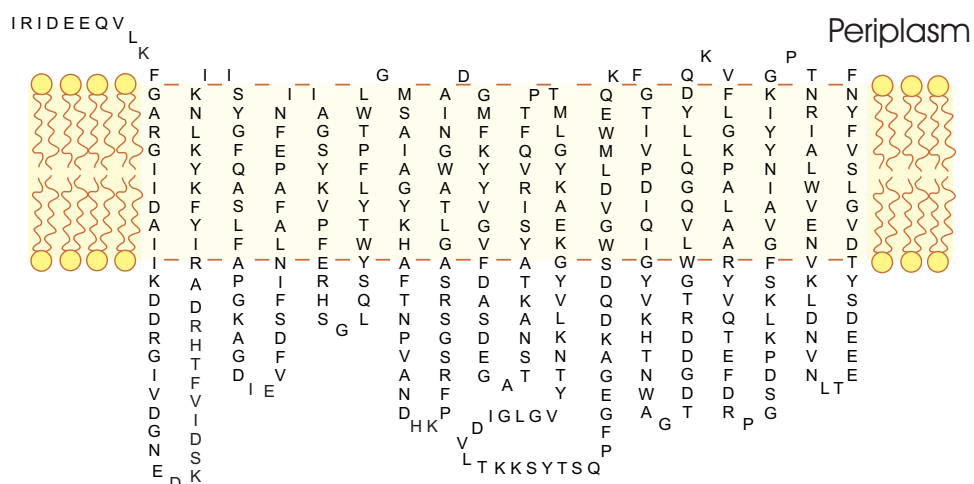

Figure 3.2-1: β-barrel topological prediction of AQ_1862 predicted using PRED-TMBB (Bagos *et al.*, 2004).

Experimentally, by means of circular dichroism, it was shown that the protein's secondary structure is mainly β-sheet without a substantial amount of α-helices or random coil, similar to other porins like OmpF (Eisele *et al.*, 1990) and PorB (Dahan *et al.*, 1996). AQ_1862 is not sensitive towards protease treatment. This feature has been attributed in the literature to the fact that the whole amino acid sequence is an integral part of the barrel and shields the protease sites from digestion. With the

exception of the porins of the OmpA class, which are considered to have a barrel forming domain and a periplasmic domain, porins are usually very stable against proteolytic digestion.

Taken together, the data provide a strong hint that AQ_1862 is a porin originating from the outer membrane. Although sequence conservation is very low, the tertiary structure model resembles the structure of the phosphate-selective porins OprP and PhoE. Its general function as a porin could be confirmed by lipid bilayer experiments. However, the similarity to the phosphate-selective porins seems to be limited to the general fold of porins, the β-barrel structure. No hint of functional similarity was found. In contrast the passage of phosphate through the protein channel is almost completely blocked: a zcp of +42 mV in the presence of a 1:10 gradient of K-phosphate suggest cation selectivity. In addition, the basic conductance of 75 pS 100 mM K-phosphate is towards the lower limit for all anions investigated here, suggesting no or very low passage of phosphate.

3.2.2 Biochemical Properties

Many porins, including the well characterised member of the O and P family, OprP from *P. aeruginosa* (Angus *et al.*, 1983; Moraes *et al.*, 2007), form trimeric channels in the outer membrane. More recently the existence of monomeric porins has become evident (Conlan *et al.*, 2000). The oligomeric state of the putative porin protein AQ_1862 was determined by various means including gel filtration, PAGE, cross-linking and lipid bilayer experiments.

A simple but rather imprecise technique to analyse the oligomeric state of membrane proteins after detergent solubilisation is gel filtration chromatography and the comparison of the elution volume with soluble standard proteins. The results agree with a porin trimer and a detergent micelle of 68 kDa and 131 kDa for the native and the refolded protein, respectively. The variation can have resulted from the presence of lipids in the native protein and a shifted apparent molecular weight.

One typical feature of many trimeric outer membrane porins is their behaviour in SDS PAGE analysis: they run at the apparent molecular weight of the trimer and are only separated to monomers upon heating in SDS sample buffer. The SDS and β-mercaptoethanol concentration necessary to separate the trimer into the monomers varies from porin to porin. AQ_1862, however, runs at the apparent molecular weight of the protein monomer: at approximately 40 kDa as a double band if not heated, (approximately 38 kDa and 42 kDa), suggesting a monomeric porin, or a very labile oligomer. Double bands of porin proteins have been described in the literature and were attributed either to partial unfolding (Conlan *et al.*, 2000) of the protein in SDS or to different amount of LPS bound to the protein (Garavito *et al.*, 1983). Refolded AQ_1862, which most likely does not contain LPS, runs as a double band as well, so it can be concluded that the electrophoretic behaviour is due to partial unfolding of the protein at low temperature and complete unfolding when the sample is boiled in SDS. However, in the case of OmpG the difference between unfolded and folded protein is larger, more than 10 kDa. Partial unfolding of the protein in the presence of SDS at low temperatures is also observed in circular dichroism spectra with and without SDS. Already at 20 °C the signal of the β-sheet structure

begins to decrease and (see Figure 2.2-9), and the refolded protein is even more sensitive toward SDS denaturation.

The result obtained from BN PAGE analysis suggests a protein size of between 132 kDa and 440 kDa, most likely close to 300 kDa. However, these values are imprecise since comparison with the soluble standard proteins might not reveal reliable values for membrane proteins.

Cross-linking with a bifunctional reagent like glutaraldehyde is a method commonly used to determine the oligomeric state of a native protein (Azem *et al.*, 1995; Locher *et al.*, 1997) and was performed to confirm the results obtained for AQ_1862 in gel filtration chromatography. The size of the cross-linked protein and the pattern of the two oligomeric forms suggest a trimer. Bands of the cross-linked native protein appear blurred or broadened most likely caused by interchain cross-linking of the single polypeptide chains and thus different electophoretic behaviour. Similar behaviour has been described for several other outer membrane porins (Azem *et al.*, 1995). The gel pattern of the refolded protein suggests a trimeric form as well but with a much lower cross-linking efficiency. The efficiency can be influenced by different factors, like protein concentration, fold and flexibility, lipid or detergent content of the proteins. Similar pattern with a very weak band of trimeric protein has been described for the *P. aeruginosa* porin OprP (Angus *et al.*, 1983) which has been shown to be a trimer in nature.

The symmetry given by the space group was also taken as indication by Nestel and co-workers that a porin forms a trimer in the crystal (Nestel *et al.*, 1989). Many trimeric porins, like AQ_1862, crystallised in a space group with a crystallographic three-fold axis that goes through the centre of the trimer (Cowan *et al.*, 1992; Forst *et al.*, 1998; Kreusch *et al.*, 1994). However, symmetry operation in the crystal is not sufficient evidence to indicate the oligomeric state of a protein.

Results from lipid bilayer experiments have to be interpreted carefully regarding the oligomeric state of the protein. A single basic and fluctuating conductance i.e. single opening and closing events were evident. Trimeric outer membrane porins like OmpF usually show a three step opening and closing behaviour (see Figure 1.1-1 B). In the case of OmpG the appearance and blocking of a single step was taken as evidence for a monomeric porin (Conlan *et al.*, 2000). However, there are also examples of trimeric porins suggesting to have only one active channel (Buehler *et al.*, 1993). Therefore, it cannot be excluded that under the conditions used for lipid bilayer experiments (lipid composition, temperature) only one channel is active.

Taken together the data suggest a trimeric form of the protein, but additional evidence has to be obtained from the three-dimensional structure.

3.2.3 Electrophysiological Properties

Porins facilitate diffusion across the outer membrane. Many outer membrane porins have been studied in detail so far. The most commonly used technique to study outer membrane porins are planar lipid bilayers. They offer an open and easy to use system and in addition results can easily be compared with other studies.

The functional properties of AQ_1862 differ from the properties of the homologous proteins OprP, OprO, or PhoE in not showing any selectivity towards phosphate. Furthermore it differs from other outer membrane porins of the general diffusion pathway by showing two different conductances with different selectivities.

The broad selectivity towards several electrolytes and the absence of appreciable amounts of other outer membrane porins suggests its function as a general porin rather than as a substrate specific porin like OprO and OprP. However, unlike most porins, AQ_1862 did not incorporate spontaneously from aqueous solution into a preformed membrane, if the mild detergent (dodecyl-β-D-maltoside) was used. The lack of spontaneous incorporation was confirmed with the detergent C_8E_4. For concentrations between 0.2 to 1.8 μg/ml and observation times between 2 to 3 hours no spontaneous incorporation was detected indicating that the behaviour in dodecyl-β-D-maltoside is not an exception. There are also studies which show, that the choice of lipid used to form the planar lipid bilayer plays an important role in the frequency of protein insertion (Rostovtseva *et al.*, 2006). Although the single unit conductance cannot be observed as a step-like process, this system offers a convenient stability over the course of the experiment and measurements can be performed for hours without interruptions.

In multi-channel experiments AQ_1862 is able to increase the conductance of a lipid membrane drastically. The conductance is anion selective at acidic and neutral pH and cation selective at basic pH. In the presence of large organic anions cation-selectivity is observed from pH 4 to pH 10. The macroscopic selectivity is dependent on the salt, varying from mainly anion-selective in KCl to mainly cation-selective in K-acetate. The macroscopic selectivity is untypical for outer membrane porins, which, in general, keep the selectivity regardless of the electrolyte (Benz *et al.*, 1985). In addition, the permeability ratios obtained in the presence of K-acetate and K-gluconate cannot be fitted with the GHK equation. This feature has not been described for porins so far. However, differential selectivity of AQ_1862 is similar to that of VDAC, which is also anion selective in KCl and cation selective in K-acetate (Benz, 1994). For simplicity results for KCl and K-acetate are discussed although similar data were obtained for other combinations of organic anions or with halides at alkaline pH-values.

More detailed insights into the properties of AQ_1862 were obtained with experiments designed to reduce the number of channels incorporated to the level of one channel. Although the incorporation of a single channel could only rarely be followed by step-like increase of conductance, it is suggested that this conductance is caused by a single molecule, because (1) 1.4 nS is the lowest stable conductance in 100 mM KCl observed in more than 20 single channel experiments; (2) a shutdown of the basic and the fluctuating conductance in one step could be observed occasionally (Figure 2.2-14) and (3) in a few experiments multiples of this conductance occurred (not shown). No voltage-dependent channel closing at conductances up to 150 mV was observed, neither for macroscopic experiments (Figure 2.2-12) nor for fluctuating or basic conductance (Figure 2.2-15 and Figure 2.2-16). Many porins (Schein *et al.*, 1978; Schindler *et al.*, 1978) tend to have a lower open probability at

voltages above 100 mV. However, it is unclear if this phenomenon is of physiological relevance since the potential across the outer membrane was estimated to be around 30 mV (Stock *et al.*, 1977).

In single channel experiments two different open states, a basic conductance and a fluctuating part with different ion-selectivity are observed.

There are several published examples for substates or "residual conductances" in addition to the defined open and closed states in porin conductances (Basle *et al.*, 2004; Danelon *et al.*, 2003; Nestorovich *et al.*, 2003; Nestorovich *et al.*, 2006). OmpF for example showed a residual conductance even if all monomers of a trimer were closed that comprises 5 % to 15 % of the monomer conductance but no detailed study on the selectivity has been carried out so far.

Mitochondrial VDACs also show channel like fluctuations between their open (= basic conductance plus fluctuations) and their closed (= basic conductance) states (Komarov *et al.*, 2005; Pavlov *et al.*, 2005). The selectivity of the mixed open state has been determined to be anion selective and of the closed state to be cation selective.

To our knowledge, very large conductance with superimposed fluctuations a have only been described for OmpA from *E. coli* (Saint *et al.*, 1993): in 250 mM KCl the large conductance was 1.2 nS and the fluctuating part 180 pS. However, the large conductance has not been addressed in detail in later publication (Arora *et al.*, 2000). Recently, the existence of two distinctly folded isoforms for the homolog OmpF from *P. aeruginosa* was suggest, one forming a barrel with the entire protein sequence and the other forming an N-terminal barrel and a C-terminal periplasmic domain (Nestorovich *et al.*, 2006; Sugawara *et al.*, 2006). The existence of two isoforms of AQ_1862 seems rather unlikely: (1) a very pure protein preparation as judged by gel filtration chromatography and (2) the data obtained from lipid bilayer strongly suggest a coexistence of the two conductances in one molecular entity.
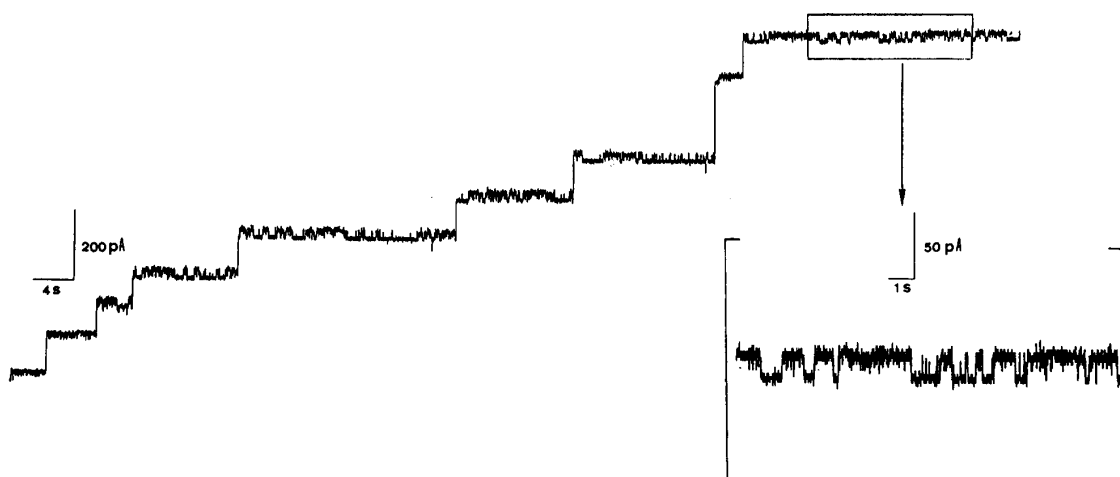

Figure 3.2-2: OmpA conductance traces in the presence of 250 mM KCl under a potential of 100 mV (Saint *et al.*, 1993).

The two conductance parts, the basic conductance and the fluctuations, are compared to mainly three well studied porins further on: OmpF and PhoE from *E. coli* and OprP from *P. aeruginosa*. All three

form trimers in the outer membrane. OmpF and PhoE show modest cation and anion selectivity, respectively, and OprP is highly selective for anions.

Table 3.2-1: Comparison of conductances, zero current potentials and permeability ratios of different porins with AQ_1862. *in 150 mM KCl; zcp is the zero current potential in presence of an ion gradient 1:10. 1: (Danelon *et al.*, 2003), 2: (Benz *et al.*, 1985), 3: (Benz *et al.*, 1984), 4: (Sukhan *et al.*, 1996), 5: (Maier *et al.*, 2001), 6: (Hodge *et al.*, 1997), 7: (Colombini *et al.*, 1996).

Porin Protein	Conductance [pS]	zcp [mV]	P_{Cl}/P_K
AQ_1862 –basic	1380	-40	9 : 1
- fluctuations	240	+40	1 : 9
OmpF [*E. coli*]	300 [1]	+26 [2]	1 : 3.6
PhoE [*E. coli*]	200 [3]	-24 [2]	3.3 : 1
OprP [*P. aeruginosa*]	160 [4]	-58 [2]	> 100 :1
Porin [*T. thermophilus*]	2500 [5]	+ 20 [5]	1 : 2.7
VDAC – open state *	800 [6]	-32 [7]	5.1 :1
- closed state *	300 [6]	+36 [7]	1 : 0.15

The fluctuating part of AQ_1862 is mostly cation selective ($P_{Cl}/P_K \approx 1:8$) with a conductivity sequence of $K^+ \approx Rb^+ > NH_4^+ > Na^+ \approx Li^+ \approx Cs^+$. The cation-selectivity of OmpF is less pronounced ($P_{Cl}/P_K \approx 1:3.6$) (Benz *et al.*, 1985). The conductivity sequence for OmpF (Benz *et al.*, 1984) and in aqueous solution (Robinson *et al.*, 1959) in the presence of Cl^- is similar ($Cs^+ \approx Rb^+ \approx K^+ > Na^+ > Li^+$), however, Cs^+ has the highest conductivity in OmpF and aqueous solution but the lowest in AQ_1862, although the other alkaline ions follow the same order. The fluctuating part of AQ_1862 has a maximal conductance of approximately 250 pS in 100 mM KCl and saturates at rather low concentrations (K_m around 10 mM). Under the same conditions the single channel conductance for the OmpF monomer is 300 pS (Danelon *et al.*, 2003), however, the conductance does not saturate but increases linearly up to at least 1 M KCl (Danelon *et al.*, 2003). The fluctuating part is independend of pH regarding selectivity, conductance or saturation, however, the open probability decreases with increasing pH (data not shown). OmpF shows the opposite effect: the open probability is decreased at acidic pH (Basle *et al.*, 2004). AQ_1862´s fluctuating part shows a voltage dependent asymmetry, in one direction the conductance is one third smaller than in the other. A similar effect was described for OmpF at lower ion concentrations (100 mM KCl) (Alcaraz *et al.*, 2004) but the cause was not addressed. Assuming that this asymmetry is an intrinsic property of the protein this result indicates a preferential orientation of the incorporated porins. Possibly this asymmetry is due to differential distribution of charges at both ends of the pore leading to different ion concentrations at opposite openings of the pore which might result in different conductances.

The basic conductance shows a variety of peculiar features. The elementary conductance is rather high: 1.4 nS at 0.1 M KCl in combination with a relatively high selectivity indicated by zero current experiments ($P_{Cl}/P_K \approx 8:1$) at neutral pH (7.5). Other porins of comparable molecular weight like the anion selective PhoE and OprP have a much lower conductance and selectivity (Benz, 1984). With respect to the selectivity for anions the basic conductance of AQ_1862 lies between PhoE and OprP

(e.g.: P_{Cl}/P_K: PhoE ≈ 3, OprP >100) but the basic conductance is much higher than that of the classical porins: 1.4 nS vs. 0.2 nS for PhoE (Benz *et al.*, 1984) or 0.1 nS to 0.16 nS for OprP (Sukhan *et al.*, 1996). The order of conductivity for anions in the presence of K^+ is identical for PhoE (Benz *et al.*, 1984), AQ_1862 and aqueous solution (Robinson *et al.*, 1959): $Cl^- ≈ Br^- > F^-$. Porins with a comparable conductance are much larger (Maier *et al.*, 2001): a porin from *T. thermophilus* with a single-channel conductance of 2.5 nS in 0.1 M KCl has an apparent molecular weight of 185 kDa and is only moderate cation selective (P_{Cl}/P_K≈1:2.7). At pH 7.5 the basic conductance increases linearly with the concentration of KCl up to at least 0.5 M. Increase of pH leads to loss of selectivity (P_{Cl}/P_K≈1:1) in parallel with a decrease in the conductance and dramatic changes of the concentration dependence. At pH 10.5 the K_m for KCl is about 20 mM, compared to a K_m >> 0.5 M at pH 7.5. Saturation at low concentrations at pH 7.5 is also observed with K-acetate or other organic anions simultaneously with cation selectivity (P_{anion}/P_{cation}=1:8) and reduced conductance. The concentration dependence of the conductance of PhoE is comparable to that of the basic conductance at neutral pH: in the presence of KCl the conductance of PhoE increases linearly up to 0.5 M. On the other hand the conductance of OprP saturates with a K_m ≈ 50 mM, which is comparable to the concentration dependence of AQ_1862's basic conductance at basic pH (10.5). With respect to the concentration dependencies observed in AQ_1862 it may be of interest that in OprP the mutation of a single lysine into glutamate relieves saturation of OprP (Sukhan *et al.*, 1996): the concentration dependence of the conductance is linear up to 3 M KCl.

Intuitively this effect of large anions or high pH on the concentration dependence of the basic conductance is not easily understood if cations as well as anions move independently in the pore. One possibility is that there is interaction between cations and anions such that slowing down of anion translocation by size and/or modified interaction with charged residues also limits cation translocation. Another hint for the interaction between anions and cations is the reduction of the conductivity observed if K^+ is replaced by NMG. In 0.1 M KCl at pH 7.5, the basic conductance is around 1.4 nS. Replacing KCl with NMG-Cl reduces this conductance sevenfold to 0.2 nS (similar results were obtained with Tris-Cl). The profound effect of the size of the cation species also hints into the direction of an interaction between the flows of cations and anions. The concentration dependencies indicate that mobile anions as well as mobile cations are a pre-requisite to enable high translocation rates i.e. high conductivity.

At the moment it is not clear how these interactions are mediated. Im and Roux (Im *et al.*, 2002) described a model for ion translocation in 1 M KCl on the cation selective porin OmpF based on structural data. Here, K^+ and Cl^- follow two well separated pathways. There are indications that interactions between cations and anions occur and play an important role in the passage of anions. In this model passage of Cl^- only occurs together with K^+ as counter ions, while isolated K^+ ions can move independently. AQ_1862´s conductance pathway formed by the basic conductance seems to be

more complicated as both, cations and anions seem to need ions of antagonistic charge to achieve high translocation rates.

Since no structures of homologous molecules are available and predictions of β-barrel structures are still challenging an experimental determination of the three-dimensional structure of AQ_1862 is required to identify the amino acid(s) responsible for this effect. In addition, the investigation of channel properties of AQ_1862 carrying mutated amino acids would be helpful to understand the properties of the different conductances.

The different selectivity of AQ_1862 and especially its pH dependence observed in macroscopic experiments can be explained by the relative conductances and selectivities of the basic conductance and the fluctuating part: in the presence of KCl at neutral pH the basic conductance is more than five times higher than the fluctuating conductance and anion selective, therefore the macroscopic anion selectivity is determined by the basic conductance. In the presence of K-acetate and KCl at alkaline pH the basic conductance is in the same range as the fluctuating conductance and non-selective while the fluctuating conductance shows cation selectivity so that a macroscopic cation selectivity is observed. The selectivity is also influenced by the different open probabilities i.e. the macroscopic selectivity is averaged over time, but this question has not been investigated in detail. In addition, a similar effect of dependences of ion selectivity on pH was reported by Alcaraz and co-workers (Alcaraz *et al.*, 2004) for OmpF. This effect cannot be explained by Benz, who suggested that the different selectivity of VDAC in KCl and K-acetate is dependent on the combination of ions and their respective mobility in aqueous solution (Benz, 1994); therefore it is more likely that the selectivity is not determined by the properties of the different ions but by the protein molecule.

Although it cannot be definitively excluded that two channels or one channel in different states are observed, there are strong arguments in favour of a single molecule with two different conductance pathways. The single molecule hypothesis is based on SDS PAGE analysis which shows only one band, as well as by the statistics of the BLM measurements which show that both conductance components occur and disappear simultaneously (see Figure 2.2-14). Nonetheless the two "pores" differ by a number of properties, i.e. the pH dependences of selectivity and conductance, and the concentration and voltage dependence of conductance. Therefore we suggest that AQ_1862 does not just posses one pore-like structure which can adopt two different states but that it rather contains two different conductance pathways.

We could show that ion flux across the outer membrane of *A. aeolicus* is tightly regulated. The bacterium is able to discriminate ions and even react to changes in the environment before ions reach the periplasm. For example, when the external milieu becomes more alkaline the pores acquire cation selectivity and allow ions like NH_4^+ to pass and to neutralise the periplasm. Porins from evolutionary ancient bacteria have not been studied in detail so far, but obviously many features of AQ_1862 resemble the mitochondrial VDAC's more than the porins of the proteobacteria.

3.2.4 Towards Structure Determination

Since the first porin structures were published in 1990 and 1991 (Cowan *et al.*, 1992; Weiss *et al.*, 1991) a number of atomic structures of outer membrane proteins have been determined, most of them by X-ray crystallography. Initially native protein originating from the bacterial outer membrane was the major source for porin proteins (Cowan *et al.*, 1992; Nestel *et al.*, 1989). However, purification from the native source has certain disadvantages: only the most abundant proteins are applicable and introduction of artificial amino acids like seleno-methionine is challenging. Crystallographic phase determination has to depend on the binding heavy metals to the protein. Today heterologous production of the target protein becomes increasingly important. Porin proteins are produced in two ways: (1) by directing production into the outer membrane using a signal peptide resulting in a functional porin or (2) by removing the signal peptide and induce production in non-soluble form i.e. inclusion bodies.

In the case of AQ_1862, native membranes yielded sufficient amounts of pure and homogenous protein to pursue crystallisation. Crystals obtained in the presence of DDM diffracted X-rays to approximately 4 to 5 Å, which is before the lower limit for structure determination by X-ray crystallography. Broad screening of chemical and physical parameters could neither control crystal growth nor improve the crystal quality.

The chemical properties of the detergent are crucial for the success of a membrane protein crystallisation attempt, i.e. the size of the detergent micelle has to be optimised to shield the hydrophobic part of the protein. At the same time the detergent micelle has to allow crystal contacts, which are made in most cases by the hydrophilic domains of the protein. Therefore a variety of detergents was screened, in order to obtain well ordered three-dimensional crystals suitable for structure determination. So far most outer membrane porins have been crystallised successfully in the presence of alkylpolyoxyethylens, in most of the cases the octyl-derivatives like octyltetraoxyethylene. Therefore the detergents C_8E_4 and C_8E_n were chosen. Other detergents with a rather short alkyl chain like OG have been successful as well. In fact, crystals of high quality were obtained in the presence of NG, OG, and OTG in the space group R3 (chapter 2.2.4.1). Crystals of high quality were obtained from a rather broad range of conditions. The pH value could be varied from pH 4.0 to pH 10.0 with an optimum around pH 4.5 and above pH 7.0.

Crystals were obtained in many different conditions, however most three-dimensional crystals appeared in the presence of polyethylene glycols of low molecular weight (PEG 400, PEG 4000, and PEG 6000), and divalent salt ions (Ca^{2+}, Mg^{2+} or Zn^{2+}) either at neutral pH or at pH 4.6. Electrophysiological studies have shown that these are conditions in which the fluctuating part is not detectable (in the presence of divalent ions and at low pH). Especially the well ordered crystals in the presence of DDM and OG diffracting to 4 Å and 2 Å, respectively were obtained in those conditions. These data suggest that crystal quality better if the protein is at least partially inactive. Protein inactivation using inhibitors is an often followed strategy to obtain crystals or to improve crystal

quality. So far no inhibitor of the protein was identified, but it might help to further improve the crystal quality.

For structure determination molecular replacement was attempted. Proteins which were identified by fold recognition programs to be similar in fold but also unrelated porins were applied using a polyalanine model (Pauptit *et al.*, 1991). However, none of these resulted in a reasonable solution. Presumably AQ_1862 is not closely enough related to porins with known three-dimensional structures in terms of sequence (and fold) conservation. Therefore phasing with heavy atom derivatives was attempted.

Phasing was challenging for many outer membrane porins. Although they were among the first membrane proteins that yielded crystals at a resolution suitable for X-ray analysis (Garavito *et al.*, 1980), it took more than a decade until the first structure was solved (Cowan *et al.*, 1992; Weiss *et al.*, 1991). The two major drawbacks were that only few heavy atom derivatives bound to the protein and the few that did all had a centrosymmetric distribution. Thus any derived electron density would contain a superposition of both enantiomers of the protein and would thus be not-interpretable (Pauptit *et al.*, 1991). Phasing and subsequent structure solution was only possible through finding of an alternative crystal form.

For phase determination of AQ_1862 a variety of heavy atom derivatives were tested in co-crystallisation and soaking experiments. The addition of heavy atom derivatives reduced the crystal quality for some derivatives, others had no influence, and some even improved the quality. However, in the cases where data collection was possible the anomalous signal was not strong enough for phase determination. In structural proteomics where protein is obtained from the native organism, the possibilities to introduce heavy atom derivatives in a protein of interest are limited. They can only be added during or after purification. The introduction of non-native amino acids is restricted to production host that are auxotroph for methionine. These are difficult to obtain, especially in case of *A. aeolicus*, which is on one hand side difficult to grow and on the other autotrophic and maybe not capable of taking up and insert non-native amino acids. Therefore it was attempted to utilise the well established *E. coli* expression system for introduction of the non-native amino acid Se-Met in order to obtain selenium labelled protein crystals suitable for structure determination by anomalous dispersion experiments. The functional production of the protein in the outer membrane under the control of the *ptac* or the arabinose promoter was attempted because both promoter systems have been used successfully to produce outer membrane porins (Stauffer *et al.*, 1990; Subbarao *et al.*, 2006; Ye *et al.*, 2004). However, this approach was not successful due in part to premature lysis of the host cell leading to low yields of the target protein and impeding protein purification, and in part to degraded protein. The toxicity of produced porins for their host cell is a general problem limiting the production level and the range i.e. the source organism and thus source of proteins to be produced (Laage *et al.*, 2001). Therefore this approach remains often unsuccessful for major outer membrane proteins (Pullen *et al.*, 1995; Schmid *et al.*, 1996) even if an *E. coli* signal sequence and a strain deficient for the most

abundant outer membrane porins OmpF, OmpC and LamB is used. Functional overproduction in *E. coli* is usually most successful with outer membrane proteins from *E. coli* (Hong *et al.*, 2006; Subbarao *et al.*, 2006; Ye *et al.*, 2004) or proteins from γ-proteobacteria (Meng *et al.*, 2006). It has been so far unsuccessful for proteins from α-proteobacteria (Schmid *et al.*, 1996).

For bacterial outer membrane porins the production of inclusion bodies and subsequent solubilisation under denaturing conditions followed by refolding into the native state has proven successful. This method has often yielded high quality crystals for structure determination by X-ray crystallography (Bannwarth *et al.*, 2003). Therefore a protocol for refolding the protein from inclusion bodies into the native state was established. In most cases, protein production was attempted under the control of the T7 promotor in the *E. coli* host strain BL21(DE3) (Schmid *et al.*, 1996; Yildiz *et al.*, 2006). Solubilisation with urea was not successful, either qualitatively (monodisperse, non-aggregated) or quantitatively. Only guanidinium hydrochloride could quantitatively solubilise the protein. In addition it was crucial to retain only the wild type number of amino acids, as an additional methionine at the N-terminus and a his-tag at the C-terminus completely abolished refolding. Possible explanations for this behaviour include structurally important interactions of the N and C-termini, as seen with the *Rhodopseudomonas blastica* porin (Kreusch *et al.*, 1994; Schmid *et al.*, 1996) or a structurally disadvantageous influence of the hexa-histidine tag that was present in the initial construct. The protein was refolded by dilution into detergent containing buffer. The second crucial step was the choice of detergent that was identified in a broad screen. Once LDAO was identified as a detergent suitable for refolding the protein into its native state, a two-step purification protocol was established. Refolding was followed by monitoring the increase in secondary structure content with CD spectroscopy and the elution volume in gel filtration. A single gel filtration step following the refolding step was not sufficient to purify the protein because refolded and misfolded protein showed a very similar elution volume in gel filtration chromatography. However the protein which eluted from the anion exchange material comprised almost exclusively the refolded protein as judged by gel filtration chromatography and CD signal (see Figure 2.2-4 B and C).

Refolded AQ_1862 resembles the native in several points: the elution volume in gel filtration chromatography is almost identical, it is as heat stable as the native protein, and the minimum in CD spectroscopy is almost identical, thus the fold seems to be similar. However, many purifications show a decreased CD signal suggesting a fraction of unfolded or incompletely folded protein, the protein is less stable towards SDS denaturation and it is free of lipids. Analysis of the lipid content revealed an additional peak with an elution volume of 6.5 ml that might originate from LDAO that remained from the refolding buffer.

It was possible to obtain three-dimensional crystals in conditions similar to those used for the native protein which diffracted X-rays up to 5 Å. However the low crystal quality impeded X-ray data collection. In order to improve crystal quality the stability of the refolded protein has to be increased. In addition a broad screen for additives and usage of different detergents has to be tested.

3.3 AQ_1558

3.3.1 Similarity

The hypothetical protein AQ_1558 was identified in the solubilised membrane fraction of *A. aeolicus*. Unlike the other proteins analysed in this work, sequence alignments of AQ_1558 did not reveal strong similarity to any protein annotated in the database and hence sequence comparison was not helpful in the determination of a putative biological function.

Interpro domain searches showed similarity to an N-acetylglucosaminyl transferase domain, a protein involved in biosynthesis of glycosylphosphatidylinositol anchors (Leidich *et al.*, 1996), however there is no evidence that GPI anchors are present in prokaryota and therefore it is rather likely that this result is an artefact. The central domain contains two structural motifs, a TRP-like motif and a coiled coil motif; three stretches have been masked out due to low complexity. Both the TRP-like domain and the coiled coil motif appear in a variety of proteins. The TRP-like motif is a 34 amino acid helix-turn-helix motif (Main *et al.*, 2005). It is present in a number of proteins that are functionally unrelated and mediate a variety of different protein-protein interactions (Blatch *et al.*, 1999; D'Andrea *et al.*, 2003). The coiled coil is a structural motif of two or more α-helices wrapping around each other in a left-handed manner to form a supercoil (Mason *et al.*, 2004). Regions of low complexity have an unusual amino acid composition which might cause problems in sequence alignment programs and are therefore masked out (Wootton *et al.*, 1996).

Sequence alignment of AQ_1558 revealed similarity with very low scores towards very few proteins. Furthermore the result varied drastically dependent on the mode of the program used. Sequence alignment using BLAST was strongly dependent on the setting of the program, namely the compositional adjustments, a tool to compensate for different amino acid composition of different proteins. In the standard mode of the NCBI BLAST program AQ_1558 is similar to several hypothetical proteins and to a band 7 protein from *Anabaena variabilis*. As mentioned above, a member of the band 7 proteins has been identified in the membrane of *A. aeolicus* as well, the erythrocyte band 7 homolog (AQ_911). However, sequence conservation among band 7 proteins is usually very high, AQ_1558 is not well conserved. Similarity to several response regulator proteins was also detected. Response regulator proteins pass on a signal dependent on phosphorylation. However, the similarity is limited to a short stretch of amino acids and therefore it is questionable that AQ_1558 can conduct a similar function.

In the "not adjusted" mode of NCBI BLAST or with the standard ExPASy BLAST program, similarity of the central domain of AQ_1558 to additional proteins was detected: bassoon proteins from several mammals. Bassoon proteins, also known as zinc-finger proteins, are cytoplasmatic proteins located in active zones of nerve terminals in the brain (rat and mouse have been investigated) and play an important role in regulated neurotransmitter release from a subset of glutamatergic synapses (Altrock *et al.*, 2003; Brandstatter *et al.*, 1999; Ohtsuka *et al.*, 2002). However the similarity is again limited to

a stretch of approximately 100 amino acids and might not be of functional relevance for the large bassoon protein. The similar region in bassoon (approximately from amino acid 970 to 1090) comprises a coiled coil domain and two regions of low complexity just like the central domain of AQ_1558. The program might be "fouled" by the coiled coil motif rather than detecting true similarity, however the regions of low complexity were masked out. AQ_1558 also contains a zinc-finger like domain (see below) but close to the C-terminus, while the zinc-finger domains of bassoon are located in the N-terminal region. Still it is tempting to suggest that AQ_1558 conducts a similar function as bassoon i.e. regulation of vesicle or membrane fusion, however no experimental evidence is yet available.

3.3.2 Heterologous Production

Low sequence conservation would also agree with a classification as species- or genus specific protein (Galperin *et al.*, 2004), although there is no clear threshold given in the literature. Although there are many genomic sequences of bacteria known today the knowledge about extremophiles is still limited by the availability and feasibility of growing the organism. Homologous sequences might be detected in species which have not been investigated so far. From bioinformatic studies on AQ_1558, no clear conclusions can yet be drawn.

AQ_1558 was produced heterologously since its abundance in the native membrane was rather low. It was produced using different sets of affinity tags, vectors, and expression hosts. Production was obtained in all different combinations; interestingly highest production levels were reached with the A-construct in pTTQ18 but using the C-construct in pBAD. Vice versa using the C-construct in pTTQ18 and the A-construct in pBAD showed very low production levels. The full-length construct was prone to aggregation that could be improved but not avoided using lower temperatures during protein production (20 °C). Crystallisation might be hampered by flexible termini that are often provided by N- or C-terminal tags and proteolytic cleavage omitting these termini might be useful. Removing the C-terminal his-tag by TEV digestion of the pTTQ18-A construct was straightforward, while is could not be removed from the N-terminus in the C construct. Obviously the TEV site at the N-terminus is blocked, may be due the vicinity to the transmembrane helices and the detergent micelle, while the one at the C-terminus is freely accessible.

Prediction programs indicated two possible transmembrane segments near the N-terminus which provide membrane localisation of AQ_1558. The first transmembrane segment is predicted to be a signal peptide but ambiguous results were obtained concerning cleavage. In AQ_1558 produced in *E. coli* the signal peptide is clearly present as shown by Western blot and mass spectrometric analysis. The architecture of domains with the transmembrane domain close to the N-terminus suggests that AQ_1558 is anchored to the membrane and consist mainly of a soluble domain. Such proteins are easiest to handle if the membrane anchor is removed by a protease or genetically by deleting the coding sequence of the membrane helix while retaining the relevant biological activity of the soluble

domain. If the fragments are stable without detergent they can be treated and crystallised like soluble protein; this approach is often followed for structural analysis of receptor proteins. Such constructs of AQ_1558 were designed in a way to start protein production at several positions after the last predicted transmembrane domain. However, all constructs were localised in the membrane fraction. The association with the membrane is as strong as in the full-length construct; the protein's association to the membrane could not be broken unless the membrane was solubilised using detergent. Furthermore the buffers had to be supplemented with detergent during purification in order to keep the protein in a non-aggregated form. These data suggest strong interaction of the truncated version with the membrane as well, although no extended stretch of amino hydrophobic acids indicative for a transmembrane helix is detected in the central or the C-terminal domain. The protein might be associated to the membrane by one or more amphipathic or membrane inserted helices or by a covalently attached hydrocarbon tail. Monotopic membrane proteins like squalene cyclase or the fatty acid amide hydrolase are typically associated with the membrane without having transmembrane helices (Balliano *et al.*, 1992; Patricelli *et al.*, 1998; Seckler *et al.*, 1986).

Another approach to obtain rigid protein suitable for crystallisation attempts is to determine the domain structure by limited proteolysis. However it was not possible to identify stable domains, the unprocessed protein was the most stable form.

The oligomeric state of AQ_1558 was estimated by gel filtration chromatography and comparison of the elution volume with standard molecular weight marker proteins. The determined molecular weight of native and heterologously produced AQ_1558 would agree with a trimeric or tetrameric form. The N-terminally truncated version does not contain transmembrane domains, however it was only stable in the presence of detergent and therefore it can be assumed that detergent was bound to the protein. Indeed the elution volume is only slightly shifted compared to native AQ_1558, suggesting a trimeric or tetrameric species with addition of a detergent micelle.

By SDS PAGE analysis a dimeric form of the protein becomes obvious, in cross-linking experiments with glutaraldehyde two additional bands of putative trimers and tetramers appear, supporting the result obtained from gel filtration. The size of the protein determined from BN PAGE analysis suggests a trimer with oligomers of the trimer. The oligomerisation is obviously not conducted by the N-terminus of the protein since the N-terminally truncated form shows a very similar behaviour. It was not possible to determine exact oligomeric state but the data suggest a trimer or tetramer.

AQ_1558 contains four cysteines close to the C-terminus arranged in a motif of two cysteines separated by two amino acids XX (CXXC). Proteins containing this motif are often involved metal binding, e.g. zinc, mercury, cupper, iron, etc., in redox-reactions by promoting the formation, reduction or isomerisation of disulphide bonds (thiol/disulphide oxidoreductases) or reduction of hydrogen peroxide (non-heme peroxidase (Poole, 2005)). The presence of metal ions in the purified protein could indeed be shown by TXRF, per monomer of protein 1.4 Zn are present.

Proteins that bind metal ions for their function (metalloproteins) make up a large fraction of any proteome. The metal ion may be needed for the catalytic function of the proteins of to stabilise the tertiary or quaternary structure. To determine a proteins´ function it could be beneficial to know the metal content. However the annotation of zinc binding domains is still challenging (Andreini *et al.*, 2006). AQ_1558 putative zinc binding domain ($CX_{(2)}CX_{(10)}CX_{(2)}C$) is very similar to Pfam annotated domains, however the linker region of 10 amino acids in AQ_1558 vary between 8 and 25 in proteins described in the literature (Andreini *et al.*, 2006). Zinc has essentially two different roles: catalytic like in metal dependent peptidases or structural like in zinc-finger. Zinc-fingers are often involved in DNA interaction, for example transcription factors. In addition to the two CXXC motifs, the C-terminal domain shows weak similarity towards DNA binding proteins, suggesting a zinc-finger like domain in AQ_1558. However the total amount of phosphorous determined by TXRF does not suggest that DNA is present in the sample. Further experiments should clarity if the protein is indeed able to bind DNA.

3.4 AQ_1760

3.4.1 Similarity and Function

In the solubilised membranes of *A. aeolicus* a homo-oligomeric complex consisting of the "hypothetical protein" AQ_1760 was identified. By sequence comparison, the Linocin_M18 domain was identified as the major component. The sequence identity to characterised proteins is rather low, below 30 %, which is most likely the reason why the protein is annotated as a "hypothetical protein" in the databases. However, bioinformatic evidence is given that it is a linocin_M18/bacteriocin homolog.

Bacteriocins are a class of peptides, proteins, and protein complexes which have an antimicrobial effect towards other organisms than the host. A subfamily of the bacteriocins is the Linocin_M18 domain containing family proteins which is present in bacteria and archaea. It originates from the protein Linocin M18 from *Brevibacterium linens* which was described by Valdes-Stauber to inhibit the growth of several Gram-positive bacteria from the same habitat.

AQ_1760 was purified from the membrane fraction of *A. aeolicus*. It is absent in the cytosol and in the culture filtrate. The localisation of the homologous proteins linocin M18 and CFP29 is in the culture filtrate and the membrane fraction, while bacteriocin and vlp are purified from the whole cell lysate. A function similarity to that of bacteriocin seems only feasible if the protein is released to the culture medium or bound to the outside of the membrane. In the case of CFP29, it was suggested that it is exported and gradually released from the cell envelope (Rosenkrands *et al.*, 1998). Such a mechanism might be applied by *A. aeolicus* as well. AQ_1760´s absence from the culture medium might be a matter of culture condition. However identifying conditions under which the organism should produce a bacteriocin is very challenging for *A. aeolicus* since not much is known about its habitat and its putative antagonistic bacteria.

One of the closest homologs, bacteriocin from *T. maritima*, has been shown to act as a protease. The proteolytic activity of AQ_1760 was determined and compared to bacteriocin. In zymograms a band was observed which fits the size of AQ_1760. AQ_1760 is obviously able to cleave gelatine even after denaturation in SDS sample buffer. These data suggest that the single subunit is able to conduct the cleavage because refolding of the denatured protein to form the complex tertiary structure within the acryl amide mesh seems rather unlikely, however it was not examined. In the peptide based assay AQ_1760 behaves very similar to bacteriocin. Bacteriocin has a much broader specificity, while activity in AQ_1760 could only be determined against one peptide, VKM-AMC. The specific activity of AQ_1760 is 8.0 nmol*min^{-1}*mg^{-1}, similar to the activity determined for bacteriocin. However, both assays are not specific for AQ_1760 but for proteases in general. Several other fractions showed activity in the peptide based assay and thus might contain other proteases. Therefore no enrichment of proteolytic activity could be observed during purification. A screen against a wider range of peptides and proteins in a solution assay should be established to determine the specificity of AQ_1760 further. The function of other homologous proteins was not studied in detail.

AQ_1760 shows homology to the carboxy-terminal domain of vlp from *P. furiosus*, whose structure was determined recently (Akita *et al.*, 2007). In the crystal structure the amino-terminal 109 amino acids are disordered and thus the structure only contains the C-terminal domain ($\approx$ 200 amino acids), that is similar to AQ_1760 and the Linocin_M18 domain. Although a conserved domain search clearly predicted a Linocin_M18 domain in vlp as well (from amino acid 160 to 345), the authors do not comment on this similarity and thus did not provide functional information related to growth inhibition or proteolytic function.

Testing functional similarity i.e. the growth inhibition is very challenging in hyperthermophilic microaerophilic bacteria and needs to be carried out in the future.

3.4.2 Towards Structure Determination

AQ_1760 resembles not only by sequence similarity and but also by morphology the Linocin_M18 and of the homologs in *T. maritima* (bacteriocin), *P. furiosus* (vlp) and *M. tuberculosis* (CFP29) (Hicks *et al.*, 1998; Namba *et al.*, 2005; Rosenkrands *et al.*, 1998; Valdes-Stauber *et al.*, 1994): all form large complexes. Electron microscopic studies showed either spherical or elongated particles: vlp and linocin M18 form spherical particles with a diameter of 30 nm (see also Figure 3.4-1) and 20 to 30 nm, respectively; bacteriocin is tube shape with dimensions of 20 nm times 60 nm to 80 nm.

Gel filtration chromatography on a Superose 6 column demonstrated that the protein forms a complex of more than 700 kDa, similar to the homologous proteins (Hicks *et al.*, 1998; Rosenkrands *et al.*, 1998; Valdes-Stauber *et al.*, 1994). From gel filtration studies on Sephacryl-500 HR the complex appears to be very large, between 0.7 and 20 MDa. The column could not be calibrated due to the lack of standard proteins of matching size, however the resin has been frequently used to separate DNA

(Gingeras *et al.*, 1987), RNA-protein complexes (Hannan *et al.*, 1999), glycoproteins or a filament-forming protein with monomers of 13 nm diameter (Weaver *et al.*, 1992).

Particles observed by electron microscopy confirm a molecular size of more than 700 kDa. At the same time the homogeneity of the sample was proven. The apparent shape of AQ_1760 resembles a spherical structure with an outer diameter of 38 nm and a particle thickness of approximately 1.5 nm to 2 nm. The surface is obviously made out of multiple small subunits. It is no smooth layer but resembles a cage or mesh of individual particles that silhouette against a hollow interior (see Figure 2.4-5).
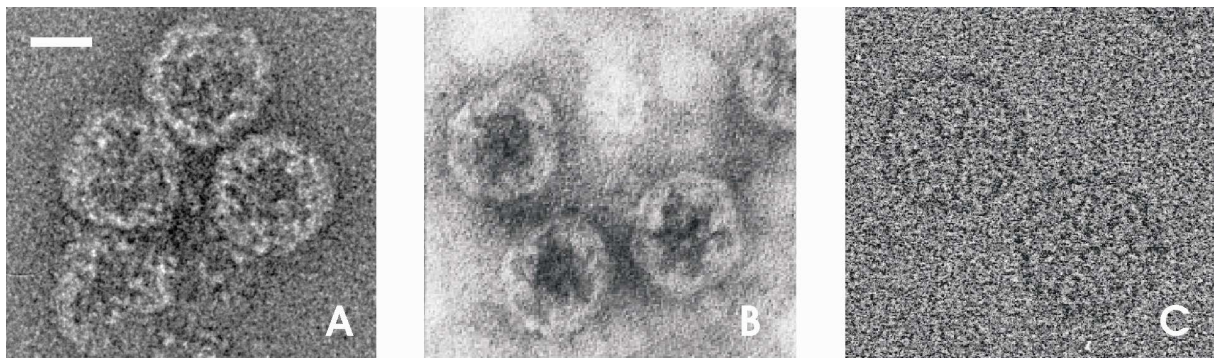
Figure 3.4-1: Comparison of the appearance in electron micrographs from negative stained samples from A: AQ_1760 from *A. aeolicus*, B: vlp from *Pyrococcus furiosus* (Namba *et al.*, 2005) and C: gp5 prohead (Conway *et al.*, 1995). The scale bar represents 20 nm.

From the particle dimension the molecular mass of the particle can be estimated. Assuming a fully assembled hollow sphere of AQ_1760, the molecular mass would be in the range of 5.1 MDa and 6.6 MDa; depending on the thickness of the wall and the partial specific volume. One particle would comprise approximately 160 to 207 subunits of AQ_1760. The homologous protein vlp is made out of 180 subunits; its molecular mass is approximately 7 MDa. In two-dimensional image reconstructions, hollow and filled particles can be distinguished (Figure 2.4-5 D). The outer shape dominates the reconstructions and detailed information on the shape of the protrusions is not present. However the particles appear to be somewhat larger than the particle on individual images, more to 40 nm in diameter. The surface appears smoother and the thicker, approximately 3 nm. The molecular mass of such particle would be 9.2 MDa and 290 subunits.

The mass determined for AQ_1760 is in a similar range but larger than the size of the smallest species determined by means of analytical ultracentrifugation. In the presence of detergent a size of 70 S ($\approx$ 3.5 MDa equivalent to $\approx$ 120 subunits) was detected. However both techniques have constraints. Size determination from electron micrographs is rather imprecise. From negative stained samples the surface of the protein can only be estimated but not directly detected, because only the heavy atom at the surface but not the protein itself is visualised. In this case it was very difficult to determine an exact value for the thickness of the surface. In addition the surface is mesh-like, so estimating the number of monomers is only a rough approximation and the free space between the individual

monomers should be taken into account. At the same time both methods are only precise if the partial specific volume is known. Analytical ultracentrifugation is hampered by the purity of the sample; inhomogeneity is the source of multiple species hampering the fitting of the c(s) distribution. More exact determination of the size can only be obtained from more homogeneous samples. These should be accomplished with sedimentation equilibrium experiments.

Single particles and particles from two-dimensional image reconstruction were variable in appearance, Figure 2.4-5 C provides an overview on the particle shape, some appear hollow, some filled and some only partially assembled. The partial assembled particles are most likely damaged particles or particle under construction. Occasionally in the electron microscopy studies particles are located very close together, suggesting the formation of oligomers. Analytical ultracentrifugation experiments in the absence of detergent supported this suggestion: complexes of higher but distinct masses were detected, suggesting an oligomerisation of the 70 S complexes. Together with the particles a non-negligible amount of amorphous material was detected in the electron micrographs; however gel filtration experiments showed no material of smaller molecular weight. Taken together the inhomogeneity of the sample impeded the exact determination of the molecular mass by means of analytical ultracentrifugation and could also be a major drawback in obtaining well-ordered three-dimensional crystals suitable for structure determination.

Most important, it was clearly shown that the oligomeric state is not a protein aggregate but an ordered structure. This was supported by the determination of the crystal structure of vlp from *P. furiosus*. The vlp is a spherical particle that resembles a virus or phage-like structure. It consists of 180 copies of the 38.8 kDa monomer in the icosahedral arrangement with a triangulation class of T=3 i.e. it is made up of 12 pentamers and 20 hexamers of the vlp protein. Most spherical viruses or phages, including HK97, exhibit an icosahedral symmetry varying in the triangulation class and thus the number of subunits. In addition to the similarity to vlp the sequence of AQ_1760 showed similarity to the fold of the phage protein gp5, the major capsid protein from the bacteriophage HK97.

Gp5 on the other hand is of viral origin, the capsid of HK97 is made out hexamers and heptamers of gp5 in the icosahedral symmetry T=7. Assuming an icosahedral arrangement of AQ_1760 monomers with T=1, T=3, T=4 or T=7 and 120, 180, 240 and 420 subunits the size of the particle should be 3.8 MDa, 5.7 MDa, 7.6 MDa or 13.3 MDa respectively. This size calculated from electron micrographs could account for particle with a triangulation class of T=3 with 180 subunits or T=4 with 240 subunits, while the size determined from analytical ultracentrifugation resembles more a particle with a triangulation class of T=1 with 120 subunits. The correct number of subunits remains to be discovered from electron microscopic image reconstructions of higher resolution or from the crystal structure. Gp5 and AQ_1760 are unrelated at the sequence level, but according to the results of fold prediction the polypeptide chains of gp5, vlp, and AQ_1760 have a similar fold. Although the crystal structure of AQ_1760 could not be elucidated and the fold compared to verify the prediction, we could

show by electron microscopy that the two proteins indeed have a very similar quaternary structure, suggesting that the fold is indeed similar.

Unlike gp5 and virus-like particle AQ_1760 cannot be expressed in a native-like form in *E. coli*. This might be due to the lack of matching assembly factors.

The function of gp5 is packaging of the phage DNA, however, it also able to assemble to its icosahedral form if gp5 is produced alone without the other viral proteins. Vlp on the other hand is encoded by the genome of *P. furiosus* and does not contain DNA (Namba *et al.*, 2005). Akita and co-workers suggested that it is a relict of a bacteriophage remaining in the archaean genomes. Like other viruses and bacteriophages, it might be able to propagate within the host in a stable carrier state and under certain conditions become lytic (Akita *et al.*, 2007).

It remains to be discovered if the two proteins are also functionally related i.e. if AQ_1760 is involved in packaging.

3.4.3 Cofactors

The AQ_1760 complex was analysed for the presence of putative cofactors. UV VIS spectra revealed an absorption peak at 420 nm suggesting the presence of heme. TXRF measurements revealed the presence of two molecules of iron per monomer of protein; all other analysed elements are present in substoichiometric amounts. Calculating the concentration of the heme revealed its substoichiometric presence. However the absorption coefficients were taken from literature for *b*-type hemes but in the case of AQ_1760 the coefficient might be considerably different. Tight association of the putative heme absorption with the protein complex could be shown by co-elution of both 280 nm and 420 nm in the gel filtration chromatograms (Figure 2.4-3) as well as the same particle size of the 280 nm and 420 nm absorbance in analytical ultracentrifugation (Figure 2.4-4). The protein might have acquired heme by attachment or by trapped free heme within the sphere.

In conserved domain search homologous sequences the virus-like particle from *P. furiosus* or the conserved hypothetical (diheme) protein from *Candidatus Kuenenia stuttgartiensis* contain an additional amino-terminal ferritin-like domain (amino acids 25 to 75 in vlp), which shows only weak homology towards AQ_1760 (Figure 2.4-1). This domain is located in the inside of the particle, however these residues are disordered in the crystal structure and no iron or metal was reported to be present in vlp (Akita *et al.*, 2007). The domain structure of vlp suggests a particle, which is formed by the carboxy-terminal domain, utilised to pack or shield the amino-terminal domain, which carried the ferritin-like domain. The protein might be used to store or transport iron. AQ_1760 might have a similar domain organisation: a carboxy-terminal particle forming domain, that is to the carboxy-terminal domain of vlp and an amino-terminal iron binding domain. However, it remains to be discovered from the crystal structure how the AQ_1760 particle is formed and how iron binds to AQ_1760.

A protein that combines the two features of AQ_1760, proteolytic activity and iron binding, is the haemoglobin protease in the pathogenic *E. coli* Strain EB1. It shows proteolytic activity and the ability to bind iron (Otto *et al.*, 1998). It was proposed to use haemoglobin as an iron source by binding and degrading the haemoglobin while the released iron in order to internalise it and make use of it for its own needs. AQ_1760 might also lyse iron containing proteins and store or transport the released iron. However, the two sequences are unrelated and haemoglobin protease does not form large particles.

3.5 Conclusion and Outlook

The membrane proteomic approach presented here, demonstrates a very simple and efficient way to analyse the membrane proteome of an organism. In order to complete the membrane proteomics approach, the fraction 20 to 40 of the MonoQ chromatography shall be investigated in detail. The number of identified membrane proteins could be increased in three ways: different detergents can be screened for their efficiency to solubilise membrane proteins, other separation methods can be applied to attain better separation, and a more efficient way in protein identification, like tandem mass spectrometry, can be utilised.

A variety of membrane proteins and protein complexes have been identified and could be selected for further detailed studies: the putative outer membrane porin AQ_1259 or other hypothetical proteins. In addition it was shown, that not only proteins of very high abundance can be purified and characterised subsequently, but also proteins of lower abundance. AQ_1760 gives a good example of a protein that was purified, characterised, and crystallised, although only microgram amounts were extracted from the membrane. If purification of the low abundance membrane protein fails, characterisation can still be achieved by heterologous production in *E. coli*. This has been shown to be successful for many *A. aeolicus* membrane proteins (Surade *et al.*, 2006). Thereby the study can be extended to proteins like the erythrocyte Band 7 homolog, AQ_389, AQ_1436 or AQ_1439.

The major advantage towards conventional proteomics is given with AQ_1862 as example: functional and structural characterisation was possible immediately after identification and selecting it as target protein because the protein was already present in almost pure form. In any conventional proteomics study no further information than bioinformatic prediction would be provided since the protein is not directly available in its pure and native form. At the same time the example of AQ_1862 demonstrates the ideal case of functional proteomics: in consecutive steps from identification via bioinformatic analysis to confirmation by functional studies and to structural characterisation. It was identified as a "hypothetical protein" of unknown function; bioinformatic analysis revealed its putative function as a porin protein originating from the outer membrane. The function was confirmed by electrophysiological studies and at the same time structural investigation was begun.

Many features of bacterial outer membrane porins can be ascribed to AQ_1862: properties derived from the sequence like secondary and tertiary structure prediction, biochemical properties like its high abundance, a mainly β-sheet structure, the resistance towards proteolytic digest and its trimeric form

and electrophysiological properties like increase of lipid bilayer conductance and single channel recordings. These data strongly suggest that AQ_1862 is an outer membrane porin, the major outer membrane porin of *A. aeolicus*. Although many outer membrane porins have been identified and functionally characterised in detail, this work reveals that unexpected properties can be attributed to bacterial outer membrane porins. Electrophysiological characterisation demonstrates the existence of two distinct conductances with very different properties regarding ion-selectivity, conductance, pH-dependence and concentration dependence, strongly suggesting two separate pathways for anions and cation in the protein.

The three-dimensional structure will be of outstanding importance in understanding these electrophysiological properties of the protein. Site-directed mutagenesis and electrophysiological characterisation of single amino acid mutants will lead to better understanding of the ion conductance and selectivity of AQ_1862. Therefore genetic manipulation of AQ_1862 is an asset. Crystallisation of AQ_1862 was successful in a variety of detergents and conditions; in the presence of OG crystals diffracted X-rays better than 2 Å. However, up to now it was not possible to elucidate the three-dimensional structure because of missing phase information, although a broad range of heavy atom derivatives has been tested in soaking and co-crystallisation. In order to obtain phase information from selenium labelled protein, a recombinant production system using *E. coli* as host was introduced. Native production into the outer membrane was not possible. However, production of non-soluble inclusion bodies and refolding into the native state was achieved. Crystals were obtained in the presence of DDM that diffracted X-rays up to 5 Å.

In order to solve the structure three approaches should be followed: screening more heavy atom derivatives using the crystal form grown in the presence of OG, improving the quality of the crystals derived from heterologously produced protein and further screening for an alternative crystal form that allows phase determination using heavy atom derivatives.

This study demonstrates that AQ_1760 is a linocin/bacteriocin of *A. aeolicus* resembling the function and appearance of other bacteriocins of the Linocin_M18 domain family described in the literature. The proteolytic activity is comparable to the activity determined for bacteriocin from *T. maritima*. The main function of homo-oligomeric bacteriocins was postulated to be the growth inhibition of other bacteria. Still the mechanisms and the targets remain unclear. Electron microscopy showed a ball or virus-like structure: a sphere of 38 nm diameter, similar to proteins that are homologous in sequence or fold, however, no function could be determined from the three-dimensional structure of the homolog vlp. A unique finding- not described for bacteriocins, is the iron content of the AQ_1760 complex.

AQ_1558, a hypothetical protein, was identified in the membrane of *A. aeolicus*. Sequence alignment suggested that it could be a species specific protein. Over-production in *E. coli* of the full-length and a truncated form was achieved. It could be shown that not only its N-terminal domain but also the remaining part is associated with the membrane; most likely it forms a trimer or tetramer; in addition,

it contains zinc, most likely complexes with four cysteines that are arranged in two CXXC motifs near the C-terminus. The biological function of both AQ_1760 and AQ_1558 remain unclear.

In conclusion this approach is valid to characterise membrane protein functionally and structurally, however for most membrane protein the native abundance is too low thus supplementation by an overproduction system is necessary to broaden the list of targets.

4 Materials and Methods

4.1 Materials

4.1.1 Suppliers

Table 4.1-1: Supplier list.

Company	Location	Web address
Aldrich, Sigma-Aldrich Laborchemikalien GmbH	Seelze, DE	http://www.sigmaaldrich.com
Amersham Biosciences	Freiburg, Germany	http://www4.amershambiosciences.com/
Anatrace	Maumee, USA	http://www.anatrace.com/
Atomika Instruments	Oberschleissheim, DE	http://www.atomika.com/
Axon Instruments (Molecular devices)	Sunnyvale, USA	http://www.moleculardevices.com/
Applied Biosystems	Framingham, MA, USA	http://www.appliedbiosystems.com/
Avanti Polar Lipids	Birmingham, USA	http://www.avantilipids.com/
Bachem	Bubendorf, CH	http://www.bachem.com/
Beckman Coulter	Krefeld, DE	http://www.beckman.com/
Bioline	Luckenwalde, DE	http://www.bioline.com
Biomers.net	Ulm, DE	http://www.biomers.net/
Biometra	Göttingen, DE	http://www.biometra.de/
Biomol	Hamburg, DE	http://www.biomol.de/
Bio-Rad Laboratories GmbH	Munich, DE	http://www.bio-rad.com/
Branson	Danbury, USA	http://www.bransonultrasonics.com/
Bruker	Rheinstetten, DE	http://www.bruker.de/
Cambrex	Rockland, USA	http://www.cambrex.com/
Cartesian Dispensing Systems,	Huntingdon, UK	http://www.cartesiantech.com/
Eppendorf	Hamburg, DE	http://www.eppendorf.com/
Fermentas	St. Leon-Rot, DE	http://www.fermentas.de/
Finnzymes	Espoo, FIN	http://www.finnzymes.com
Fluka, Sigma-Aldrich Laborchemikalien GmbH	Seelze, DE	http://www.sigmaaldrich.com
GE Bayer Silicones	Bergen op Zoom, NL	http://www.gesilicones.com/
GE Healthcare	Munich, DE	http://www.gehealthcare.com/dede/index.html
GERBU Biochemical Mart	Gaiberg, DE	http://www.gerbu.de/
Gilson / ABIMED, Gilson	Bad Camberg, DE	http://www.gilson.com/
Glycon Biochemicals	Luckenwalde, DE	http://www.glycon.de/
Hampton Research	Aliso Viejo, USA	http://www.hamptonresearch.com/
Heraeus	Hanau, DE	http://www.wc-heraeus.com/
Jena Bioscience GmbH	Jena, DE	http://www.jenabioscience.com/
IKA Werke	Staufen, DE	http://www.ika.de/
Infors AG	Bottmingen-Basel, CH	http://www.infors-ht.com/
Invitrogen	Karlsruhe, DE	http://www.invitrogen.com/
Knick	Egelsbach, DE	http://www.knick.de/
Koch-Light	Haverhill, GB	http://www.chemicalregister.com/Koch-Light_Ltd/Supplier/sid1639.htm
MAGV	Rabenau, DE	http://www.magv-gmbh.de/
Memmert	Schwabach, DE	http://www.memmert.com/
Merck/VWR International	Darmstadt, DE	http://de.vwr.com/app/Home
Merck-Hitachi Ltd.	Tokyo, JP	http://www.hitachi.com/
Messer Griesheim	Krefeld, DE	http://www.messergroup.com/
Mettler-Toledo GmbH	Giessen,DE	http://www.mt.com/
Millipore	Schwalbach, DE	http://www.millipore.com
Molecular Dimensions Ltd.	Soham, UK	http://www.moleculardimensions.com/
New England Biolabs GmbH	Frankfurt/Main, DE	http://www.neb.com/
Novagen/ Merck Chemicals Ltd	Darmstadt, DE	http://www.merckbiosciences.de
NUNC GmbH	Wiesbaden, DE	http://www.nunc.de/
Olympus	Hamburg, DE	http://www.olympus.de/
Pierce Biotechnology	Rockford, USA	http://www.piercenet.com/
Perbio Science D. GmbH)	Bonn, DE	http://www.perbio.com/
Promega	Madison, USA	http://www.promega.com
Qiagen	Hilden, DE	http://www1.qiagen.com/
Rainin Instrument LLC (Mettler Toledo)	Giessen, DE	http://www.rainin.com/
Roche Diagnostics GmbH,	Mannheim, DE	http://www.roche.de/
Applied Science	Mannheim, DE	http://www.roche.de/
Riedel De Haen AG	Seelze, DE	http://www.riedeldehaen.com/
Roth, Carl Roth GmbH + Co. KG	Karlsruhe, DE	http://www.carl-roth.de/
SANOclav	Bad Überkingen-Hausen, DE	wolf-sanoclav-macryl@t-online.de
Samou GmbH	Radolfzell, DE	http://www.moulinex.de/
Sartorius AG	Göttingen, DE	http://www.sartorius.de/
Stanford Research Systems	Sunnyvale, USA	http://www.thinksrs.com/
SeqLab	Goettingen, DE	http://www.seqlab.de/
SERVA Electrophoresis GmbH	Heidelberg, DE	http://www.serva.de/

Sigma-Aldrich Laborchemikalien GmbH	Seelze, DE	http://www.sigmaaldrich.com
Stratagene	La Jolla, USA	http://www.stratagene.com/
TOSOH Bioscience	Montgomeryville, USA	http://www.tosohbioscience.com/
Vivascience	Hannover, DE	http://www.vivascience.com/
Zeiss, Carl Zeiss AG	Oberkochen, DE	http://www.zeiss.de/

4.1.2 Chemicals

The chemicals which have been used are listed in the following table.

Table 4.1-2: Chemical list.

Name	Supplier
acetic acid (CH_3COOH)	Roth
acetonitrile	Roth
acrylamid/ bisacrylamid	Roth
agar agar	Roth
Agarose *SeaKem®LE Agarose*	Cambrex
Al´s oil	Hampton
ammonium carbonate $(NH_4)_2CO_3$	Fluka
ammonium chloride NH_4Cl	Merck
ammonium hydrogencarbonate NH_4HCO_3	Fluka
ammonium peroxodisulphatate (APS)	Koch-Light
ampicillin	Gerbu
BCA™ *Protein Assay Reagent* A+B	Pierce
Benzamindine	Sigma
boric acid	Merck
5-bromo-4-chloro-3-indolyl phosphate (BCIP)	Roth
bromphenolblue	Roth
caesium acetate (CH_3COOCs)	Sigma
caesium chloride (CsCl)	Biomol
calcium chloride ($CaCl_2$)	Merck
casein hydrolysat	Sigma
di-chloro-di-methylsilane	Aldrich
tri-chloroethane	Aldrich
chloroform	Roth
Coomassie Brilliant Blue R250 and G250	Serva
α-cyano-4-hydroxy-cinamicacid (CHCA)	Bruker
N-cyclohexyl-3-aminopropanesulfonic acid (CAPS)	Gerbu
n-decane	Roth
n-decyl-β-D-maltoside (DM)	Glycon
diphytanoyl phosphatidylcholine	Avanti Polar Lipids
dNTP-Mix (2,5 mM; 10 mM)	Bioline
n-dodecyl-β-D-maltoside (DDM)	Glycon
ethylene diamine tetra-acetic acid, di-sodium salt (EDTA)	Gerbu
ethanol (p. a.)	Roth
ethyl acetate	Fluka
formic Acid (HCOOH)	Merck
FOS-CHOLINE-12 (Fos-12)	Anatrace
FOS-CHOLINE-16 (Fos-16)	Anatrace
glycerol (99.5 %)	Gerbu
glycin	Gerbu
guanidinium Hydrochloride	Fluka
4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES)	Gerbu
heptane-1,2,3-triol (>98 %)	Fluka
hexane-1,2,3-triol (>98 %)	Fluka
n-hexane	Roth
hydrochloric acid (HCl)	Roth
imidazole	Merck
isopropanol (p.a.)	Roth
isopropyl-β-D-thiogalactopyranosid (IPTG)	Gerbu
kanamycin	Gerbu
n-lauryl-N,N-dimethylamine-N-oxide (LDAO)	Fluka

lithium acetate (CH$_3$COOLi)	Sigma
lithium chloride (LiCl)	Sigma
2-mercaptoethanol	Fluka
methanol (p. a.)	Roth
2-methly-2,4-pentandienol (MPD)	Aldrich
methyl amine tungstate	Nanoprobes
2-(N-morpholine)-ethane sulphonic acid (MES)	Gerbu
Nanovan	Nanoprobes
nickel chloride (NiCl$_2$*6H$_2$O)	Riedel De Haen
nitro blue tetrazolium (NBT)	Biomol
Nitrocellulose Membrane	Millipore
nitrogen (liquid)	Messer
octyl-2-hydroxy ethylsulfoxide (C$_8$HESO)	Bachem
n-nonyl-β-D-glucopyranoside (NG)	Biomol
n-octyl-β-D-glucopyranoside (OG)	Glycon
1-S-n-octyl-β-D-thioglucopyranoside (OTG)	Glycon
octyltetraoxyethylene (C$_8$E$_5$)	Bachem
octylpolyoxyethylene (C$_8$E$_n$)	Bachem
PEG 400 – PEG10000	Fluka
phosphotunstic acid	Ted Pella
potassium acetate (CH$_3$COOK)	Gerbu
potassium bromide (KBr)	Sigma
potassium chloride (KCl)	Merck
potassium di-hydrogenphosphate trihydrate (KH$_2$PO$_4$*3H$_2$O)	Roth
di-potassium hydrogenphosphate trihydrate (K$_2$HPO4*3H$_2$O)	Roth
potassium fluoride (KF)	Sigma
potassium formiate (COOK)	Sigma
potassium gluconate	Sigma
potassium hydroxid (KOH)	Roth
potassium propionate (C$_2$H$_5$COOHK)	ABCR
potassium sulphate (K$_2$SO$_4$)	Sigma
propanol (p. a.)	Roth
rubidium acetate (CH$_3$COORb)	Sigma
rubidium chloride (RbCl)	Gerbu
silicon oil	GE Bayer Silicones
SimplyBlue SafeStain,	Invitrogen
sodium dodecyl sulfat (SDS)	Gerbu
sodium acetate (CH$_3$COONa)	Merck
sodium azide (NaN$_3$)	Sigma
tri-sodium citrate dihydrate	Merck
sodium chloride (NaCl)	Gerbu
sodium gluconate	Sigma
sodium hydroxid (NaOH)	Gerbu
sulfuric acid (H$_2$SO$_4$)	Roth
n-tetradecyl-β-D-maltoside (C$_{14}$M)	Glycon
N,N,N,N-tetramethyl-p-ethylendiamid (TEMED)	Roth
trifluoric acid (CF$_3$COOH, TFA)	Merck
tris-hydroxymethyl-9-aminomethane (Tris)	Roth
Triton X-100	Gerbu
Tryptone	Merck
Tween 20	Gerbu
uranyl acetate	Ted Pella, USA
urea	Merck
yeast extract	Merck

4.1.3 Equipment

The equipment which has been used is listed in the following table.

Table 4.1-3: Equipment list.

Device	Type	Supplier
Agarose gel imaging station	DNA gel documentation system (UV) (+ PC maxdata)	Bio-Rad
Äkta	Purifier 10/ Purifier 100	Amersham Biosciences
Amplifier (BLM)		home-made, MPIBP
Analytical Balance	R180D -*D1	Sartorius, Göttingen
Autoclave	V150	Systec
Autoclave	La-VA-MCS	SANOclav
Autoclave	DSL 676-1-FD	Holzner
Balance	PM 4600 Delta Range	Mettler
CD-Spectrometer		
Centrifuge	Sigma 4K10	Sigma
Centrifuge (6x1 liter)	Avanti J20 XPI	Beckman Coulter
Centrifuge	Centrifuge 5415 R	Eppendorf
Centrifige tubes 1 l		Beckman Coulter
Concentrators	Amicon cells, 50 ml and 250 ml capacity	Millipore
Concentrators	Centriprep	Millipore
Concentrators	Vivaspin, 500 µl tube	Vivascience
Cuvette (BLM)		home-made, MPIBP
Crystallization Robot	Synquad Dispensing System	Cartesian Technologies
Digitizer	MiniDigi 1A	Axon Instruments
Electron microscope	Tecnai1 2	Philips
Elektrophorese chamber		home-made, MPIBP
Electrophoresis chamber	XCell SureLock	Invitrogen
Electrophoresis power supply		Bio-Rad
Electrophoresis plates	Glass Plates for Hoefer	Pharmacia Biotech
Elisa reader	PowerWave X	Bio-Tek Instruments
French-press (+ chamber)	SLM-AMINCO	SLM Instrum., Inc.
Flat bed scanner		
Gel drier		Biometra
Heating block	TBI Thermoblock	Biometra
Heating block-Mixer	Thermomixer Compact	Eppendorf
Homogeniser	T 50 basic ULTRATURRAX®	IKA Werke, Staufen
Incubator		MEMMERT
Incubator	BK-600	Heraeus
Incubators		Infors
Laminar flow bench	Microflow / Laminar Flow	NUNC GmbH
Luminescence Spectrometer	LS 50 B	Perkin-Elmer
Magnetic stirrer	RET IKAMAG	IKA Werke, Staufen
MALDI-TOF MS	Voyager STR	Applied biosystems
MALDI-TOF MS	Omni Flex®	Bruker
Mikrowelle	Moulinex	Samou, Radolfzell
Multipipette		Rainin
Optical microscope SZ40		Olympus
Oscilloscope		Nicolet
Peristaltic pump	Miniplus 2	Gilson / ABIMED
pH-Meter	pH-Meter 765 Calimatic	Knick
Photometer	Ultrospec 2100 pro	Pharmacia Biotech
Photometer	UV/VIS Spectrometer Lampda 40	Perkin-Elmer
Pipettes	P10- P5000	Gilson
R-Axis		Bruker
Rotors	70.1Ti, 70Ti, 60Ti, 45Ti, JLA-8.1000	Beckman Coulter,
Rotors	SLA3000, GS3	Sorvall (GMI, Inc)
SDS-PAGE-Gels	NuPAGE 4-12 % Bis-Tris Gel	Invitrogen
Sonifier	SONIFIER 250	Branson

Sonification bath	RK 255 S	Bandelin Sonorex
Speedvacuum	Concentrator 5301	Eppendorf
Thermostatic room 18°		
Thermocycler (x PCR)	T3	Biometra
TXRF-spectrometer	EXTRA IIA	ATOMIKA Instruments, Oberschleissheim
Ultracentrifuge	L8-70M	Beckman
Ultracentrifuge	Optima MAX UC	Beckman
UV-Monitor	Single Path Monitor UV-1	Pharmacia Biotech
UV-Monitor Writer	BD 41	Kipp & Zonen
Vacuum pumps		Biometra
Vaccum centrifuge	Concentrator 5301	Eppendorf
Vortexer	Vortex-Genie 2	Scientific Industries

4.1.4 Software and Databases

Software that has been used is the following:

- Unicorn v4.10 and v5.01 control system, for the control of the Äkta chromatography system (Amersham Biosciences)
- Flex Control, Bruker MALDI-TOF
- Mascot for database comparison of the peptide mass fingerprints (Matrix Science, Boston, USA, http://www.matrixscience.com/search_intro.html)
- BLASTP and PSI-BLAST for database comparison of the protein sequences (NCBI BLAST server (Altschul *et al.*, 1997) and Max Planck Institute for Developmental Biology server (http://protevo.eb.tuebingen.mpg.de/toolkit/index.php?view=search)
- FASTA3 for database comparison of the protein sequences (EMBL-EBI server (http://www.ebi.ac.uk/fasta33/, (Pearson *et al.*, 1988)),
- SignalP (http://www.cbs.dtu.dk/services/SignalP/, (Bendtsen *et al.*, 2004)) for prediction of signal sequence
- InterPro Scan (http://www.ebi.ac.uk/InterProScan/, (Zdobnov *et al.*, 2001)) for database search of domains
- SMART (http://smart.embl-heidelberg.de/ (Schultz *et al.*, 1998)) for database search of domains
- ProtParam (http://www.expasy.org/tools/protparam.html, (Gasteiger *et al.*, 2005)) for calculation of the theoretical isolelectric point and the hydrophobicity
- TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/(Krogh *et al.*, 2001)) for prediction of transmembrane helices
- 3D-PSSM (http://www.sbg.bio.ic.ac.uk/~3dpssm/, (Kelley *et al.*, 2000)), Phyre (http://www.sbg.bio.ic.ac.uk/~phyre/) and HHPred (http://protevo.eb.tuebingen.mpg.de/toolkit/index.php?view=hhpred, (Söding *et al.*, 2005)) for database comparison of the protein fold and 3 D model creation
- PRED-TMBB (http://biophysics.biol.uoa.gr/PRED-TMBB/, (Bagos *et al.*, 2004)) for transmembrane beta-barrel recognition
- pCLAMP 9 program package for recording and analysing of lipid bilayer experiments (Axon Instruments, Sunnyvale, USA)
- HKL package (XDISPLAY, DENZO, and SCALEPACK, HKL research, Charlottesville, USA), for data set processing

Data processing was performed by Dr. Jürgen Köpke

Other software:

Office 2003 (Microsoft), CorelDraw 11 (Corel Cooperation), Photoshop 6.0 (Adobe), Acrobat 7.0 (Adobe), Endnote X (Thomson ISI ResearchSoft), Sophos Anti-Virus (Sophos).

Databases that have been used are the following:

- PEDANT (http://pedant.gsf.de/)
- NCBI (http://www.ncbi.nlm.nih.gov/)
- TransportDB (http://www.membranetransport.org/, (Ren *et al.*, 2004))
- Transport Classification Database (http://www.tcdb.org/)

4.1.5 Columns and Chromatographic Matrices

The columns and column material that has been used is listed in the following table.

Table 4.1-4: Chromatographic matrices

Column material	Supplier
CHT Ceramic Hydroxyapatite	BIO-RAD, Munich
Cibacron Blue 3GA	Sigma-Aldrich, Seelze
Reactive Blue 4	Sigma-Aldrich, Seelze
Reactive Blue 72	Sigma-Aldrich, Seelze
Reactive Brown 10	Sigma-Aldrich, Seelze
Reactive Green 5	Sigma-Aldrich, Seelze
Reactive Green 19	Sigma-Aldrich, Seelze
Reactive Red 120	Sigma-Aldrich, Seelze
Reactive Yellow 3	Sigma-Aldrich, Seelze
Reactive Yellow 86	Sigma-Aldrich, Seelze
HisTrap™ 1 ml	GE Healthcare, Munich
Nickel-nitriloacetat-(Ni-NTA)-Agarose	QIAGEN, Hilden
Mini Q 4.6/50 PE	GE Healthcare, Munich
Mini Q PC 3.2/3	GE Healthcare, Munich
Mono Q HR 10/10	GE Healthcare, Munich
HiTrap™ Q-Sepharose™ *Fast Flow*	GE Healthcare, Munich
HiTrap™ DEAE-Sepharose™ *Fast Flow*	GE Healthcare, Munich
HiTrap™ CM-Sepharose™ *Fast Flow*	GE Healthcare, Munich
HiTrap™ SP-Sepharose™ *Fast Flow*	GE Healthcare, Munich
HiTrap™ Phenyl™ *High Performance*	GE Healthcare, Munich
HiTrap™ Butyl™ *Fast Flow*	GE Healthcare, Munich
HiTrap™ Octyl™ *Fast Flow*	GE Healthcare, Munich
Disposable PD-10 Desalting Columns	GE Healthcare, Munich
Superdex 200 HR 10/30	GE Healthcare, Munich
Superdex 200 PC 3.2/30	GE Healthcare, Munich
Superose 12 HR 10/30	GE Healthcare, Munich
Superose 6 HR 10/30	GE Healthcare, Munich
Superose 6 PC 3.2/30	GE Healthcare, Munich
TSK-GEL G4000SW	Tosho Bioscience, Montgomeryville
Sephacryl-500 HR	GE Healthcare, Munich

4.1.6 Genomic DNA

A. aeolicus VF5 genomic DNA were purchased from Archaeenzentrum Regensburg. The sequence was determined from NCBI.

4.1.7 Primer

The oligonucleotide primers that were used for analytical and preparative PCR as well as for sequencing were synthesised by Biomers.net and are shown in the following table.

Table 4.1-5: Oligonucleotide primers.

Name	Sequence
A1760.For	gcgggatcctgatggaatttcttcagagagaccaa
A1760.Rev	gcgcaattggtggaggagaagatgaggatcgc
A1862.For	gcgagatcttgatgaagaaaggattgctcctcgca
A1862.Rev	gcggaattcgtgaagttgtagaatacggaaagtcc
A1862_NSS.for	gcgagatcttggtaaccataaggattgacgaagag
AQ1862pET.For	atacatatggtaaccataaggattgacgaagag
AQ1862pET.Rev	ccgctcgaggaagttgtagaatacggaaagtcc

AQ1862pET_MTIR.For	gggaattccatatgaccataaggattgacgaagagcaa
AQ1862pET_stop.Rev	gcg gaattcgtttagaagttgtagaatacggaaagtcc
A1760pET.For	gggaattccatatggaatttcttcagagagaccaa
A1760pET.Rev	ccgctcgagggaggagaagatgaggatcgc
AQ1558.1.For	gcgggatcctgggggataataaaacttctcttc
AQ1558.67.For	gcg ggatcctgaagagagaagaaaagattctccag
AQ1558.85.For	gcg ggatcctgggaggggcctttgacctattg
AQ1558.101.For	gcg ggatcctgatgcccttttacataaggacgaaa
AQ1558.115.For	gcg ggatcctgctggacgttcaactcgacctctac
AQ1558.Rev	gcg caattggtaagtggtctattatcatattcctccac

4.1.8   Enzymes, Proteins and Standards

For preparative PCR Phusion Polymerase (Finnzymes) was used; where as for analytical purposes BioTaq Red (Bioline). Restriction enzymes (EcoRI, BamHI, MfeI, BglII, NdeI), alkaline phosphatase, T4 Ligase was purchased from Fermentas. Trypsin (Sequencing Grade Modified Trypsin) was purchased from Promega. "Monoclonal Anti-poly Histidine-Alkaline Phosphatase" and "Streptavidine-Alkaline Phosphatase" was purchased from Qiagen.

For molecular weight estimation of protein in SDS PAGE gels and Western blot SeeBlue® Plus2 *Pre-Stained Standard* from Invitrogen or Prestained Protein Marker, Broad Range (6-175 kDa) from New England Biolabs was used. For molecular weight estimation of DNA fragments in agarose gels 1 kb DNA standard from New England Biolabs was used.

4.2   Methods

4.2.1   Molecular Biology

4.2.1.1   Standard Methods

Agarose gels for DNA electrophoresis were prepared using LE agarose. Plasmid extraction (mini- or maxi-prep), PCR purification and DNA extraction from agarose gels and genomic DNA preparation (DNeasy Tissue Kit) was carried out using Qiagen Kits. DNA ligations were performed at 20 °C over night, restriction digestion according to supplier's instruction. DNA sequencing was performed according to Sanger and co-workers by SeqLab (Sanger *et al.*, 1977).

4.2.1.2   Protein Production System

The *E. coli* expression system is the best established system for heterologous production of soluble and membrane proteins. Here the vectors pTTQ18 and pBAD were utilised. In pTTQ18 a moderately strong *tac* promoter (Stark, 1987) controls the genes expression. Addition of IPTG induced protein production. The pBAD (Invitrogen) vector uses the arabinose promoter, which tightly repressed protein production before the addition of arabinose. Both vectors were modified as described by Surade and co-workers 2006. In short two different sets of tags were introduced. Proteins expressed in the first, the A version contains a TEV protease cleavage site at the C-terminus followed by a deca-histidine tag to facilitate purification. Proteins produced in the second, the C version contain a deca-

histidine tag at the N-terminus, followed by a TEV protease cleavage site, as well as a C-terminal strep-tag II to allow a second affinity purification step. (Surade *et al.*, 2006)

The pTTQ18 constructs were tested for expression in *E. coli* NM554 and BL21(DE3) cells, the pBAD constructs in TOP10 because this stem lacks the *ara* operon and is therefore not able to consume arabinose. However, pBAD derivates have been successfully used expression hosts containing the arabinose operon like *E. coli* C43(DE3), suggesting that arabinose consumption is not severely interfering with protein production.

The pTTQ18 constructs were tested for expression in *E. coli* NM554 and BL21(DE3) cells, the pBAD constructs in *E. coli* TOP10 and in BL21(DE3).

In the pET-vector system genes are expressed under the control of the strong T7 promotor (Buchanan, 1999). The vector pET-24a(+) (Novagene) was used. DNA was inserted using the NdeI and EcoRI restriction site.

4.2.1.3 Competent Cells and Transformation

Competent cells were prepared according a modified protocol based on (Mandel *et al.*, 1970). For transformation plasmid DNA is mixed with competent cells, incubated on ice for 10 min, at 37 °C for 2 min, and kept on ice for 10 min. The cells are incubated with LB (5 g/l yeast extract, 10 g/l tryptone, 5 g/l NaCl) for 30 min at 37 °C and spread on LB agar (LB, 15 g/l agar agar) plate with the according antibiotic.

4.2.1.4 Protein Production Test

A 4 ml LB culture was inoculated ≈ 1:100 with over night culture. The cells were grown at 37 °C and samples were taken when $OD_{600\,nm}$ reached 0.4 and 0.8. Expression was induced by adding 0.5 mM IPTG to pTTQ18 and pQE contructs or with 0.02 % arabinose for pBAD constructs. Cultures were kept at 37 °C for four hour and at 20 °C over night. 1 ml culture is spun down and the cell pellet resuspended in 10 % (w/v) SDS. Expression was checked by western-blot analysis.

4.2.1.5 Preparative Protein Production

For preparative protein purification the *E. coli* strains BL21(DE3), C43(DE3) or NM554 are transformed with the plasmids derived from pTTQ18 of pET26, the strain TOP10 is transformed with plasmids derived from pBAD. A 24 l culture (LB or TB (24 g/l yeast extract, 12 g/l tryptone, 0.4 % (v/v) glycerol, 2.31 g KH_2PO_4, 12.54 g K_2HPO_4) is inoculated with 200 ml of over night culture, grown until on $OD_{600\,nm}$ of 0.5 to 1.0 is reached and induced with 0.2 mM to 1.0 mM IPTG in the case of pTTQ18 derivatives, 0.002 % to 0.2 % arabinose in the case of pBAD derivatives and grown without induction in the case of pET26 derivatives. The cells are harvested by centrifugation after 3 h to 24 h.

4.2.2 Protein Biochemistry

4.2.2.1 SDS PAGE

Protein samples were mixed 1:5 with SDS-sample buffer (0.25 M Tris-Cl pH 8.0, 10 % (w/v) SDS, 12.5 % (w/v) β-mercaptoethanol, 0.002 % (w/v) bromophenol blue, 25 % (v/v) glycerol) and incubated for 3 min at 95 °C or for 30 min at room temperature.

NuPAGE 4-12 % Bis-Tris gels (Invitrogen) were used in an "XCell SureLock" (Invitrogen) gel system with the MES-SDS running buffer and run according to manufactures protocol.

SDS PAGE gels were prepared with 15 % separating gel (0.125 M Tris-Cl pH 8.0, 0.1 % (w/v) SDS, 15 % (w/v) acryl amid/ bisacrylamid (37.5 %/ 1 %)) and 4 % stacking gel (0.125 M Tris-Cl pH 6.8, 0.1 % (w/v) SDS, 4 % (w/v) acryl amid/ bisacrylamid (37.5 %/ 1 %)). A home-made gel chamber was used. The gels are run in a Tris-glycin buffer system (50 mM Tris, 190 mM glycin, 0.1 % (w/v) SDS). Gels were applied to constant current of 20 mA for approx. 60 min.

Detection of covalently bound FAD was carried out on a SDS PAGE gel containing approximately 5 µg of protein. After electrophoresis the gel was treated for 10 min with 10 % (v/v) acetic acid and washed with water. The fluorescence was excited by UV light (Unden *et al.*, 1980).

All acryl amide gels stained either with colloidal Coomassie stain (SimplyBlue SafeStain, Invitrogen) according to manufacturer's protocol or in 0.15 % (w/v) Coomassie in 25 % (v/v) ethanol, 8 % (v/v) acetic acid. Silver stain was performed according to a mass spectrometry modified protocol (Shevchenko *et al.*, 1996).

4.2.2.2 Western blot

Proteins separated by SDS PAGE gels were transferred to a nitrocellulose membrane (Millipore) using a semidry method (25 mM Tris, 100 mM glycin, 10 % (v/v) methanol) with a constant current of 0.8 mA/cm^2 of gel for 1.5 h. The blot was incubated either with the "Monoclonal Anti-poly Histidine-Alkaline Phosphatase" (Qiagen) or "Streptavidine-Alkaline Phosphatase" (Qiagen). The activity of the alkaline phosphatase was detected with BCIP and NBT.

4.2.2.3 Native PAGE

BN PAGE is a "charge shift" method performed to determine the molecular mass and the oligomeric state of membrane proteins and membrane protein complexes. BN PAGE gels were performed according to Schägger and co-workers (Schägger *et al.*, 1994; Schägger *et al.*, 1991).

4.2.2.4 Membrane Preparation

Small scale membrane preparations were carried out to check the localisation of the overproduced protein in *E. coli*. Cells grown in a 50 ml LB culture are pelleted by centrifugation for 10 min at 5000 g, resuspended in 50 mM Tris-Cl pH 8.5, 100 mM NaCl, 1 mM EDTA, and disrupted by a French Press at 120 mPa. Cell debris and (if present) inclusion bodies were removed by centrifugation

at 10.000 g for 30 min and membranes sedimented from the supernatant (=cytosol) by ultracentrifugation for 1 h at 100000 g. 0.1 μg to 20 μg of membrane, cytosol and pellet fraction were tested for protein content with a western blot.

For preparative scale *E. coli* cells were harvested by centrifugation. *A. aeolicus* cells were stored at -20 °C and thawed. Membrane preparation is carried out according to Peng 2003. Cells were resuspended in 50 mM Tris-Cl pH 8.5, 100 mM NaCl, 1 mM EDTA, and by a microfluidizer at 1000 MPa. Cell debris were removed by centrifugation at 10.000 g for 30 min and membranes sedimented from the supernatant by ultracentrifugation (100000 g for 1 h). The membranes were washed and resuspended in the same buffer and stored at -20 °C.

4.2.2.5   Membrane Protein Solubilisation

A. aeolicus membranes were solubilised as described by (Peng *et al.*, 2003). The heterologously produced protein was applied to a solubilisation screen using various detergents (DDM, DM, OG, OTG, NG, LDAO, Fos12), temperatures (4 °C, 20 °C, 40 °C) and pH-values (pH 6.5, pH 7.5, pH 8.5). In the quantitative scale *E. coli* membranes were thawed, mixed 1:1 with detergent and slowly stirred. Non solubilised material was removes as described above.

4.2.2.6   Ion Exchange Chromatography

For preparative protein purification the solubilised membrane proteins were loaded onto a Mono Q HR 10/10 column (Amersham Bioscience), preequilibrated with buffer A (20 mM Tris-Cl pH 7.4, 0.05 % (w/v) NaN$_3$, and 0.05 % (w/v) DDM) using an Äkta Purifier 10. The bound proteins were eluted with a segmented gradient of buffer B (buffer A and 1 M NaCl). Fractions were analysed by SDS PAGE gels and combined according to the protein content. The samples were concentrated using a pressure dialysis concentrator (Amicon) with a membrane cut-off of 30 to 100 kDa.

For analytical purposes MiniQ, MiniS, MonoS, HiTrap Q-Sepharose, HiTrap DEAE-Sepharose, HiTrap CM-Sepharose, and HiTrap SP-Sepharose, columns were used either on Äkta purifier 10 or on Smart system according to manufacturer protocol. In short the columns were equilibrated with buffer at the appropriate pH and detergent proteins loaded and eluted with a gradient of 0 M to 1.0 M NaCl.

4.2.2.7   Dye Affinity Chromatography

Cibacron Blue 3GA, Reactive Blue 4, Reactive Blue 72, Reactive Brown 10, Reactive Green 5, Reactive Green 19, Reactive Red 120, Reactive Yellow 3 and Reactive Yellow 86 was equilibrated in 20 mM Tris-Cl pH 7.4, 0.05 % (w/v) DDM, the protein loaded, washed with five column volumes of the equilibration buffer and the proteins eluted with a step gradient of 1.5 M NaCl.

4.2.2.8   Hydrophobic Interaction Chromatography

HiTrap Phenyl HP, HiTrap Butyl FF, and HiTrap Octyl FF columns were used according to manufactures protocol with addition of detergent.

4.2.2.9   Immobilized Metal-Ion Affinity Chromatography

Heterologously produced proteins that contain a deca-histidine tag were purified by immobilized metal-ion affinity chromatography (IMAC) using a HisTrap™, GE Healthcare or nickel-nitriloacetat-agarose (Ni-NTA), Qiagen. Solubilised membrane proteins were loaded onto a HisTrap 1 ml column preequilibrated with buffer (50 mM Tris-Cl pH 8.0, 100 mM to 500 mM NaCl, 0.03 % (w/v) DDM, if not noted differently) and eluted with a gradient of 0 - 0.5 M Imidazole. Protein fractions were pooled according to positive Western blot and concentrated using a Centricon with a cut-off 10 kDa.

4.2.2.10   Gel Filtration Chromatography

Gel filtration chromatography was performed in order to ensure the monodispersity of the sample, to improve the purity and to exchange the buffer. For preparative purposes a Superose 6 HR 10/30, Superdex 200 10/300 GL, Superose 12 10/300 GL or TSK-GEL G4000SW (Tosoh) on an Äkta Purifier 10 or Äkta Purifier 100, for analytical purposes Superdex 200 PC 3.2/30, Superose 6 PC 3.2/30 or Superose 12 PC 3.2/30 was performed. After equilibration with at least one column volume columns were run according to manufacturer protocol.

4.2.2.11   Inclusion Body Preparation, Refolding and Purification

Cells were thawed, resuspended in 50 mM Tris-Cl pH 8.5, 100 mM NaCl, 1 mM EDTA and disrupted by sonification. Unbroken cells were removed by centrifugation at 3000 g for 10 min. Inclusion bodies were collected by centrifugation at 8000 g at 4 °C and washed by resuspension in 50 mM Tris-Cl pH 8.5, 100 mM NaCl, 1 mM EDTA and centrifugation at 8000 g at 4 °C.

The inclusion bodies were solubilised in 6 M guanidinum hydrochloride in 20 mM Tris-Cl pH 7.4, 150 mM NaCl for 2 hours at room temperature. Non-soluble material was removed by centrifugation at 20.000 g at 4 °C. Refolding is induced by adding the protein drop-wise into 3 % (w/v) LDAO in 20 mM Tris-Cl pH 7.4, 150 mM NaCl (1:10) under stirring. Non-soluble protein was removed by centrifugation for 10 min at 5000 g and the supernatant concentrated. Buffer is exchanged to 0.08 % (w/v) LDAO in 20 mM Tris-Cl on a PD-10 column (GE Healthcare) and applied onto a MonoQ HR10/10 column equilibrated in 20 mM Tris-Cl, 0.03 % (w/v) DDM. Protein is eluted with a linear gradient on 0 to 1.0 M NaCl. Protein containing fractions are pooled and concentrated using a Centricon protein concentrator with a cut-off 10 kDa.

4.2.2.12 Protein Extraction from the Peptidoglycan Layer

Peptidoglycan associated proteins were purified according to (Lakey *et al.*, 1985). In short, crude total membrane were shaken with 2 % SDS in 10 mM Tris-Cl pH 7.4 at 55 °C and then centrifuged at 100000 g for 45 min at 20 °C. This step is repeated three times. The pellet contains the peptigoglycan associated proteins. The proteins were freed from peptidoglycan by mixing the pellet with extraction buffer (2 % SDS, 0.5 M NaCl, 0.7 M 2-mercaptoethanol in 10 mM Tris-Cl pH 7.4) followed by centrifugation at 100000 g for one hour at 20 °C. The supernatant was dialysed twice for 12 h against 2 % SDS in 20 mM Tris-Cl and precipitated with 90 % acetone for 2 h at 4 °C. The pellet is washed twice with acetone and resuspended in 2 % Triton X 100, 5 mM EDTA, 50 mM Tris-Cl pH 7.4. Insoluble material was removed by centrifugation at 16000 g for 30 min at 20 °C and resuspended in 0.1 % SDS in 20 mM Tris-Cl pH 7.4.

4.2.3   Protein Characterisation

4.2.3.1   Mass Spectrometry of Peptides

SDS PAGE gels were Coomassie or silver stained. Bands were excised from the gels and subjected to in-gel digestion by trypsin as described elsewhere (Shevchenko *et al.*, 1996). In short, the gel pieces are washed in 100 mM ammonium hydrogencarbonate / 50 % (v/v) acetonitrile, dried, and incubated with trypsin (20 µg/ml) over night at 37 °C. The supernatant and the acetonitrile extract of the gel pieces is analysed separately. The peptides are dissolved in 70 % acetonitrile / 0.1 % trifluoric acid (TFA) and mixed 1:1 with matrix (saturated CHCA, Bruker) directly on a MALDI-TOF target (Applied Biosystems or Bruker).

MALDI-TOF mass spectra were recorded either on a Voyager-DE STR instrument (Applied Biosystems) by Dr. Isam Rais in the department of Prof. Michael Karas, JW Goethe University, Frankfurt am Main, or on an Omnifex instrument (Bruker).

Spectra were calibrated internally by trypsin autoproteolysis peptides 842.51 Da and 2211.1046 Da or by external standard (Peptide calibration standard, Bruker). Masses of monoisotopic peaks were searched on the mascot peptide mass fingerprint data base (Mascot). The following search parameters were used: taxonomy "Bacteria (Eubacteria) → other bacteria", trypsin as enzyme, oxidised methionine as variable modification, 100 ppm mass tolerance. A protein was considered as identified when the MOWSE score returned by Mascot was above the threshold of the program. Furthermore the protein should have been identified at least in two independent preparations.

4.2.3.2   Mass Spectrometry Proteins

The preparation of proteins for mass spectrometry was modified from (Cadene *et al.*, 2000). Matrix (CHCA, Bruker) was prepared as a saturated solution in a 3:2:1 (v/v/v) mixture of water, formic acid and isopropanol. 10 to 50 µM of protein in 0.03 % DDM were diluted 1:20 with matrix solution.

A saturated matrix solution was prepared from 1:2 water/ acetonitrile, further diluted 4-fold with isopropanol and applied to the target. The drops were allowed to dry and the matrix wiped with tissue paper until only a faint layer of CHCA is left. The protein/matrix mixture is spotted on the pretreated target, dried and washed with 0.1 % TFA.

4.2.3.3 Lipid analysis

Lipids were extracted according to the methods described by Folch and co-workers (Folch *et al.*, 1957). Lipid analysis was carried out by Sebastian Richers.

4.2.3.4 Similarity Searches and Sequence Alignment

A BLASTP and PSI-BLAST searches were performed using the National Center for Biotechnology Information (NCBI) BLAST server (Altschul et al., 1997), the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (http://www.expasy.org/tools/blast/) or the Max Planck Institute for Developmental Biology server (http://protevo.eb.tuebingen.mpg.de/toolkit/index.php?view=search). FASTA searches were carried out using the EMBL-EBI server Fasta3 (http://www.ebi.ac.uk/fasta33/), conserved domain search with Prosite Scan (Bucher *et al.*, 1996) or Smart (Letunic *et al.*, 2006; Schultz *et al.*, 1998). Secondary structure prediction (β-barrel recognition) was carried out with PRED-TMBB (Bagos *et al.*, 2004). Tertiary structure comparison was carried out using 3D-PSSM (Kelley *et al.*, 2000) or HH-PRED (Söding *et al.*, 2005). Putative homologous sequences were compared using ClustalW (Thompson *et al.*, 1994). The program ProtParam (Gasteiger *et al.*, 2003) was used for calculating isoelectric points, the program TMHMM (Moller *et al.*, 2001) for prediction of transmembrane helices.

4.2.3.5 Lipid Bilayer Experiments

Painted lipid bilayers (Müller *et al.*, 1963) were formed across an aperture separating two half-cells (volume 1.4 ml) of a Teflon cuvette (see Figure 1.1-1). The diameter of the aperture was 1 mm for multi-channel experiments or 0.1 mm for single channel experiments. The hole between the two compartments was pretreated with 0.5 % (w/v) diphytanoyl phosphatidylcholine in hexane and lipid bilayers were prepared from 1.5 % (w/v) diphytanoyl-phosphatidylcholine (Avanti Polar lipids, Birmingham, USA) in n-decane. The cuvette was connected to the measuring circuit with Ag/AgCl electrodes via salt bridges (3 M KCl, 5 % agar). One compartment was connected to a variable voltage source while the other one was connected to a current to voltage converter (Stanford Research Systems, Sunnyvale, USA or custom made: amplification 10^9 A/V). For voltage measurements the current amplifier was replaced by a voltage amplifier with high input impedance (> 10 GΩ, home-made). Signals were digitised (MiniDigi 1A, Axon Instruments, Sunnyvale, USA; sample-rate 1000 Hz) and recorded on a PC. For recording and analysis of the data the pCLAMP 9 program package (Axon Instruments, Sunnyvale, USA) was used.

If not indicated otherwise all salt solutions (Sigma) were buffered at pH 7.5 with K-HEPES or Tris-Cl. The temperature was 20 °C to 22 ˚C. Protein (0.4 mg/ml in 20 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.01 % DDM, final concentrations 0.5 - 2.0 µg/ml for multi-channel measurements, 0.05 - 0.2 µg/ml for single channel experiments) was added to the compartment connected to the amplifier, (cis-side) before forming the membrane. Usually membranes were formed in 10 mM salt solutions (both sides). Concentration gradients were established by adding small amounts of concentrated (1 M or 3 M) salt solutions to the compartment connected to the amplifier (cis-side) while stirring. Equilibration reached its final value after 10 - 15 minutes of stirring. The conductance of membranes (multi-channel experiments) containing protein was 2 nS/mm^2 to 0.2 µS/mm^2 in 100 mM KCl. In control membranes without protein the conductance was 0.15 pS/mm^2.

4.2.3.6   Protease Activity Assay

Proteolytic activity against N-terminally blocked peptides (Bachem) was measured by the release of 7-amido-4-methylcoumarin (AMC) from the carboxy-termini of the peptides as described earlier (Zimmerman *et al.*, 1978). The peptides (Suc-Ala-Ala-Phe-AMC, Z-Val-Lys-Met-AMC and Bz-Phe-Val-Arg-AMC) were dissolved in dimethylsulfoxide. Enzyme assays were conducted at 20 °C or 30 °C with 0.5 to 10 mM of peptide in 20 mM Tris-Cl pH 7.4. The final volume was 40 µl. 1 µg to 20 µg of protein were used per assay. Fluorescence of the AMC was measured using a Perkin-Elmer Luminescence Spectrometer LS 50 B. Excitation and emission wavelength (383 nm and 455 nm respectively) were chosen according to Zimmerman and co-workers. Fluorescence units were converted to AMC concentration by standard curve determined from a known concentration of free AMC. Free AMC was achieved by digestion of peptide with proteinase K for 1 h at room temperature. The selectivity was adjusted so that one fluorescence unit equals 0.25 µM AMC (10 pmol in 40 µl assay).

4.2.3.7   Zymograms

The proteolytic activity against gelantine was determined using zymograms. 0.1 % (w/v) gelatine was co-polymerised in a standard SDS PAGE gel. Samples were incubated in SDS sample buffer at different temperatures. The zymogram was prepared like a standard SDS PAGE gels. After electrophoresis the polypeptides of the zymogram were renaturated for 2 hours at room temperature in 10 mM Tris-Cl pH 7.4, 2 % Triton X 100. The proteolytic reaction was stopped after 24 hours at 80 °C in 10 mM Tris-Cl by coomassie staining of the zymogram. Proteolytic active samples were detected as white bands on a blue background.

4.2.3.8   Cross-linking Experiments

The protein was kept in standard buffer with the difference that the Tris-Cl buffer was exchanged for Na-HEPES or Na-phosphate buffer at neutral pH. The protein concentration was kept at 0.5 mg/ml.

Different amounts of glutaraldehyde (0.01 % to 0.05 %) were tested. Protein samples were taken after different time points and boiled with SDS-sample buffer for 3 min.

4.2.3.9 Total reflection X-ray Fluorescence

Total reflection X-ray Fluorescence (TXRF) measurements were performed by Claudia Rittmeyer and Steffen Metz in the department of Prof. Bernd O. Kolbesen (Institute of Inorganic and Analytical Chemistry, Johann Wolfgang Goethe University of Frankfurt, Frankfurt am Main, Germany).

All measurements were performed on a TXRF-spectrometer EXTRA IIA (ATOMIKA Instruments, Oberschleissheim Germany) equipped with Mo- and W-excitation tubes, a Si(Li) solid state detector (80 mm^2 area), an automated sample changer and a computer controlled multichannel analyser system. The samples were excited by tungsten Lα and molybdenum Kα radiation. The tungsten tube was operated at 25 kV and 60 mA with attenuation filter (30 µm Cu) whereas in the case of molybdenum excitation the tube was acted at 50 kV and 40 mA (attenuation filter: 200 µm Mo, aluminium foil). A measurement time of 1000 s was chosen.

Sulphur and metal concentrations were determined by TXRF in samples of the protein and the buffer (Tris-acetate pH 7.4). The protein stock solutions contained typically 20 µM protein. To 18 µL of these solutions 2 µL Cr standard (10 mg/L) was added. A sample volume of 4 µL was pipetted onto an unsiliconized quartz glass carrier and air dried at room temperature under clean bench conditions. In a similar way the Tris-acetate buffer samples were prepared. Each sample was measured four times.

4.2.3.10 Electron Microscopy

Electron microscopy was performed by Todd A. Clason and Teresa Ruiz in the department of Prof. Michael Radermacher (Department of Molecular Physiology and Biophysics, College of Medicine, University of Vermont, Burlington, USA).

5 µl of the sample was (0.005 mg/ml) was applied to 400 mesh copper grids coated with a thin carbon film. Four different stain solutions were used to establish the best staining conditions and particle preservation. Three drops (8 µl each) of either 1 % uranyl acetate (Ted Pella), 2 % phosphotungstic acid, pH 7 with NaOH (Ted Pella , 2 % methyl amine tungstate (Nanoprobes) or Nanovan (Nanoprobes) were applied. The last drop was left on the grid for 30 s. Finally, the excess liquid was wicked off and the grids fast air dried.

The grids were observed on a Tecnai12 Philips electron microscope (FEI, Holland) equipped with a LaB6 cathode (Kimball, USA), and a 14 µm 2048*2048 CCD camera (TVIPS, Germany). Images were recorded at an accelerating voltage of 100 kV and a nominal magnification of 42000 X under low dose conditions on S0-163 Kodak film.

Micrographs were scanned on an Intergraph SCAI scanner (Intergraph Corp., Madison, AL), at a raster size of 7 µm. The images were binned by a factor of three resulting in a final pixel size corresponding to 21 µm or 0.517 nm on the specimen scale. Particles were extracted from these

micrographs, pre-centered, and aligned against one of the images selected as reference, using the alignment algorithms based in the cross-correlation of two-dimensional Radon transforms (Radermacher, 1994; Radermacher, 1997). The data set was analysed further by correspondence analysis (Frank *et al.*, 1982; van Heel *et al.*, 1981), followed by classification by moving centers (Diday, 1971) combined with hierarchical ascendant classification. To obtain our final results, we carried out one additional multireference alignment and a final correspondence-analysis and classification step.

4.2.4 Protein Crystallisation

After purification to homogeneity the proteins were applied to standard crystallisation conditions. The proteins were concentrated to a concentration of 5 mg/ml and 10 mg/ml in a buffer of 20 mM Tris-Cl, 150 mM NaCl and 0.03 % DDM. The protein concentration was determined by bicinchoninic acid (BCA) assay using bovine serum albumine (BSA) as protein standard. For standard crystallisation screening a crystallisation robot (Cartesian) was used to mix the protein with the crystallisation buffer. Crystallisation buffers were mainly purchased from Sigma, Jena Bioscience and Qiagen. 300 nl of protein solution was mixed 1 to 1 with crystallisation buffer, sealed and kept at 18 °C. As a control buffer without protein is mixes with crystallisation buffer in the same way. Crystallisation is done by vapour diffusion in a sitting drop device (CrystalQuick 96-well standard profile - round bottom). Conditions which show crystals or crystalline precipitate are reproduced with self-made buffers. Buffers were prepared from Millipore water and filtered (0.2 µm cut-off). 1 µl of protein solution is mixed manually 1 to 1 with the self-made crystallisation buffer on silanised cover slides (2 % di-chloro-di-methylsilane in tri-chloroethane and washed in ethanol) and kept in a 24-well plate hanging drop device at 18 °C. Protein, precipitant and salt concentration, pH and kind of salt and precipitant are varied in order to obtain three-dimensional crystals. Sitting drop trials were also carried out in 24-well plate on micro bridges (round bottom, 12 µl capacity) glued with ethyl acetate. For micro batch crystallisation the crystallisation buffer was mixed with the protein under a layer of oil (paraffin oil or paraffin oil and silicon oil, 1:1) in 72-well microbatch plates.

5 References

Adam, G., Läuger, P., and Stark, G. (1995) Biophysikalische Chemie und Biophysik, 3. Ed., Springer-Verlag, Berlin Heidelberg New York

Aivaliotis, M., Haase, W., Karas, M., and Tsiotis, G. (2006) Proteomic analysis of chlorosome-depleted membranes of the green sulfur bacterium *Chlorobium tepidum Proteomics* **6**, 217-232

Akita, F., Chong, K. T., Tanaka, H., Yamashita, E., Miyazaki, N., Nakaishi, Y., Suzuki, M., Namba, K., Ono, Y., Tsukihara, T., and Nakagawa, A. (2007) The Crystal Structure of a Virus-like Particle from the Hyperthermophilic Archaeon *Pyrococcus furiosus* Provides Insight into the Evolution of Viruses *J Mol Biol* **368**, 1469-1483

Alcaraz, A., Nestorovich, E. M., Aguilella-Arzo, M., Aguilella, V. M., and Bezrukov, S. M. (2004) Salting out the ionic selectivity of a wide channel: The asymmetry of OmpF *Biophys J* **87**, 943-957

Altrock, W. D., tom Dieck, S., Sokolov, M., Meyer, A. C., Sigler, A., Brakebusch, C., Fassler, R., Richter, K., Boeckers, T. M., Potschka, H., Brandt, C., Loscher, W., Grimberg, D., Dresbach, T., Hempelmann, A., Hassan, H., Balschun, D., Frey, J. U., Brandstatter, J. H., Garner, C. C., Rosenmund, C., and Gundelfinger, E. D. (2003) Functional Inactivation of a Fraction of Excitatory Synapses in Mice Deficient for the Active Zone Protein Bassoon *Neuron* **37**, 787-800

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W., and Lipman, D. J. (1990) Basic Local Alignment Search Tool *J Mol Biol* **215**, 403-410

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

Andreini, C., Banci, L., Bertini, I., and Rosato, A. (2006) Zinc through the Three Domains of Life *J Proteome Res* **5**, 3173-3178

Angus, B. L., and Hancock, R. E. W. (1983) Outer-Membrane Porin Protein-F, Protein-P, and Protein-D1 of *Pseudomonas aeruginosa* and PhoE of *Escherichia coli* - Chemical Cross-Linking to Reveal Native Oligomers *J Bacteriol* **155**, 1042-1051

Arora, A., Abildgaard, F., Bushweller, J. H., and Tamm, L. K. (2001) Structure of outer membrane protein A transmembrane domain by NMR spectroscopy *Nat Struct Biol* **8**, 334-338

Arora, A., Rinehart, D., Szabo, G., and Tamm, L. K. (2000) Refolded Outer Membrane Protein A of *Escherichia coli* Forms Ion Channels with Two Conductance States in Planar Lipid Bilayers *J Biol Chem* **275**, 1594-1600

Azem, A., Shaked, I., Rosenbusch, J. P., and Daniel, E. (1995) Cross-Linking of Porin with Glutardialdehyde - a Test for the Adequacy of Premises of Cross-Linking Theory *Biochim Biophys Acta* **1243**, 151-156

Bagos, P. G., Liakopoulos, T. D., Spyropoulos, I. C., and Hamodrakas, S. J. (2004) PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins *Nucleic Acids Res* **32**, W400-W404

Balliano, G., Viola, F., Ceruti, M., and Cattel, L. (1992) Characterization and partial purification of squalene-2,3-oxide cyclase from *Saccharomyces cerevisiae Arch Biochem Biophys* **293**, 122-129

Bannwarth, M., and Schulz, G. E. (2003) The expression of outer membrane proteins for crystallization *Biochim Biophys Acta* **1610**, 37-45

Basle, A., Iyer, R., and Delcour, A. H. (2004) Subconductance states in OmpF gating *Biochim Biophys Acta* **1664**, 100-107

Basle, A., Qutub, R., Mehrazin, M., Wibbenmeyer, J., and Delcour, A. H. (2004) Deletions of single extracellular loops affect pH sensitivity, but not voltage dependence, of the *Escherichia coli* porin OmpF *Protein Eng* **17**, 665–672

Baymann, F., Tron, P., Schoepp-Cothenet, B., Aubert, C., Bianco, P., Stetter, K. O., Nitschke, W., and Schutz, M. (2001) Cytochromes c555 from the Hyperthermophilic Bacterium *Aquifex aeolicus* (VF5). 1. Characterization of Two Highly Homologous, Soluble and Membranous, Cytochromes c555 *Biochemistry* **40**, 13681-13689

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0 *J Mol Biol* **340**, 783-795

Benz, R. (1984) Structure and Selectivity of Porin Channels *Current Topics in Membranes and Transport* **21**, 199-219

Benz, R. (1994) Permeation of Hydrophilic Solutes through Mitochondrial Outer Membranes - Review on Mitochondrial Porins *Biochim Biophys Acta* **1197**, 167-196

Benz, R., Darveau, R. P., and Hancock, R. E. W. (1984) Outer-membrane protein PhoE from *Escherichia coli* forms anion-selective pores in lipid-bilayer membranes *Eur J Biochem* **140**, 319-324

Benz, R., Egli, C., and Hancock, R. E. W. (1993) Anion Transport through the Phosphate-Specific OprP-Channel of the *Pseudomonas aeruginosa* Outer-Membrane - Effects of Phosphate, Di-Basic and Tribasic Anions and of Negatively-Charged Lipids *Biochim Biophys Acta* **1149**, 224-230

Benz, R., and Hancock, R. E. W. (1987) Mechanism of Ion Transport through the Anion-selective Channel of the *Pseudomonas aeruginosa* Outer Membrane *J Gen Physiol* **89**, 275-295

Benz, R., Janko, K., Boos, W., and Läuger, P. (1978) Formation of large, ion-permeable membrane channels by the matrix protein (porin) of *Escherichia coli Biochim Biophys Acta* **511**, 305-319

Benz, R., Schmid, A., and Hancock, R. E. W. (1985) Ion Selectivity of Gram-Negative Bacterial Porins *J Bacteriol* **162**, 722-727

Benz, R., Schmid, A., Maier, C., and Bremer, E. (1988) Characterization of the Nucleoside-Binding Site inside the Tsx Channel of *Escherichia coli* Outer-Membrane - Reconstitution Experiments with Lipid Bilayer-Membranes *Eur J Biochem* **176**, 699-705

Blatch, G. L., and Lässle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions *BioEssays* **21**, 932-939

Brandstatter, J. H., Fletcher, E. L., Garner, C. C., Gundelfinger, E. D., and Wassle, H. (1999) Differential expression of the presynaptic cytomatrix protein bassoon among ribbon synapses in the mammalian retina *Eur J Neurosci* **11**, 3683-3693

Brenner, S. E., Chothia, C., and Hubbard, T. J. P. (1998) Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships *Proc Natl Acad Sci U S A* **95**, 6073-6078

Brugna-Guiral, M., Tron, P., Nitschke, W., Stetter, K. O., Burlat, B., Guigliarelli, B., Bruschi, M., and Giudici-Orticoni, M. T. (2003) [NiFe] hydrogenases from the hyperthermophilic bacterium *Aquifex aeolicus*: properties, function, and phylogenetics *Extremophiles* **V7**, 145-157

Buchanan, S. K. (1999) beta-Barrel proteins from bacterial outer membranes: structure, function and refolding *Curr Opin Struct Biol* **9**, 455-461

Bucher, P., Karplus, K., Moeri, N., and Hofmann, K. (1996) A flexible motif search technique based on generalized profiles *Comp. & Chem.* **20**, 3-23

Buehler, L. K., and Rosenbusch, J. P. (1993) Single Channel Behavior of Matrix Porin of *Escherichia coli Biochem Biophys Res Commun* **190**, 624-629

Bunai, K., Nozaki, M., Hamano, M., Ogane, S., Inoue, T., Nemoto, T., Nakanishi, H., and Yamane, K. (2003) Proteomic analysis of acrylamide gel separated proteins immobilized on polyvinylidene difluoride membranes following proteolytic digestion in the presence of 80% acetonitrile *Proteomics* **3**, 1738–1749

Butt, A., Davison, M. D., Smith, G. J., Young, J. A., Gaskell, S. J., Oliver, S. G., and Beynon, R. J. (2001) Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis *Proteomics* **1**, 42-53

Cadene, M., and Chait, B. T. (2000) A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins *Anal Chem* **72**, 5655-5658

Cambillau, C., and Claverie, J. M. (2000) Structural and Genomic Correlates of Hyperthermostability *J Biol Chem* **42**, 32383–32386

Colombini, M., Blachly-Dyson, E., and Forte, M. (1996) VDAC, a channel in the outer mitochondrial membrane, Plenum Press, New York

Conlan, S., Zhang, Y., Cheley, S., and Bayley, H. (2000) Biochemical and Biophysical Characterization of OmpG: A Monomeric Porin *Biochemistry* **39**, 11845-11854

Conway, J. F., Duda, R. L., Cheng, N., Hendrix, R. W., and Steven, A. C. (1995) Proteolytic and Conformational Control of Virus Capsid Maturation: The Bacteriophage HK97 System *J Mol Biol* **253**, 86-99

Cowan, S. W., and Rosenbusch, J. P. (1994) Folding Pattern Diversity of Integral Membrane Proteins *Science* **264**, 914-916

Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A., Jansonius, J. N., and Rosenbusch, J. P. (1992) Crystal structures explain functional properties of two *E. coli* porins *Nature* **358**, 727-733

D'Andrea, L. D., and Regan, L. (2003) TPR proteins: the versatile helix *Trends Biochem Sci* **28**, 655-662

Dahan, D., Srikumar, R., Laprade, R., and Coulton, J. W. (1996) Purification and refolding of recombinant *Haemophilus influenzae* type b porin produced in *Bacillus subtilis FEBS Lett* **392**, 304-308

Danelon, C., Brando, T., and Winterhalter, M. (2003) Probing the orientation of reconstituted maltoporin channels at the single-protein level *J Biol Chem* **278**, 35542-35551

Danelon, C., Suenaga, A., Winterhalter, M., and Yamato, I. (2003) Molecular origin of the cation selectivity in OmpF porin: single channel conductances vs. free energy calculation *Biophys Chem* **104**, 591-603

DeAngelis, A. A., Howell, S. C., Nevzorov, A. A., and Opella, S. J. (2006) Structure Determination of a Membrane Protein with Two Trans-membrane Helices in Aligned Phospholipid Bicelles by Solid-State NMR Spectroscopy *J Am Chem Soc* **128**, 12256-12267

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., and Swanson, R. V. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus Nature* **392**, 353-358

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

Diday, E. (1971) La methode de nuees dynamiques *Revue Statistique Appliquée* **19**, 19-34

Dmitriev, O., Jones, P. C., Jiang, W., and Fillingame, R. H. (1999) Structure of the Membrane Domain of Subunit b of the *Escherichia coli* F_0F_1 ATP Synthase *J Biol Chem* **274**, 15598-15604

Eisele, J. L., and Rosenbusch, J. P. (1990) In vitro folding and oligomerization of a membrane protein. Transition of bacterial porin from random coil to native conformation *J Biol Chem* **265**, 10217-10220

Essen, L. O., Siegert, R., Lehmann, W. D., and Oesterhelt, D. (1998) Lipid patches in membrane protein oligomers: Crystal structure of the bacteriorhodopsin-lipid complex *Proc Natl Acad Sci U S A* **95**, 11673–11678

Everberg, H., Leiding, T., Schioth, A., Tjerneld, F., and Gustavsson, N. (2006) Efficient and non-denaturing membrane solubilization combined with enrichment of membrane protein complexes by detergent/polymer aqueous two-phase partitioning for proteome analysis *J Chromatogr A* **1122**, 35-46

Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K., and Welte, W. (1998) Siderophore-mediated iron transport: Crystal structure of FhuA with bound lipopolysaccharide *Science* **282**, 2215-2220

Fernandez, C., Hilty, C., Wider, G., Guntert, P., and Wüthrich, K. (2004) NMR Structure of the Integral Membrane Protein OmpX *J Mol Biol* **336**, 1211-1221

Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J., and Iwata, S. (2004) Architecture of the Photosynthetic Oxygen-Evolving Center *Science* **303**, 1831-1838

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocyne, J. D., Scott, J., Shirley, R., Liu, L. I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995) Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd *Science* **269**, 496-512

Folch, J., Lees, M., and Stanley, G. H. S. (1957) A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues *J Biol Chem* **226**, 497-509

Forst, D., Welte, W., Wacker, T., and Diederichs, K. (1998) Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose *Nat Struct Biol* **5**, 37-46

Fountoulakis, M., Takács, M. F., Berndt, P., Langen, H., and Takács, B. (1999) Enrichment of low abundance proteins of *Escherichia coli* by hydroxyapatite chromatography *Proteomics* **20**, 2181-2195

Frank, J., and van Heel, M. (1982) Correspondence analysis of aligned images of biological particles *J Mol Biol* **161**, 134-137

Gallagher, P. G., and Forget, B. G. (1995) Structure, Organization, and Expression of the Human Band 7.2b Gene, a Candidate Gene for Hereditary Hydrocytosis *J Biol Chem* **270**, 26358–26363

Galperin, M. Y. (2001) Conserved 'hypothetical' proteins: new hints and new puzzles *Comp Funct Genomics* **2**, 14-18

Galperin, M. Y., and Koonin, E. V. (2004) Conserved hypothetical' proteins: prioritization of targets for experimental study *Nucleic Acids Res* **32**, 5452–5463

Garavito, R. M., Jenkins, J., Jansonius, J. N., Karlsson, R., and Rosenbusch, J. P. (1983) X-ray Diffraction Analysis of Matrix Porin, an Integral Membrane Protein from *Escherichia coli* Outer Membranes *J Mol Biol* **164**, 313-327

Garavito, R. M., and Rosenbusch, J. P. (1980) Three-dimensional crystals of an integral membrane protein: an initial x-ray analysis *J Cell Biol* **86**, 327-329

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

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker, J. M. (ed). *The Proteomics Protocols Handbook, Humana Press*

Gingeras, T. R., Kwoh, D. Y., and Davis, G. R. (1987) Hybridization properties of immobilized nucleic acids *Nucleic Acids Res* **15**, 5373-5390

Goldman, D. E. (1944) Potential, impedance, and rectification in membranes *J Gen Physiol* **27**, 37-60

Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C., and Walz, T. (2005) Lipid-protein interactions in double-layered two-dimensional AQP0 crystals *Nature* **438**, 633-638

Gonen, T., Sliz, P., Kistler, J., Cheng, Y., and Walz, T. (2004) Aquaporin-0 membrane junctions reveal the structure of a closed water pore *Nature* **429**, 193-197

Goodman, M. B., Ernstrom, G. G., Chelur, D. S., O'Hagan, R., Yao, C. A., and Chalfie, M. (2002) MEC-2 regulates C. elegans DEG/ENaC channels needed for mechanosensation *Nature* **415**, 1039-1042

Graham, R. L. J., O'Loughlin, S. N., Pollock, C. E., Ternan, N. G., Weatherly, D. B., Jackson, P. J., Tarleton, R. L., and McMullan, G. (2006) A Combined Shotgun and Multidimensional Proteomic Analysis of the Insoluble Subproteome of the Obligate Thermophile, *Geobacillus thermoleovorans* T80 *J Proteome Res* **5**, 2465-2473

Griffiths, E., and Gupta, R. S. (2004) Signature sequences in diverse proteins provide evidence for the late divergence of the Order *Aquificales Int Microbiol* **7**, 41-52

Guiral, M., Tron, P., Aubert, C., Gloter, A., Iobbi-Nivol, C., and 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

Hannan, R. D., Cavanaugh, A., Hempel, W. M., Moss, T., and Rothblum, L. (1999) Identification of a mammalian RNA polymerase I holoenzyme containing components of the DNA repair/replication system *Nucleic Acids Res* **27**, 3720-3727

Hantke, K. (1976) Phage-T6 - Colicin-K Receptor and Nucleoside Transport in *Escherichia coli FEBS Lett* **70**, 109-112

Helling, S., Schmitt, E., Joppich, C., Schulenborg, T., Mullner, S., Felske-Muller, S., Wiebringhaus, T., Becker, G., Linsenmann, G., Sitek, B., Lutter, P., Meyer, H. E., and Marcus, K. (2006) 2-D differential membrane proteome analysis of scarce protein samples *Proteomics* **16**, 4506-4513

Henikoff, S., and Henikoff, J. G. (1992) Amino Acid Substitution Matrices from Protein Blocks *Proc Natl Acad Sci U S A* **89**, 10915-10919

Hicks, P. M., Rinker, K. D., Baker, J. R., and Kelly, R. M. (1998) Homomultimeric protease in the hyperthermophilic bacterium *Thermotoga maritima* has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria *FEBS Lett* **440**, 393-398

Hodge, T., and Colombini, M. (1997) Regulation of Metabolite Flux through Voltage-Gating of VDAC Channels *J Membr Biol* **157**, 271–279

Hodgkin, A. L., and Katz, B. (1949) The Effect of Sodium Ions on the Electrical Activity of the Giant Axon of the Squid *J Physiol* **108**, 37-77

Hong, H., Patel, D. R., Tamm, L. K., and van den Berg, B. (2006) The Outer Membrane Protein OmpW Forms an Eight-stranded beta-Barrel with a Hydrophobic Channel *J Biol Chem* **281**, 7568-7577

Huber, R., Wilharm, T., Huber, D., Trincone, A., Burggraf, S., Konig, H., Rachel, R., Rockinger, I., Fricke, H., and Stetter, K. O. (1992) *Aquifex pyrophilus* Gen-Nov Sp-Nov Represents a Novel Group of Marine Hyperthermophilic Hydrogen-Oxidizing Bacteria *Syst Appl Microbiol* **15**, 340-351

Huynen, M., Doerks, T., Eisenhaber, F., Orengo, C., Sunyaev, S., Yuan, Y., and Bork, P. (1998) Homology-based fold predictions for *Mycoplasma genitalium* proteins *J Mol Biol* **280**, 323-326

Hwang, P. M., Choy, W. Y., Lo, E. I., Chen, L., Forman-Kay, J. D., Raetz, C. R. H., Privé, G. G., Bishop, R. E., and Kay, L. E. (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR *Proc Natl Acad Sci U S A* **99**, 13560–13565

Im, W., and Roux, B. (2002) Ions and Counterions in a Biological Channel: A Molecular Dynamics Simulation of OmpF Porin from *Escherichia coli* in an Explicit Membrane with 1 M KCl Aqueous Salt Solution *J Mol Biol* **319**, 1177-1197

Jaenicke, R., and Bohm, G. (1998) The stability of proteins in extreme environments *Curr Opin Struct Biol* **8**, 738-748

Jahnke, L. L., Eder, W., Huber, R., Hope, J. M., Hinrichs, K. U., Hayes, J. M., Des Marais, D. J., Cady, S. L., and Summons, R. L. (2001) Signature Lipids and Stable Carbon Isotope Analyses of Octopus Spring Hyperthermophilic Communities Compared with Those of *Aquificales* Representatives *Appl Environ Microbiol* **67**, 5179-5189

Jensen, S. O., Thompson, L. S., and Harry, E. J. (2005) Cell Division in *Bacillus subtilis*: FtsZ and FtsA Association Is Z-Ring Independent, and FtsA Is Required for Efficient Midcell Z-Ring Assembly *J Bacteriol* **187**, 6536–6544

Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., and MacKinnon, R. (2003) X-ray structure of a voltage-dependent K+ channel *Nature* **423**, 33-41

Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krausz, N. (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution *Nature* **411**, 909-917

Jung, H., Windhaber, R., Palm, D., and Schnackerz, K. D. (1995) NMR and circular dichroism studies of synthetic peptides derived from the third intracellular loop of the β-adrenoceptor *FEBS Lett* **358**, 133-136

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

Khademi, S., O'Connell, J., Remis, J., Robles-Colmenares, Y., Miercke, L. J. W., and Stroud, R. M. (2004) Mechanism of Ammonia Transport by Amt/MEP/Rh: Structure of AmtB at 1.35 A *Science* **305**, 1587-1594

Kimata, N., Nishino, T., Suzuki, S., and Kogure, K. (2004) *Pseudomonas aeruginosa* Isolated from Marine Environments in Tokyo Bay *Microb Ecol* **47**, 41-47

Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search *Acta Crystallogr D* **55**, 484-491

Klein, C., Garcia-Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., Siedler, F., and Oesterhelt, D. (2005) The membrane proteome of *Halobacterium salinarum Proteomics* **5**, 180-197

Komarov, A. G., Deng, D. F., Craigen, W. J., and Colombini, M. (2005) New insights into the mechanism of permeation through large channels *Biophys J* **89**, 3950-3959

Kopperschläger, G., Freyer, R., Diezel, W., and Hofmann, E. (1968) Some kinetic and molecular properties of yeast phosphofructokinase *FEBS Lett* **1**, 137-141

Kreusch, A., and Schulz, G. E. (1994) Refined Structure of the Porin from *Rhodopseudomonas blastica* - Comparison with the Porin from *Rhodobacter capsulatus J Mol Biol* **243**, 891-905

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes *J Mol Biol* **305**, 567-580

Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydrophobic character of a protein *J Mol Biol* **157**, 105-132

La Rocca, P., Biggin, P. C., Tieleman, D. P., and Sansom, M. S. P. (1999) Simulation studies of the interaction of antimicrobial peptides and lipid bilayers *Biochim Biophys Acta* **1462**, 185-200

Laage, R., and Langosch, D. (2001) Strategies for Prokaryotic Expression of Eukaryotic Membrane Proteins *Traffic* **2**, 99-104

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

Lai, E. M., Nair, U., Phadke, N. D., and Maddock, J. R. (2004) Proteomic screening and identification of differentially distributed membrane proteins in *Escherichia coli Mol Microbiol* **52**, 1029-1044

Lakey, J. H., Watts, J. P., and Lea, E. J. A. (1985) Characterisation of channels induced in planar bilayer membranes by detergent solubilised *Escherichia coli* porins *Biochim Biophys Acta* **817**, 208-216

le Maire, M., Champeil, P., and Möller, J. V. (2000) Interaction of membrane proteins and lipids with solubilizing detergents *Biochim Biophys Acta* **1508**, 86-111

Leidich, S. D., and Orlean, P. (1996) Gpi1, a *Saccharomyces cerevisiae* Protein that Participates in the First Step in Glycosylphosphatidylinositol Anchor Synthesis *J Biol Chem* **271**, 27829-27837

Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006) SMART 5: domains in the context of genomes and networks *Nucleic Acids Res* **34**, D257-D260

Li, N., Shaw, A. R. E., Zhang, N., Mak, A., and Li, L. (2004) Lipid raft proteomics: Analysis of in-solution digest of sodium dodecyl sulfate-solubilized lipid raft proteins by liquid chromatography-matrix-assisted laser desorption/ionization tandem mass spectrometry *Proteomics* **4**, 3156-3166

Locher, K. P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J. P., and Moras, D. (1998) Transmembrane signaling across the ligand-gated FhuA receptor: Crystal structures of free and ferrichrome-bound states reveal allosteric changes *Cell* **95**, 771-778

Locher, K. P., and Rosenbusch, J. P. (1997) Oligomeric states and siderophore binding of the ligand-gated FhuA protein that forms channels across *Escherichia* coli outer membranes *Eur J Biochem* **247**, 770-775

Lubec, G., Afjehi-Sadat, L., Yang, J. W., and John, J. P. P. (2005) Searching for hypothetical proteins: Theory and practice based upon original data and literature *Prog Neurobiol* **77**, 90-127

Lunin, V. V., Dobrovetsky, E., Khutoreskaya, G., Zhang, R. G., Joachimiak, A., Doyle, D. A., Bochkarev, A., Maguire, M. E., Edwards, A. M., and Koth, C. M. (2006) Crystal structure of the CorA Mg2+ transporter *Nature* **440**, 833-837

Macfarlane, D. E. (1983) Use of benzyldimethyl-n-hexadecylammonium chloride ("16-BAC"), a cationic detergent, in an acidic polyacrylamide gel electrophoresis system to detect base labile protein methylation in intact cells *Anal Biochem* **132**, 231-235

MacKenzie, K. R., Prestegard, J. H., and Engelman, D. M. (1997) A Transmembrane Helix Dimer: Structure and Implications *Science* **276**, 131-133

Maes, D., Zeelen, J. P., Thanki, N., Beaucamp, N., Alvarez, M., Thi, M. H. D., Backmann, J., Martial, J. A., Wyns, L., Jaenicke, R., and Wierenga, R. K. (1999) The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: A comparative thermostability structural analysis of ten different TIM structures *Proteins* **37**, 441-453

Maier, E., Polleichtner, G., Boeck, B., Schinzel, R., and Benz, R. (2001) Identification of the outer membrane porin of *Thermus thermophilus* HB8: the channel-forming complex has an unusually high molecular mass and an extremely large single-channel conductance *J Bacteriol* **183**, 800-803

Main, E. R. G., Stott, K., Jackson, S. E., and Regan, L. (2005) Local and long-range stability in tandemly arrayed tetratricopeptide repeats *Proc Natl Acad Sci U S A* **102**, 5721–5726

Mandel, M., and Higa, A. (1970) Calcium-Dependent Bacteriophage DNA Infection *J Mol Biol* **53**, 159-162

Mason, J. M., and Arndt, K. M. (2004) Coiled coil domains: stability, specificity, and biological implication *Chembiochem* **5**, 170-176

Meng, G., Surana, N. K., St Geme, J. W., and Waksman, G. (2006) Structure of the outer membrane translocator domain of the *Haemophilus influenzae* Hia trimeric autotransporter *EMBO J* **25**, 2297–2304

Mills, J., Wyborn, N. R., Greenwood, J. A., Williams, S. G., and Jones, C. W. (1997) An outer-membrane porin inducible by short-chain amides and urea in the methylotrophic bacterium *Methylophilus methylotrophus Microbiology* **143**, 2373-2379

Miroux, B., and Walker, J. E. (1996) Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels *J Mol Biol* **260**, 289–298

Miyazawa, A., Fujiyoshi, Y., and Unwin, N. (2003) Structure and gating mechanism of the acetylcholine receptor pore *Nature* **423**, 949-955

Moller, S., Croning, M. D. R., and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions *Bioinformatics* **17**, 646-653

Moraes, T. F., Bains, M., Hancock, R. E. W., and Strynadka, N. C. J. (2007) An arginine ladder in OprP mediates phosphate-specific transfer across the outer membrane *Nat Struct Mol Biol* **14**, 85-87

Müller, P., Rudin, D. O., Tien, H., and Westcott, W. (1963) Methods for the Formation of Single Bimolecular Lipid Membranes in Aqueous Solution *J Phys Chem* **67**, 534-535

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A. M., and Fujiyoshi, Y. (2000) Structural determinants of water permeation through aquaporin-1 *Nature* **407**, 599-605

Murzin, A. G., and Patthy, L. (1999) From sequence to structure to function - Editorial overview *Curr Opin Struct Biol* **9**, 359-362

Mushegian, A. R., Bassett, D. E., Boguski, M. S., Bork, P., and Koonin, E. V. (1997) Positionally cloned human disease genes: Patterns of evolutionary conservation and functional motifs *Proc Natl Acad Sci U S A* **94**, 5831-5836

Namba, K., Hagiwara, K., Tanaka, H., Nakaishi, Y., Chong, K. T., Yamashita, E., Armah, G. E., Ono, Y., Ishino, Y., Omura, T., Tsukihara, T., and Nakagawa, A. (2005) Expression and molecular characterization of spherical particles derived from the genome of the hyperthermophilic euryarchaeote *Pyrococcus furiosus J Biochem (Tokyo)* **138**, 193-199

Nestel, U., Wacker, T., Woitzik, D., Weckesser, J., Kreutz, W., and Welte, W. (1989) Crystallization and preliminary X-ray analysis of porin from *Rhodobacter capsulatus FEBS Lett* **242**, 405-408

Nestorovich, E. M., Rostovtseva, T. K., and Bezrukov, S. M. (2003) Residue ionization and ion transport through OmpF channels *Biophys J* **85**, 3718-3729

Nestorovich, E. M., Sugawara, E., Nikaido, H., and Bezrukov, S. M. (2006) *Pseudomonas aeruginosa* porin OprF - Properties of the channel *J Biol Chem* **281**, 16230-16237

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites *Protein Eng* **10**, 1-6

Nikaido, H. (2003) Molecular basis of bacterial outer membrane permeability revisited *Microbiol Mol Biol Rev* **67**, 593-656

Nübel, T., Klughammer, C., Huber, R., Hauska, G., and Schütz, M. (2000) Sulfide:quinone oxidoreductase in membranes of the hyperthermophilic bacterium *Aquifex aeolicus* (VF5) *Arch Microbiol* **V173**, 233-244

Ohtsuka, T., Takao-Rikitsu, E., Inoue, E., Inoue, M., Takeuchi, M., Matsubara, K., Deguchi-Tawarada, M., Satoh, K., Morimoto, K., Nakanishi, H., and Takai, Y. (2002) Cast: a novel protein of the cytomatrix at the active zone of synapses that forms a ternary complex with RIM1 and munc13-1 *J Cell Biol* **158**, 577-590

Ostermeier, C., Iwata, S., Ludwig, B., and Michel, H. (1995) Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase *Nat Struct Mol Biol* **2**, 842-846

Otto, B. R., van Dooren, S. J. M., Nuijens, J. H., Luirink, J., and Oudega, B. (1998) Characterization of a hemoglobin protease secreted by the pathogenic *Escherichia coli* strain EB1 *J Exp Med* **188**, 1091-1103

Oxenoid, K., and Chou, J. J. (2005) The structure of phospholamban pentamer reveals a channel-like architecture in membranes *Proc Natl Acad Sci U S A* **102**, 10870-10875

Pace, C. N. (1992) Contribution of the hydrophobic effect to globular protein stability *J Mol Biol* **226**, 29-35

Pace, C. N. (2000) Single surface stabilizer *Nat Struct Biol* **7**, 345 - 346

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Trong, I. L., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor *Science* **289**, 739-745

Patricelli, M. P., Lashuel, H. A., Giang, D. K., Kelly, J. W., and Cravatt, B. F. (1998) Comparative Characterization of a Wild Type and Transmembrane Domain-Deleted Fatty Acid Amide Hydrolase: Identification of the transmembrane Domain as a Site for Oligomerization *Biochemistry* **37**, 15177-15187

Paul, C., and Rosenbusch, J. P. (1985) Folding patterns of porin and bacteriorhodopsin *EMBO J* **4**, 1593 - 1597

Pauptit, R. A., Schirmer, T., Jansonius, J. N., Rosenbusch, A. P., Parker, M. W., Tucker, A. D., Tsernoglou, D., Weiss, M. S., and Schulzt, G. E. (1991) A common channel-forming motif in evolutionarily distant porins *J Struct Biol* **107**, 136-145

Pauptit, R. A., Zhang, H., Rummel, G., Tilman, S., Jansonius, J. N., and Rosenbusch, J. P. (1991) Trigonal crystals of porin from *Escherichia coli J Mol Biol* **218**, 505-507

Pavlov, E., Grigoriev, S. M., Dejean, L. M., Zweihorn, C. L., Mannella, C. A., and Kinnally, K. W. (2005) The mitochondrial channel VDAC has a cation-selective open state *Biochim Biophys Acta* **1710**, 96-102

Pearson, W. R., and Lipman, D. J. (1988) Improved Tools for Biological Sequence Comparison *Proc Natl Acad Sci U S A* **85**, 2444-2448

Peng, G. H., Bostina, M., Radermacher, M., Rais, I., Karas, M., and Michel, H. (2006) Biochemical and electron microscopic characterization of the F_1F_0 ATP Synthase from the hyperthermophilic eubacterium *Aquifex aeolicus FEBS Lett* **580**, 5934–5940

Peng, G. H., Fritzsch, G., Zickermann, V., Schägger, H., Mäntele, R., Lottspeich, F., Bostina, M., Radermacher, M., Huber, R., Stetter, K. O., and 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

Perl, D., Müller, U., Heinemann, U., and Schmid, F. X. (2000) Two exposed amino acid residues confer thermostability on a cold shock protein *Nat Struct Biol* **7**, 380 - 383

Perutz, M. F. (1978) Electrostatic effects in proteins *Science* **201**, 1187-1191

Perutz, M. F., and Raidt, H. (1975) Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2 *Nature* **255**, 256-259

Phale, P. S., Schirmer, T., Prilipov, A., Lou, K. L., Hardmeyer, A., and Rosenbusch, J. P. (1997) Voltage gating of *Escherichia coli* porin channels: Role of the constriction loop *Proc Natl Acad Sci U S A* **94**, 6741-6745

Poole, L. B. (2005) Bacterial defenses against oxidants: mechanistic features of cysteine-based peroxidases and their flavoprotein reductases *Arch Biochem Biophys* **433**, 240-254

Pullen, J. K., Liang, S. M., Blake, M. S., Mates, S., and Tai, J. Y. (1995) Production of *Haemophilus influenzae* type-b porin in *Escherichia coli* and its folding into the trimeric form *Gene* **152**, 85-88

Radermacher, M. (1994) 3-Dimensional Reconstruction from Random Projections - Orientational Alignment via Radon Transforms *Ultramicroscopy* **53**, 121-136

Radermacher, M. (1997) Radon transform techniques for alignment and 3D reconstruction from random projections *Scanning Microsc* **11**, 171-177

Rais, I., Karas, M., and Schägger, H. (2004) Two-dimensional electrophoresis for the isolation of integral membrane proteins and mass spectrometric identification *Proteomics* **4**, 2567-2571

Razvi, A., and Scholtz, J. M. (2006) Lessons in stability from thermophilic proteins *Protein Sci* **15**, 1569-1578

Ren, Q. H., Kang, K. H., and Paulsen, I. T. (2004) TransportDB: a relational database of cellular membrane transport systems *Nucleic Acids Res* **32**, D284-D288

Renault, L., Chou, H. T., Chiu, P. L., Hill, R. M., Zeng, X., Gipson, B., Zhang, Z. Y., Cheng, A., Unger, V., and Stahlberg, H. (2006) Milestones in electron crystallography *J Comput Aided Mol Des* **20**, 519-527

Robinson, R. A., and Stokes, R. H. (1959) Electrolyte Solutions, 2nd Ed., Butterworth Scientific Publications, London

Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005) NMR Structure of Mistic, a Membrane-Integrating Protein for Membrane Protein Expression *Science* **307**, 1317-1321

Rosenbusch, J. P. (1974) Characterization of Major Envelope Protein from *Escherichia coli* - Regular Arrangement on Peptidoglycan and Unusual Dodecyl-Sulfate Binding *J Biol Chem* **249**, 8019-8029

Rosenkrands, I., Rasmussen, P. B., Carnio, M., Jacobsen, S., Theisen, M., and Andersen, P. (1998) Identification and characterization of a 29-kilodalton protein from *Mycobacterium tuberculosis* culture filtrate recognized by mouse memory effector cells *Infect Immun* **66**, 2728-2735

Rostovtseva, T. K., Kazemi, N., Weinrich, M., and Bezrukov, S. M. (2006) Voltage Gating of VDAC is Regulated by Nonlamellar Lipids of Mitochondrial Membranes *J Biol Chem* **281**, 37496-37506

Russell, R. J. M., Ferguson, J. M. C., Hough, D. W., Danson, M. J., and Taylor, G. L. (1997) The Crystal Structure of Citrate Synthase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* at 1.9 A Resolution *Biochemistry* **36**, 9983-9994

Saier, M. H. (2000) A functional-phylogenetic classification system for transmembrane solute transporters *Microbiol Mol Biol Rev* **64**, 354–411

Saint, N., De, E., Julien, S., Orange, N., and Molle, G. (1993) Ionophore Properties of OmpA of *Escherichia coli Biochim Biophys Acta* **1145**, 119-123

Sanchez-Barrena, M. J., Martinez-Ripoll, M., Galvez, A., Valdivia, E., Maqueda, M., Cruz, V., and Albert, A. (2003) Structure of Bacteriocin AS-48: From Soluble State to Membrane Bound State *J Mol Biol* **334**, 541-549

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA Sequencing with Chain-Terminating Inhibitors *Proc Natl Acad Sci U S A* **74**, 5463-5467

Santoni, V., Molloy, M., and Rabilloud, T. (2000) Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* **21**, 1054-1070

Schägger, H., Cramer, W. A., and 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

Schägger, H., and von Jagow, G. (1991) Blue Native Electrophoresis for Isolation of Membrane-Protein Complexes in Enzymatically Active Form *Anal Biochem* **199**, 223-231

Schein, S. J., Kagan, B. L., and Finkelstein, A. (1978) Colicin-K Acts by Forming Voltage-Dependent Channels in Phospholipid Bilayer Membranes *Nature* **276**, 159-163

Schindler, H., and Rosenbusch, J. P. (1978) Matrix Protein from *Escherichia coli* Outer Membranes forms Voltage-Controlled Channels in Lipid Bilayers *Proc Natl Acad Sci U S A* **75**, 3751-3755

Schirmer, T. (1998) General and specific porins from bacterial outer membranes *J Struct Biol* **121**, 101-109

Schlüsener, D., Fischer, F., Kruip, J., Rögner, M., and Poetsch, A. (2005) Mapping the membrane proteome of *Corynebacterium glutamicum Proteomics* **5**, 1317-1330

Schmid, B., Kromer, M., and Schulz, G. E. (1996) Expression of porin from *Rhadopseudomonas blastica* in *Escherichia coli* inclusion bodies and folding into exact native structure *FEBS Lett* **381**, 111-114

Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: Identification of signaling domains *Proc Natl Acad Sci U S A* **95**, 5857-5864

Schütz, M., Schoepp-Cothenet, B., Lojou, E., Woodstra, M., Lexa, D., Tron, P., Dolla, A., Durand, M. C., Stetter, K. O., and Baymann, F. (2003) The Naphthoquinol Oxidizing Cytochrome bc_1 Complex of the Hyperthermophilic Knallgasbacterium *Aquifex aeolicus*: Properties and Phylogenetic Relationships *Biochemistry* **42**, 10800-10808

Schwartz, R., Ting, C. S., and King, J. (2001) Whole Proteome pI Values Correlate with Subcellular Localizations of Proteins for Organisms within the Three Domains of Life *Genome Res* **11**, 703-709

Seckler, B., and Poralla, K. (1986) Characterization and Partial Purification of Squalene Hopene Cyclase from *Bacillus acidocaldarius Biochim Biophys Acta* **881**, 356-363

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels *Anal Chem* **68**, 850-858

Singer, S. J., and Nicolson, G. L. (1972) The Fluid Mosaic Model of the Structure of Cell Membranes *Science* **175**, 720-731

Smith, T. F., and Waterman, M. S. (1981) Identification of Common Molecular Subsequences *J Mol Biol* **147**, 195-197

Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction *Nucleic Acids Res* **33**, W244-W248

Stark, M. J. R. (1987) Multicopy Expression Vectors Carrying the Lac Repressor Gene for Regulated High-Level Expression of Genes in *Escherichia coli Gene* **51**, 255-267

Stauffer, K. A., Page, M. G. P., Hardmeyer, A., Keller, T. A., and Pauptit, R. A. (1990) Crystallization and preliminary X-ray characterization of maltoporin from *Escherichia coli J Mol Biol* **211**, 297-299

Stock, J. B., Rauch, B., and Roseman, S. (1977) Periplasmic space in *Salmonella typhimurium* and *Escherichia coli J Biol Chem* **252**, 7850-7861

Struyve, M., Moons, M., and Tommassen, J. (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein *J Mol Biol* **218**, 141-148

Subbarao, G. V., and van den Berg, B. (2006) Crystal structure of the monomeric porin OmpG *J Mol Biol* **360**, 750-759

Sugawara, E., Nestorovich, E. M., Bezrukov, S. M., and Nikaido, H. (2006) *Pseudomonas aeruginosa* Porin OprF Exists in Two Different Conformations *J Biol Chem* **281**, 16220-16229

Suhre, K., and Claverie, J. M. (2003) Genomic Correlates of Hyperthermostability, an Update *J Biol Chem* **278**, 17198-17202

Sukhan, A., and Hancock, R. E. W. (1996) The role of specific lysine residues in the passage of anions through the *Pseudomonas aeruginosa* porin OprP *J Biol Chem* **271**, 21239-21242

Surade, S., Klein, M., Stolt-Bergner, P. C., Muenke, C., Roy, A., and Michel, H. (2006) Comparative analysis and "expression space" coverage of the production of prokaryotic membrane proteins for structural genomics *Protein Sci* **15**, 2178-2189

Svensson, H. (1961) Isoelectric Fractionations, Analysis, and Characterisation of Ampholytes in natural pH gradients. I. The Differential Equation of Solute Concentrations at a Steady State and its Solution for Simple Cases. *Acta Chem Scand* **15**, 325-341

Svensson, H. (1962) Isoelectric Fractionations, Analysis, and Characterisation of Ampholytes in natural pH gradients. II. Buffering Capicity and Conductance of Isoionic Ampholytes. *Acta Chem Scand* **16**, 456-466

Szilágyi, A., and Závodszky, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey *Structure* **8**, 493-504

Szmelcman, S., and Hofnung, M. (1975) Maltose transport in *Escherichia coli* K-12 - Involvement of bacteriophage lambda receptor *J Bacteriol* **124**, 112-118

Szponarski, W., Sommerer, N., Boyer, J. C., Rossignol, M., and Gibrat, R. (2004) Large-scale characterization of integral proteins from *Arabidopsis* vacuolar membrane by two-dimensional liquid chromatography *Proteomics* **4**, 397-406

Tanoue, E., Nishiyama, S., Kamo, M., and Tsugita, A. (1995) Bacterial membranes: possible source of a major dissolved protein in seawater *Geochim Cosmochim Acta* **59**, 2643-2648

Thompson, J. D., Higgins, D. G., and 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

Thompson, M. J., and Eisenberg, D. (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability *J Mol Biol* **290**, 595-604

Unden, G., Hackenberg, H., and Kröger, 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

Valdes-Stauber, N., and Scherer, S. (1994) Isolation and Characterization of Linocin M18, a Bacteriocin Produced by *Brevibacterium linens Appl Environ Microbiol* **60**, 3809-3814

van Heel, M., and Frank, J. (1981) Use of multivariate statistics in analysing the images of biological macromolecules *Ultramicroscopy* **6**, 187-194

Vieille, C., Hess, J. M., Kelly, R. M., and Zeikus, J. G. (1995) xylA cloning and sequencing and biochemical characterization of xylose isomerase from *Thermotoga neapolitana Appl Environ Microbiol* **61**, 1867–1875

Vitkup, D., Melamud, E., Moult, J., and Sander, C. (2001) Completeness in structural genomics *Nat Struct Biol* **8**, 559-566

Vogel, H., and Jahnig, F. (1986) Models for the structure of outer-membrane proteins of *Escherichia coli* derived from raman spectroscopy and prediction methods *J Mol Biol* **190**, 191-199

von Heijne, G. (1986) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology *EMBO J* **5**, 3021–3027

von Heijne, G. (1999) Recent advances in the understanding of membrane protein assembly and structure *Q Rev Biophys* **32**, 285-307

Wallin, E., and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms *Protein Sci* **7**, 1029-1038

Wang, D. N., Safferling, M., Lemieux, M. J., Griffith, H., Chen, Y., and Li, X. D. (2003) Practical aspects of overexpressing bacterial secondary membrane transporters for structural studies *Biochim Biophys Acta* **1610**, 23-36

Weaver, D. J., and Viancour, T. A. (1992) A crustacean neuronal cytoskeletal protein with characteristics of neurofilaments and microtubule-associated proteins *J Comp Neurol* **320**, 110 - 120

Weiss, M. S., Abele, U., Weckesser, J., Welte, W., Schiltz, E., and Schulz, G. E. (1991) Molecular Architecture and Electrostatic Properties of a Bacterial Porin *Science* **254**, 1627-1630

Wimley, W. C. (2003) The versatile β-barrel membrane protein *Curr Opin Struct Biol* **13**, 404-411

Wootton, J. C., and Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases *Comput Chem* **17**, 149-163

Wootton, J. C., and Federhen, S. (1996) Analysis of compositionally biased regions in sequence databases *Methods Enzymol* **266**, 554-571

Wu, S., Wan, P., Li, J., Li, D., Zhu, Y., and He, F. (2006) Multi-modality of pI distribution in whole proteome *Proteomics* **6**, 449-455

Wüthrich, K. (1998) The second decade — into the third millenium *Nat Struct Biol* **5**, 492-495

Yakunin, A. F., Yee, A. A., Savchenko, A., Edwards, A. M., and Arrowsmith, C. H. (2004) Structural proteomics: a tool for genome annotation *Curr Opin Chem Biol* **8**, 42-48

Yamashita, A., Singh, S. K., Kawate, T., Jin, Y., and Gouaux, E. (2005) Crystal structure of a bacterial homologue of Na$^+$/Cl$^-$-dependent neurotransmitter transporters *Nature* **437**, 215-223

Ye, J., and Van den Berg, B. (2004) Crystal structure of the bacterial nucleoside transporter Tsx *EMBO J* **23**, 3187-3195

Yildiz, O., Vinothkumar, K. R., Goswami, P., and Kühlbrandt, W. (2006) Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation *EMBO J* **25**, 3702-3713

Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R., and Kim, S. H. (1998) Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics *Proc Natl Acad Sci U S A* **95**, 15189-15193

Zdobnov, E. M., and Apweiler, R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro *Bioinformatics* **17**, 847-848

Zhang, N., Chen, R., Young, N., Wishart, D., Winter, P., Weiner, J. L., and Li, L. (2007) Comparison of SDS- and methanol-assisted protein solubilization and digestion methods for *Escherichia coli* membrane proteome analysis by 2-D LC-MS/MS *Proteomics* **7**, 484-493

Zhang, N., Li, N., and Li, L. (2004) Liquid Chromatography MALDI MS/MS for Membrane Proteome Analysis *J Proteome Res* **3**, 719-727

Zimmerman, M., Quigley, J. P., Ashe, B., Dorn, C., Goldfarb, R., and Troll, W. (1978) Direct Fluorescent Assay of Urokinase and Plasminogen Activators of Normal and Malignant Cells - Kinetics and Inhibitor Profiles *Proc Natl Acad Sci U S A* **75**, 750-753

List of Figures

List of Tables

List of Formulas

Acknowledgements

Special thanks to Paolo Lastrico, Helga Volk, Barbara Schiller, and Dr. Lutz Kampmann for their support.

Hanno and Nicole for showing me what I can do best.

All the people who made Frankfurt a place to be and everybody at home for true friendship throughout the years.

Especially I would like to thank my parents and my family love, patience for constant support. My grandma who sadly cannot read this thesis anymore.

Michael for more than I can say.

Curriculum Vitae

Personal Data

Name	Ulrike Wedemeyer
Date of birth	17th May 1978
Place of birth	Hannover, Germany

Education

1984 – 1988	Elementary School Albert-Schweizer Grund- und Hauptschule, Wunstorf
1988 – 1990	Elementary School Orientierungsstufe Süd, Wunstorf
1990 – 1997	High School Hölty-Gymnasium Wunstorf
10 1997 – 05 2003	Diploma Studies in Biochemistry University of Hannover
09 1999	Vordiplom Biochemie
07 2002 – 01 2003	Diploma Thesis at the Hannover Medical School, Institute of Physiological Chemistry, Hannover
	Internal supervisor: Prof. Matthias Gaestel External supervisor: Prof. Walter Müller
	"Characterization of the proline-rich N-terminal domain of MAPKAP-Kinase 2. Analysis of subcellular localization of proteins, protein-protein interactions"
05 2003	Diplom Biochemie
10 2003 – 09 2007	Doctoral Studies in Biochemistry, Johann Wolfgang Goethe University Frankfurt am Main
06 2003 – 06 2007	Doctoral Thesis at the Max Planck Institute of Biophysics, Molecular Membrane Biology, Frankfurt am Main
	Internal supervisor: Prof. Robert Tampé External supervisor: Prof. Hartmut Michel
	„Structural and Functional Membrane Proteomics using *Aquifex aeolicus*"