

**Heterologous Production and Characterization of  
Selected Secondary Active Transporters from the CDF,  
KUP, MOP, FNT, RhtB and SulP Families**

**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**

**Devrishi Goswami  
aus Chandannagar  
-Indien-**

**Frankfurt am Main  
2010**

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

Dekan: Prof. Dr. Dieter Steinhilber

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

Datum der Disputation:

Diese Doktorarbeit wurde vom 20. Juni 2006 bis zum 20. April 2010 unter Leitung von Prof. Dr. Hartmut Michel in der Abteilung für Molekulare Membranbiologie am Max-Planck-Institute 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.

Devrishi Goswami  
Frankfurt am Main

*Dedicated to my parents*

**TABLE OF CONTENTS**

<b>Kurzfassung und Zusammenfassung</b>	i-vii
<b>Abstract and Summary</b>	viii-xiv
<b>Abbreviations</b>	a-c

**1. Introduction**

1.1 Biological membrane	1
1.2 Integral membrane proteins (IMP) and their biogenesis	1
1.3 Secondary active transporters	2
1.4 Represented families of secondary active transporters	3
1.4.1 Cation diffusion facilitator family (CDF)	3
1.4.2 Formate nitrite transporter (FNT) family	6
1.4.3 Multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily	7
1.4.4 Potassium uptake permease (KUP) family	8
1.4.5 Resistance to homoserine/threonine (RhtB) family	9
1.4.6 The sulfate permease (SulP) family	9
1.5 Heterologous overproduction of membrane protein	10
1.5.1 Choice of target: source organism	11
1.5.2 Expression vectors and tags	12
1.5.3 Choice of the the expression host: <i>E. coli</i>	13
1.5.4 Bottlenecks affecting heterologous overexpression and remedies	13
1.5.4.1 Overexpression, targeting and folding	13
1.5.4.2 Membrane space and accommodation of foreign structures	14
1.5.4.3 Lipid composition	14
1.5.4.4 Stability of messenger RNA	15
1.5.4.5 Toxicity of target protein	15
1.5.4.6 Poor expression and N terminal fusion partner; MBP fusion	15
1.6 Cell-free production of membrane protein	16
1.6.1 Development of coupled transcription translation system	17
1.6.2 Types and modes of reaction	17

## TABLE OF CONTENTS

1.6.3 <i>E. coli</i> S30 extract	19
1.6.4 Plasmid construct and template quality	19
1.6.5 Mg <sup>2+</sup> , phosphate and K <sup>+</sup>	20
1.6.6 Energy source	20
1.6.7 Temperature	21
1.6.8 Detergents: for DCF mode and solubilization in PCF mode	21
1.7 Biophysical methods to study membrane protein	22
1.7.1 X-ray crystallography	22
1.7.2 Electron crystallography	24
1.7.3 Differential scanning calorimetry (DSC)	24
1.8 Scope of this thesis	26

## 2. Results

2.1 Cation diffusion facilitator (CDF) family	30
2.1.1 Target selection and <i>in silico</i> analysis	30
2.1.2 Expression screening of 4 members of CDF family	31
2.1.3 Functional complementation of 4 CDF transporters	31
2.1.4 Production, isolation and characterization of Aq_2073	33
2.1.4.1 Solubilization screening for choosing the right detergents	33
2.1.4.2 Affinity purification and homogeneity of Aq_2073	34
2.1.4.3 Removal of His tag using TEV digestion	34
2.1.4.4 Stability and monodispersity of Aq_2073	35
2.1.4.5 Thermal unfolding studies using DSC	37
2.1.4.6 Oligomerization studies	38
2.1.4.7 Substrate binding assay with DSC	40
2.1.4.8 Generation of Aq_2073 constructs for crystallization	41
2.1.4.9 Isolation, purification and characterization Aq_2073 constructs	43
2.1.4.9.1 Aq_2073 C1	43
2.1.4.9.2 Aq_2073 CTD1 and Aq_2073 CTD2	45
2.1.5 Two-dimensional (2D) crystallization	45
2.1.6 Three-dimensional (3D) crystallization	47

## **TABLE OF CONTENTS**

2.2 Potassium uptake permease (KUP) family	48
2.2.1 Expression screening of different constructs	48
2.2.2 Solubilization screening and large scale purification	48
2.2.3 Characterizing stability in different conditions	49
2.2.4 Reconstitution and freeze fracture	51
2.2.5 Solid supported membrane experiment to check functionality	51
2.3 Cell-free production of selected transporters	53
2.3.1 PF0708 of the MOP family	53
2.3.1.1 PCF mode production of PF0708	53
2.3.1.2 DCF production of PF0708	55
2.3.2 STM3476 of the FNT family	56
2.3.2.1 Detergent screening in the PCF and DCF mode	56
2.3.2.2 Effect of temperature on expression and solubilization	56
2.3.2.3 Large scale production and purification	57
2.3.3 STM1781 of the SulP family	58
2.3.4 STM3959 of the RhtB family	60
2.4 MBP fusion and effect on production	61
<b>3. Discussion</b>	
3.1 Cation diffusion facilitator (CDF) family	64
3.1.1 The proteins	64
3.1.2 Expression screening	64
3.1.3 Functional complementation	66
3.1.4 Purification and stability of Aq_2073	67
3.1.5 Oligomeric state of Aq_2073	69
3.1.6 Substrate(s) of Aq_2073	71
3.1.7 Construct design and crystallization of Aq_2073	72
3.2 Potassium uptake permease (KUP) family	73
3.2.1 STM3880: expression, purification and stability	73
3.2.2 Solid supported membrane based electrophysiological studies	74
3.3 Cell-free production	75
3.3.1 Cell-free production and characterization of PF0780	75

## **TABLE OF CONTENTS**

3.3.1.1 PCF mode production and characterization	76
3.3.1.2 Soluble production in DCF mode	77
3.3.2 Cell-free production and characterization of STM3476	77
3.3.2.1 Detergent screening in PCF and DCF mode	77
3.3.2.2 Effect of temperature	78
3.3.2.3 Large scale PCF production, purification and characterization	78
3.3.3 Production and characterization of SulP and RhtB proteins	79
3.4 MBP fusion and its effect on production	80
3.5 Overall discussion	81

## **4. Materials and methods**

4.1 Materials	83
4.1.1 Chemicals	83
4.1.2 Detergents	83
4.1.3 Lipids	84
4.1.4 Protease inhibitors	84
4.1.5 Chromatographic matrices and pre-packed columns and instruments	84
4.1.6 Enzymes	84
4.1.7 Antibodies	85
4.1.8 Kits	85
4.1.9 Marker probes	85
4.1.10 Buffers, solutions and culture- media composition	85
4.1.11 Apparatus and consumables	87
4.1.12 Microorganisms	88
4.1.13 Plasmids	88
4.1.14 Components for cell-free productions	88
4.2 Methods	90
4.2.1 Selection of families and targets	90
4.2.2 General molecular biological techniques	90
4.2.2.1. DNA isolation	90
4.2.2.2. DNA restriction digestion	90



---

## TABLE OF CONTENTS

4.2.2.3. Ligation	90
4.2.2.4. Vector modification	90
4.2.2.5 Cloning for MBP fusion work	92
4.2.3 General cell culture techniques for <i>E. coli</i>	92
4.2.3.1. Preparation of chemically competent <i>E. coli</i> cells	92
4.2.3.2. Transformation of competent <i>E. coli</i> cells	93
4.2.4 Detection of protein production, and protein visualization	93
4.2.4.1 Protein production screening in <i>E. coli</i>	93
4.2.4.2 Western blot procedure	94
4.2.4.3 Coomassie staining procedure	94
4.2.5 Protein Purification	95
4.2.5.1. Protein production in large scale cultures	95
4.2.5.2. Membrane isolation in large scale	95
4.2.5.3. Solubilization screen of membrane proteins with various detergents	96
4.2.5.4. Protein purification with affinity chromatography	96
4.2.5.5. TEV protease digestion	96
4.2.5.6. Gel filtration chromatography	97
4.2.6 Cell-free production of membrane proteins	97
4.2.6.1 S30 extract preparation	97
4.2.6.2 T7 RNA polymerase preparation	99
4.2.6.3 Cell-free expression reaction mixture	100
4.2.6.4 Purification of cell-free produced protein	101
4.2.7 Determination of the oligomerization state	102
4.2.7.1 Cross-linking of proteins with glutaraldehyde	102
4.2.7.2 Blue native (BN) gel electrophoresis	102
4.2.8 Reconstitution of membrane protein into liposome	104
4.2.8.1. Preparation of the lipid stock	104
4.2.8.2. Reconstitution procedure	104
4.2.9 Functional and biophysical characterization	105
4.2.9.1 Functional complementation of GG48 strain with CDF transporters	105
4.2.9.2 Solid supported membrane experiments	105

---

**TABLE OF CONTENTS**

4.2.9.3 Differential scanning calorimetry	106
4.2.10 2D crystallization, negative staining and screening by EM	106
4.2.11 3D crystallization	106
<b>References</b>	108
<b>Appendix</b>	122
<b>Acknowledgements</b>	
<b>Curriculum vitae</b>	

**Kurzfassung**

In den meisten Organismen kodieren 25%-30% des Genoms für Membranproteine. Sie spielen eine wichtige Rolle für die Zellfunktion, und ihre Bedeutung wird verstärkt durch die Tatsache, dass viele Medikamente an Membranproteinen angreifen. Paradoxe Weise entfallen lediglich 1,7% der hinterlegten Proteinstrukturen in der *protein databank* (PDB) auf experimentell ermittelte Strukturen von Membranproteinen. Dies ist hauptsächlich auf die schwierige Handhabung, bedingt durch ihren amphipatischen Charakter, zurückzuführen. Das geringe Vorkommen in natürlichen Geweben macht die heterologe Überexpression dieser Gene zu einer Notwendigkeit. Diese Arbeit zielte darauf ab, verschiedene sekundär aktive Transportproteine für strukturelle und funktionelle Untersuchungen heterolog zu produzieren und alternative Strategien zu etablieren, um die Hürden im Zusammenhang mit der heterologen Überexpression zu überwinden.

Vier Mitglieder der Schwermetall-transportierenden *cation diffusion facilitator* (CDF) Familie aus *S. typhimurium* und *A. aeolicus* wurden heterolog in *E. coli* produziert und durch einen *in vivo* Komplementationassay unter Zuhilfenahme des Zink-Transport defizienten *E. coli* Stammes GG48 funktionell charakterisiert. Von diesen vier Proteinen konnte Aq\_2073 aus *A. aeolicus* im präparativen Maßstab mit einer beachtlichen Ausbeute und Reinheit hergestellt werden, um strukturelle Untersuchungen durchzuführen. Nach umfangreichen Stabilitätsuntersuchungen mit verschiedenen Detergenzien, pH-Werten und Temperaturen wurden 2D sowie 3D-Kristallisationsversuche unternommen. Zusätzlich wurden mehrere C-terminal verkürzte Konstrukte hergestellt, welche ebenfalls für Kristallisationsversuche genutzt wurden. Diese ergaben in 3D-Kristallisationsversuchen erste nadelähnliche Kristalle oder in 2D-Kristallisationsversuchen Vesikel mit kristallinen Stellen, aber keine offensichtlichen Kristalle. Das Protein besaß eine signifikant höhere Schmelztemperatur in Gegenwart von Cadmiumionen wie durch *differential scanning calorimetry* (DSC) gezeigt wurde.

Ein weiterer Transporter, STM3880 aus der *potassium uptake permease* (KUP) Familie aus *S. typhimurium*, wurde in *E. coli* heterolog exprimiert, durch Affinitätschromatographie gereinigt, in künstliche Liposomen rekonstituiert und durch

elektrophysiologische Untersuchungen mit Festphasen-gestützten Membranen (*solid supported membranes*; *SSM*) funktionell charakterisiert.

Um alternative Expressionsstrategien zu etablieren, wurden Proteine von vier verschiedenen Familien zellfrei unter kontinuierlichem Austausch (*continuous exchange cell free expression*; *CECF*) produziert. Diese Methode schien geeignet, um den zellbasierten Ansatz zu komplementieren. Zielproteine der *resistence to homoserine/threonine* (RhtB) Familie, die *in vivo* keine Expression zeigten, konnten so hergestellt und gereinigt werden. STM1781 aus der *sulfate permease* (SuIP) Familie konnte zellfrei hergestellt, gereinigt und hinsichtlich seiner Stabilität charakterisiert werden, wohingegen die zellbasierte Produktion in starkem Abbau resultierte. PF0780 der *multidrug/oligosaccharidyl-lipid/polysaccharide flippase* (MOP) Familie wurde ebenfalls zur Homogenität gereinigt und zeigte eine vergleichbare Stabilität zum *in vivo* produzierten Protein. Des Weiteren wurde der Effekt einer Fusion des Maltose bindenden Proteins (MBP) am N-Terminus auf die Produktion und Membranintegration bei drei verschiedenen Proteinen untersucht. Die Analyse ergab eine verringerte Ausbeute in Anwesenheit des MBPs, wenn sich beide Termini des Proteins im Cytoplasma befanden.

In dieser Arbeit konnten verschiedene sekundär aktive Transporter erfolgreich hergestellt und entsprechende Reinigungsprotokolle etabliert werden, die darauf abzielen strukturelle und funktionelle Untersuchungen im Rahmen eines *structural genomics* Projektes durchzuführen. Es zeigte sich, dass die Einbindung alternativer Strategien wie die Anwendung zellbasierter und zellfreier Expressionssysteme, die Gesamtanzahl herstellbarer Proteine steigert und damit die Erfolgsaussichten derart angelegter Projekte verbessert.

## **Zusammenfassung**

Membranproteine sind oft sehr komplex, haben viele verschiedene Funktionen und spielen eine Schlüsselrolle bei Prozessen wie Signaltransduktion, Zellwachstum und Differenzierung, Transport und Stoffwechsel. Zusätzlich wird ihre Bedeutung dadurch verstärkt, dass viele der heutigen Medikamente an Membranproteinen angreifen. Trotz ihrer steigenden Bedeutung entfallen lediglich 1,7% der in der *protein databank* (PDB) hinterlegten Proteinstrukturen auf experimentell ermittelte Strukturen von Membranproteinen. Auf eine ähnliche Situation stieß man in den 1970er Jahren bei löslichen Proteinen. Technologischer Fortschritt im Bereich der Methoden zur Strukturaufklärung führte zu einer fast exponentiell steigenden Zahl gelöster Strukturen. Jedoch ist die Handhabung von Membranproteinen, bedingt durch ihre Topologie, die den ineffizienten Transport und die Insertion in zelluläre Membranen beeinflusst, die toxischen Effekte der heterologen Expression für den Wirtsorganismus und ihre Instabilität, schwierig. Da die Menge von Membranproteinen die in natürlichen Geweben bis auf wenige Ausnahmen vorkommen, nicht ausreicht, um sie aufzureinigen, ist eine heterologe Überexpression notwendig. In den vergangenen Jahren legten verschiedene *structural genomics* Projekte ihren Schwerpunkt nicht nur darauf Membranproteine zu produzieren, sondern auch darauf, effiziente Methoden zur Überexpression zu entwickeln.

Diese Arbeit zielte darauf ab, verschiedene sekundär aktive Transportproteine für strukturelle und funktionelle Untersuchungen heterolog zu produzieren und alternative Strategien, wie die zellfreie Expression und Maltose-Bindeprotein-Fusionen (MBP), zu etablieren, um die Hürden im Zusammenhang mit der heterologen Überexpression zu überwinden. Zu diesem Zweck wählte ich, auf Grundlage verfügbarer bioinformatischer Daten der *Transporter protein analysis database* (*TransportDB*; [www.membranetransport.org](http://www.membranetransport.org)), 14 sekundär aktive Transporter aus sieben verschiedenen Familien der drei Organismen *Salmonella typhimurium*, *Aquifex aeolicus* und *Pyrococcus furiosus*, aus. Es sollten solche Transporter ausgewählt werden, von denen keine atomare Proteinstruktur mit atomarer Auflösung verfügbar war und die noch wenig untersucht wurden.

## ZUSAMMENFASSUNG

Ich began mit vier Proteinen der Schwermetall-transportierenden *cation diffusion facilitator* (CDF) Familie, von denen jeweils zwei aus *Salmonella typhimurium* und *Aquifex aeolicus*, stammten. Die Klonierungs- und Expressionsversuche mit drei verschiedenen Vektoren, die jeweils in zwei verschiedenen Versionen vorlagen, welche sich in der Position des *tags* unterschieden, ergab, dass die CDF-Familie eine gut exprimierbare Familie darstellt. Alle Konstrukte exprimierten gut in den entsprechenden *E. coli* Wirtsorganismen. Funktionelle Komplementationsstudien mit dem Zink-Export defizienten *E. coli* Stamm GG48 zeigten, dass STM0758 aus *S. typhimurium* die Zink-Defizienz komplementieren kann. Jedoch erfordern strukturelle Untersuchungen ausreichende Mengen an stabilem und gereinigtem Protein. Diesbezüglich stellte sich Aq\_2073 aus *A. aeolicus* als vielversprechendster Kandidat heraus.

Aq\_2073 wurde in den araBAD Operon basierten Vektor pBAD, der einen N-terminalen deca-His *tag* und einem C-terminalen STREP-II *tag* enthält kloniert, in *E. coli* Top10 Zellen exprimiert und mittels immobilisierter Metallionen-Affinitätschromatographie (IMAC) zur Homogenität gereinigt. Das finale Reinigungsprotokoll ergab eine Ausbeute von 2,5 mg Protein pro einem Liter Kulturvolumen.

Der nächste Schritt war die Charakterisierung des Proteins hinsichtlich seiner Stabilität, da Membranproteine auf Grund ihrer amphiphilen Natur oft unter unvorteilhaften Bedingungen aggregieren, was die Kristallisation ausschließt. Aq\_2073 wurde umfangreich in verschiedenen Detergenzien und bei unterschiedlichen pH-Werten untersucht, wobei das Gelfiltrationsprofil als Maßstab für eine korrekte Faltung und die Stabilität diente. Es konnte gezeigt werden, dass das Protein in einer Vielzahl von Detergenzien und Detergenz-Gemischen, einschließlich bestimmter harter Detergenzien wie N-dodecylphosphocholine (FOS12) und N,N-dimethyldodecylamin-N-oxid (LDAO), homogen war. Interessanterweise zeigte sich, dass in einer Reihe verschiedener Maltoside die Stabilität des Proteins mit sich verkürzender Kettenlänge der Kohlenwasserstoffkette abnahm. Allerdings war das Protein im Wesentlichen stabil von pH 4 bis pH 8, selbst nach Lagerung über 15 Tage bei Raumtemperatur. Außerdem konnte für Aq\_2073 eine hohe Schmelztemperatur ( $T_m$ ) um 96 °C in einem physiologischen pH-Bereich mit *differential scanning calorimetry* (DSC) gemessen

## ZUSAMMENFASSUNG

werden. Diese Eigenschaften zusammen machten Aq\_2073 zu einem attraktiven Kandidaten für strukturelle Untersuchungen.

Kristallisationsversuche wurden mit kommerziell erhältlichen Screens durchgeführt und resultierten in ersten nadelähnlichen Kristallen. Jedoch ergaben umfangreiche Optimierungen keine weiteren vielversprechenden Bedingungen. Daran anschließend entwarf ich sechs N- und C-terminal verkürzte Versionen von Aq\_2073, basierend auf Homologie-Modellen, einschließlich zweier Konstrukte, die lediglich aus der vorhergesagten löslichen C-terminalen Domäne bestanden. Drei dieser Konstrukte wurden zur Homogenität gereinigt, wobei die Kristallisationsversuche nicht erfolgreich waren. Die Bemühungen zur Strukturaufklärung von Aq\_2073 waren nicht nur auf 3D-Kristallisation beschränkt, sondern wurden auch auf 2D-Kristallisation erweitert mit Hilfe von cryo-EM. Um hierbei Kristalle zu erhalten wurden verschiedene Detergenzien, Lipide, Lipid-zu-Protein Verhältnisse (*lipid protein ratio; LPR*), Salzkonzentrationen, Temperaturen und Additive mit dem Vollängenkonstrukt und mit einem C-terminal verkürzten Konstrukt (C1) von Aq\_2073 getestet. Das beste Ergebnis, Vesikel optimaler Größe mit kristallinen Bereichen, wurde bei einem LPR von 0.5 in Gegenwart von *E. coli total lipid extract*, einer minimalen Salzkonzentration und Dialyse bei 30 °C erhalten. Die oligomere Struktur von Aq\_2073 wurde durch komplementäre Techniken untersucht, wie *blue native* Elektrophorese, Größenausschlußchromatographie und chemischer Vernetzung. Alle Ergebnisse deuten auf einen oligomeren Zustand von Aq\_2073 in Detergenzlösung hin, möglicherweise als Dimer und einen höheren Oligomerezustand. Bindungsstudien mit einer Vielzahl möglicher Substrate (Zink, Cadmium, Nickel und Eisen) durch DSC zeigte, dass Cadmium die Schmelztemperatur ( $T_m$ ) um 10 °C steigerte.

Die nächste Familie sekundär aktiver Transporter, mit der ich mich beschäftigte, war die *potassium uptake permease* (KUP) Familie. STM3880 aus *S. typhimurium* wurde im pBAD Vektor mit einem C-terminalen deca-His-tag exprimiert. Mit Hilfe einer IMAC konnte eine ausreichende Menge an Protein für strukturelle Untersuchungen aufgereinigt werden. Jedoch blieb STM3880 offenkundig hinsichtlich seiner Stabilität problematisch. Weitreichende Optimierungen von Detergenzien, Salzen, pH-Werten, Additiven und Kultivierungsbedingungen genügten nicht, um die Langzeitstabilität des Proteins zu verbessern, was eine Voraussetzung für die Kristallisation ist. Die einzige Lösung war

## ZUSAMMENFASSUNG

zügig nach der Reinigung weiter zu arbeiten. STM3880 wurde in Liposomen rekonstituiert und durch SSM-Experimente funktionell charakterisiert. Lediglich bei pH 6 führte ein Konzentrationsprung des Substrats (KCl) von 1 mM zu einem messbaren Strom, was darauf schließen lässt, dass das Protein wahrscheinlich nur bei niedrigerem pH-Wert aktiv ist. Während der Arbeit stieß ich auf Membranproteine, die bedingt durch ihre toxischen Effekte, unzureichender Ausbeute und proteolytischen Abbau in zellbasierten Systemen schwierig heterolog herzustellen waren. Um diese Proteinen trotzdem untersuchen und kristallisieren zu können, entschied ich mich für das *E. coli* S30-Extrakt basierte zellfreie Expressionssystem, das nicht nur die Möglichkeit bietet, Probleme mit der zellbasierten heterologen Expression zu umgehen, sondern auch die Reaktionsbedingungen zu kontrollieren. Der *in vivo* hergestellte Transporter PF0780 der MOP-Transporter Familie aus *P. furiosus* lieferte gut beugende Kristalle. Ich etablierte zellfreie Produktionsprotokolle, sowohl in Abwesenheit (PCF-Modus), als auch in Gegenwart von Detergenz (DCF-Modus). Das im PCF-Modus hergestellte Protein wurde zur Homogenität gereinigt und wies eine gute Langzeitstabilität auf, wie durch analytische Größenausschlußchromatographie gezeigt wurde. Kristallisation wurde mit PF0780, welches im PCF-Modus produziert wurde, versucht. *Blue native* Gelelektrophorese zeigte, dass das zellfrei produzierte Protein in einem oligomeren Zustand vorliegt. Da jedoch das im DCF-Modus hergestellte Protein eine geringere Reinheit und Stabilität zeigte, ist der PCF-Modus als Expressionsmethode für weitere strukturelle und funktionelle Untersuchungen vorzuziehen.

Erste funktionelle Charakterisierungen von STM3476, einem Transporter der *formate nitrite transporter* (FNT) Familie aus *S. typhimurium*, zeigten Hinweise für einen Nitrittransport. Ich versuchte STM3476 zellfrei für funktionelle und strukturelle Untersuchungen zu produzieren. Sowohl der PCF- als auch der DCF-Modus wurden mit verschiedenen Detergenzien für eine maximale Ausbeute getestet. Mit Hinblick auf die endgültige Ausbeute wurde der Einfluss der Temperatur im PCF-, im DCF-Modus und während der Solubilisierung des Präzipitats untersucht. Sowohl im PCF- als auch im DCF-Modus verringerte sich die Ausbeute mit sinkender Temperatur nahezu linear. Im Gegensatz dazu stellte sich bei der Solubilisierung des Präzipitats bei verschiedenen Temperaturen (4 °C, 20 °C, 30 °C) mit zwei verschiedenen Detergenzien heraus, dass der Art des Detergenz eine wichtigere Rolle zukam als der Temperatur.



## ZUSAMMENFASSUNG

Zwei weitere Transporter, STM1781 aus der *sulfate permease* (SulP) Familie und STM3959 aus der Aminosäuretransportierenden *resistance to homoserine/threonine* (RhtB) Familie, beide aus *S. typhimurium*, wurden im PCF-Modus hergestellt. Im zellbasierten Expressionssystem wurde STM1781 stark abgebaut, während STM3959 überhaupt nicht produziert wurde. Das im PCF-Modus hergestellte STM1781 wurde ohne Abbau aufgereinigt und an einem Smart-System hinsichtlich seiner Stabilität durch analytische Größenausschlußchromatographie mit einer Superdex 200 Säule in Gegenwart der Detergenzien *n*-Dodecyl- $\beta$ -D-Maltosid ( $\beta$ -LM) und FOS12 untersucht. Der Transporter STM3959 hingegen wurde im PCF-Modus hergestellt und in Gegenwart von  $\beta$ -LM aufgereinigt. Jedoch besaß dieses Protein keine Langzeitstabilität.

An drei ausgewählten Transportern wurde der Einfluss von MBP-Fusionen auf die Expression und Membranintegration von Membranproteinen getestet. STM0365 und STM3959 wurden, sowohl mit als auch ohne MBP-Fusion, unter dem Detektionslimit gehalten. Andererseits führte die MBP-Fusion bei STM0758, einem Mitglied der CDF Familie, dessen beide Termini im Cytoplasma erwartet werden, zu einer 3-5 fachen Verringerung in der Produktion und Membranintegration, je nach Position des *His-tags*. Eventuell behinderte die Positionierung des MBPs am N-Terminus eines Proteins, das beide Enden im Cytoplasma hat, die korrekte Insertion in die Membran und führte somit zu einer verringerten Produktion. Diese Beobachtung unterstützt ebenfalls die für STM0758 vorhergesagte Topologie mit beiden Termini im Cytoplasma.

In dieser Arbeit konnten einige sekundär aktive Transporter erfolgreich heterolog produziert und entsprechende Reinigungsprotokolle etabliert werden, um strukturelle und funktionelle Untersuchungen im Rahmen eines *structural genomics* Projektes durchzuführen. Es zeigte sich ebenfalls, dass die Einbindung alternativer Strategien, wie die Anwendung zellbasierter und zellfreier Expressionssysteme, die Gesamtanzahl herstellbarer Proteine steigert und damit die Erfolgsaussichten derart angelegter Projekte verbessert.

**Abstract**

Genes coding for membrane proteins make up 25%-30% of the genome in most organisms. Membrane proteins play an important role in cell functioning and their importance is enhanced by the fact that a large number of drugs are targeted at membrane proteins. Paradoxically, experimentally determined structures of membrane protein correspond to only about 1.7% of protein structures deposited in the protein data bank (PDB). This is largely due to the fact that membrane proteins are difficult to deal with owing to their amphipathic nature. The low abundance of membrane proteins in native tissue makes heterologous overexpression of these genes a necessity. This thesis work aimed at heterologous production of several secondary active transporter proteins for structural and functional characterizations and establishing alternative strategies to overcome the obstacles associated with heterologous overproduction.

Four members of the heavy metal transporting cation diffusion facilitator (CDF) family from *S. typhimurium* and *A. aeolicus* were heterologously overproduced in *E. coli* and functionally characterized by an *in vivo* complementation assay using the zinc transport deficient *E. coli* GG48 strain. Out of these four, Aq\_2073 from *A. aeolicus* was produced in large scale with substantial yield and purity sufficient to carry out structural studies. After extensive stability studies with different detergents, pHs and temperatures, the protein was subjected to 3D and 2D crystallization trials. Several C- terminal truncated constructs were made and the simultaneous crystallization screenings were carried out. These resulted in initial needle like crystals in 3D crystallization trials or optimum sized vesicles with crystalline patches in 2D crystallization trials but no obvious crystal. The protein showed significant increase in melting temperature in the presence of cadmium, when tested by differential scanning calorimetry.

Another transporter, STM3880 of the potassium uptake permease (KUP) family from *S. typhimurium*, was heterologously overproduced in *E. coli*, purified by affinity chromatography, reconstituted into artificial liposome and functionally characterized by solid supported membrane based electrophysiology.

In order to establish alternative expression strategies, continuous exchange cell free expression (CECF) of proteins from four different families was carried out. This method

## **ABSTRACT**

found to be aptly complementing the cell-based production approach. Targets from resistance to homoserine/threonine (RhtB) family not expressing *in vivo* could be expressed and purified using CECF. STM1781 of the sulfate permease (SulP) family was expressed, purified and characterized for stability while the cell-based production resulted in extensive degradation. PF0780 of multidrug/oligosaccharidyl-lipid/polysaccharide flippase (MOP) family was also purified to homogeneity and the stability was comparable to *in vivo* produced protein. Moreover, the effect of maltose binding protein (MBP) fusion at N-terminus on production and membrane integration was tested with three selected targets. The analysis revealed decreased yields in the presence of MBP if the protein had both termini in the cytoplasm.

This work succeed in heterologously overproducing and establishing purification protocols for several secondary active transporters aiming at structural and functional characterization in a structural genomics framework. It also showed that integration of alternative strategies, like employing both cell-based and cell-free heterologous expression systems, expands the overall expression space coverage and in turn increases the chance of success of a structural genomics styled project.

## Summary

Membrane proteins are often very complex, function in many different ways and play key roles in processes including signal transduction, cell growth and differentiation, transport and metabolism. In addition, their importance is enhanced by the fact that a large number of today's drugs are targeted at membrane proteins. Despite this high significance, experimentally determined structures of membrane protein correspond to only about 1.7% of protein structures deposited in the protein data bank (PDB). A similar situation was encountered for soluble proteins in 1970s. Technological advancement in the methods for structure determination led to an almost exponential increase in the number of solved structures. However, membrane proteins are difficult to deal with due to their topology which affects the inefficient transport and insertion in cellular membranes, the toxic effect of recombinant membrane proteins on host cells, and their instability. As the abundance of membrane proteins in native tissue with few exceptions is insufficient for purification, heterologous overexpression is a necessity. In recent years, several structural genomics projects put emphasis not only to overproduce membrane proteins but also to develop efficient methods for membrane protein production.

This thesis work aimed at heterologously overproducing several secondary active transporter proteins for structural and functional characterizations and establishing alternative strategies, namely, cell-free expression and Maltose binding protein (MBP) fusion to overcome the obstacles associated with heterologous overproduction. For this purpose, I selected 14 secondary active transporter proteins of 7 different families from 3 different source organisms, namely, *Salmonella typhimurium*, *Aquifex aeolicus* and *Pyrococcus furiosus* based on available bioinformatics information provided by Transporter protein analysis database (TransportDB; [www.membranetransport.org](http://www.membranetransport.org)). The criteria were to select those secondary active transporter families which lack an atomic resolution structure and are not sufficiently studied as evident from the literature.

I started off with four heavy metal transporting cation diffusion facilitator (CDF) family proteins, two each from *Salmonella typhimurium* and *Aquifex aeolicus*. Cloning and expression screening with three different vectors having two different versions based on tag positions revealed that the CDF family is a well expressing family. All the constructs

## SUMMARY

expressed well in respective *E. coli* host cells. Functional complementation studies with zinc export deficient *E. coli* strain GG48 showed STM0758 from *S. typhimurium* complementing the zinc deficiency function. However, structural studies with membrane proteins need sufficient amount of stable and purified protein. In this regard, Aq\_2073 from *A. aeolicus* turned out to be the most suitable one.

Aq\_2073, cloned in araBAD operon based pBAD vector with N-terminal deca His-tag and C-terminal STREP-II tag, was expressed in *E. coli* Top10 cells and purified to homogeneity by immobilized metal affinity chromatography (IMAC). The purification protocol for Aq\_2073 was established providing a final yield around 2.5 mg from 1 L culture volume.

The next step was to characterize the protein for its stability as membrane proteins often aggregate in unfavorable conditions precluding crystallization owing to their amphiphilic nature. Aq\_2073 was characterized extensively in different detergents and pHs taking the gel filtration profile as a benchmark for correct folding and stability. The protein was found to be homogeneous in a range of detergents and detergent mixtures including certain strong detergents like N-dodecylphosphocholine (FOS12) and N,N-dimethyldodecylamin-N-oxid (LDAO). Interestingly, when tested in a series of maltoside detergents differing in alkyl chain length, the stability went down with decreasing carbon chain length. However, the protein was substantially stable at pH 4 to pH 8 and even after 15 days of storage at room temperature. In addition, Aq\_2073 showed a high melting temperature ( $T_m$ ), around 96 °C, at a physiological pH range in differential scanning calorimetry (DSC) measurement. These characteristics altogether made Aq\_2073 an attractive target for structural studies.

Crystallization trials were performed with commercially available screens and resulted in initial needle like crystals. However, extensive optimization of this initial condition failed to generate any further promising crystals. Subsequently, I designed six N- and C-terminally truncated versions of Aq\_2073 based on homology alignment, including two constructs made only of the predicted soluble C-terminal domain. Three of these were purified to homogeneity and subjected to crystallization trials without any success. The structural endeavor of Aq\_2073 was not only limited to 3D crystallization screens but

## SUMMARY

also extended to 2D crystallization using cryo-EM. Different lipids, detergents, lipid protein ratios (LPR), salt concentrations, temperatures and additives were screened, both with full length and a C-terminal truncated version (C1) of Aq\_2073, to generate 2D crystals. The best result, an optimum sized vesicle with crystalline patches, was obtained with *E. coli* total lipid extract at LPR 0.5 with minimum salt concentration and dialysis at 30 °C.

The oligomeric structure of Aq\_2073 was studied by complementary techniques like blue native electrophoresis, size exclusion chromatography and chemical crosslinking. All results indicated that Aq\_2073 exists as an oligomer in detergent solution, possibly in equilibrium between a dimer and a higher oligomeric state. Binding studies with a range of putative substrates (zinc, cadmium, nickel and ferrous) by DSC showed that cadmium could increase the melting temperature ( $T_m$ ) by 10 °C.

The next family of secondary active transporters I worked on was the potassium uptake permease family (KUP). STM3880 of KUP family from *S. typhimurium* was expressed in the pBAD vector with a C-terminal deca His-tag. The protein was purified using IMAC to sufficient quantity for structural studies. However, STM3880 remained notoriously difficult in terms of stability. Extensive optimization with detergents, salts, pHs, additives and culturing conditions were not enough to increase the long term stability of this target which is the prerequisite for crystallization. The only solution was to work fast after purification. Purified STM3880 was reconstituted in liposomes and functionally characterized by solid supported membrane (SSM) experiments. Only at pH 6, a substrate (KCl) concentration jump of 1 mM produced a detectable current which indicated that the protein is probably active only at low pH.

In due course of my thesis work, I came across with membrane protein targets which were difficult to heterologously overproduce in cell-based systems because of their toxic effect, insufficient yield, or extensive proteolytic degradation. I opted for *E. coli* S30 extract based cell-free expression system which not only offers the possibilities to overcome the problems associated with cell-based heterologous overexpression but also to control the reaction conditions. In addition, it was also necessary to establish alternative production protocols for attractive targets which already generated either

diffracting crystals or functional data and test the crystallizability and functional reproducibility of the cell-free produced protein.

*In vivo* produced PF0780, the multidrug/oligosaccharidyl-lipid/polysaccharide flippase (MOP) family transporter from *P. furiosus* generated well diffracting crystals. I established a cell-free production protocol of PF0780, both in the absence of detergent (PCF mode) and in the presence of detergent (DCF mode). The PCF produced protein was purified to homogeneity and showed long term stability when analyzed by analytical gel filtration chromatography. Crystallization was attempted with PCF produced PF0780. Blue native gel electrophoresis showed the cell-free produced protein to be oligomeric in nature. However, DCF mode production resulted in decreased purity and yield, establishing PCF mode production as the preferred method for further structural and functional studies.

Initial functional characterization of STM3476, the formate nitrite transporter (FNT) family protein from *S. typhimurium* showed evidence of nitrite transport. I tried to produce STM3476 by cell-free expression for functional and structural characterizations. Both PCF and DCF mode production were carried out with detergent screening for maximum yield. The effect of the reaction temperature in terms of final yield was studied on PCF production, DCF production and the solubilization of PCF produced protein precipitates in presence of detergents. Both in PCF and DCF mode production, the yield went down linearly with the decrease in temperature. However, solubilization of precipitates (PCF mode) at different temperatures (4 °C, 20 °C and 30 °C) with two different detergents revealed that the detergent type played a more prominent role than the temperature.

Two more transporters, STM1781 of sulfate permease (SulP) family and STM3959 of amino acid transporting resistance to homoserine/threonine (RhtB) family, both from *S. typhimurium*, were produced in PCF mode. In the cell-based heterologous expression system, STM1781 was heavily degraded while STM3959 was not expressed at all. PCF produced STM1781 protein was purified without degradation and characterized for stability using *n*-dodecyl- $\beta$ -D-maltoside ( $\beta$ LM) and FOS12 detergents by analytical gel filtration using superdex 200 column in SMART system. Similarly, PCF produced

## SUMMARY

STM3959 protein of RhtB family was purified in the presence of  $\beta$ LM. However, the protein lacks long term stability.

The effect of MBP fusion on production and membrane integration of the membrane protein was tested with three selected transporters. STM0365 and STM3959 of the RhtB family were expressed under the detection limit both with and without MBP fusion. On the other hand, production and membrane integration upon MBP fusion of STM0758, a CDF family protein predicted to have both termini inside the cytoplasm, was lowered by 3 to 5 times depending on the His tag positions. Probably, the placement of MBP on the N-terminus of a protein having its ends in the cytoplasm interfered with correct insertion into the membrane and resulted in reduced production. This observation also supported the predicted topology of STM0758 with both ends in the cytoplasm.

This work succeed in heterologously overproducing and establishing purification protocols for several secondary active transporters aiming at structural and functional characterization in a structural genomics framework. It also showed that integration of alternative strategies, like employing both cell-based and cell-free heterologous expression systems, expands the overall expression space coverage and in turn increases the chance of success of a structural genomics styled project.



**Abbreviations**

2D	2 Dimensional
3D	3 Dimensional
$\beta$ ME	$\beta$ -Mercaptoethanol
ABC	ATP Binding Cassette
AP	Alkaline phosphatase enzyme
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BCA	Biscinchonic acid
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BNPAGE	Blue native polyacrilamide gel electrophoresis
BSA	Bovine serum albumin
CBB	Commassie brilliant blue dye
CECF	Continuous exchange cell free
CMC	Critical micellar concentration
DCF	Detergent cell free reaction mode
DSC	Differential scanning calorimetry
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothritol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ETL	<i>E. coli</i> lipid total extract
GFP	Green fluorescent protein
GPCR	G protein coupled receptor
HEPES	N-(2-Hydroxyethyl)-piperazine-N'-2-ethanesulfonic acid
His	Histidine
IMAC	Immobilized metal ion chromatography
IMP	Integral membrane protein
IPTG	isopropyl- $\beta$ -1-thiogalactoside
kD	kilo Dalton
LPR	Lipid protein ratio

## ABBREVIATIONS

MME	mono methyl ether
NBT	Nitrobluetetrazolium
NMR	Nuclear magnetic resonance
ORF	Open reading frame
PCF	Precipitate cell free reaction mode
PDB	The protein data bank
PEG	Polyethylene glycol
polyHis	Poly histidine
PVDF	Polyvinylidene difluoride
RPM	Revolutions per minute
RT	Room temperature
SEC	Size exclusion chromatography
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrilamide gel electrophoresis
SSM	Solid supported membrane
TEV protease	Tobacco Etch Virus protease
TEMED	N,N,N,'N,' - tetramethylethylenediamine
T <sub>m</sub>	Melting temperature
TRIS	Trishydroxymethylaminoethane

### **Transporter protein families**

CDF	Cation diffusion facilitator family
FNT	Formate nitrite transporter family
MOP	Multidrug/oligosaccharidyl-lipid/polysaccharide flippase superfamily
KUP	Potassium uptake permease family
SulP	Sulfate permease family
RhtB	Resistance to homoserine/threonine family

**Detergents**

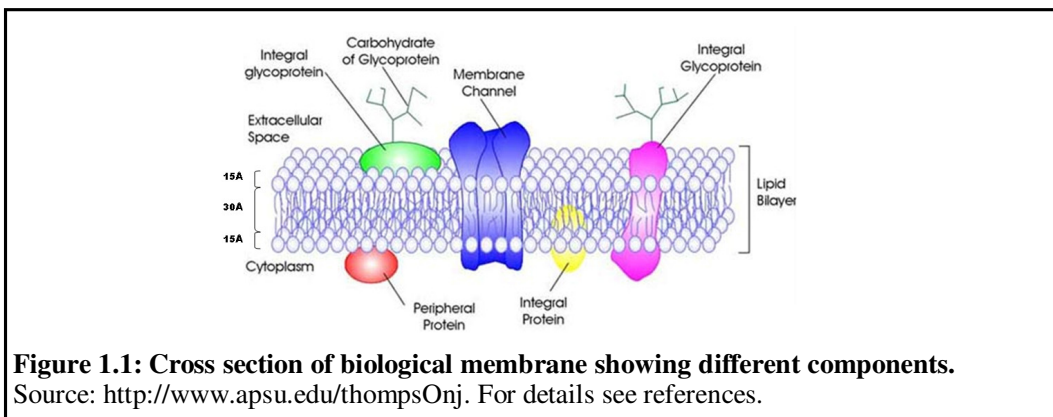
$\beta$ LM	<i>n</i> -dodecyl- $\beta$ -D-maltoside
DM	<i>n</i> -decyl- $\beta$ -D-maltoside
UM	<i>n</i> -undecyl- $\beta$ -D-maltoside
NM	<i>n</i> -nonyl- $\beta$ -D-maltoside
Fos-12	N-dodecylphosphocholine
Fos-11	N-undecylphosphocholine
C <sub>10</sub> E <sub>5</sub>	Penteethylene glycolmonodecyl ether
C <sub>12</sub> E <sub>9</sub>	dodecyl nonaoxyethylene ether
OG	<i>n</i> -octyl- $\beta$ -D-glucoside
OTG	<i>n</i> -octyl- $\beta$ -D-thioglucoside
NG	<i>n</i> -nonyl- $\beta$ -D-glucoside
LDAO	N,N-dimethyldodecylamin-N-oxid
LMPC	1-myristoyl-2-hydroxy- <i>sn</i> -glycero-3-phosphocholine
LMPG	1-myristoyl-2-hydroxy- <i>sn</i> -glycero-3-[phospho- <i>rac</i> -(1-glycerol)]
LPPG	1-palmitoyl-2-hydroxy- <i>sn</i> -glycero-3-[phospho- <i>rac</i> -(1-glycerol)]
Brij35	Polyoxyethylene-(23)-lauryl-ether (C12/23)
Brij-56	Polyoxyethylene-(10)-cetyl-ether (C16/10)
Brij-58	Polyoxyethylene-(20)-cetyl-ether (C16/20)
Brij-72	Polyoxyethylene-(2)-stearyl-ether (C18/2)
Brij-78	Polyoxyethylene-(20)-stearyl-ether (C18/20)
Brij-98	Polyoxyethylene-(20)-oleyl-ether (C18-1/20)

## *Introduction*

---

## **1.1 Biological membrane**

Membrane formation was a key step towards the formation of individual life. All living organisms are confined to a limited volume and surrounded by a surface border. The construction of that border represents one of the most fundamental considerations in biological organization. This outer shell is built to keep the interior contents from leaking out into the surrounding environment. The chemical processes of cellular life generally take place in an aqueous solution and the intracellular constituents of cells are largely molecules which are readily soluble in water. Similarly, the environment surrounding cells is an aqueous one. A biological membrane composed of fatty molecules separate inside and outside (Alberts et al. 1994). This biological membrane or plasma membrane is almost invariably a bilayer consisting of amphiphilic phospholipid molecules (Fig. 1.1). However, membranes are not merely a passive barrier. They include arrays of proteins specialized for facilitating various cellular processes, such as transporting metabolites and ions, sensing extracellular signals and energy transduction (Nelson and Cox 2000).



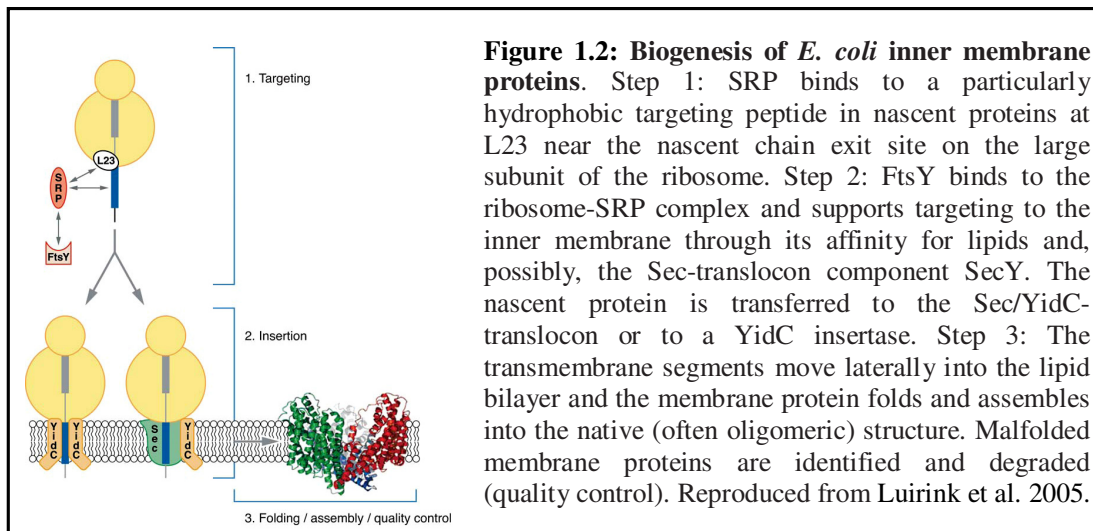
## **1.2 Integral membrane proteins (IMP) and their biogenesis**

Integral membrane proteins (IMPs) hold a critical role in the cell factory. They are of paramount importance for a wide variety of cellular processes. In the completely sequenced genomes, the fraction of genes coding for membrane proteins is estimated to be about 25% (Lundstrom 2006). IMPs are composed of either helices (helical protein) or  $\beta$  sheets ( $\beta$  barrel protein). Though the helical types are predominant among IMPs,  $\beta$  barrel proteins are mostly found in outer membranes of gram negative bacteria. Apart

## INTRODUCTION

from that, membranes of cyanobacteria, mycobacteria, chloroplasts and mitochondria also have  $\beta$  barrel proteins. The firm attachment of IMPs to membranes is the result of mainly hydrophobic interactions between membrane lipids and hydrophobic domains of the proteins.

Biogenesis of most  $\alpha$  helical proteins appears to follow a partly conserved co-translational pathway (reviewed in details by White and von Heijne 2004; Luirink et al. 2005). In *E. coli* targeting involves a relatively simple signal recognition particle (SRP) and its cognate receptor SR (SRP Receptor). Insertion into the plasma membrane occurs via the Sec translocon, often with the involvement of some other associated proteins like YidC, Oxa1 and Alb3 (Fig. 1.2)



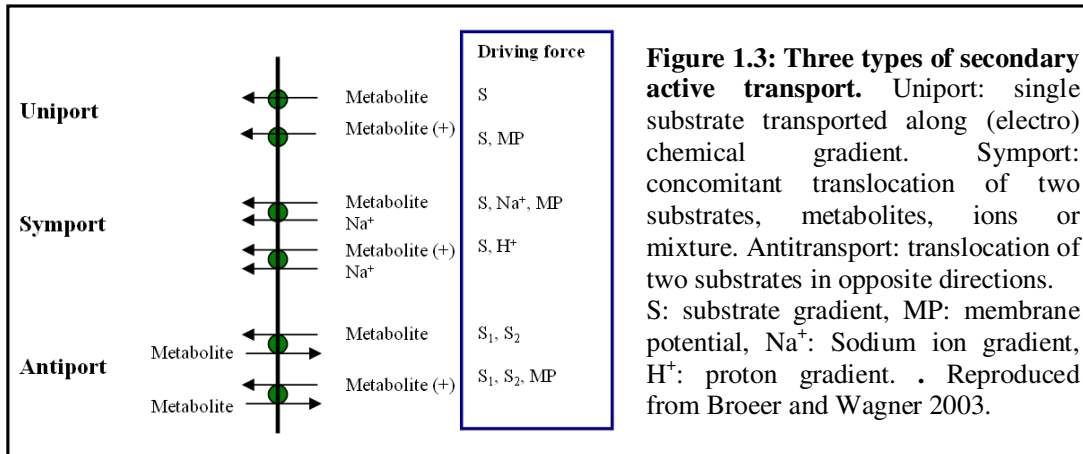
**Figure 1.2: Biogenesis of *E. coli* inner membrane proteins.** Step 1: SRP binds to a particularly hydrophobic targeting peptide in nascent proteins at L23 near the nascent chain exit site on the large subunit of the ribosome. Step 2: FtsY binds to the ribosome-SRP complex and supports targeting to the inner membrane through its affinity for lipids and, possibly, the Sec-translocon component SecY. The nascent protein is transferred to the Sec/YidC-translocon or to a YidC insertase. Step 3: The transmembrane segments move laterally into the lipid bilayer and the membrane protein folds and assembles into the native (often oligomeric) structure. Malfolded membrane proteins are identified and degraded (quality control). Reproduced from Luirink et al. 2005.

### 1.3 Secondary active transporters

In terms of energy coupling mechanisms, two classes of membrane transporters, the primary active and secondary active transporters are abundant in all known species across different domains of life. Primary active transporters convert light or chemical energy into electrochemical energy, such as solute concentration gradients plus an electric voltage across membranes. The majority of the primary systems belong to the ATP binding cassette (ABC) superfamily. On the other hand secondary active transporters use the free energy stored in ion and/or solute gradients to drive the transport of another ion or solute across the cytoplasmic or internal membranes of biological cells. They are widely spread through all kingdoms of life; they are found in all biological cells and can probably be found for every low molecular weight compound in nature (Sobczak

## INTRODUCTION

and Lolkema, 2005). Three different mechanisms of secondary active transports are recognized (Fig. 1.3), uniport, symport and antiport. In uniport a single substrate is transported along the electrochemical gradient while in symport two different substrates are transported in the same direction. The antiporters transport two different substrates in the opposite direction. However, some transporters may actually obey a mixture of these modes, such as the glutamate transporter which translocates glutamate together with three  $\text{Na}^+$  ions and a proton and also in exchange of  $\text{K}^+$  (Broer and Wagner 2003).



The high abundance of secondary active transporters is reflected in the great diversity of their coding sequences. The transporter classification system (TC system) developed in the Saier lab (Saier 2000) is based on sequence homology and lists some 550 transporter families based on 3000 protein sequences (<http://www.tcdb.org/>).

### 1.4 Represented families of secondary active transporters

#### 1.4.1 Cation diffusion facilitator family (CDF)

The members of the cation diffusion facilitator (CDF) family serve as major metal efflux proteins. The primary substrate of the CDF transporters is zinc, but they also transport cobalt, nickel, cadmium and ferrous ions along with a proton or potassium ion in antiport fashion (Guffanti et al. 2002; Nies 2003). CzcD of *Ralstonia metallidurans* CH34, the archaetype of the family, was reported to mediate resistance to cobalt, cadmium and zinc in *R. metallidurans* while only zinc and cadmium in *E. coli* (Anton et al. 1999; 2004). The *E. coli* protein ZitB was reported to function as an antiporter *in vivo*, exchanging zinc for potassium ion, while YiiP showed evidence of an iron efflux function along with zinc (Grass et al. 2001; Lee et al. 2002; Grass et al. 2005). The human homologues of CDF transporters are called ZnT (SLC30). There are at least 9 ZnT transporters (Znt1-

## INTRODUCTION

ZnT9) found in human cells which exhibit unique tissue specific expression, differential responsiveness to dietary zinc deficiency and excess and differential responsiveness to physiologic stimuli via hormones and cytokines (Liuzzi 2004). Most CDF transporters have 6 putative transmembrane segments and function as antiporters. Fig. 1.4 depicts the sequence alignment of known CDF transporters with the selected targets for this thesis work. The CDF members show the highest level of amino acid residue conservation in transmembrane helix II. Paulsen and Saier (1997) proposed a CDF family specific signature sequence which begins with a fully conserved serine and continues just past a fully conserved aspartate.

S X (ASG) (LIVMT)<sub>2</sub> (SAT) (DA) (SGAL) (LIVFYA) (HDN) X<sub>3</sub>DX<sub>2</sub>(AS)

[X = any residue; alternative residues at any one position are in parentheses]

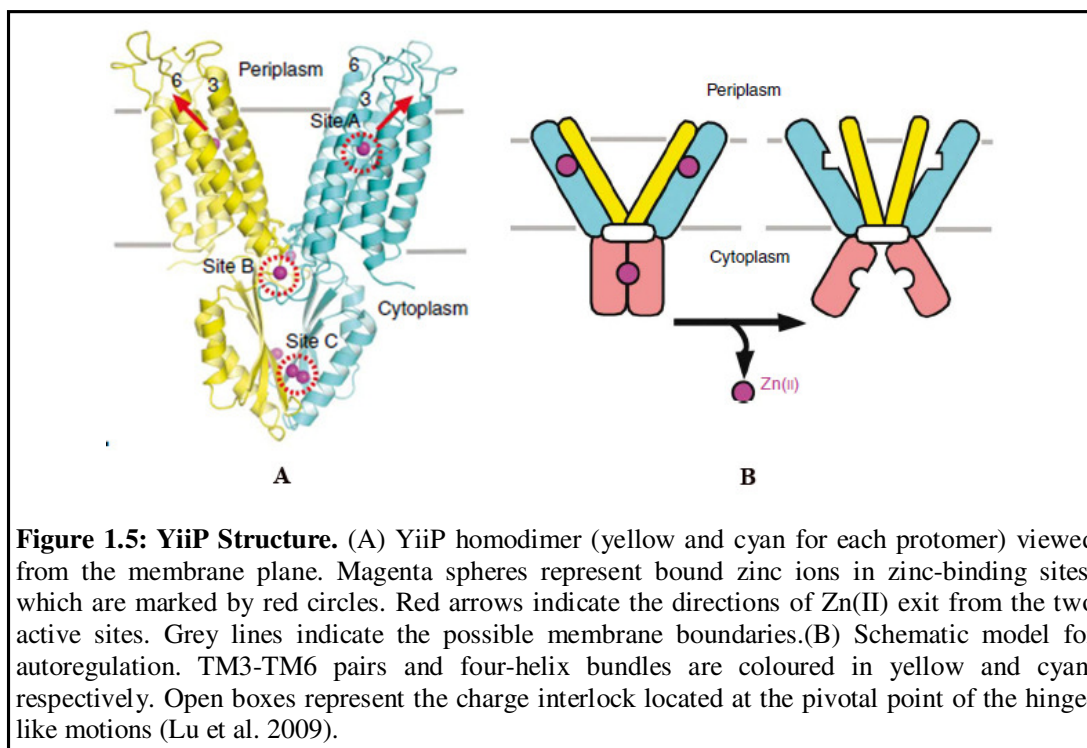
Recently, an atomic structure of YiiP, CDF protein from *E. coli*, has been published (Lu and Fu 2007, Lu et al. 2009). It is a homodimer held together in a parallel orientation through four zinc ions at the interface of the cytoplasmic domains (CTD), whereas the two transmembrane segments (TMS) swing out to yield a Y-shaped structure. In each protomer, the CTD adopts a metallochaperone like protein fold; the TMS features a bundle of six transmembrane helices and a tetrahedral zinc binding site located in a cavity that is open to both, the membrane outer leaflet and the periplasm (Fig. 1.5). Site directed fluorescence resonance energy transfer (FRET) and mutation activity analysis suggested that zinc binding triggers hinge movements of two transmembrane domains. Highly conserved salt bridges interlock the transmembrane helices at the dimer interface, where they are well positioned to transmit zinc induced interdomain movements to reorient transmembrane helices, thereby modulating coordinating geometry of the active site for zinc transport.

The active site of zinc transport (Site A, one of three sites in each YiiP protomer) consists of four coordinating residues which facilitate preferred tetrahedral coordination for Zn(II) and Cd(II), but not for other divalent cations, such as Fe(II), Ni(II) and Co(II). This failed to explain the iron tolerance of a mutant *E. coli* strain overexpressing YiiP (Grass et al. 2005). Moreover, the role of protons in zinc transport by CDF transporters still remains obscure. Clearly more structural and functional efforts are required to find out how exactly these broad substrate specific CDF transporters work.



Yiip	-----MNQSYGRVLSRAAIAATAMASLLLLIKIFAWWYTG	SVSILAALVDSLVDIGA	52
<b>STM0758</b>	MAHSHSHADSHLPKDNARRLLFAFIVTAGFMLEVVGILSG	SLALLADAGHMLTAAA	60
ZitB (YbgR)	MAHSHSHTSSHLPEDNNARRLLYAFVGTAGFMLEVVGFLSG	SLALLADAGHMLTAAA	60
<b>Aq_1073</b>	-----MEREKSLKVLAFSFLLI FLFAFIEFLGGLLTNS	SLALLSDAGHMLTAAVS	49
CzcD	MGAGHSHD----HPGNERSLKIALALTGTFLIAEVVGVMTK	SLALISDAAHMLDTVA	56
ZnT8	HSGSKPTEKGANEYAYAKWKLCSASAICFIFMIAEVVGGHIAG	SLAVVTDAAHLLIDLTS	113
ZnT3	HRDPLPPPGLTPERLHARRQLYAACAVCFVFMAGEVVGGLAHS	SLAIMTDAAHLLADVGS	115
<b>STM4061</b>	-----MNQTYGRVLSRAAIAATAMASLLLLIKIFAWWYTG	SVSILAALVDSLVDIAA	52
<b>Aq_2073</b>	-----MKKHHWALVVSFGFNIFQSLIKLVGGLLTG	SLSLIGDATHSLSDATA	46
Yiip	SLTNLLVVRYSLQPADNHSFGHGKAESLAAALQSMFISGSALFL	FLFTGTGIQHLSIPTMT	112
<b>STM0758</b>	LLFALLAVQFSRRPPTVRHTFGWLRRLTTLAAFVNIAIALVVI	TLLIVWEAIERFYTPR-PV	119
ZitB (YbgR)	LLFALLAVQFSRRPPTIRHTFGWLRRLTTLAAFVNIAIALVVI	TLLIVWEAIERFYTPR-PV	119
<b>Aq_1073</b>	LSIALVAQYLALKVKTKRTTYGLYRLEVALALVNGVFLGLG	IYIILEAIHRFENPE-PV	108
CzcD	LAIALAATAIAKRPAKRRKRTFGYRFEILAAAFNALLLFGVA	IYIILEAYLRKLSPP-QI	115
ZnT8	FLLSLFLSLWLSKPPSKRLTFGWHRAEILGALLSILCIVWVT	GLVLYLACERLLYVDYQI	173
ZnT3	MMGSLFLSLWLSRTPATRTMTFGWHRSETLGALASVVSLWM	VTGILLYLAFVRLHSDYHI	175
<b>STM4061</b>	SLTNLLVVRYSLQPADDEHTFGHGKAESLAAALQSMFISGSAL	FLFTSIQNLIKPTPMN	112
<b>Aq_2073</b>	SLIAFLSIKFS-EIKSERFPYGLYKLENIGAVIAFFLLFTAW	EIQRALKGEININ-FE	104
Yiip	DPGVGVIVTIVALICTIILVSRQWVVRRT-----	QSQAVRAD	150
<b>STM0758</b>	AGNLMVIAVAGLANLFAFWILHRGSEK-----	NLNVRAA	156
ZitB (YbgR)	EGGMMMAIAVAGLANILSFWLLHHGSEEK-----	NLNVRAA	156
<b>Aq_1073</b>	KP-QMIYAFAGLIVNLVVGVIILKHSEE-----	NINI KSA	143
CzcD	ESTCMFVAVLGLIINLISMRMLSSGQSS-----	SLNVKGA	151
ZnT8	QATVMIIVSSCAVAANIVTLVVLHQRCLGHNHKEVQAN-----	ASVRAA	217
ZnT3	EGGAMLLTASIAVCANLLMAFVLHQAGPPHSHGSRGAEYAP	LEEGPEEPLPLGNTSVRAA	235
<b>STM4061</b>	DPGVGIGVTIVALICTIILVTFQRWVVRKT-----	QSQAVRAD	150
<b>Aq_2073</b>	NLPIGIGVTIVLSLVLSTLSFLERRAGKKN-----	SPTLIAD	142
Yiip	MLHYQSDVMMNGAILLALGLSWYG--WHRADALFALGIGIYI	LYSALRMGYEAVQSLDDR	208
<b>STM0758</b>	ALHVMGDLGSLVGAIVAALII IWT-GWTPADPILSILVSVL	VLRSARLLKDSVNELLEG	215
ZitB (YbgR)	ALHVLGDLGSLVGAIIAALII IWT-GWTPADPILSILVSVL	VLRSARLLKDSVNELLEG	215
<b>Aq_1073</b>	LLHVATDTLGSVAIIAGIAIVFW-KFYLADPILSVAVALLI	PLSAYSVIKETVNVLLEV	202
CzcD	YLEVWSDLLGSLVSVIAGAI IIRFT-GWAWVDSIAI AVL	IGLWVLPRTWILLKSLNVLEGG	210
ZnT8	FVHALGDLFQSI SVLISALIIYFKPEYKIADPICTFIFSI	LVLASTITILKDFSILLMGG	277
ZnT3	FVHVLGDLQSFVGLAASILIYFKPQYKAADPISTFLFSI	CALGSTAPTLRDVLRLMGG	295
<b>STM4061</b>	MLHYQSDVMMNGAILLALGLSWYG--WHRADALFALGIGIY	LYSALRMGYEAVQSLDDR	208
<b>Aq_2073</b>	SYHTLTDAFSSFLVLSLSSYFYG---INIERVAVAVALI	IVYTAFELLKEQIGAILDI	199
Yiip	ALPDEERQEI I DIVT-SWPGVSGAHDLRTRQSGP-TRF	IQIHLEMEDSLPLVQAHMVDQ	266
<b>STM0758</b>	APVSLDINALQRHLSREIPEVRNVHVVHVMVGE-KPVM	TLHAQVIP--PHDHD-ALLER	271
ZitB (YbgR)	APVSLDIAELKRRMCREIPEVRNVHVVHVMVGE-KPVM	TLHVQVIP--PHDHD-ALLDQ	271
<b>Aq_1073</b>	APSHINTEELEKELL-NLQGVKGVHDLHVWISITPGTEV	LTVHVVED--TSICN-DILKE	258
CzcD	VPDDVDLAEVEKQIL-ATPGVKSFDLHLI WALTS	GKASLTVHVVNDT--AVNPEMEVLPE	267
ZnT8	VPKSLNYSGVKELIL-AVDGVLVSHSLHIWLSLTMNQV	ILSAHVATAA--SRDSQ-VVRRE	333
ZnT3	TPRNVGFEPVRDILL-SVPGVRATHELHLWALTLTYH	VASAHLAIDS--TADPE-AVLAE	351
<b>STM4061</b>	ALPDAERQEI I DIVT-SWPGVSGAHDLRTRQSGP-TRF	IQIHLEMEDNPLVQAHFVADQ	266
<b>Aq_2073</b>	SADKETVEKIKRIIL-SFPEVSEVKRLLVRNAGG-RLF	IDAVITINT-DDFKSHAIAD	256
Yiip	VEQAILRRFP-GSDVI IHQDPCSVVPREGRKMSLSS-----		301
<b>STM0758</b>	IQDFLMHEYH-IAHATIQMEYQVCHGPDCH--LN-QTSS	GHVHHH-----	312
ZitB (YbgR)	IQHLYLMDHYQ-IEHATIQMEYQVCHGPDCH--LN-EG	VSGHSHHHH-----	313
<b>Aq_1073</b>	VEK-IAHKYG-IKHTTVQLEKEGYACAECPLLSPOGL	KFHSHHHHHGHEHEH-----	308
CzcD	LKQMLADKFD-I THVTIQFELAPCEQADAAQHFN	ASPALVGSKSLAAGGN-----	316
ZnT8	IAKALSKSFT-MHSLTIQMEYQVCHGPDCH--LN-EG	VSGHSHHHH-----	369
ZnT3	ASSRLYSRFG-FSSCTLQVEQYQPEMAQCCLRCQEP	PQA-----	388
<b>STM4061</b>	VEQAILQRFPGSDVI IHQDPCSVVPREGRKFE	LV-----	300
<b>Aq_2073</b>	IERKIMKEIPNVDMVFIHYEPCCVKGT	SVAVLAKDGVVARSFKDVKI IIFK	ENEGNPE 316

Figure 1.4: Sequence alignment of selected CDF transporters for this thesis work and other known transporters. Zinc binding sites of YiiP are marked with black dots and CDF signature sequence with a box (Lu and Fu 2007, Lu et al. 2009).



### 1.4.2 Formate nitrite transporter (FNT) family

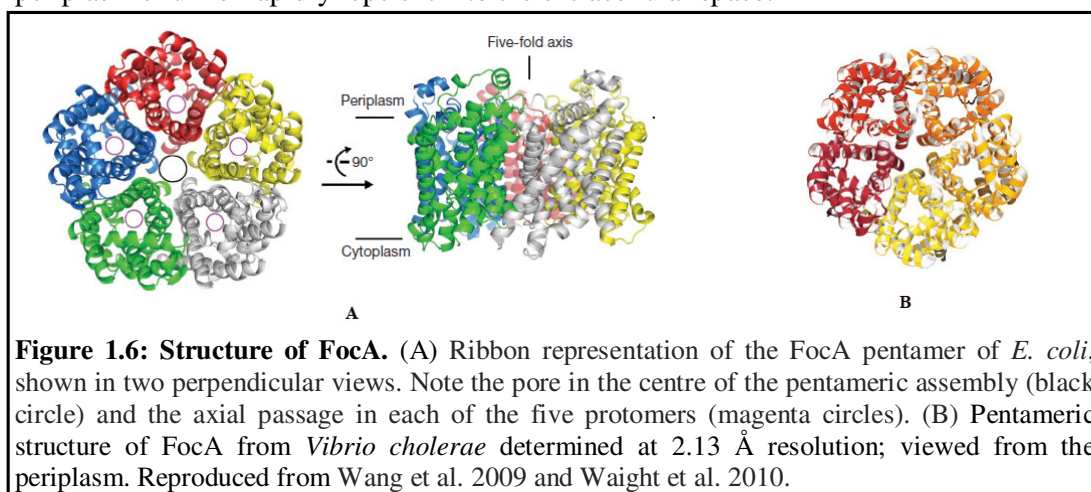
The FNT family is a group of integral membrane proteins that are, on average, between 256 and 285 residues in length and are predicted to possess 6–8 transmembrane segments. They are found widely in prokaryotes. In addition, 64 species of eukaryotes, including yeast, fungi, and protists, possess genes encoding FNT proteins but no clear homologues have been observed in the genomes of higher plants or metazoans. Although very few of these transporters have been characterized biochemically, FNT proteins are believed to enable the specific transport of formate and/or nitrite. Prokaryotic FNT proteins that have been characterized include FocA and NirC of *E. coli* that have been implicated in the transport of formate and nitrite, respectively (Suppman and Sawers 1994, Saier et al. 1999; Clegg et al. 2002).

Several reports suggested FNT proteins to function as higher oligomers (Falke et al. 2009; Beckham et al. 2009). Very recently, two atomic structures of FocA from *E. coli* and *Vibrio cholerae*, have been published back to back. These structures revealed FocA to form a symmetric pentamer, with each protomer consisting of 6 transmembrane segments (Fig. 1.6). Despite a lack of sequence homology, the overall structure of the FocA protomer closely resembles that of aquaporin and strongly argues that FocA is a

## INTRODUCTION

channel, rather than a transporter. Unlike aquaporin, FocA is impermeable to water but allows the passage of formate (Wang et al. 2009; Waight et al. 2010).

According to Waight et al. (2010), the positively charged surface of the cytoplasmic funnel helps to concentrate formate ions in a preselection process. The selectivity filter of the formate channel begins at the cytoplasmic slit. At the inner site, both oxygen atoms of the formate form a hydrogen bond with the N atom of histidine 208, and the formate makes van der Waals contact with the slit, suggestive of a coin in a slot, except that the coin cannot tilt or rotate due to the hydrogen bonds. The binding of the inner formate ion in turn facilitates the binding of the outer formate ion by forming a hydrogen bond with the latter. Once the first formate ion passes the cytoplasmic slit, it goes through the remaining part of the selectivity filter fairly smoothly. There is probably little electrostatic interaction with the periplasmic half of the pore due to its hydrophobic surface; the size of the 2.3 Å central restriction ring fits with the dimensions of the formate. After the formate ion reaches the exit of the filter, the negatively charged periplasmic funnel rapidly repels it into the extracellular space.



**Figure 1.6: Structure of FocA.** (A) Ribbon representation of the FocA pentamer of *E. coli*, shown in two perpendicular views. Note the pore in the centre of the pentameric assembly (black circle) and the axial passage in each of the five protomers (magenta circles). (B) Pentameric structure of FocA from *Vibrio cholerae* determined at 2.13 Å resolution; viewed from the periplasm. Reproduced from Wang et al. 2009 and Waight et al. 2010.

### 1.4.3 Multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily

The MOP exporter superfamily consists of four previously recognized families: (a) the ubiquitous multi-drug and toxin extrusion (MATE) family; (b) the prokaryotic polysaccharide transporter (PST) family; (c) the eukaryotic oligosaccharidyl-lipid flippase (OLF) family and (d) the bacterial mouse virulence factor family (MVF). These transporters are all homologous and are, therefore, related by common descent. Since, I had chosen a MATE family transporter from *Pyrococcus furiosus* for this thesis work,

## **INTRODUCTION**

only this family will be discussed in details. MATE transporters typically possess 12 transmembrane segments. NorM from *Vibrio parahaemolyticus* was the first characterized MATE family transporter. It functions as a sodium drug antiporter (Morita et al. 1998). One of the most important properties of a MATE transporter is its substrate specificity. From the published data regarding drug susceptibility tests, almost all MATE-family transporters can recognize fluoroquinolones as transport substrates, such as norfloxacin. Among cationic dyes, acriflavine and ethidium are pumped out via several MATE transporters. Substrates for the MATE transporters identified up to now are various and have unrelated chemical structures (Kuroda and Tsuchiya 2009). Though initially believed to be solely a sodium drug antiporter, MATE family proteins are also reported to function as H<sup>+</sup>/drug antiporter (He et al. 2004; Su et al. 2005). Bacterial species that have developed clinical resistance to antimicrobial agents are increasing in numbers and have become a serious problem in hospitals. MATE family transporters could easily become an important drug target. The most desired step towards this, is a three dimensional structure. Only an atomic structure could answer the broad substrate specificity and ubiquity of this scanty characterized MATE family transporter.

### **1.4.4 Potassium uptake permease (KUP) family**

Proteins of the KUP family include the KUP (TrkD) protein of *E. coli* and homologues in both Gram-positive and Gram-negative bacteria. High affinity (20 μM) K<sup>+</sup> uptake systems (Hak1) of the yeast *Debaryomyces occidentalis* as well as of *Neurospora crassa*, and several homologues in plants have been characterized. *Arabidopsis thaliana* and other plants possess multiple KUP family paralogues. The *E. coli* protein is 622 amino acid residues long and has 12 putative transmembrane spanners (440 residues) with a requisite hydrophilic, C-terminal domain of 182 residues, localized at the cytoplasmic side of the membrane. Deletion of most of the hydrophilic domain reduces but does not abolish KUP transport activity. The function of the C-terminal domain is not known. The *E. coli* KUP protein is believed to be a secondary active transporter. Uptake is blocked by protonophores such as CCCP (but not arsenate), and evidence for a proton symport mechanism has been presented (Zakharyan and Trchounian 2001). The *N. crassa* protein was earlier shown to be a K<sup>+</sup>:H<sup>+</sup> symporter, establishing that the KUP family consists of secondary active transporters. The yeast high affinity (K<sub>M</sub> = 1 μM) K<sup>+</sup>-transporter Hak1

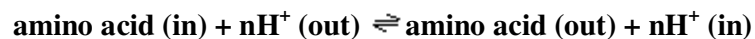
## INTRODUCTION

is 762 amino acid residues long with 12 putative transmembrane segments. Like the *E. coli* KUP protein, it possesses a C-terminal hydrophilic domain, probably localized at the cytoplasmic side of the membrane. Hak1 may be able to accumulate  $K^+$   $10^6$ -fold against a concentration gradient. The plant high and low affinity  $K^+$  transporters can complement  $K^+$  uptake defects in *E. coli*. The generalized transport reaction for members of the KUP family is:



### 1.4.5 Resistance to homoserine/threonine (RhtB) family

About 100 sequenced proteins, derived from Gram-negative and Gram-positive bacteria as well as from archaea, comprise the RhtB family, but few of these proteins are functionally characterized (Vrljic et al. 1999). *E. coli* possesses five paralogues, and a large region of one of them (YahN of *E. coli*) exhibits significant sequence similarity to YggA of *E. coli*, an established member of the LysE family. The PSI-BLAST program groups the L-Lysin exporter (LysE) family, the RhtB family and the cadmium resistance (CadD) family together. These proteins are all of about the same size and possess apparently the same topology, further suggesting a common evolutionary origin (www.tcdb.org). The first two members of the RhtB family to be characterized functionally were the RhtB and RhtC permeases of *E. coli* (Aleshin et al. 1999; Zakataeva et al. 1999). YfiK of *E. coli* exports cysteine, O-acetylserine and azaserine (Franke et al. 2003). The YeaS (LeuE) homologue exports leucine and several other neutral, hydrophobic amino acids (Kutukova et al. 2005). Aleshin et al. (1999) reported a partial alignment of recognized bacterial and archaeal members of the RhtB and LysE families, but not the CadD family. Vrljic et al. (1999) reported phylogenetic trees for all three families of the LysE superfamily (LysE, RhtB and CadD). The transport reaction presumably catalyzed by members of the RhtB family is:



### 1.4.6 The sulfate permease (SulP) family

The SulP family is a large and ubiquitous family with over 200 sequenced members derived from archaea, bacteria, fungi, plants and animals. Many organisms including

## **INTRODUCTION**

*Bacillus subtilis*, *Synechocystis sp.*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Caenorhabditis elegans* possess multiple SulP family paralogues. Many of these proteins are functionally characterized, and most are inorganic anion uptake transporters or anion:anion exchange transporters. Some transport their substrate(s) with high affinities, while others transport it or them with relatively low affinities. Many function by  $\text{SO}_4^{2-}:\text{H}^+$  symport, but  $\text{SO}_4^{2-}:\text{HCO}_3^-$ , or more generally, anion:anion antiport has been reported for several homologues. The bacterial proteins vary in size from 434 residues to 573 residues with only a few exceptions. The eukaryotic proteins vary in size from 611 residues to 893 residues with a few exceptions. Thus, the eukaryotic proteins are usually larger than their prokaryotic homologues. These proteins exhibit 10-13 putative transmembrane  $\alpha$ -helical segments (TMSs) depending on the protein. The generalized transport reactions catalyzed by SulP family proteins are:

### **1.5 Heterologous overproduction of membrane protein**

Transport proteins are usually present at low levels and constitute 0.1% of the total cell membrane protein (Lundstrom 2006). This is a key bottleneck when considering the structural studies of transporters which demands milligram quantities of purified protein. Moreover, use of natural resources excludes the possibility of genetically modifying proteins to improve their stability and facilitate their easy detection and purification. With the advance of genome sequencing, a whole lot of other transporter targets of potential therapeutic importance, from diverse domains of life, are queuing up. But one could imagine how tedious and costly it might be to isolate transporters from natural sources. So, the obvious alternative is to heterologously overproduce the recombinant target protein from diverse source organisms in a suitable expression host. The heterologously overproduced membrane proteins are more homogeneous compared to proteins isolated from the native source. Both the necessity and feasibility of overexpression are evident from the steadily growing number of high resolution structures of helical membrane proteins obtained through overexpression ([http://blanco.biomol.uci.edu/Membrane\\_Proteins\\_xtal.html](http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html)). In the last few years, there were several reports of heterologous overproduction of membrane proteins and many of

them were in highthroughput fashion (Dobrovetsky and Lu 2005; Eshaghi et al. 2005; Korepanova et al. 2005; Surade et al. 2006; Psakis et al. 2007; Gordon et al. 2008).

### 1.5.1 Choice of target: source organism

Sequence data of genomics projects and the availability of genomic DNA have made it possible to clone almost any transporter protein. As a result, multiple members of a protein family of interest can be selected from different source organisms for overexpression and crystallization (Goeddel 1990; Chang and Roth 2001; Locher et al. 2002; Surade et al. 2006). Among the members of the chosen protein family, transporters that have been biochemically and functionally characterized are worth considering, because such information may help crystallization experiments.

The work for this thesis is also a part of structural genomics project where more than 240 transporters belonging to 40 different families were selected from three different source organisms. The detailed characteristics and the reasons for the choice of these organisms are tabulated below.

**Table 1.1 Details of the three selected source organisms for this work**

Parameter	<i>Aquifex aeolicus</i> VF5	<i>Pyrococcus furiosus</i> DSM 3638	<i>Salmonella typhimurium</i> LT2
First isolation	Underwater volcanoes, hot springs	Heated marine sediments	Pig intestinal fluid
Ambient temperature (°C)	85-95	70-103 (100)	27-40 (35.5)
Size (µm)	2-6	0.8-1.5	0.5-3
Sequencing center	Diversa corporation & University of Illinois at UC	University of Utah and University of Maryland	Washington University Consortium
Year of annotation at CMR	January-December 2000	June 2002	November 2001
Genome size(Mbp)	1.55	1.91	4.95
GC content (%)	43.47	40.76	52.23
Primarily annotated genes	1522	2065	4553
Structures in PDB <sup>#</sup> (with genus)	312	749	204
Reason for selection	Hyperthermophilic bacterium	Hyperthermophilic archaeon	Pathogenic, mesophilic bacterium

CMR : Comprehensive microbial resource, PDB : Protein data bank, # as of Feb. 18,2010

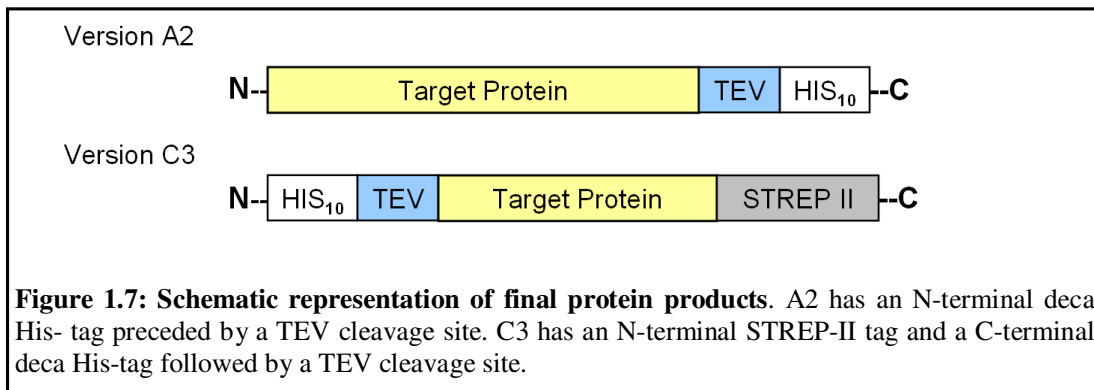
Source : <http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>

**1.5.2 Expression vectors and tags**

Heterologous overexpression of membrane protein genes/cDNAs greatly depends on the choice of expression vectors. An ideal vector should have a tightly regulated, moderately strong promoter and should have a wide range of usable inducer concentrations. Tight regulation prevents leaky expression, which can lead to *in vivo* proteolysis or even cell death when a toxic membrane protein is produced. To meet the needs of this project, three vectors, pBAD, pTTQ18 and pQE were chosen to overproduce membrane transporters *in vivo*.

The pBAD vectors from Invitrogen, containing the arabinose araBAD operon, are particularly suitable for overproduction of membrane proteins (Guzman et al. 1995; Surade et al. 2006). The promoter is repressed 1200 fold and can be induced by L-arabinose at concentrations upto 0.2%. This vector was used to overproduce a wide range of membrane proteins (Auer et al. 2001; Li et al. 2001; Surade et al. 2006), showing its general applicability. The pTTQ18 vector uses the moderately strong *tac* promoter. It has been used to successfully overproduce membrane proteins, with a yield of upto 1-2mg /L by Ward et al. (1999; 2000; 2001). Finally the pQE vector (Qiagen), which uses the stronger T5 promoter, has two lac operator binding sites, which prevent expression without inducer.

For the convenience of highthroughput cloning, all vectors were modified to have the identical multiple cloning sites as well as one of two sets of tags (A2 or C3). Fig.1.7 depicts the final protein product with different tags.





### **1.5.3 Choice of the expression host: *E. coli***

By far, *E. coli* is the preferred expression host for heterologous overexpression of membrane protein genes, where the production level could go up to 50% of the total cytoplasmic protein (Ward et al. 1999). Convenient features, like well studied genetics, low cost, growth to relatively high density and availability of a large number of cloning and expression vectors, make *E. coli* invaluable. Moreover, the availability of different promoters for expression provides a wide range of choices for heterologous production. Because of the well studied genetics and physiology, a range of *E. coli* strains are available for expression. As changing strains could affect the expression by 2-5 fold (Auer et al. 2001), one should try different strains at the beginning. Several strains with protease deficiencies or with extra codons are available. *E. coli* host strain carrying plasmids (pRARE) encoding rare tRNAs are used to express genes from thermophilic archaea (Wakagi et al. 1998; Kanaya et al. 1999). Auxotrophic mutants of *E. coli*, like B834 (methionine auxotroph), allow efficient labeling of target protein with selenomethionine, a feature that facilitates structure determination by the multiwavelength anomalous diffraction method. While pET and pQE based vectors could be expressed in BL21(DE3) strain or C43(DE3) strain or NM554 strain, pBAD vectors require the host cell to be deficient in arabinose metabolism, like TOP10 cell. For this work, NM5554 and BL21(DE3) for pTTQ18 and pQE vectors while TOP10 for pBAD vector were chosen.

### **1.5.4 Bottlenecks affecting heterologous overexpression and remedies**

#### **1.5.4.1 Overexpression, targeting and folding**

Prokaryotic membrane proteins are usually overexpressed in *E. coli* while eukaryotic membrane proteins have been overexpressed in *Lactococcus lactis* and various eukaryotic systems. Heterologous overexpression of target protein can be hampered by different synthesis, targeting, insertion and folding characteristics in the overexpression host (Tate 2001). Incompatibility of factors involved in the processing of heterologously expressed membrane proteins (discussed earlier in the membrane protein biogenesis section) indicates a general problem (Goder et al. 2004; Bowie 2005). In addition, the

availability of endogenous factors necessary for biogenesis of membrane protein could turn out to be the limiting factor for heterologous overexpression (Valent et al. 1995).

Optimizing the overexpression condition by lowering the culture temperature or inducer concentration or choosing a different promoter might be beneficial. A recent study in yeast indicates that culture condition not only influence the amount of produced protein but also its membrane integration (Bonander et al. 2005). Wang and co-workers (Chen et al. 2003) reported increased overproduction of the *E. coli* CorA protein in *E. coli* when coexpressed with cytoplasmic DnaK-DnaJ chaperone system. Moreover, yields of functionally overproduced membrane proteins can also be increased by adding ligands to the culture medium which assists folding and stabilization of the overproduced protein (Weiss et al. 1998; Andre et al. 2006).

### **1.5.4.2 Membrane space and accommodation of foreign structures**

The membrane space required to accommodate overproduced proteins represents another bottleneck. The accumulation of foreign structures in a membrane can induce stress responses and activate proteolytic systems of the overexpressing hosts (Kihara et al. 1995; Griffith et al. 2003). An attractive remedy is to use an expression system which has a larger membrane surface area or by nature has a high biogenesis capacity for membrane proteins and sufficient space in the membrane to accommodate the overexpressed materials (Eroglu et al. 2002). Roy et al. (2008) from our lab reported functional overproduction of several GPCRs in *Rhodobacter sphaeroides*, employing the fact that *R. sphaeroides* provides more membrane surface per cell compared to other expression hosts.

### **1.5.4.3 Lipid composition**

Membranes from prokaryotes, yeasts or higher eukaryotes differ in their lipid composition, which might lead to problems with the heterologous overproduction of functional membrane proteins (Opekarova and Tanner 2003). For example, the mammalian presynaptic serotonin transporter is functionally obtained only in the presence of cholesterol (Tate 2001), while the *E. coli* protein LacY absolutely requires the lipid phosphatidylethanolamine for proper folding (Bogdanov et al. 1996; 1999).

### **1.5.4.4 Stability of messenger RNA**

Gene expression levels are mainly determined by the efficiency of transcription, mRNA stability (rate of decay) and the frequency of mRNA translation. The average half life of mRNA in *E. coli* at 37 °C ranges from seconds to maximally 20 minutes and the expression rate depends directly on the inherent mRNA stability (Rauhut and Klug 1999; Regnier and Arraiano 2000). A mRNA fragment encoding the C-terminal region of an *E. coli* F<sub>0</sub> ATPase subunit was stabilized by fusion to the sequence encoding green fluorescent protein (GFP) (Arechaga et al. 2003). *E. coli* strains containing a mutation in the gene encoding RNaseE (*rne131* mutation) are available for the enhancement of mRNA stability in recombinant expression systems (Invitrogen BL21 star strain).

### **1.5.4.5 Toxicity of target protein**

Toxicity of heterologously produced protein in the host cell poses a major hindrance in highthroughput protein production, especially in a structural genomics set up. *E. coli* strain C42(DE3) and C43(DE3), derived from *E. coli* strain BL21(DE3), are best known for their ability to mitigate the toxic effect of overexpressing membrane protein and also increasing the protein level (Miroux et al. 1996; Eshaghi et al. 2005; Drew et al. 2006). Additionally, the newly developed cell-free expression system for membrane protein, discussed in details in the following pages, is an attractive alternative.

### **1.5.4.6 Poor expression and N terminal fusion partner; MBP fusion**

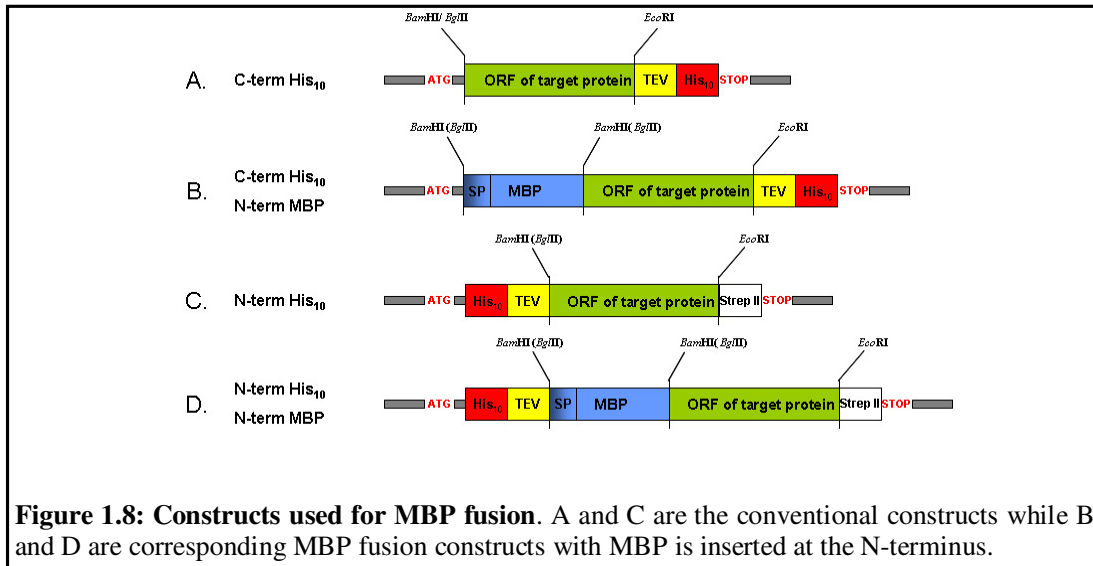
While the N-terminal ends of membrane proteins show a strong preference for the cytoplasm (Wallin et al. 1995; Daley et al. 2005), their translocation efficiency during biogenesis hampers the overexpression, specially the ones that have to translocate across the membrane (Monne et al. 2005). N-terminal soluble protein fusion partners, such as glutathione *S*-transferase, NusA, green fluorescent protein (GFP) and maltose binding protein (MBP) are routinely tested for their ability to improve the expression of poorly expressed membrane proteins. MBP fusion strategy was employed for some of the targets in the present study to check its effect on production level and also to gain insight of their topology.

MBP is localized in the periplasmic space of *E. coli* and possesses a signal peptide which is cleaved off during translation of MBP across the membrane. The C-terminal fusion of MBP is usually performed without the periplasmic signal peptide (first 26 amino acid

## INTRODUCTION

residues), while the N-terminal fusion is described for both periplasmic full-length MBP and cytoplasmic MBP lacking the signal sequence (Spangfort et al. 1996; Korepanova et al. 2007). The successful application of MBP fusion for the expression of 22 low molecular weight (less than 4 transmembrane helices) membrane proteins from *Mycobacterium tuberculosis* was recently demonstrated (Korepanova et al. 2007). Several human G-protein coupled receptors (GPCR) were functionally overproduced in *E. coli* as fusions with full length MBP (Grisshammer et al. 1994; Grisshammer et al. 2002; Weiss and Grisshammer 2002; White et al. 2004; Yelisseev et al. 2007). In contrast to many other solubility tags, MBP has the additional benefit of being an affinity tag (Esposito and Chatterjee 2006; Cabanne et al. 2009).

To evaluate the effect of full length MBP (with signal peptide) on the heterologous expression of secondary active transporters, MBP was fused to the N-terminus of three prokaryotic membrane transporters belonging to 2 different families. N- and C-terminal His-tag fusions were tested, in parallel, to compare the effect of tags. The constructs are schematically represented in Fig 1.8.



## 1.6 Cell-free production of membrane protein

Low natural abundance is the biggest obstacle for working with membrane proteins. Moreover, *in vivo* overexpression is notoriously problematic, often resulting in low yield, cell toxicity, protein aggregation and misfolding. Isolation and purification of the protein

from the membrane and subsequent reconstitution are not easy. Cell-free expression systems are increasingly being considered as a viable alternative for overcoming the above obstacles. The cell-free systems offer, in principle: (i) the possibility to eliminate toxic effects of recombinant protein to the cell, (ii) flexibility, as conditions like pH, redox potential or buffer system could easily be altered, (iii) open nature, as nearly any compound can be added at any time point directly to the reaction chamber, (iv) reduced proteolytic degradation and enhanced functional folding by addition of protease inhibitor or chaperones, respectively, (v) stabilization by addition of substrates or ligands, (vi) generation of an artificial hydrophobic environment in the cell-free reaction for the production of membrane proteins (Spirin and Swartz 2008).

### **1.6.1 Development of coupled transcription translation system**

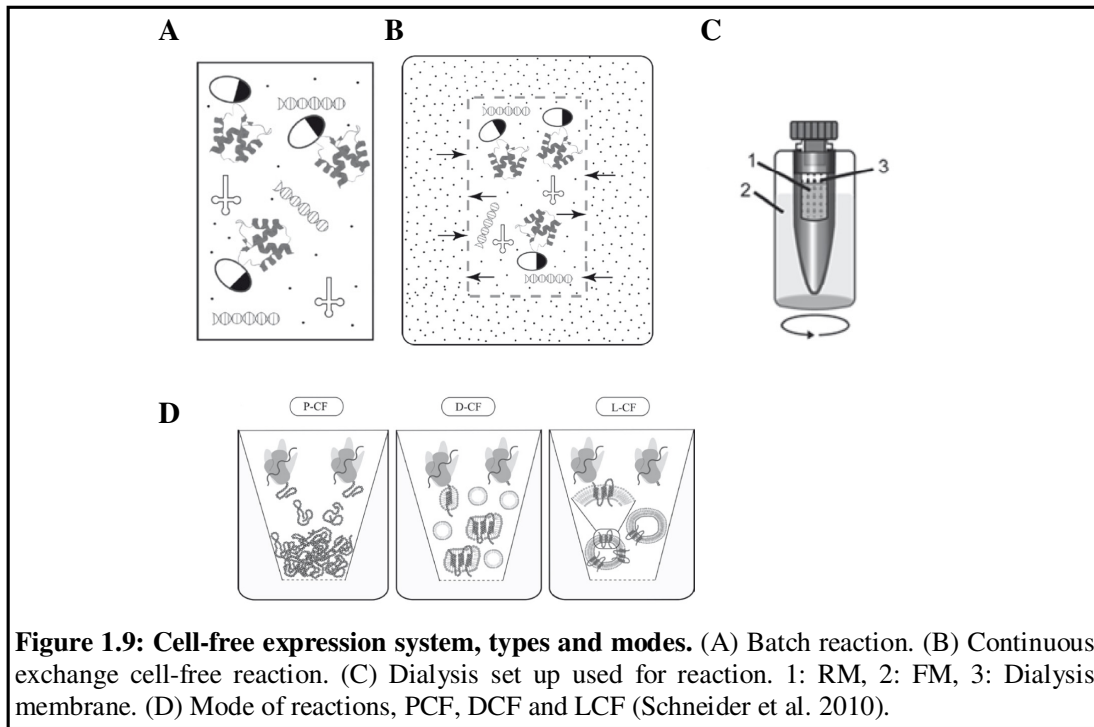
Cell-free expression systems are reconstituted reactions based on cellular extracts that recapitulate the transcription-translational capabilities of a cell *in vitro*. The first demonstration of DNA-dependent incorporation of amino acids into synthesized proteins as well as the first evidence for the coupled transcription-translation process in bacterial cell-free systems were made by Nirenberg's group in the beginning of the 1960s (Matthaei and Nirenberg 1961; Byrne et al. 1964). In 1967, Zubay and colleagues made significant improvements (DeVries and Zubay 1967; Lederman and Zubay 1967) and introduced an efficient bacterial coupled transcription-translation system for expression of exogenous DNA. The next important step in the development of cell-free gene expression was the combination of a cell extract with a specific bacteriophage RNA polymerase (T7 or SP6) that used a phage-specific promoter for transcription. According to Spirin and Swartz (2008), this system could rather be called "combined" than "coupled" transcription-translation system, due to lack of temporal and/or spatial coupling between the transcription and translational machinery.

### **1.6.2 Types and modes of reaction**

In batch mode, performed in a fixed volume of a test tube, the conditions change during incubation as a result of the consumption of substrates and the accumulation of products. Translation stops as soon as any essential substrate is exhausted or any product or by-

## INTRODUCTION

product reaches an inhibiting concentration, usually after 20–60 minutes of incubation (Fig 1.9A). In the continuous exchange cell free (CECF) system, the reaction is performed under conditions of persistent supply of the consumable substrates, such as, amino acids, nucleoside triphosphates and energy-regenerating compounds and with removal of the reaction products, mainly inorganic phosphates and nucleoside monophosphates (Fig. 1.9B).



**Figure 1.9: Cell-free expression system, types and modes.** (A) Batch reaction. (B) Continuous exchange cell-free reaction. (C) Dialysis set up used for reaction. 1: RM, 2: FM, 3: Dialysis membrane. (D) Mode of reactions, PCF, DCF and LCF (Schneider et al. 2010).

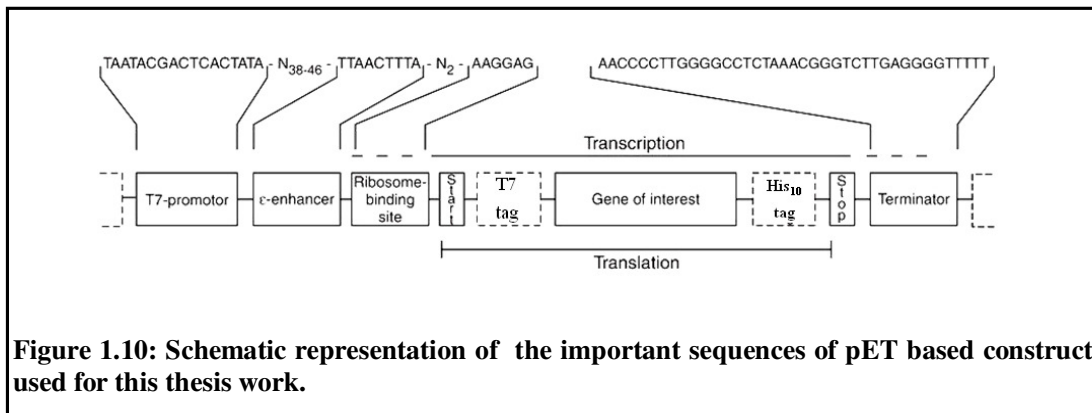
For this purpose, a porous (dialysis) membrane is used to retain the high-molecular-weight components of the protein-synthesizing machinery (ribosomes, mRNA, etc.) within a defined reaction compartment. The membrane separates the reaction compartment (RM: reaction mix) from another compartment containing a feeding solution (FM: feeding mix) with a reservoir of low-molecular-weight components (substrates) for the reaction (Fig. 1.9C). Cell-free production of membrane proteins can be performed in three different modes: (i) protein can be obtained as precipitate (PCF) and then resolubilized and purified in a suitable detergent, (ii) produced in soluble form with detergent micelles in the reaction mix (DCF) or (iii) synthesized in the presence of preformed liposomes (LCF) and isolated from the reaction mix by density gradient centrifugation (Fig. 1.9D).

**1.6.3 *E. coli* S30 extract**

The cell extract preparation is one of the key steps in obtaining successful and reproducible results using the cell-free protein synthesis. Currently up to 10mg of protein per mL of reaction mix by CECF mode is reported (Spirin and Swartz 2008). The most important aspect in making the extract is to choose the right strain. The *E. coli* strains that lack major RNases, such as MRE600, A19 were routinely used (Zubay 1973; Pratt 1984). Several BL21 derived strains, containing extra copies of genes for minor tRNAs, such as BL21 codon-plus, Rosetta or lacking RNase, such as, BL21-Star(DE3) are also used routinely (Ahn et al. 2005; Chumpolkulwong et al. 2006).

**1.6.4 Plasmid construct and template quality**

T7 RNA polymerase is mostly used for transcription because of its high specificity and strong activity. So, standard T7 promoter based vectors such as pET (Novagen) or pIVEX (Roche Bioscience) are used commonly. An efficient T7 terminator should be included to reduce non productive consumption of nucleotides while the low yield can be countered with an efficiently translated expression tag, such as T7 tag at the N-terminal end. An extended poly (His)<sub>10</sub> tag at the C-terminus not only helps for efficient binding during purification but also to verify the full length expression. The schematic drawing of the pET construct used for this thesis work is shown in the Fig. 1.10. A high quality of template DNA at a concentration of  $\geq 0.15 \mu\text{g } \mu\text{l}^{-1}$  is of extreme importance.



**Figure 1.10: Schematic representation of the important sequences of pET based construct used for this thesis work.**

### **1.6.5 Mg<sup>2+</sup>, phosphate and K<sup>+</sup>**

Divalent magnesium cations are essential for many biological reactions, particularly those involving nucleotides. Consequently, the total added Mg<sup>2+</sup> concentration is one of the most influential factors in the cell-free reaction mixture. Because the free Mg<sup>2+</sup> interacts with many cell-free reaction components, it is often beneficial to optimize Mg<sup>2+</sup> for each lot of cell extract and cell-free reagents to obtain maximal protein synthesis. For most prokaryotic reactions, total Mg<sup>2+</sup> concentrations range from about 8 to 20 mM. Stabilizing free Mg<sup>2+</sup> is particularly challenging when high energy phosphate bond donors such as phosphoenol pyruvate (PEP) and creatine phosphate are used (Spirin and Swartz 2008). One final consideration that is often taken for granted is the necessity of having sufficient phosphate available. Typically, excess phosphate comes either from the energy source or the nucleotides. However, with systems using nonphosphorylated energy sources (such as glucose), 10 mM phosphate is required for maximal protein production (Calhoun and Swartz, 2005). Additional reaction components for prokaryotic systems include potassium and ammonium acetate, potassium glutamate at high concentrations (>200 mM) (Pratt 1984; Kim et al. 1996), and acetyl phosphate as an energy source (Ryabova et al. 1995). One common theme is that of using potassium as the dominant cation, an approach that mimics the intracellular milieu (Jewett and Swartz 2004).

### **1.6.6 Energy source**

Protein synthesis is the most energy intensive metabolic process. During this reaction, two major types of substrates are being used: (i) amino acids as monomers for the synthesis of polypeptide chains, (ii) nucleoside triphosphates (NTPs) as the energy supply for translation factors, for aminoacyl-tRNA synthetases and also as monomers for the synthesis of mRNA. In CECF systems, the amino acids and energy substrates are constantly supplied and the products, including inorganic phosphate are removed by diffusion exchange to maintain a steady-state concentration. Nevertheless, even in the CECF systems, a supply and removal problem may still exist if the rate of exchange is not sufficient to keep up with both the coupled and uncoupled consumption of substrates and production of side products in the reaction mixture. Owing to NTPase and



phosphatase activities in cell extracts, rapid and uncoupled hydrolysis of NTPs occurs in the incubation mixture (Kim and Swartz 2000; Kim and Choi 2001) in addition to their productive consumption during protein synthesis. So, an *in situ* NTP regeneration is highly advantageous, if not essential for most of the cell-free systems. In bacterial systems, phosphoenol pyruvate is commonly used, but acetyl phosphate (AcP) (Ryabova et al. 1995) and creatine phosphate (Kigawa et al. 1999) are other options. According to Swartz and Spirin (2008), the combination of PEP and AcP ensured higher activity than PEP or AcP alone.

### **1.6.7 Temperature**

The temperature of the reaction mixture for protein synthesis is one of the most important parameters determining the correct folding of newly synthesized polypeptides. The temperature optimum for co-translational protein folding in cell-free systems is not necessarily the optimum for the rate of synthesis. Plausible explanations for the temperature dependence of correct protein folding are: (i) T7 RNA polymerase, being 8 times faster than the intrinsic transcriptase, breaks the transcription translation coupling. (ii) For co-translational protein folding, the elongation of a growing polypeptide at a higher temperature might be too fast, and the local secondary structure of the chain does not have sufficient time to form and thus might interfere with following section (Kommer et al. 2005; Iskakova et al. 2006).

### **1.6.8 Detergents: for DCF mode and solubilization in PCF mode**

Detergents have to be added at concentrations above their critical micellar concentrations (CMCs) to become effective for the solubilization of IMPs. Several detergents such as sodium deoxycholate, sodium cholate, *N*-laurylsarcosine, FOS12 and  $\beta$ -OG severely inhibit CF systems already at low concentrations and they are not suitable for the cell-free production of soluble IMPs. However, most commonly employed relatively mild detergents, specially the Brij group of detergents, appear to be tolerated by CF systems at concentrations that exceed the proposed specific CMCs several times (Berrier et al. 2004; Klammt et al. 2005; Ishihara et al. 2005). Evaluation of the most effective detergent for

the CF production of a specific IMP should be one of the primary tasks of initial optimization screens.

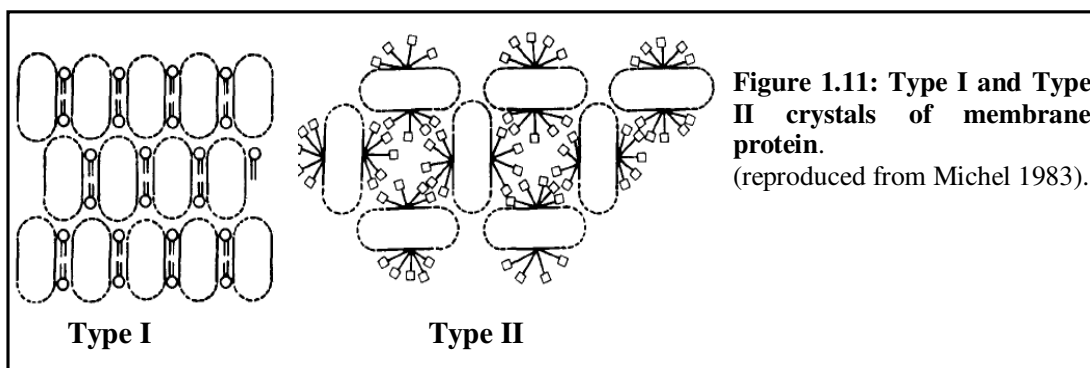
Protein precipitates obtained by PCF reactions usually solubilize rapidly upon addition of relatively mild detergents at a final concentration of 1–5% with gentle shaking at room temperature. In particular, LMPG, LPPG and FOS12 exhibit outstanding properties in the efficient solubilization of CF-produced precipitates of structurally different membrane proteins (Klammt et al. 2005).

### **1.7 Biophysical methods to study membrane protein**

#### **1.7.1 X ray crystallography**

X-ray crystallography is a method of determining the arrangement of atoms within a crystal, in which a beam of X-rays strikes a crystal and diffracts into many specific directions. From the angles and intensities of these diffracted beams, a crystallographer can produce a three-dimensional picture of the density of electrons within the crystal. Three dimensional crystallization of membrane proteins is extremely difficult, as evident from the number of solved structures in PDB. This is largely due to their amphipathic nature. Deisenhofer and Michel (1985) first solved the crystal structure of the bacterial photosynthetic reaction center crystallized in the presence of a detergent. Since then, slow but steady development in the membrane protein crystallography can be observed. Apart from its amphipathic nature, obtaining sufficient amount of pure and homogeneous protein and finding the right detergent, pH, precipitant and temperature, add complexity to the whole process.

Based on crystal formation, two types of membrane protein crystals are observed (Michel 1983). Type I crystals are in principal ordered stacks of 2D crystals, which contain the ordered protein molecules within the membrane plane. In type II, 3D crystals can be formed by membrane proteins integrated in detergent micelles similar to soluble proteins. In these crystals contacts are made by the polar surfaces of the protein, which protrude from the detergent micelle. Mixed type I and II crystals are possible, although most membrane protein crystals belong to type II (Fig 1.11).



There are several modifications one could try to get a well diffracting membrane protein crystal. In order to decrease the flexibility and to increase the hydrophilic protein- protein contacts, antibodies or their fragments against an epitope have been successfully used to generate crystals (Ostermeier et al. 1995; Dutzler et al. 2003; Hunte and Michel 2002). Engineering proteins by N or C-terminal truncation or generating a stable crystallizable core construct by controlled proteolysis, often reported to yield well diffracting crystals (Wang et al. 2009; Ressler et al. 2009; Lu et al. 2009). Size and shape of detergent micelles play an important role in crystal contact and careful optimization helps to improve the crystal quality (Kuo et al. 2003).

The role of lipid in membrane protein crystallization is widely recognized. Lipids can facilitate crystallization not only by stabilizing the protein fold and the association between subunits or monomers, but also by lipid mediated lattice contacts. The importance of lipids in crystal formation is well known for cytochrome  $b_6f$  and LacY structures (Zhang and Cramer 2004; Guan et al. 2006). In recent years, lipidic phase crystallization is gaining popularity. These approaches, including lipidic cubic phase, lipidic sponge phase, and bicelle crystallization methods, all immerse purified membrane protein within a lipid rich matrix before crystallization. This environment is hypothesized to contribute to the proteins long-term structural stability and thereby favour crystallization. Spectacular recent successes include the high-resolution structures of the  $\beta_2$ -adrenergic G-protein-coupled receptor, the  $A_{2A}$  adenosine G-protein-coupled receptor, and the mitochondrial voltage dependent anion channel (see Johansson et al. 2009 for review).

### **1.7.2 Electron crystallography**

Electron crystallography of membrane proteins involves the study of 2D crystals by electron cryo-microscopy (cryo-EM) and image processing. In cryo-EM, samples are studied at cryogenic temperature (liquid nitrogen temperature) without being stained or fixed and hence in a more native environment. The limiting factor in obtaining structural information by electron crystallography is the 2D crystallization. This is accomplished by mixing detergent-solubilized membrane proteins with detergent-solubilized lipids followed by removal of the detergent, using only little lipid to induce the formation of 2D crystals upon reconstitution. Removal of the detergent can be achieved by dialysis of the protein/lipid mixture against detergent-free buffer though other methods exist. A number of factors critical for the formation of 2D crystals include the lipid used for reconstitution, the lipid-to-protein ratio (LPR), and the composition and pH of the dialysis buffer. Less clear is the influence of other factors, such as the initial detergent concentration of the sample, a preincubation of the protein with the lipid prior to detergent removal, and temperature cycles during reconstitution (Schmidt-Krey 2007; Raunser and Walz 2009).

Compared to X-ray crystallography, electron crystallography used to be a low-resolution technique, as the resolution of density maps were typically limited to approximately 3.5 Å, although recently, the structure of aquaporin-0 was determined to 1.9 Å resolution (Gonen et al. 2005). Other problems associated with 2D crystallizations are poor reproducibility and lack of automation. However, requirement of less protein compared to 3D crystallization and the observation of the structure in a native like lipidic environment make 2D crystallization an attractive option along with 3D.

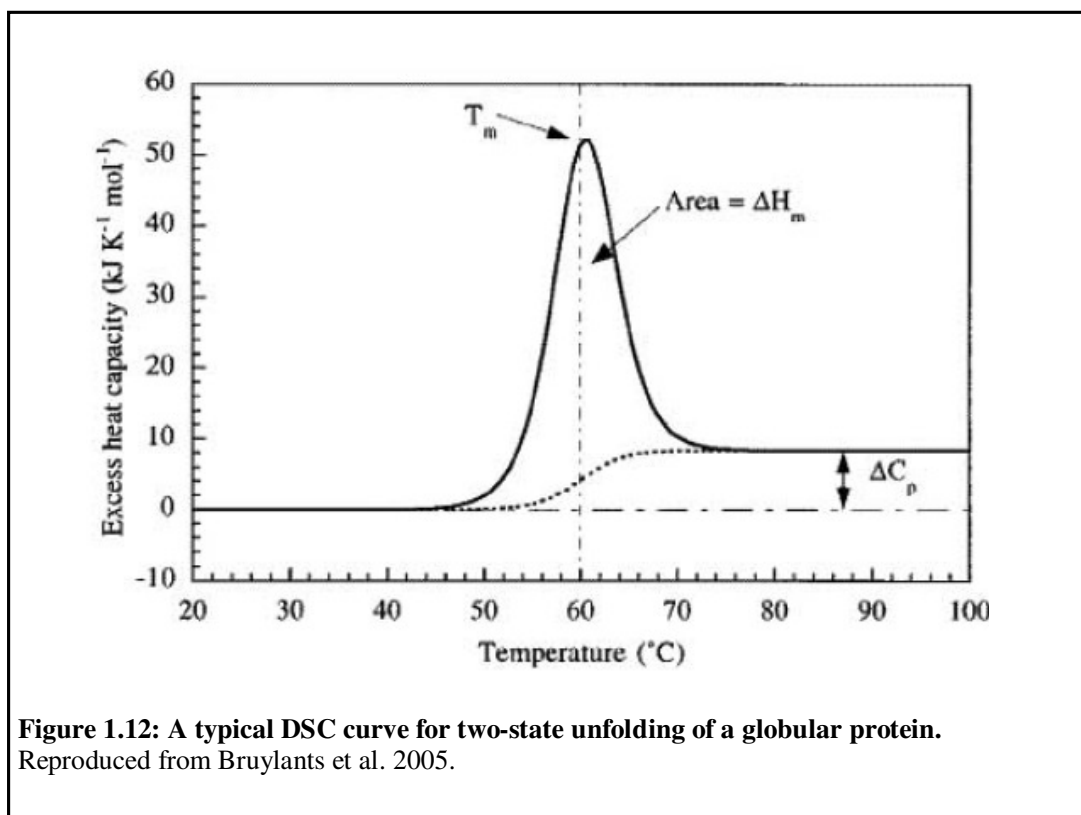
### **1.7.3 Differential scanning calorimetry (DSC)**

DSC is a technique to study thermally induced conformational transitions of biological macromolecules. Regarding instrumentation, the differential scanning calorimeter basically consists of two cells, the reference and sample cell. Both cells can be heated at a constant rate programmed in such a way as to maintain an equal temperature in both cells throughout the scan. When a temperature induced equilibrium process takes place in the sample cell (like unfolding of a macromolecule), the DSC control system supplies the

## INTRODUCTION

necessary additional heat to this cell by means of appropriate heaters in order to maintain the same temperature as the other cell. The data output from the DSC is usually this excess heat as a function of temperature. This excess heat in turn, is proportional to the difference of heat capacity ( $C_p$ ) between the two cells and their contents (Sánchez-Ruiz and Mateo 1987).

The transition is recognized as a sharp endothermic peak centered at  $T_m$  (melting temperature) and the maximum in  $C_p$  occurs directly at  $T_m$ . Integration of the  $C_p$  versus  $T$  curve yield the transition enthalpy ( $\Delta H^0$ ) and the shift in the baseline yields the  $\Delta C_p$  (Fig. 1.12). The sharpness of the transition peak can be measured as the width at half-peak height, and is an index of the cooperative nature of the transition (Bruylants et al. 2005; Wright et al. 1977).



Though scantily studied, there are few reports of membrane protein stability and ligand interaction, analyzed by DSC (Halsey et al. 1977; Kresheck et al. 1985; Rigell et al. 1985; Epand et al. 2001).

## **1.8 Scope of this thesis**

Membrane transporters are involved in diverse functions, from nutrient uptake, antibiotic efflux, protein secretion, toxin production to oxidative phosphorylation and environment sensing. There are commercial and biomedical interests in inhibiting the activities of some membrane transporters, optimizing the activities of others and employing them as transducers of electrical/chemical/mechanical energy for nanotechnology. A detailed structural and functional insight of membrane transporters is essential for these endeavors. However membrane transporters are mostly low abundant in their native membrane. Hence, heterologous overproduction of membrane proteins in sufficient amounts is the absolute necessity for structural and functional studies. In recent years, structural genomics approaches of heterologously overproducing large number of proteins in medium or high-throughput manner for crystallographic studies have been promising. Moreover, it is also necessary to develop and establish new expression and purification protocols for membrane proteins to enrich the crystallization pipeline.

The work of this thesis is part of a structural genomics project already running in our group. A set of 210 transporters from *A. aeolicus*, *S. typhimurium* and *P. furiosus* were selected based on available bioinformatics data. An initial pilot scale expression screening with 37 transporters was carried out to evaluate the use of orthologous targets, various expression vectors and different host systems in order to find out the appropriate conditions for maximum protein production. This study revealed the following (Surade et al. 2006): (i) *A. aeolicus* was the most suitable source organism for production of hyperthermophilic proteins, (ii) *E. coli* was able to express a higher number of heterologous prokaryotic transporters at higher levels than *L. lactis*, (iii) the pBAD vector was superior to others in terms of expressing the targets in higher amounts.

For my thesis work, I had chosen targets from the original set of 210 secondary active transporters as described in the table 1.2. The main criteria for selecting the target family were lack of a published structure and established protocols for heterologous overproduction of at least one member of the family. However, in the course of my doctoral work structures from two of the families were published.

**Table 1.2: Details of the selected targets**

S.No.	Protein ID <sup>a</sup>	Source <sup>b</sup>	Family <sup>c</sup>	Function	nAA	MW (kD)	nTM <sup>d</sup>
1	STM0758	ST	CDF	Heavy metal transport	312	34.4	5
2	STM4061	ST			300	32.97	6
3	Aq_1073	AA			308	34	6
4	Aq_2073	AA			385	43.01	5
5	STM3880	ST	KUP	Potassium transport	622	69.18	12
6	STM4189	ST	PNaS	Inorganic phosphate transport	543	59.4	8
7	STM0365	ST	RhtB	Amino acids transport	210	23.03	6
8	STM1270	ST			212	22.94	6
9	STM2645	ST			195	21.29	6
10	STM3959	ST			206	22.49	6
11	STM3960	ST			206	22.32	6
12	STM3476	ST	FNT	Nitrate assimilation, transport	269	28.6	6
13	STM1781	ST	SulP	Sulfate transport	553	58.6	11
14	PF0708	PF	MOP	Multidrug transport	461	49.23	12

a : according to [www.membranetransport.org](http://www.membranetransport.org)

b: ST- *Salmonella typhimurium* LT2, AA- *Aquifex aeolicus* VF5, PF- *Pyrococcus furiosus* DSM3638

c: **CDF**- Cation Diffusion Facilitator, **KUP**- Potassium Uptake permease, **PNaS**- Phosphate Sodium Symporter, **RhtB**- Resistance to Homoserine Threonine, **FNT**- Formate-Nitrite Transporter, **SulP**- Sulfate Permease, **MOP**- Multidrug/Oligosaccharidyl-lipid/Polysaccharide flippase

d: Predicted by TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>)

S.No. Serial number, nAA Number of Amino Acids, MW (kD) Molecular weight in Kilodalton, nTM number of transmembrane helices

The overall aim of this work was to heterologously overproduce selected secondary active transporters from three different source organisms, within the structural genomics framework already established in our group. After successful overproduction, they were subjected to characterization in terms of detergent compatibility and stability which in turn could facilitate the crystallization of the selected targets. The idea behind choosing multiple targets from a single family was to take at least one target up to the crystallization stage. Alternative expression strategies, such as cell-free expression and MBP fusion, were also tried to generate sufficient amount of homogeneous protein for functional and structural studies.

The specific aims of my project were:

- (i) Successful expression of chosen transporters and purification to homogeneity
- (ii) Characterization of the purified proteins in terms of detergent compatibility and stability
- (iii) Crystallization of the purified protein
- (iv) Functional characterization of the transporters
- (v) Establishing a cell-free production and purification protocol for selected transporter targets, especially for those which were difficult to produce *in vivo*
- (vi) Analysis of crystallizability and functionality of cell-free produced protein

The flow chart of Fig. 1.13, describes the road map followed for this thesis work. After initial expression screening, I tried to heterologously overproduce the target proteins in cell based system. A quality control step to ensure the stability, purity, integrity and homogeneity of the protein was followed by functional and structural characterizations. Upon encountering difficulties like no expression, degradation, low yield, poor stability or poor crystal in crystallization trials, alternative approaches like cell-free expression, MBP fusion or truncated construct designing were attempted to check whether these could overcome the obstacles.



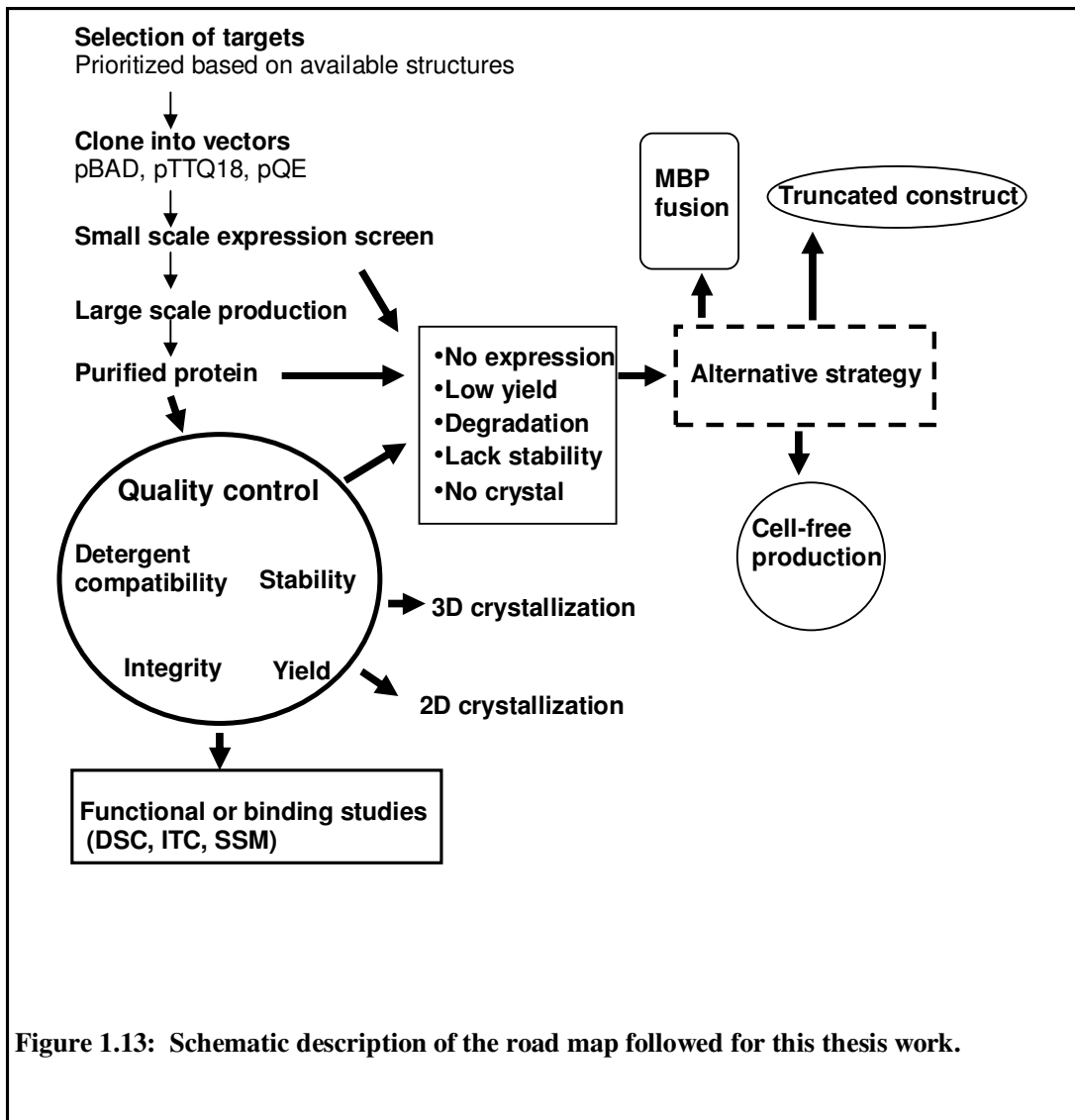


Figure 1.13: Schematic description of the road map followed for this thesis work.

## *Results*

---

## 2.1 Cation diffusion facilitator (CDF) family

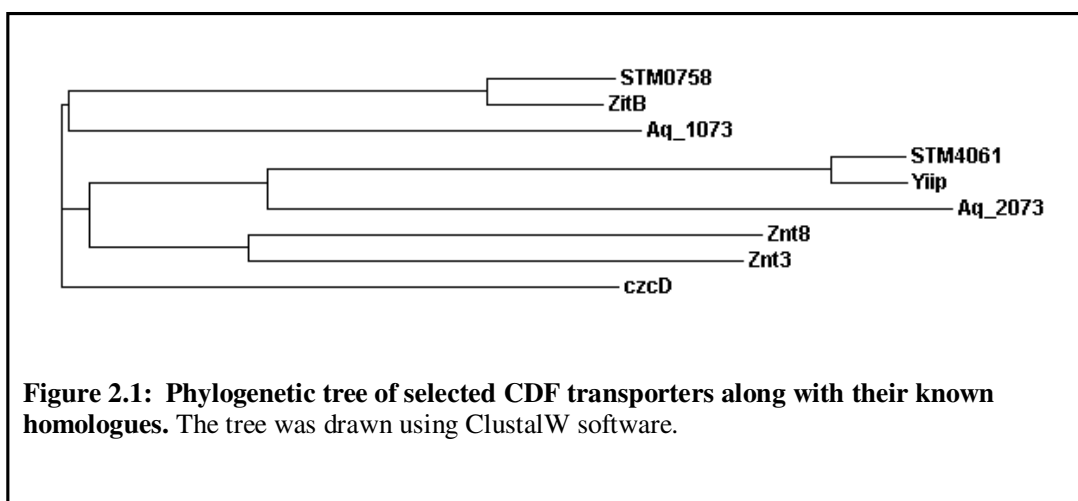
### 2.1.1 Target selection and *in silico* analysis

The heavy metal transporting CDF family members typically possess 5-6 transmembrane segments with both N- and C-terminus oriented towards the cytoplasm. Both *A. aeolicus* and *S. typhimurium* have two members each of the CDF family, Aq\_1073, Aq\_2073 and STM0758, STM4061, respectively. The results of an *in silico* analysis of these selected four targets are summarized in Table 2.1 and their phylogenetic tree in Fig. 2.1.

**Table 2.1: *In silico* analysis of selected CDF transporters**

Protein	Source organism	TM <sup>a</sup>	MW (kD)	% Identity with <sup>b</sup>				
				ZitB <sup>EC</sup>	YiiP <sup>EC</sup>	CzcD <sup>RM</sup>	Znt3 <sup>HS</sup>	Znt8 <sup>HS</sup>
STM0758	ST	5	34.4	86	27	37	31	30
STM4061	ST	6	32.9	24	92	25	19	21
Aq_1073	AA	6	34	39	25	41	28	26
Aq_2073	AA	5	43.01	25	25	34	N*	28

ST: *Salmonella typhimurium*, AA: *Aquifex aeolicus*; <sup>a</sup> Number of transmembrane segments as predicted by TMHMM server; <sup>b</sup> Identity in percentage (%), as calculated by NCBI-BLAST Align tool with default settings; N\* No significant similarity  
<sup>EC</sup> *Escherichia coli*, <sup>RM</sup> *Ralstonia metallidurans*, <sup>HS</sup> *Homo sapiens*



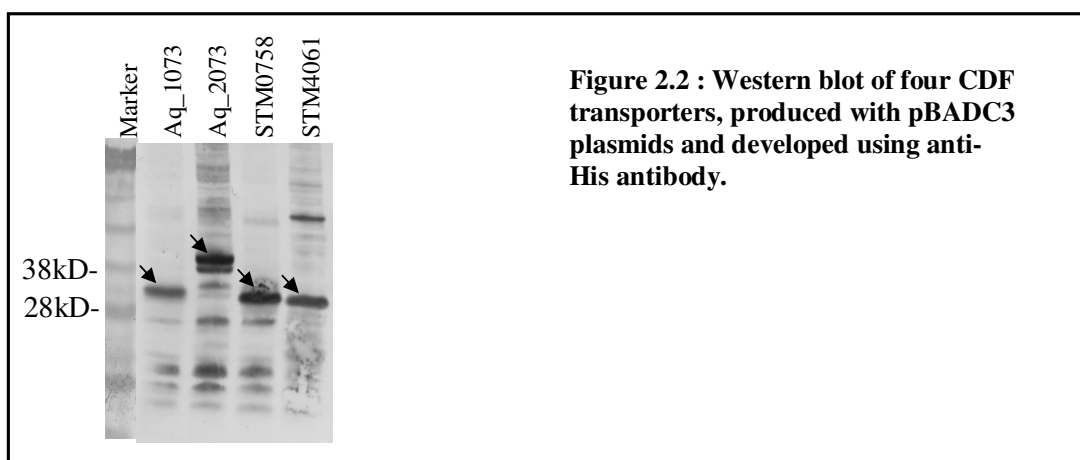
**2.1.2 Expression screening of 4 members of CDF family**

The gene sequence of the four selected transporters of *A. aeolicus* and *S. typhimurium* are obtained from TransportDB (<http://www.membranetransport.org>). The genes were cloned into the pBAD, pTTQ18 and pQE vectors, each having 2 different constructs (A2 and C3), resulting in 6 constructs altogether. These plasmids were transformed into the respective expression host cells and subjected to whole cell lysate expression screening and visualization by Western blot (Table 2.2 and Fig. 2.2). All the constructs could express the targets equally well, as determined by whole cell lysate screening.

**Table 2.2: Expression screening of CDF transporters, determined by Western blot**

Constructs		Tag position	Host cell	Aq_1073	Aq_2073	STM0758	STM4061
pBAD	A2	C-His	TOP10	++	++	++	++
	C3	N-His, C-Strep		++	+++	+++	++
pTTQ18	A2	C-His	NM554	++	++	++	++
	C3	N-His, C-Strep		++	+++	+++	++
pQE	A2	C-His	NM554	++	++	++	++
	C3	N-His, C-Strep		++	++	++	++

++ and +++ represents relative expression level



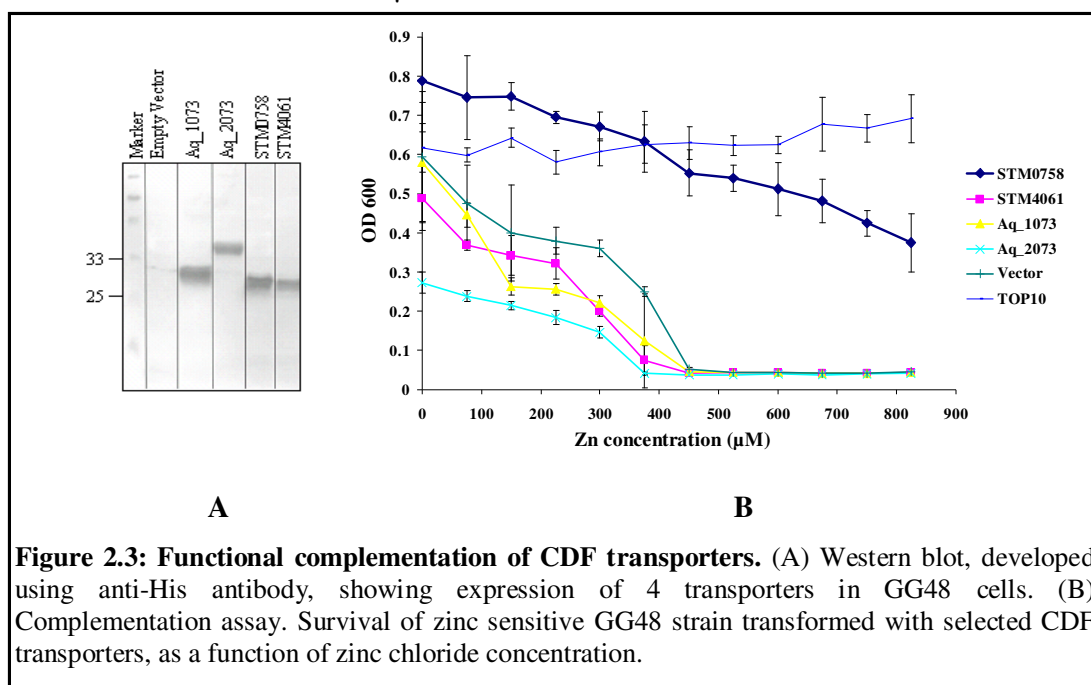
**2.1.3 Functional complementation of 4 CDF transporters**

Heterologous overproduction of membrane proteins often resulted in a nonfunctional inactive state. To check the functionality of the expressed CDF transporters, the *E. coli*

## RESULTS

strain GG48 ( $\Delta$ zitB::CmzntA::Km) was employed for complementation studies. In GG48, the Zn translocating P-type ATPase ZntA and the CDF protein ZitB have been deleted, rendering it highly Zn sensitive (Grass et al. 2001). All four constructs were expressed in GG48 cells and could be detected migrating at the correct position in SDS-PAGE as indicated by Western blot, developed using an anti-His antibody (Fig. 2.3A). For functional complementation, GG48 cells were transformed with pTTQ18C3 based constructs of STM0758, STM4061, Aq\_1073 and Aq\_2073. The GG48 strain transformed with empty vector (pTTQ18C3) was taken as a negative control, while TOP10 cells were employed as positive control. The cells were then allowed to grow in the presence of different zinc ion concentrations (0-900  $\mu$ M). The absorbance at 600 nm was measured after 12 hrs as an indicator of cell survival and plotted against the zinc ion concentration as described in details in materials and methods.

Out of four transporters tested, only STM0758 could markedly enhance the zinc tolerances of the *E. coli* GG48 strain (Fig. 2.3B). The OD<sub>600</sub> value of the negative control (GG48 transformed only with empty vector pTTQ18C3) rapidly decreased beyond 300  $\mu$ M ZnCl<sub>2</sub> concentration and became almost zero at 450  $\mu$ M concentration. Only the GG48 strain transformed with the STM0758 containing plasmid could tolerate zinc concentrations even above 800  $\mu$ M.



**Figure 2.3: Functional complementation of CDF transporters.** (A) Western blot, developed using anti-His antibody, showing expression of 4 transporters in GG48 cells. (B) Complementation assay. Survival of zinc sensitive GG48 strain transformed with selected CDF transporters, as a function of zinc chloride concentration.

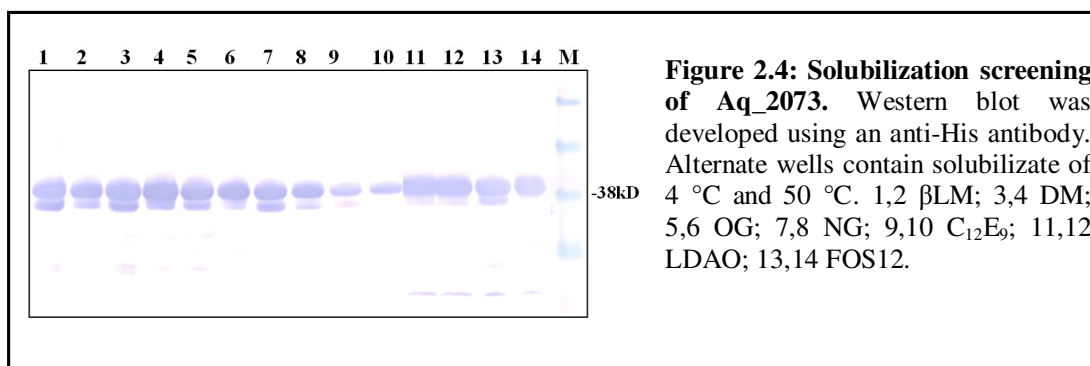
However, crystallization of membrane proteins requires sufficient amounts of pure and stable protein. When tried for large scale production, Aq\_2073 out of the four selected targets of CDF transporters turned out to be the protein of choice in terms of yield and stability.

#### **2.1.4 Production, isolation and characterization of Aq\_2073**

Aq\_2073 showed a good expression level in whole cell lysate expression screening with all the constructs tested. In a pilot scale structural genomics studies, Surade et al. (2006) reported pBAD vector to heterologously overproduce the maximum number of targets. So, to work further with Aq\_2073 in similar context, the pBAD constructs (A2 and C3) were chosen.

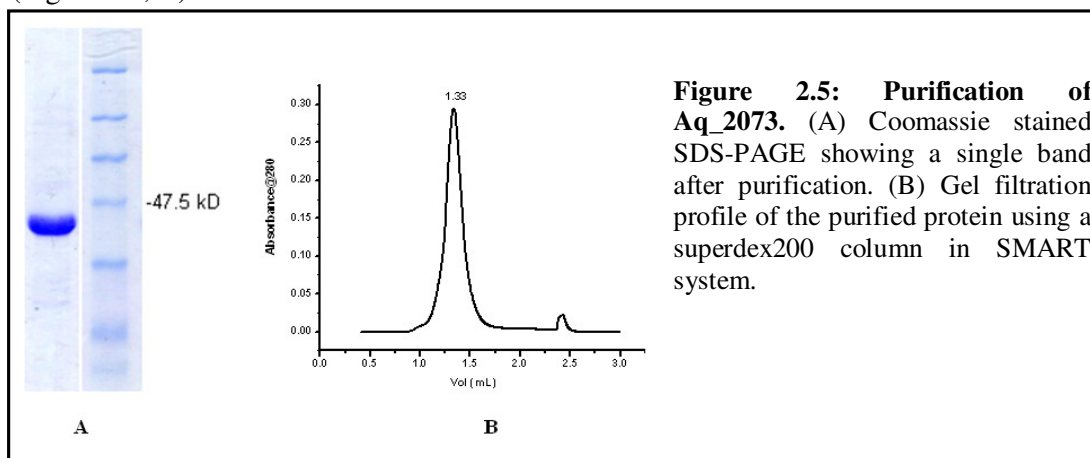
##### **2.1.4.1 Solubilization screening for choosing the right detergents**

After fresh transformation into *E. coli* TOP10 cells, a single colony of Aq\_2073 - pBADC3 was inoculated in 400 ml LB for overnight culture. This was then used to inoculate 24 liter LB media for large scale production. Cell harvest and membrane preparation was done according to materials and methods. Membranes were stored in a -80 °C freezer. For solubilization screening, aliquots of membranes were thawed on ice, and subjected to solubilization screening. Membranes were solubilized in 7 different detergents) with 1% final detergent concentration and at two different temperatures (4 °C and 50 °C. Detergents  $\beta$ LM, DM, LDAO and FOS12 turned out to be the most suitable ones. However, the temperature did not have any significant effect on solubilization (Fig. 2.4).



### 2.1.4.2 Affinity purification and homogeneity of Aq\_2073

In the presence of a His-tag, immobilized metal affinity chromatography (IMAC) using Ni-NTA matrix, is a simple and one step procedure to purify the target protein. After solubilization and ultracentrifugation, the solubilize was allowed to bind to the Ni-NTA matrix either in batch method or using the HisTrap-HP columns in Aekta Prime system. The protein was eluted with either the 0-500 mM gradient of imidazole or with 350 mM of imidazole in Aekta Prime or batch method respectively. There was no notable difference in yield with both methods tested. The protein was more than 95% pure as judged on a Coomassie stained gel. Size exclusion chromatography of the purified protein, using a Superdex 200 column in the SMART system, showed that the protein was homogeneous in the detergent  $\beta$ LM as indicated by the presence of a single peak (Fig. 2.5 A, B).

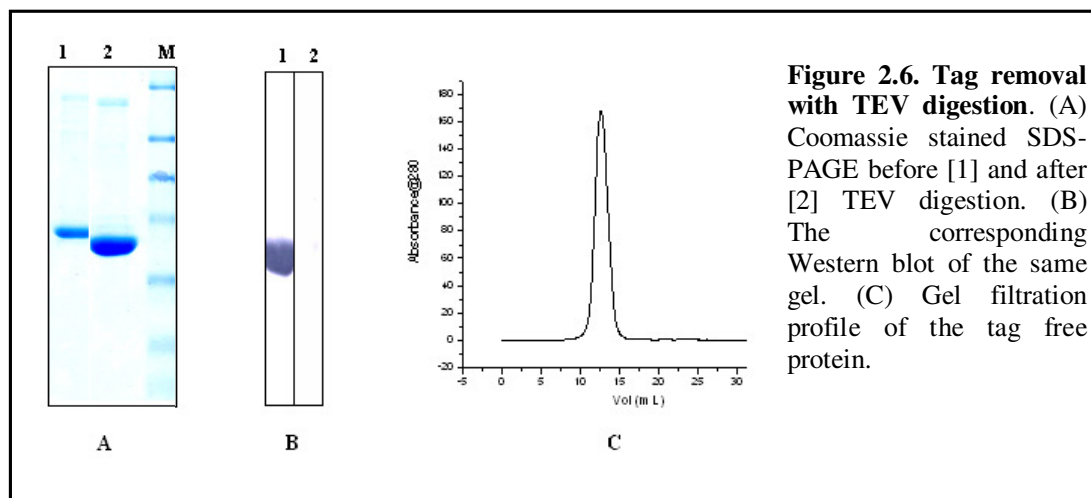


**Figure 2.5: Purification of Aq\_2073.** (A) Coomassie stained SDS-PAGE showing a single band after purification. (B) Gel filtration profile of the purified protein using a superdex200 column in SMART system.

### 2.1.4.3 Removal of His-tag using TEV digestion

Removal of the His-tag, located at the C-terminus (A2 version) or N-terminus (C3 version) of Aq\_2073, was necessary not only to characterize the protein's binding and transport properties with heavy metals (zinc, cadmium, nickel, iron), but also for crystallization attempts with the tag free protein. TEV protease recognizes a specific sequence of amino acids and carries out a proteolytic cleavage of the substrate (Kapust et al. 2001). Both A2 and C3 versions have a TEV cleavage site before and after the His-tag, respectively. The impurities in the sample after IMAC are supposed to have a high non-specific affinity for the IMAC-resin or metal. Hence, if the sample is treated with the TEV protease, it should cleave at the TEV site and separate the His-tag from the protein itself. When this reaction mixture is passed over an IMAC resin once again, all impurities with non-specific binding tendency and the His-tagged TEV protease should bind to the

IMAC resin whereas the protein of interest can be recovered in the flow-through (Fig. 2.6). A mutant version of TEV protease S219V (Kapust et al. 2001) with a higher stability was used for these experiments.



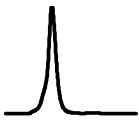
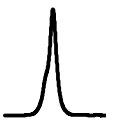

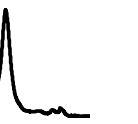
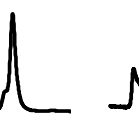
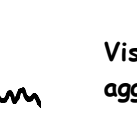



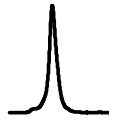
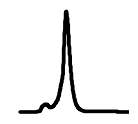
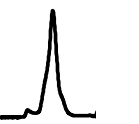


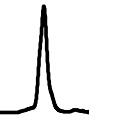
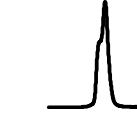
#### 2.1.4.4 Stability and monodispersity of Aq\_2073

Stability and monodispersity of a membrane protein, prerequisites for crystallization, depend on a multitude of factors like pH, salt concentration, detergents etc.. Using analytical size exclusion chromatography (SEC), a large set of these factors can be screened efficiently. Conditions that are beneficial in terms of stability and monodispersity, can then be applied for purification and crystallization.

Aq\_2073 was subjected to such screenings of variables like detergents, pHs and temperatures and analyzed by superdex 200 column chromatography using a SMART system. Aq\_2073 was stable and demonstrated monodispersity in a wide range of detergents and combination of detergents. Stability studies with the maltoside series of detergents demonstrated that monodispersity decreases with the decrease in carbon chain length of the detergent in the order LM>UM>DM>NM. The protein was less stable or not stable at all in the glucosidic detergents NG and OG, respectively. However, the protein was found to be stable and monodispersed in relatively harsh detergents like FOS12 and LDAO. A detailed description of stability in different detergents with corresponding gel filtration profiles is presented in Table 2.3.



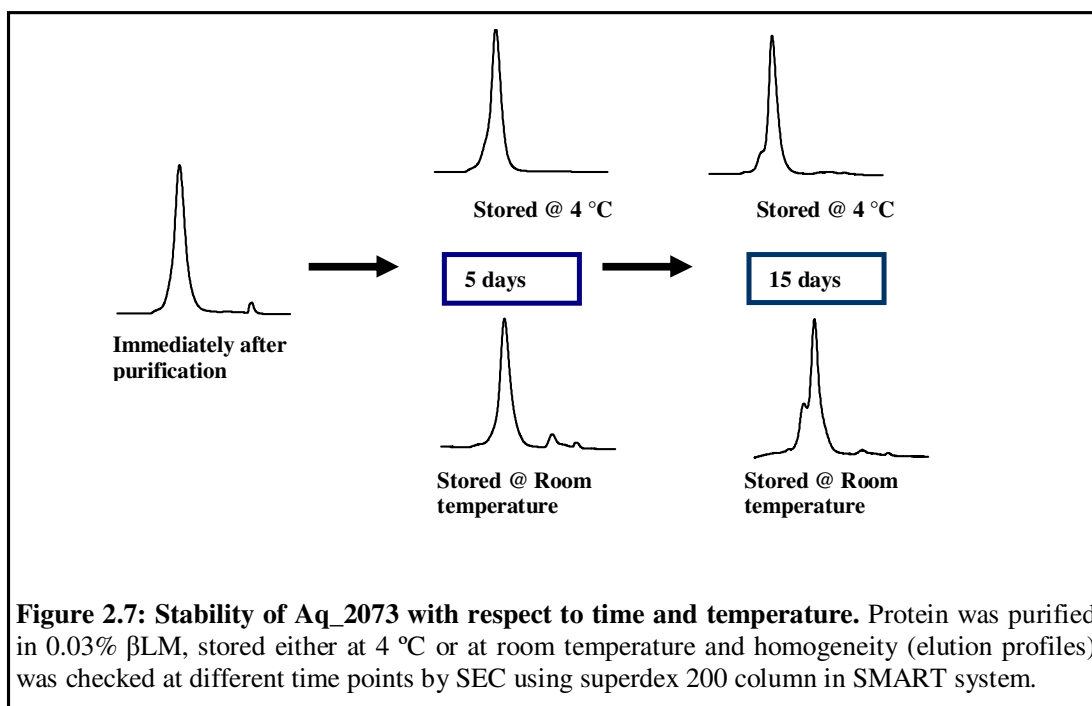
Table 2.3: Stability of Aq\_2073 in different detergents analyzed by size exclusion chromatography (SEC)

Detergent	Stability, SEC profile							Comments
Maltosides and Glucosides							<b>Visible aggregate</b> OTG (0)	$\beta$ LM, UM and DM purified proteins used for crystallization trials
Polyoxyethylene alkyl ether								$C_{10}E_5$ purified protein used for crystallization trials
Others								Used for crystallization trials
Mix								FOS11 decreased stability; others used for crystallization trials

+++ most stable; ++ intermediate stable; + less stable; 0 Visible aggregates

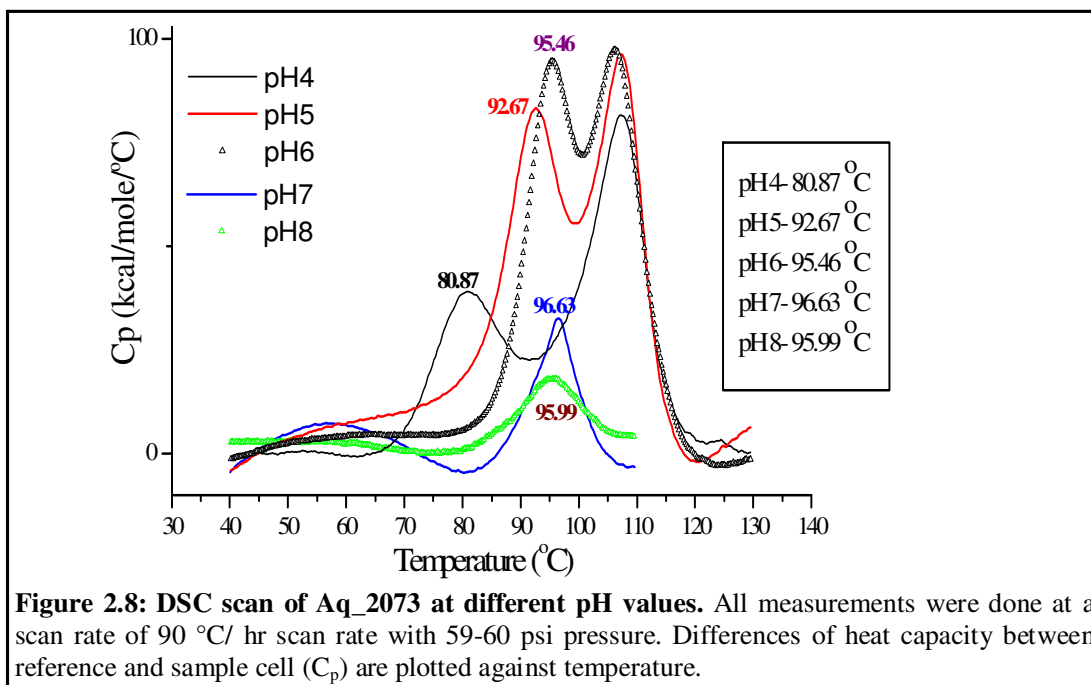
All proteins were purified in 20 mM HEPES-NaOH, pH 8

For further characterization of the stability, the protein purified in  $\beta$ LM detergent was analyzed by SEC after storage at two different temperatures (4 °C or room temperature) for different time intervals. After 15 days of storage, protein stored at 4 °C showed better homogeneity compared to protein stored at RT which eluted with a small hump before the actual peak (Fig. 2.7). Aq\_2073 showed stability in pH values ranging from 4 to 9. However, pH7 and pH 8 demonstrated maximum monodispersity with a single peak without any hump or shoulder (Appendix 3).



#### 2.1.4.5 Thermal unfolding studies using DSC

Differential scanning calorimetry is a useful method to study protein stability besides the phenomenon associated with protein unfolding. The melting point or  $T_m$  determined by a DSC experiment is the direct reflection of the tested protein's stability. Aq\_2073 was subjected to DSC scans at 5 different pH values. All measurement buffers of respective pH values (Acetate buffer pH 4 and 5; MES pH 6; HEPES pH 7 and 8: all at 50 mM concentration) additionally contained 150 mM NaCl and 0.02%  $\beta$ LM. After affinity purification, the protein samples were dialyzed overnight with a change against respective buffers, the concentration was adjusted to 0.5-1 mg/ml and the sample was filtered before injection into the sample cell. The corresponding buffers were used in the reference cell.



**Figure 2.8: DSC scan of Aq\_2073 at different pH values.** All measurements were done at a scan rate of 90 °C/ hr scan rate with 59-60 psi pressure. Differences of heat capacity between reference and sample cell ( $C_p$ ) are plotted against temperature.

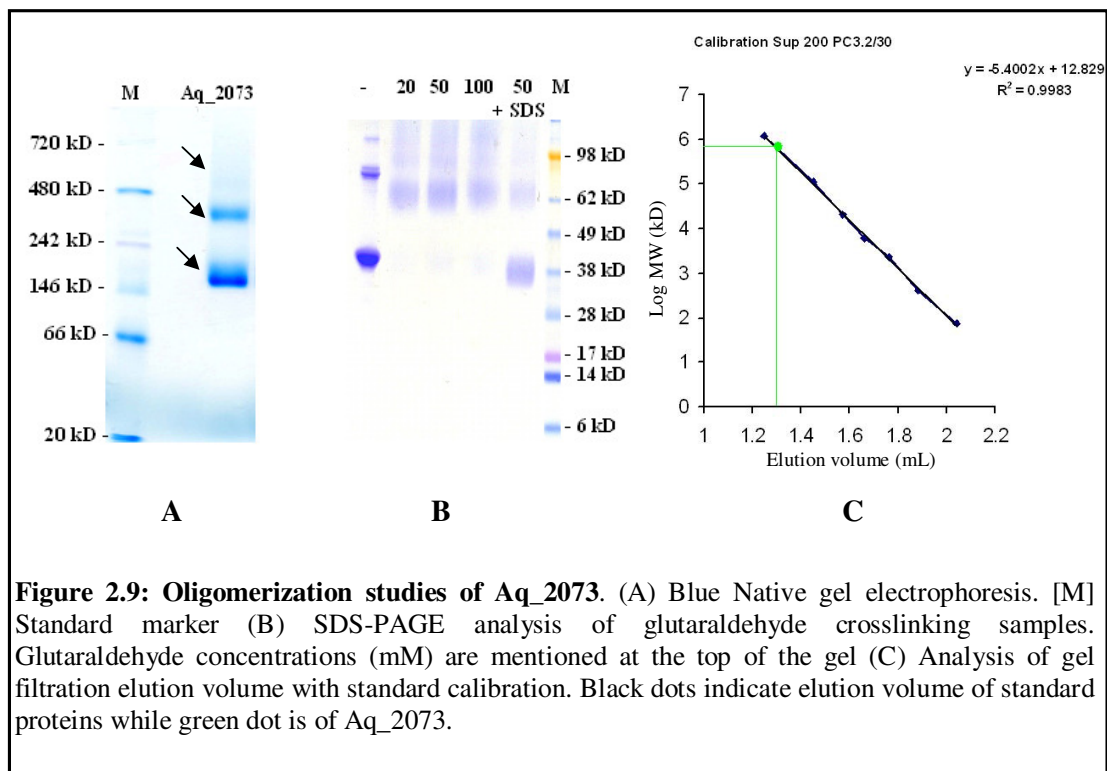
All measurements were done at a scan rate of 90 °C/ hr scan rate with 59-60 psi pressure. As observed from the combined thermogram in Fig. 2.8, the melting temperature ( $T_m$ ) started to increase from pH 4 to pH 8. Moreover, the scans at pH 4, 5 and 6 yielded an additional peak at a higher temperature. This could be due to the presence of a small fraction of higher oligomeric or aggregated protein present in the sample at these pH values which unfold separately. The gel filtration profiles of the same sample at pH 4, 5 and 6 showed a small hump before the main peak corresponding to higher oligomeric states (Appendix 3). On the other hand, scans at pH 7 and 8 yielded a single peak with a  $T_m$  of approximately 96 °C.

#### 2.1.4.6 Oligomerization studies

The oligomerization state of Aq\_2073 was tested by three different methods, namely, blue native PAGE (BN-PAGE), crosslinking studies and analytical gel filtration chromatography. In BN-PAGE, Aq\_2073 migrated as two distinct bands, one just above 146 kD and the other between 242 kD and 480 kD (Fig. 2.9A). As reported by Heuberger et al. (2002), using the conversion factor of 1.8 for bound CBB dye, these two bands correspond to a dimer and a pentamer, respectively. The presence of further higher oligomers could also be observed as a faint band above 480 kD.

## RESULTS

Glutaraldehyde crosslinking experiments were performed with 20, 50 and 100 mM of glutaraldehyde with protein at a concentration of 0.1 mg/ml for 1hr at room temperature. A control experiment was also set up with 1% final SDS and 50 mM glutaraldehyde. At all concentrations, the monomeric band faded away and a prominent band above 62 kD started appearing when analyzed by SDS-PAGE. This band corresponds to a possible dimer. Moreover, one could also see bands corresponding to higher oligomers. Even the sample without glutaraldehyde and the control sample with glutaraldehyde and SDS also showed additional bands which possibly are SDS resistant higher oligomers (Fig. 2.9B). SEC analysis using a superdex 200 PC 3.2/30 column, calibrated with a range of standard soluble proteins in sample buffer (20 mM HEPES, pH8, 150 mM NaCl, 0.02%  $\beta$ LM), was used to calculate the mass of Aq\_2073 in detergent solution. Aq\_2073 solubilized and purified in  $\beta$ LM elutes at 1.33 ml which corresponds to roughly 283 kD (Fig. 2.9C). Thus, the eluted mass of the protein detergent complex is as big as 280 kD. This definitely indicates a higher oligomeric state. However, the determination of the exact oligomeric state from gel filtration profile is rendered difficult as the amount of bound detergent is unknown.



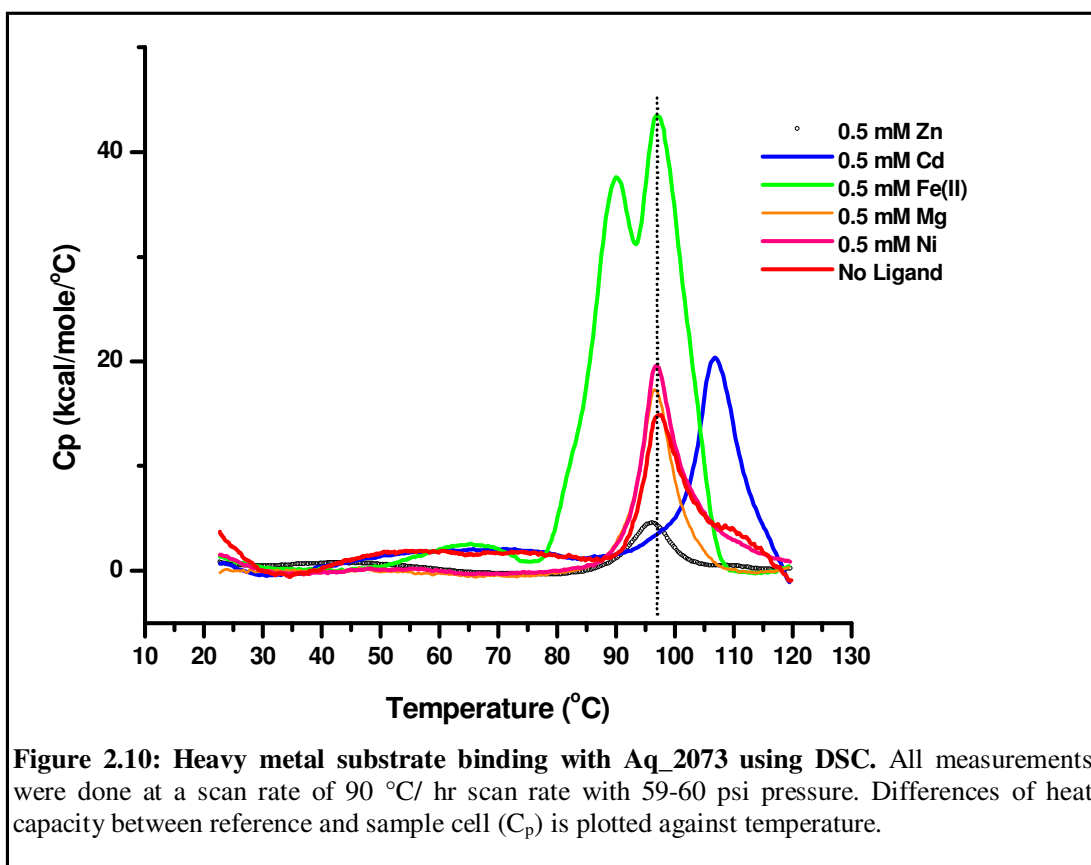
**Figure 2.9: Oligomerization studies of Aq\_2073.** (A) Blue Native gel electrophoresis. [M] Standard marker (B) SDS-PAGE analysis of glutaraldehyde crosslinking samples. Glutaraldehyde concentrations (mM) are mentioned at the top of the gel (C) Analysis of gel filtration elution volume with standard calibration. Black dots indicate elution volume of standard proteins while green dot is of Aq\_2073.

**2.1.4.7 Substrate binding assay with DSC**

Though primarily used for studying stability and folding, DSC can also be used for binding studies. When a small molecule ligand or substrate preferentially binds to the native form of a protein, the ligand stabilizes the protein and the  $T_m$  of the protein-ligand complex is higher than that of the protein in the absence of the ligand. If the ligand preferentially binds to the denatured protein, the  $T_m$  decreases in the presence of the ligand.

Aq\_2073 was purified as mentioned before. The polyhistidine tag was removed by TEV digestion as the tag might interfere with the binding of putative heavy metal substrates (Zn, Ni, Fe(II), Cd etc.). The polyhistidine tag free protein was incubated on ice with 10 mM EDTA for 30 minutes to remove traces of divalent metal ions. Before subjecting to the DSC scan, the sample was dialyzed against the measurement buffer (HEPES-KOH, pH 6.8, KCl 150 mM,  $\beta$ LM 0.02%) with one change after 6 hrs. The chloride salts of zinc, cadmium, nickel, magnesium (as negative control) and the sulfate salt of iron (Fe II) were dissolved in the same buffer at 10 mM stock concentration. Ferrous sulfate (II) solution was stabilized by 1% (W/V) ascorbic acid. Protein samples without substrates and with substrates (each at 500  $\mu$ M final concentration) were scanned separately. A reference scan with buffer in both the cells (sample and reference) was performed and subtracted from the protein scan. All scans were done at 90  $^{\circ}$ C/hr scan rate. Fig. 2.9 shows the combined thermogram after baseline normalization. Aq\_2073 at pH 6.8 showed a  $T_m$  around 97  $^{\circ}$ C. When  $\text{Ni}^{+2}$ ,  $\text{Zn}^{+2}$  or  $\text{Mg}^{+2}$  were added to the sample no significant change in  $T_m$  could be observed. In case of Fe(II), two overlapping peaks were evident, one with  $T_m$  90.82  $^{\circ}$ C and the other around 97.6  $^{\circ}$ C. This might indicate towards interaction of Fe (II) with a fraction of the protein or to a certain domain of the protein. In case of  $\text{Cd}^{+2}$ , a significant increase of  $T_m$  to 106.9  $^{\circ}$ C was observed (Fig. 2.10).

The binding studies showed that only cadmium ions can significantly stabilize the protein structure as evident from the increased  $T_m$ . However, further experiments like isothermal titration calorimetry are necessary to precisely determine the parameters associated with protein-substrate binding.



**Figure 2.10: Heavy metal substrate binding with Aq\_2073 using DSC.** All measurements were done at a scan rate of 90 °C/ hr scan rate with 59-60 psi pressure. Differences of heat capacity between reference and sample cell ( $C_p$ ) is plotted against temperature.

#### 2.1.4.8 Generation of Aq\_2073 constructs for crystallization

One general aspect of protein crystallization that is universally recognized is that rigid, stable proteins are much more likely to crystallize than proteins that are flexible or have dynamic surfaces. If we analyze Aq\_2073 using secondary structure prediction software (PredictProtein), many unstructured regions in the C-terminal domains can be observed (Fig. 2.11). To minimize the flexibility rendered by these unstructured regions at the C-terminus and hence to facilitate crystallization, different constructs with varying length of C-terminal domains were made based on homology alignment. Moreover, two different constructs of the C-terminal domain only were generated in order to check whether it could fold and function independently (Table 2.4). All constructs were cloned into the C3 version of the pBAD vector. Out of the 5 constructs checked for expression, 4 constructs were found to be expressed by whole cell lysate screening (Table 2.4).

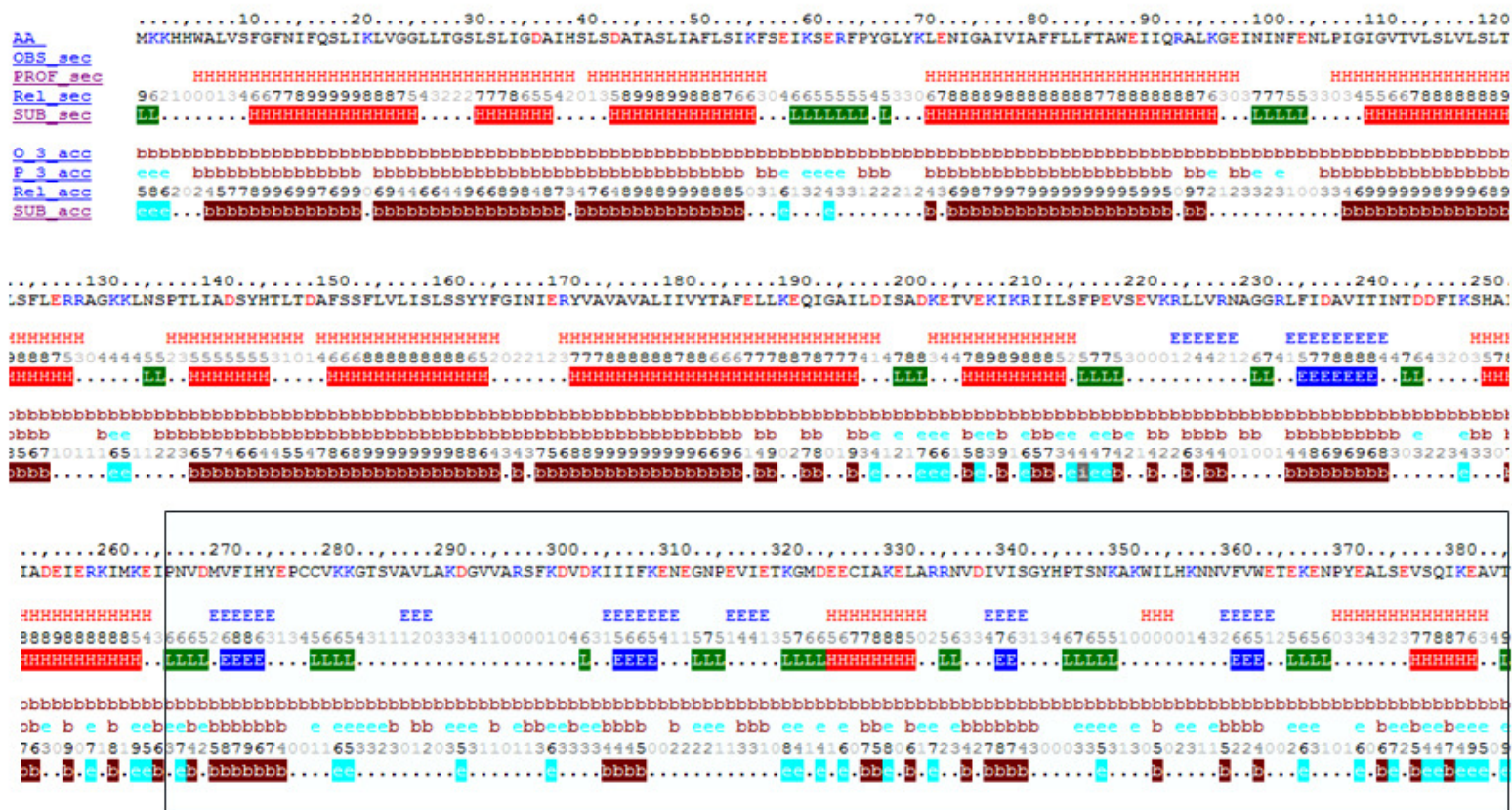









Figure 2.11: Secondary structure prediction of Aq\_2073 using the server PredictProtein ([www.predictprotein.org](http://www.predictprotein.org)). The C terminal region is marked with a box. H denotes predicted helix region, L denotes predicted loop region, E denotes predicted extended sheet region while only dot (.) denotes regions with no possible prediction because of low reliability. These dot regions represent possible unstructured regions.

Table 2.4: Different constructs of Aq\_2073

CONSTRUCTS		Expression	Purification	Crystallization
Full (1-385) →		X	X	X
C1 (1-276) →		X	X	X
C2 (1-250) →		X	X	
C3 (1-220) →		0		
C4 (4-280) →		-		
CTD1 (189-385) ->		X	X	X
CTD2 (203-385) ->		X	X	X

X denotes expression, purification and crystallization trials at corresponding stages; 0 denotes no expression; - denotes not tried

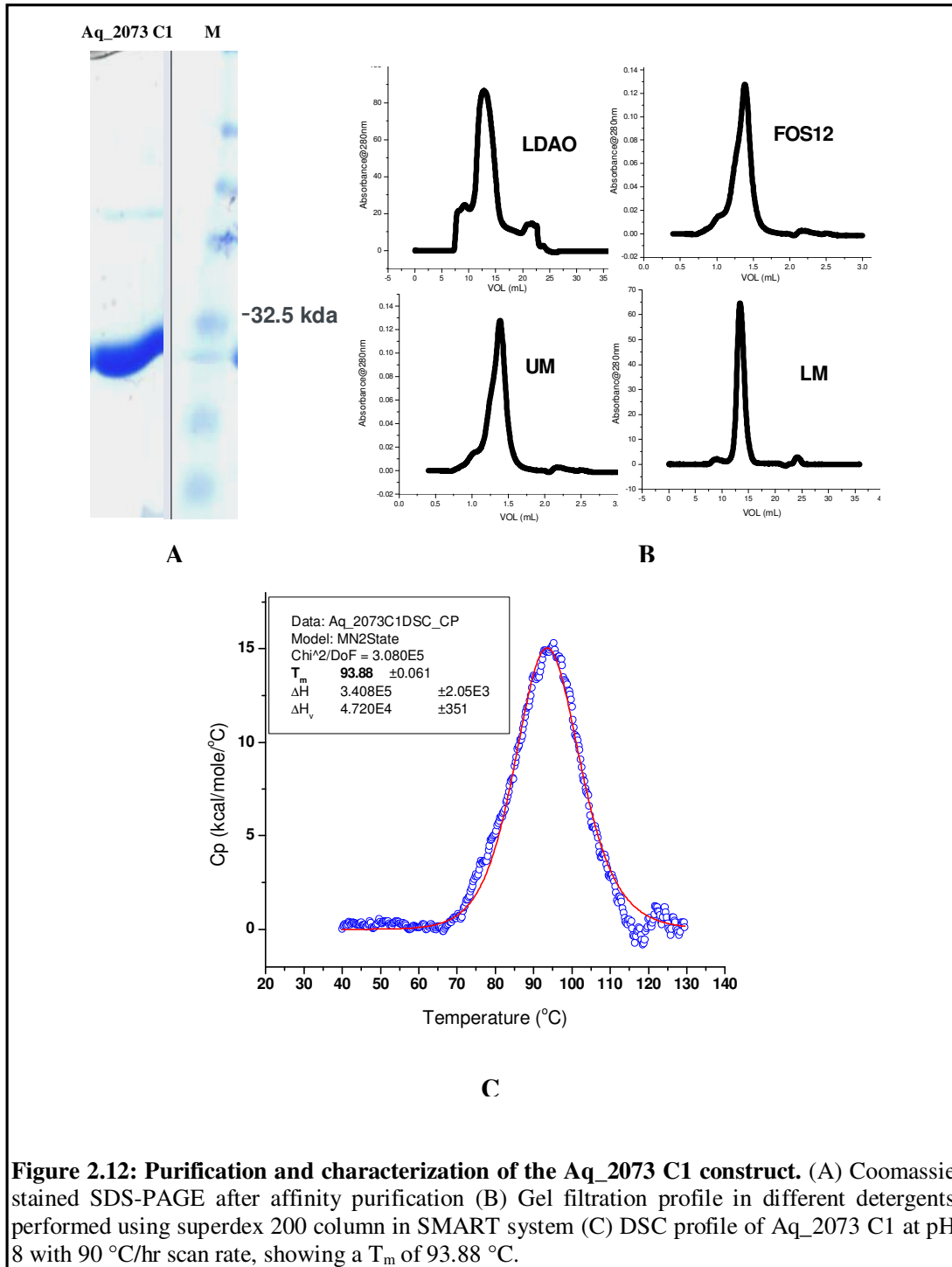
Constructs were generated based on homology alignment using NCBI BLAST tool. The transmembrane and cytoplasmic region boundary was determined using TMHMM prediction

#### 2.1.4.9 Isolation, purification and characterization Aq\_2073 constructs

##### 2.1.4.9.1 Aq\_2073 C1

After large scale membrane preparation and solubilization with 1%  $\beta$ LM, the Aq\_2073 C1 construct can be purified in 0.02%  $\beta$ LM using one step Ni-NTA affinity purification (Fig. 2.12A). Aq\_2073 C1 was further characterized for its stability in different detergents as described in Fig. 2.12B. It was found that like the full length construct, C1 was also stable in a range of detergents ( $\beta$ LM, UM, LDAO, FOS12). The thermal stability of the construct was determined by DSC as described in materials and methods. The DSC scan at pH 8 revealed its  $T_m$  to be 93.88 (Fig. 2.12C).

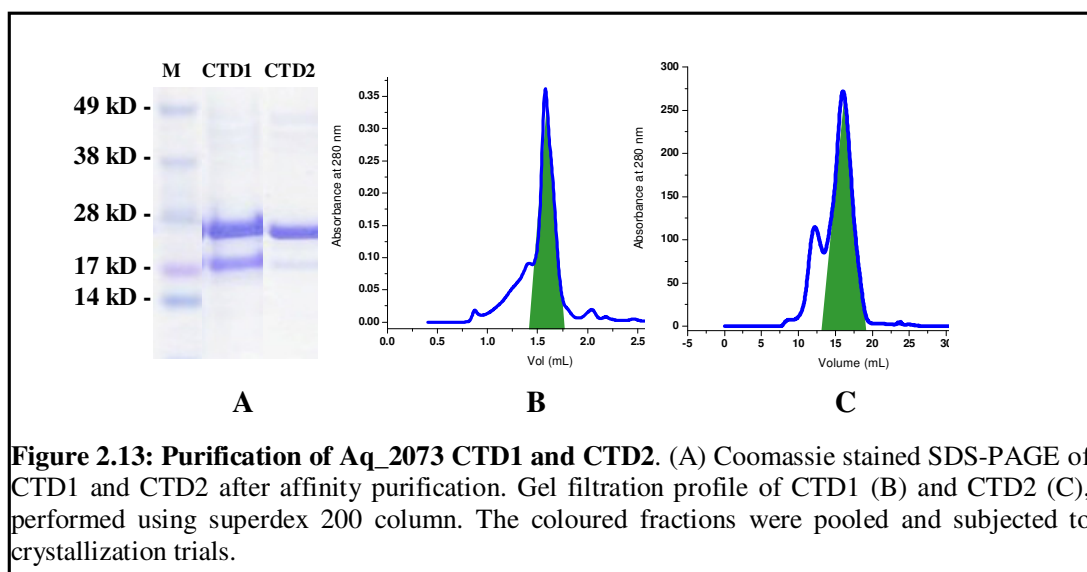




#### 2.1.4.9.2 Aq\_2073 CTD1 and Aq\_2073 CTD2

Aq\_2073 has 197 out of 385 (51%) amino acids as a C-terminal soluble domain (predicted by TMHMM server). As reported in literature, the C-terminal domain can fold independently and bind substrate (Cherezov et al. 2008). To check if this holds true for Aq\_2073, two different constructs of the C-terminal domain, CTD1 (189-385) and CTD2 (203-385) were generated based on TMHMM prediction.

The purification was carried out as described in materials and methods. Both constructs could be purified to an acceptable level of purity. After one step Ni-NTA affinity purification, the protein was further polished by size exclusion chromatography using superdex 200 column. CTD1 was found to be more stable compared to CTD2 as determined by SEC (Fig. 2.13). Both constructs were subjected to crystallization after preparative gel filtration.



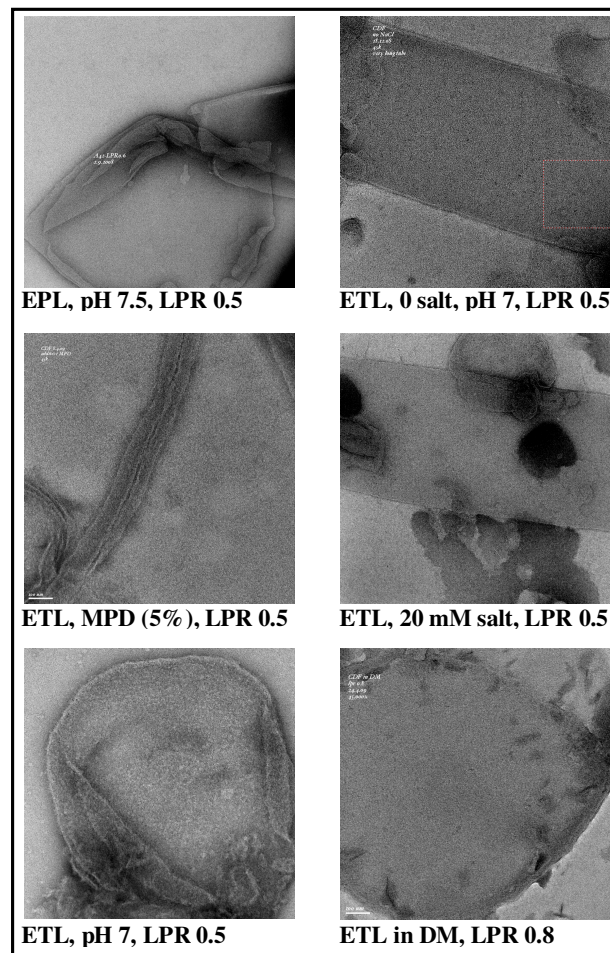
**Figure 2.13: Purification of Aq\_2073 CTD1 and CTD2.** (A) Coomassie stained SDS-PAGE of CTD1 and CTD2 after affinity purification. Gel filtration profile of CTD1 (B) and CTD2 (C), performed using superdex 200 column. The coloured fractions were pooled and subjected to crystallization trials.

#### 2.1.5 Two-dimensional (2D) crystallization

The 2D crystallization experiments were carried out using dialysis bags (Roth). Lipid, detergent, LPR, pH and salt concentrations were varied as described in Table 2.5. *E. coli* lipids were tried (total extract and polar lipid) as the protein was expressed well in *E. coli* host cells. Aq\_2073 and Aq\_2073 C1, both were subjected to 2D crystallization trials. Protein incorporation into vesicle was only observed with *E. coli* total extract lipids.

Table 2.5. Details of 2D crystallization conditions tested

Parameter	Variation	Comment
Construct	Aq_2073 and Aq_2073 C1	Both yielded vesicles with protein incorporation.
Lipid	<i>E. coli</i> total extract and <i>E. coli</i> polar lipid	Only <i>E. coli</i> total extract incorporates protein.
Lipid in detergent	OG, $\beta$ LM, DM	OG is best, LM needs more dialysis time.
Protein in detergent	$\beta$ LM	Protein is most stable in $\beta$ LM.
LPR	0.2, 0.5, 0.8, 1.1, 1.4 (Initial screening) 0.2, 0.3, 0.4, 0.5, 0.6 (Fine screening)	LPR 0,5 produce optimally sized vesicle.
pH	4, 5, 6, 7, 8	pH 7 is most suitable.
Salt (mM)	50, 200, 350, 500, 750 (Initial screening) 0, 10, 20, 40, 60, 80, 100, 150 (Fine screening)	20 to 50 mM salt yielded optimally sized vesicle with protein incorporation.
Temperature	21 °C, 30 °C, 37 °C	30 °C turned out to be optimal.
Additives	MPD (5%), Glycerol (5%). All in 5 mM concentration- ZnCl <sub>2</sub> , CaCl <sub>2</sub> , MgCl <sub>2</sub> , CdCl <sub>2</sub> , Cs <sub>2</sub> SO <sub>4</sub>	No significant effect of any additives.
Time	7 days to 1 month	10 days are optimal when using OG. For $\beta$ LM, longer period is necessary for complete detergent removal.



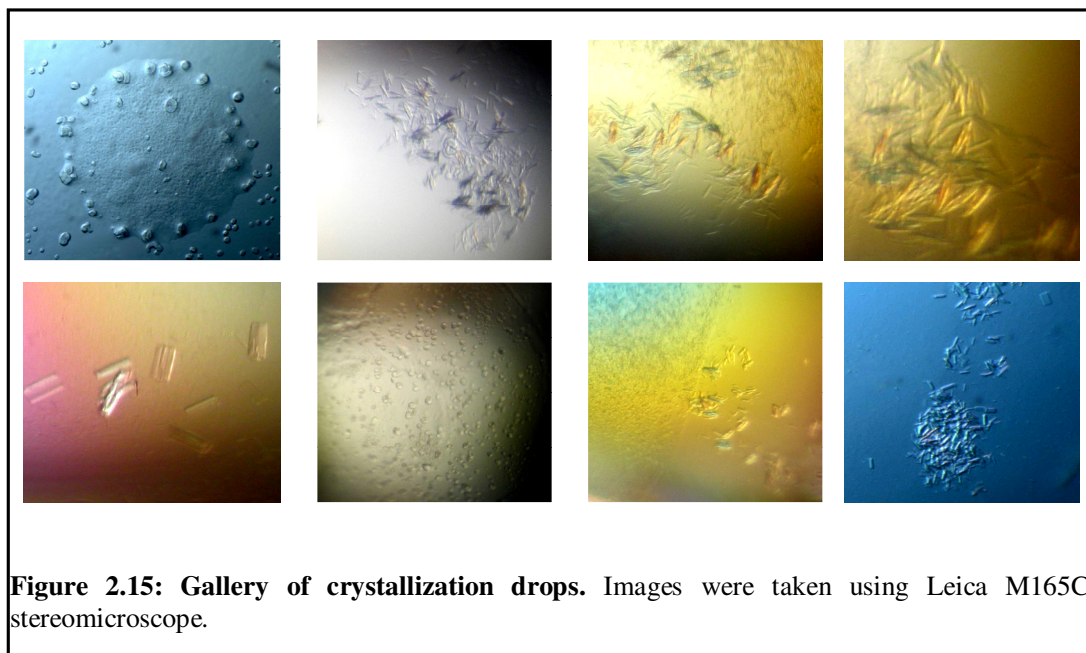
**Figure 2.14: Gallery of 2D crystallization vesicles.**  
 ETL: *E. coli* total lipid; EPL: *E. coli* polar lipid;  
 LPR: Lipid protein ratio.

## RESULTS

Crystallization with a LPR of 0.5 at pH 7 with minimum salt (29-50 mM NaCl) yielded optimally sized vesicles with maximum protein incorporation. Though vesicles with several crystallization patches were obtained, there was no sign of well formed 2D crystals in any of the conditions tested (Fig. 2.14).

### 2.1.6 Three-dimensional (3D) Crystallization

Aq\_2073 and its modifications (Aq\_2073 C1, Aq\_2073 CTD1 and Aq\_2073 CTD2) were extensively tried for crystallization. All the constructs were tried with polyhistidine tags and also after removal of the tag by TEV protease digestion. Commercially available crystallization screens, namely, Hampton, Jena bioscience, Sigma and Qiagen crystallization screens were used in 96 well format using a robot. Drop volumes were 100nL to 1uL with a protein to reservoir buffer ratio of typically 1:1. Initially, Aq\_2073 produced small needle like crystals (Fig. 2.15), which did not diffract properly. Extensive optimization of the initial condition did not improve the shape or size of the crystal. Different detergents and combination of detergents were tried for both Aq\_2073 and Aq\_2073 C1, but without yielding promising crystals. A gallery of crystallization drops is depicted in Fig. 2.15.



**Figure 2.15: Gallery of crystallization drops.** Images were taken using Leica M165C stereomicroscope.

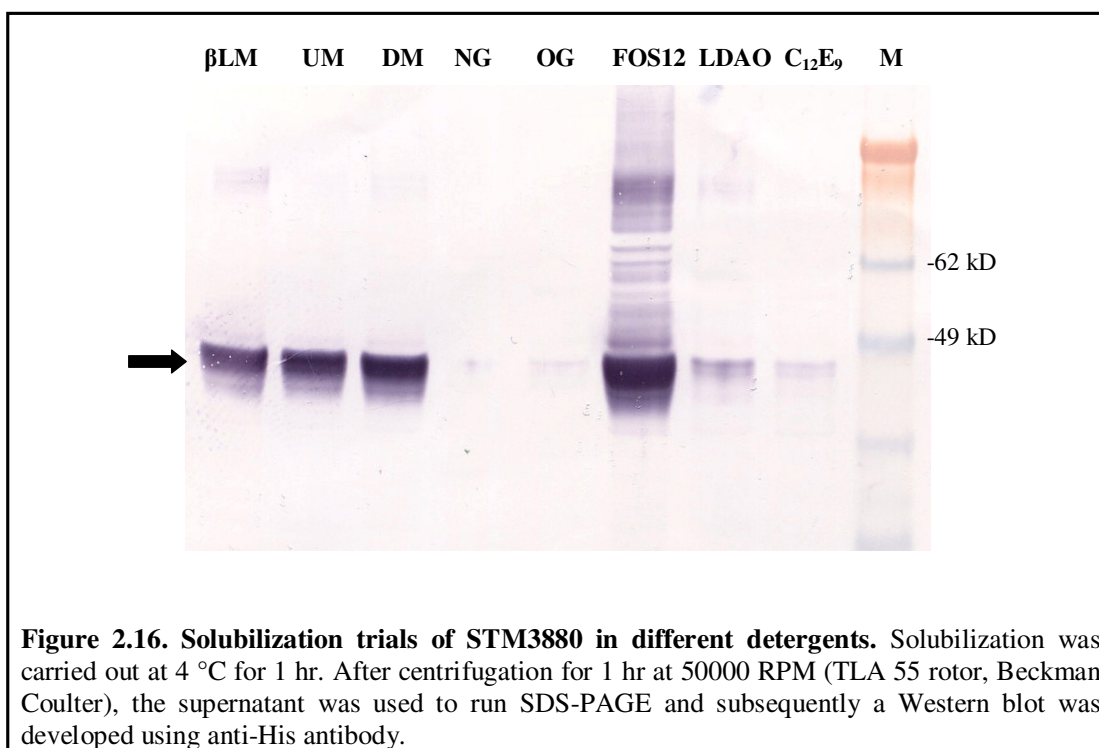
## 2.2 Potassium uptake permease (KUP) family

### 2.2.1 Expression screening of different constructs

STM3880 of *Salmonella typhimurium* is a 69 kD protein with 12 predicted transmembrane segments. The gene was amplified from *S. typhimurium* genomic DNA by PCR and subsequently cloned into pBAD, pTTQ18 and pQE vectors with each having 2 different constructs, A2 and C3. Next, all the six constructs were screened for whole cell lysate expression as described in materials and methods. Only the pBADA2 construct having a C-terminal deca His-tag allowed expression of the protein.

### 2.2.2 Solubilization screening and large scale purification

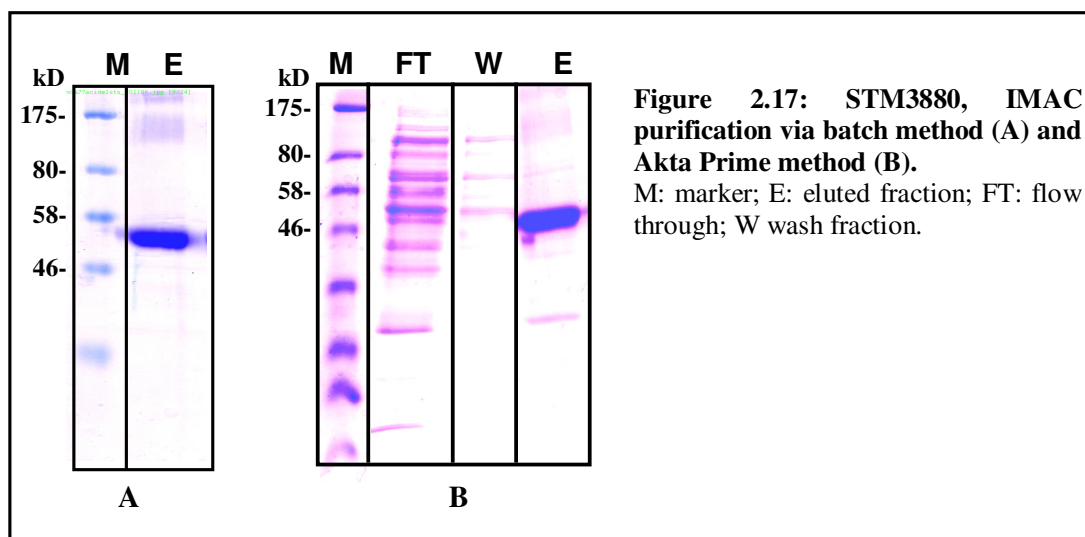
After a large scale membrane preparation (24 L culture), solubilization screening was carried out with 8 selected detergents. The SDS-PAGE and Western blot analysis of the supernatant after solubilization revealed that maltosides and FOS12 can efficiently extract the protein from the membrane (Fig. 2.16).



**Figure 2.16. Solubilization trials of STM3880 in different detergents.** Solubilization was carried out at 4 °C for 1 hr. After centrifugation for 1 hr at 50000 RPM (TLA 55 rotor, Beckman Coulter), the supernatant was used to run SDS-PAGE and subsequently a Western blot was developed using anti-His antibody.

## RESULTS

Large scale production and affinity purification using IMAC resulted in pure STM3880 as determined by using Coomassie stained gels. The yield was roughly 1mg of purified protein from 1 L culture volume as judged by BCA protein assay. STM3880 migrated between 46 and 58 kD which is normal for a 59 kD membrane protein. STM3880 purified in 0.03%  $\beta$ LM lacked stability and tended to aggregate when kept at 4 °C overnight. Even in SDS-PAGE, higher molecular weight bands close to 175 kD could be observed. These are probably SDS resistant aggregates (Fig. 2.17).



### 2.2.3 Characterizing stability in different conditions

STM3880 showed severe stability problems as mentioned before. The protein tends to form visible aggregates when stored at 4 °C overnight. To improve the stability and homogeneity, different detergents, pHs, salt concentrations and additives were screened and corresponding size exclusion chromatography profiles of STM3880 were recorded as a measure of stability. The results are presented in Table 2.6.

The protein showed a relatively higher stability at pH 6 than at the other pH values and with detergent FOS12.

Table 2.6: Variables tested to improve purification and stability of STM3880

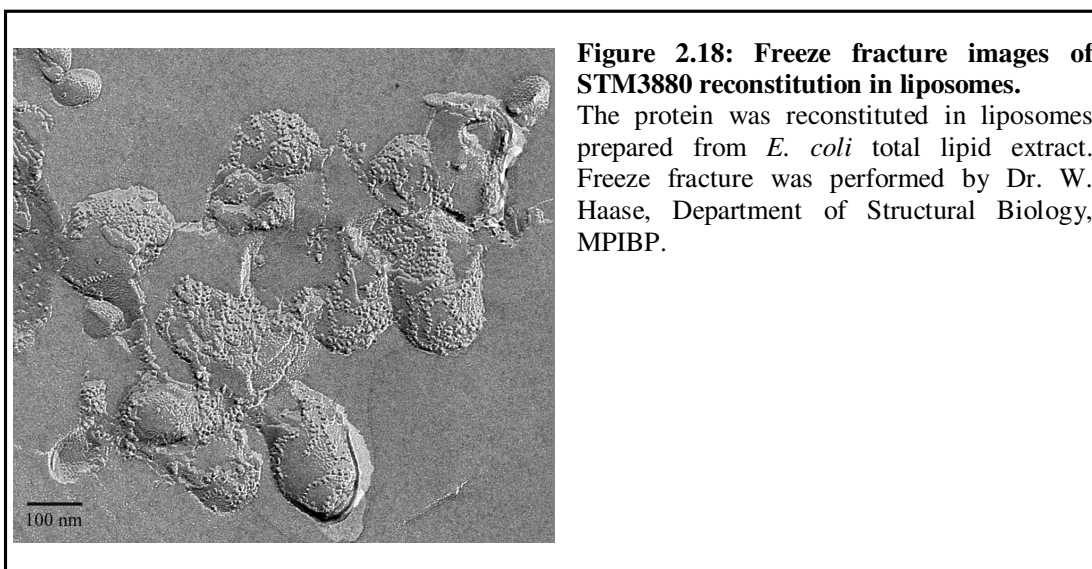
Parameter	Variation	Comment	Gel filtration profile
pH	4, 5, 6, 7, 8, 9	Protein eluted at pH 6 is relatively more stable.	
Salt concentration	50 mM to 1 M	No effect of salt.	
Additives	<ul style="list-style-type: none"> <li>• KCl up to 100 mM</li> <li>• Glycerol up to 20%</li> <li>• Sucrose up to 20%</li> <li>• β ME up to 10 mM</li> <li>• Arginine, proline, aspartate (upto 20 mM)</li> </ul>	No effect of additives.	
Detergents	LDAO, FOS12, OG, NG CYMAL6, C <sub>12</sub> E <sub>9</sub> βLM, βLM +ETL*	Relatively stable in FOS12 and βLM. No effect of lipids.	
Method	Aekta prime (GE health care), Batch method (manual)	No effect.	
Others	Low temperature induction	No effect.	

\* E. coli total lipid extracts with a final lipid concentration of 0.25 mg/ml  
 Red arrows indicate the void volume of the column used (Superdex 200)

### 2.2.4 Reconstitution and freeze fracture

Reconstitution of STM3880 was performed immediately after purifying the protein in either 0.02%  $\beta$ LM or 0.14% FOS12. *E. coli* total lipid extract was dried under nitrogen gas, dissolved in reconstitution buffer (50 mM HEPES-Tris, pH 7.5, 100 mM NaCl) at a concentration of 10 mg/ml and extruded through the liposoFast extruder (Avestin) as mentioned in materials and methods. Reconstitution was performed by the rapid dilution method with a lipid to protein ratio of 1:20.

In order to check the incorporation of STM3880 into the liposomes, freeze fracturing of proteoliposome vesicles was carried out. The analysis revealed that the protein purified with  $\beta$ LM was reconstituted into the liposomes but not protein purified with FOS12 (Fig. 2.18).



### 2.2.5 Solid supported membrane experiment to check functionality

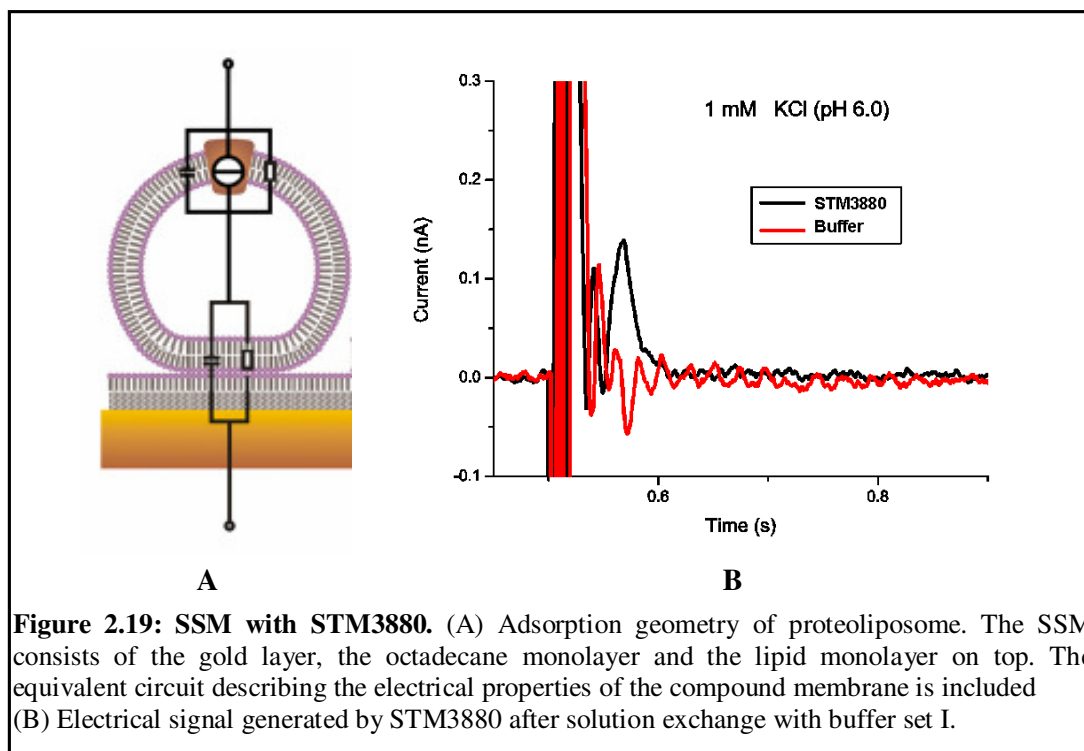
Electrophysiological measurements based on solid supported membranes (SSM) were used to functionally characterize STM3880. In this method, proteoliposomes are adsorbed to a SSM and are activated using a rapid substrate concentration jump. The charge translocation is measured via capacitive coupling of the supporting membrane



(Schulz et al. 2008). Two sets of activating (A) and non activating (NA) buffer solutions were used as described in table 2.7.

**Table 2.7: Buffers used for SSM experiment**

Buffer set I	Non-activating (NA)	50 mM HEPES-Tris, pH 7.5, 1 mM NaCl, 1 mM DTT
	Activating (A)	50 mM HEPES-Tris, pH 7.5, 1 mM KCl, 1 mM DTT
Buffer set II	Non-activating (NA)	50 mM HEPES-Tris, pH 6, 1 mM NaCl, 1 mM DTT
	Activating (A)	50 mM HEPES-Tris, pH 6, 1 mM KCl, 1 mM DTT



**Figure 2.19: SSM with STM3880.** (A) Adsorption geometry of proteoliposome. The SSM consists of the gold layer, the octadecane monolayer and the lipid monolayer on top. The equivalent circuit describing the electrical properties of the compound membrane is included (B) Electrical signal generated by STM3880 after solution exchange with buffer set I.

Fig. 2.19B shows the current induced by a 1 mM concentration jump of the substrate potassium ion at pH 6. In artifact, a current in opposite direction measured with non activating solution in both chambers, was observed. No significant current compared to the artifact was observed when the experiment was repeated with buffer set II (pH 7.5). This observation indicates that the protein is probably active only at low pH (6).

## **2.3 Cell-free production of selected transporters**

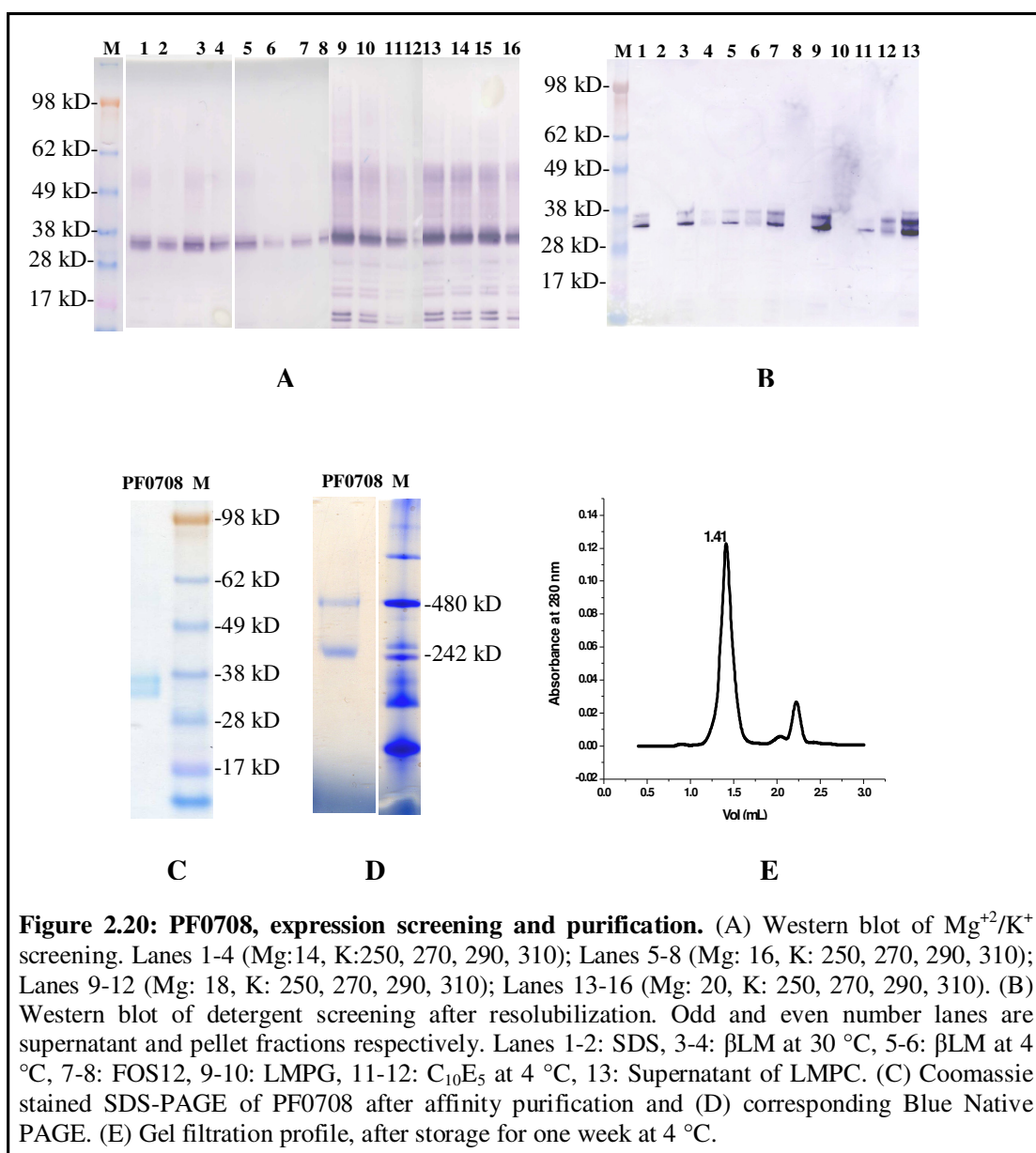
In recent years, cell-free expression has emerged as a viable alternative to the most limiting step of membrane protein structural endeavor, namely overexpressing the membrane protein targets. My objectives were not only to produce difficult targets but also to compare the stability and folding with that of *in vivo* produced protein. I selected several secondary active transporters from four different families for cell-free expression based on *E. coli* S30 extracts. All reactions were performed in continuous exchange cell-free (CECF) mode with either detergent in the reaction mix (DCF) or without (PCF) at 30 °C, if not stated otherwise.

### **2.3.1 PF0708 of the MOP family**

PF0708 of the multidrug/oligosaccharidyl-lipid/polysaccharide flippase superfamily (MOP) was cloned into a modified pET21a vector encoding for a N-terminal T7 tag and a C-terminal deca His-tag. After initial small scale (70 µl) expression check using the continuous exchange (CECF) method, screening for optimal K<sup>+</sup> and Mg<sup>+2</sup> concentrations was performed. It is worth mentioning that, this K<sup>+</sup>/Mg<sup>+2</sup> screening for optimal concentrations was performed with every new batch of an S30 extract. A typical K<sup>+</sup>/Mg<sup>+2</sup> screening result is shown in Fig. 2.20A, which demonstrates that the expression is maximum at a Mg<sup>+2</sup> concentration of 18-20 mM with K<sup>+</sup> in the range of 270-310 mM.

#### **2.3.1.1 PCF mode production of PF0708**

In the PCF mode, the protein is produced without any detergent in the reaction mixture as a precipitate. This precipitated protein or pellet is then solubilized with a suitable detergent, typically at 30 °C for 2 hrs. Fig. 2.20B shows the result of detergent screening for solubilization. The Western blot revealed that beside SDS, which worked as a positive control, FOS12, LMPG and LMPC could solubilize the protein effectively. On the other hand, mild detergents like βLM or C<sub>10</sub>E<sub>5</sub> failed to do so. Overnight solubilization with βLM at 4 °C yielded less solubilized product than at 30 °C. LMPG was chosen for further large scale production.

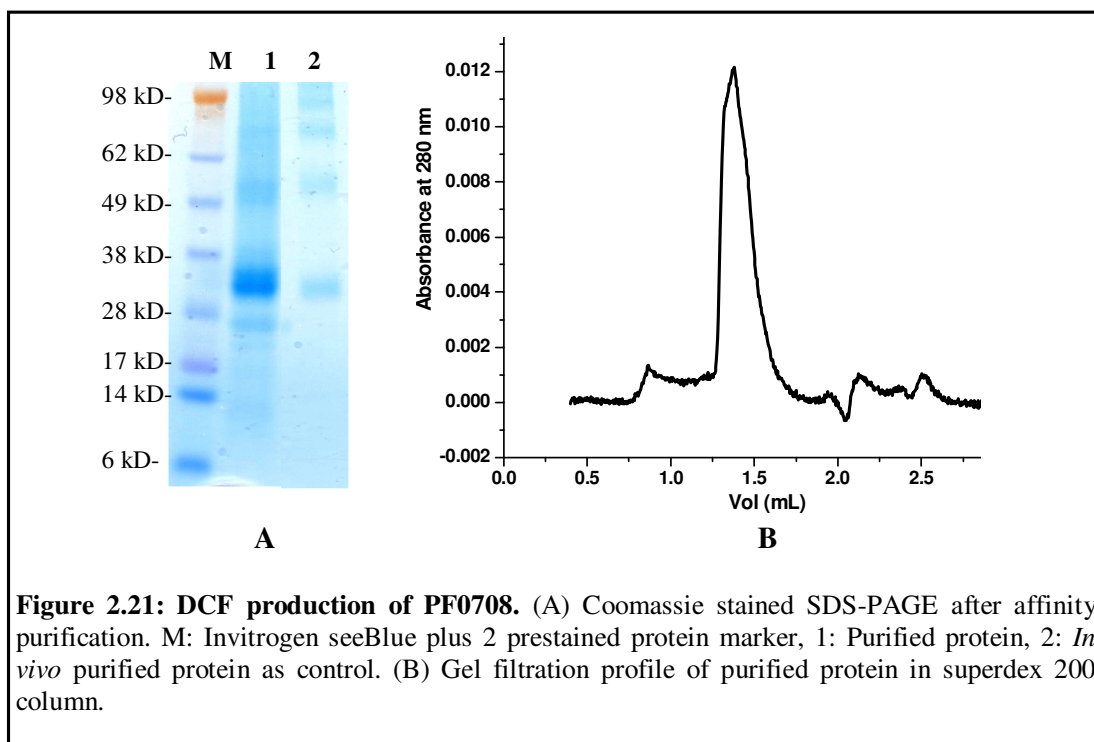


Large scale production, typically in 2 ml or its multiple, was carried out at 30 °C as described in materials and methods. The pellet was solubilized in a final concentration of 1% LMPG and after centrifugation the supernatant was applied to a 1 ml HisTrap-HP column (GE healthcare). The detergent was exchanged from LMPG to  $C_{10}E_5$  during the purification. The protein was eluted with an imidazole gradient of 0-500 mM. Coomassie stained SDS-PAGE revealed the protein to be more than 95% pure and migrating just below 38 kD. Two bands close to each other were observed in SDS-PAGE which could be caused by two differently folded states of the protein. The same sample was subjected

to blue native gel electrophoresis, which revealed two bands migrating at 242 kD and 480 kD (Fig. 2.20 C and D). The gel filtration profile after superdex 200 column showed homogeneity of the protein, even after a week of storage at 4 °C (Fig. 2.20E). The peak fractions of a gel filtration run of another batch of purification from a 10 ml reaction mix were pooled and concentrated to 2 mg/ml and crystallization trials were set.

### 2.3.1.2 DCF production of PF0708

In DCF mode, the protein is produced in the presence of detergents and recovered from the reaction mix as a detergent soluble product. Prior to the large scale production, PF0708 was screened for the most suitable detergent. Brij35 and Digitonin were the best among 5 detergents tried. The large scale production of 2 ml reaction mix was carried out in the presence of Brij35 as described in materials and methods. The reaction mix was centrifuged and the protein was purified from the supernatant using IMAC as described in PCF production. The protein was less pure than protein produced using the PCF mode production. The yield was also lower by at least 20-30%. However, gel filtration resulted in a single peak (Fig. 2.21).



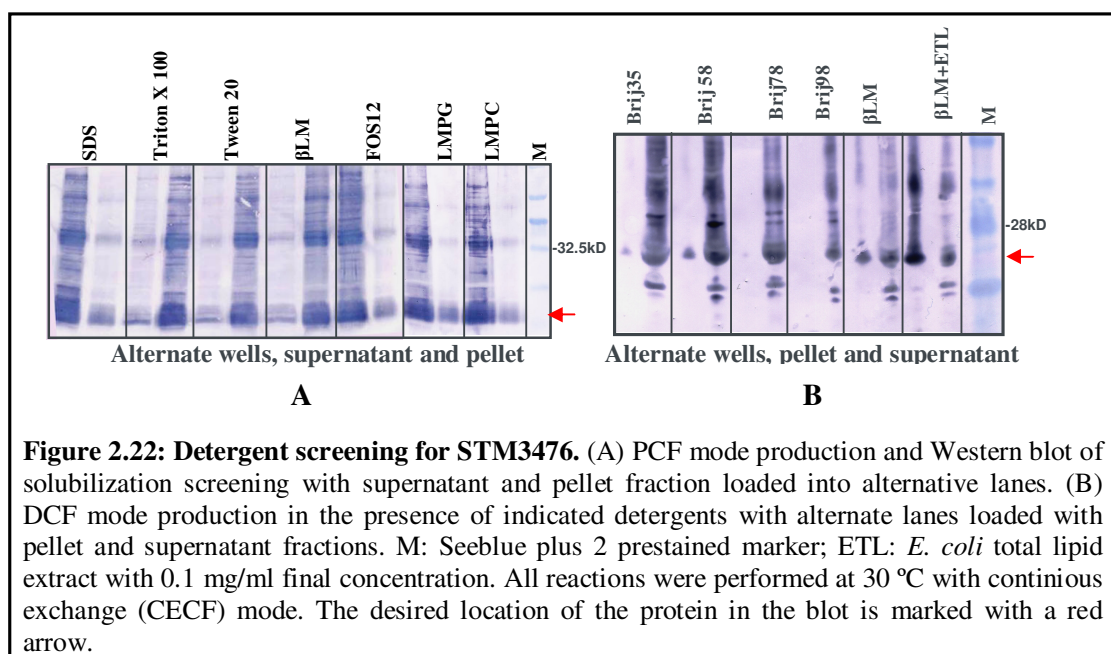
**Figure 2.21: DCF production of PF0708.** (A) Coomassie stained SDS-PAGE after affinity purification. M: Invitrogen seeBlue plus 2 prestained protein marker, 1: Purified protein, 2: *In vivo* purified protein as control. (B) Gel filtration profile of purified protein in superdex 200 column.

### 2.3.2 STM3476 of the FNT family

#### 2.3.2.1 Detergent screening in the PCF and DCF mode

STM3476 was cloned into the pET21a vector encoding for a N-terminal T7 tag and a C-terminal deca His-tag. Expression screening was carried out both in PCF and DCF mode at 30 °C. Out of 7 detergents screened in the PCF mode, LMPC, LMPG and FOS12 emerged as the preferred detergents for solubilizing STM3476 (Fig. 2.22A).

For DCF screening, the protein production was best in Brij 35, Brij58 and Brij78. Addition of *E. coli* total lipid extract (0.1 mg/ml final concentration) to  $\beta$ LM in the reaction mix could not increase the expression level as evident from the Western blot (Fig. 2.22B).



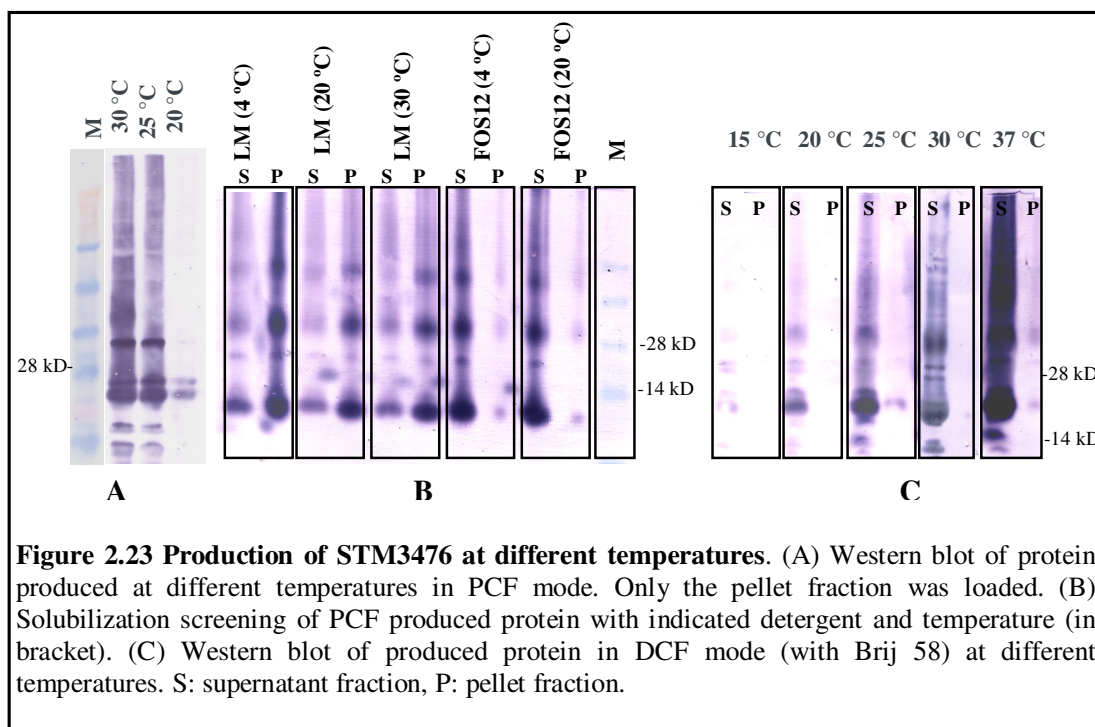
**Figure 2.22: Detergent screening for STM3476.** (A) PCF mode production and Western blot of solubilization screening with supernatant and pellet fraction loaded into alternative lanes. (B) DCF mode production in the presence of indicated detergents with alternate lanes loaded with pellet and supernatant fractions. M: Seebblue plus 2 prestained marker; ETL: *E. coli* total lipid extract with 0.1 mg/ml final concentration. All reactions were performed at 30 °C with continuous exchange (CECF) mode. The desired location of the protein in the blot is marked with a red arrow.

#### 2.3.2.2 Effect of temperature on expression and solubilization

As temperature is known to influence not only protein production but also its folding, both PCF and DCF mode production was carried out at different temperatures. Typically the production level goes down with decreasing temperature but this lowering of the temperature is often reported to enhance native folding. On the other hand in the PCF mode production, the resolubilization temperature is more important when actually the protein refolding takes place from insoluble aggregates. My trials with STM3476 revealed that the production levels go down when the temperature is lowered in both the DCF and PCF mode. In PCF mode, the difference was significant between 30 °C and 20

## RESULTS

°C but not between 30 °C and 25 °C (Fig. 2.23A). However, considering the overall low protein yield in the cell-free expression system, 30 °C was chosen as the preferred temperature for further large scale production of STM3476 using the PCF mode. When the resolubilization of PCF produced protein was carried out at 4 °C, 20 °C and 30 °C with two different detergents, a prominent effect of detergent rather than of temperature was observed. FOS12 was equally good for resolubilization at all temperatures compared to  $\beta$ LM (Fig. 2.23B). In the DCF mode, the difference was significant at a 10 °C interval with the maximal production observed at 37 °C (Fig. 2.23C).

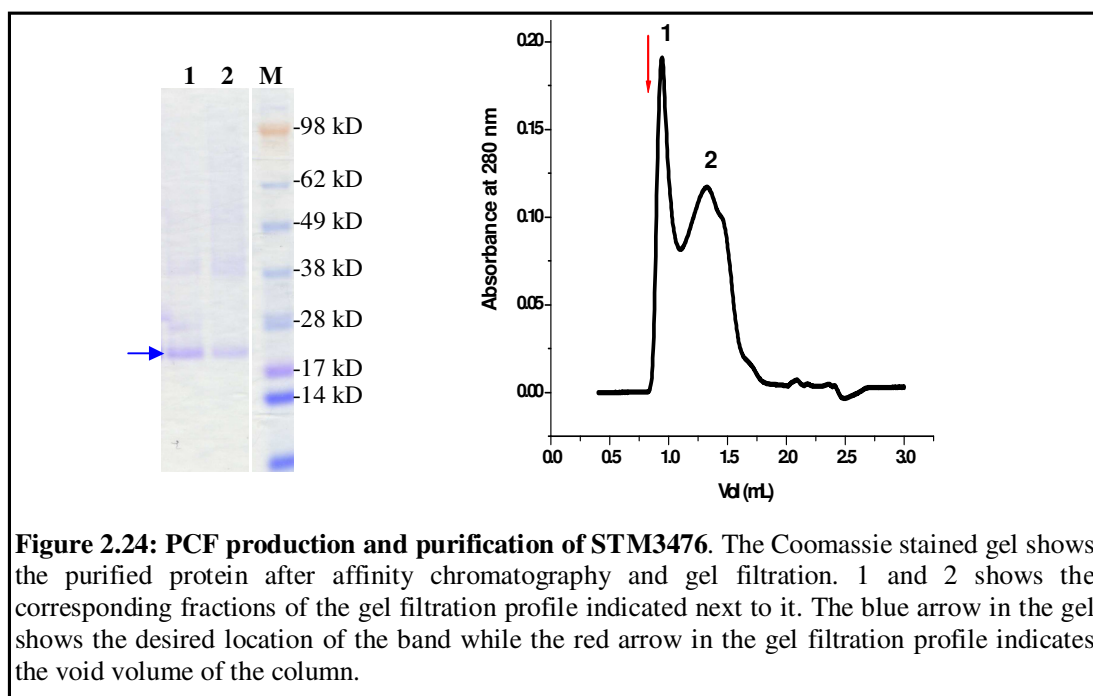


**Figure 2.23 Production of STM3476 at different temperatures.** (A) Western blot of protein produced at different temperatures in PCF mode. Only the pellet fraction was loaded. (B) Solubilization screening of PCF produced protein with indicated detergent and temperature (in bracket). (C) Western blot of produced protein in DCF mode (with Brij 58) at different temperatures. S: supernatant fraction, P: pellet fraction.

### 2.3.2.3 Large scale production and purification

Large scale production of the STM3476 protein, in 2 ml reaction mix, was carried out in the PCF mode. The pellet was washed once with washing buffer (20 mM HEPES, pH8, 150 mM NaCl, 2 mM  $\beta$ ME) and solubilized with final concentration of FOS12 at 1% in the same buffer for 1.5-2 hrs at 30 °C. After centrifugation at 15000 RPM (Eppendorf centrifuge 5417R) for 15 minutes, the supernatant was diluted 5-8 times with the same buffer but without  $\beta$ ME and with 300 mM NaCl. The supernatant was then allowed to bind to the Ni-NTA matrix for 2 hrs and the protein was eluted with 350 mM of imidazole. As depicted in Fig. 2.24, the protein migrated between 28 and 17 kD which was confirmed to be STM3476 by peptide mass fingerprinting (PMF) using mass

spectrometry (MS). The other higher bands visible in the Coomassie stained gel could be caused by different oligomeric states. MS confirmed the presence of STM3476 in the bands migrating around 38 kD and 49 kD. The purified protein showed aggregation, when concentrated. This was evident from the aggregation peak observed in the analytical gel filtration chromatography using superdex 200.



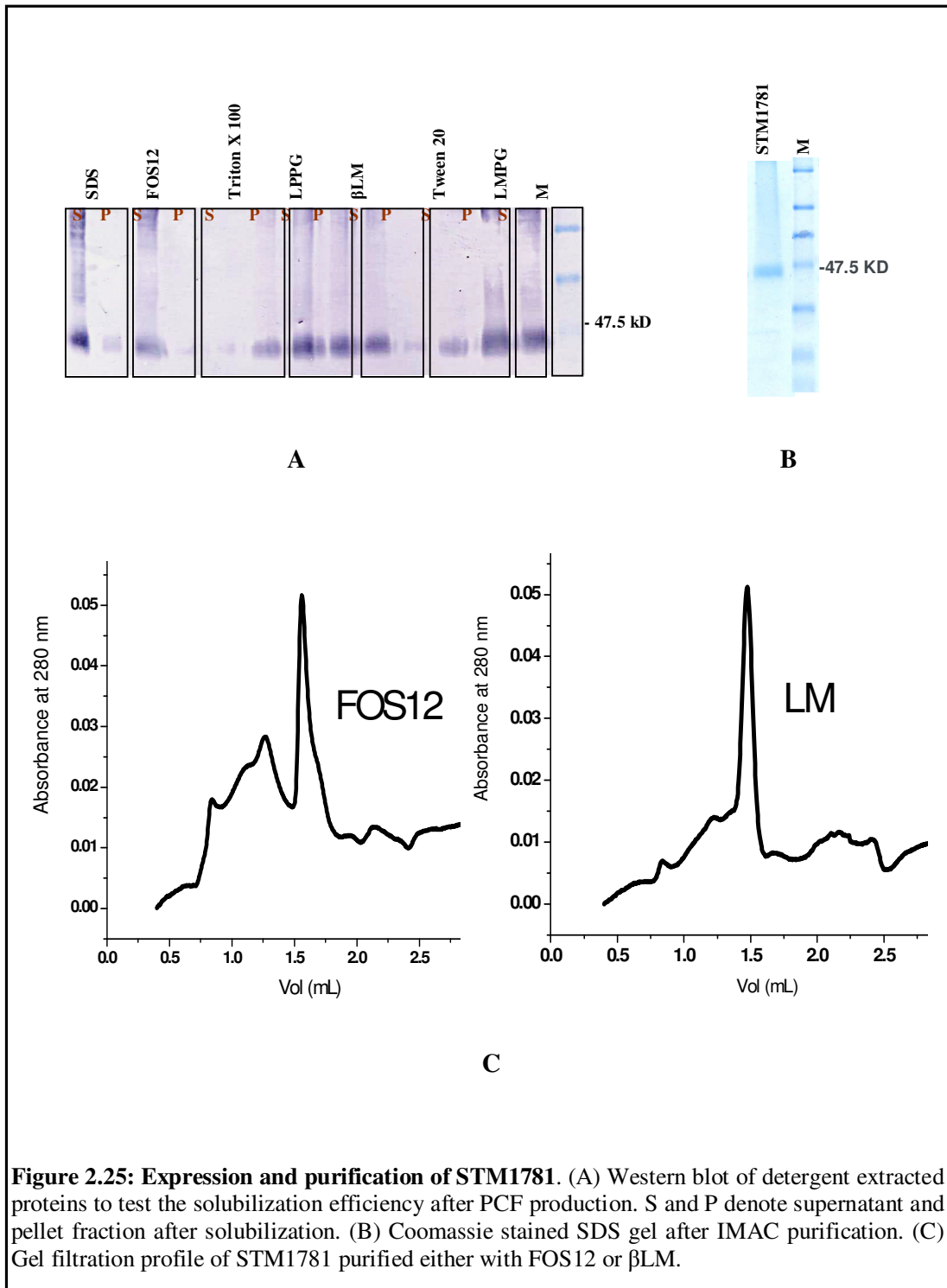
**Figure 2.24: PCF production and purification of STM3476.** The Coomassie stained gel shows the purified protein after affinity chromatography and gel filtration. 1 and 2 shows the corresponding fractions of the gel filtration profile indicated next to it. The blue arrow in the gel shows the desired location of the band while the red arrow in the gel filtration profile indicates the void volume of the column.

### 2.3.3 STM1781 of the SulP family

STM1781, belonging to the Sulphate permease (SulP) family, was expressed in PCF mode. After initial detergent screening for solubilization (Fig. 2.25A), FOS12 was chosen to resolubilize the protein in large scale PCF production. The protein was purified by Ni-NTA affinity chromatography as described earlier. The Coomassie stained SDS-PAGE showed a pure protein band (Fig. 2.25B). The final yield was below 100  $\mu\text{g}$  of purified protein from 1 ml of reaction mix as judged by Western blot. However, in SDS-PAGE the purified protein did not show any sign of degradation unlike the *in vivo* produced STM1781. The analytical gel filtration chromatography profile, which is the benchmark standard for protein homogeneity, was improved when the detergent was

## RESULTS

exchanged from FOS12 to  $\beta$ LM in the course of purification along with the addition of 1 mM sodium sulfate in the elution buffer (Fig. 2.25C ).

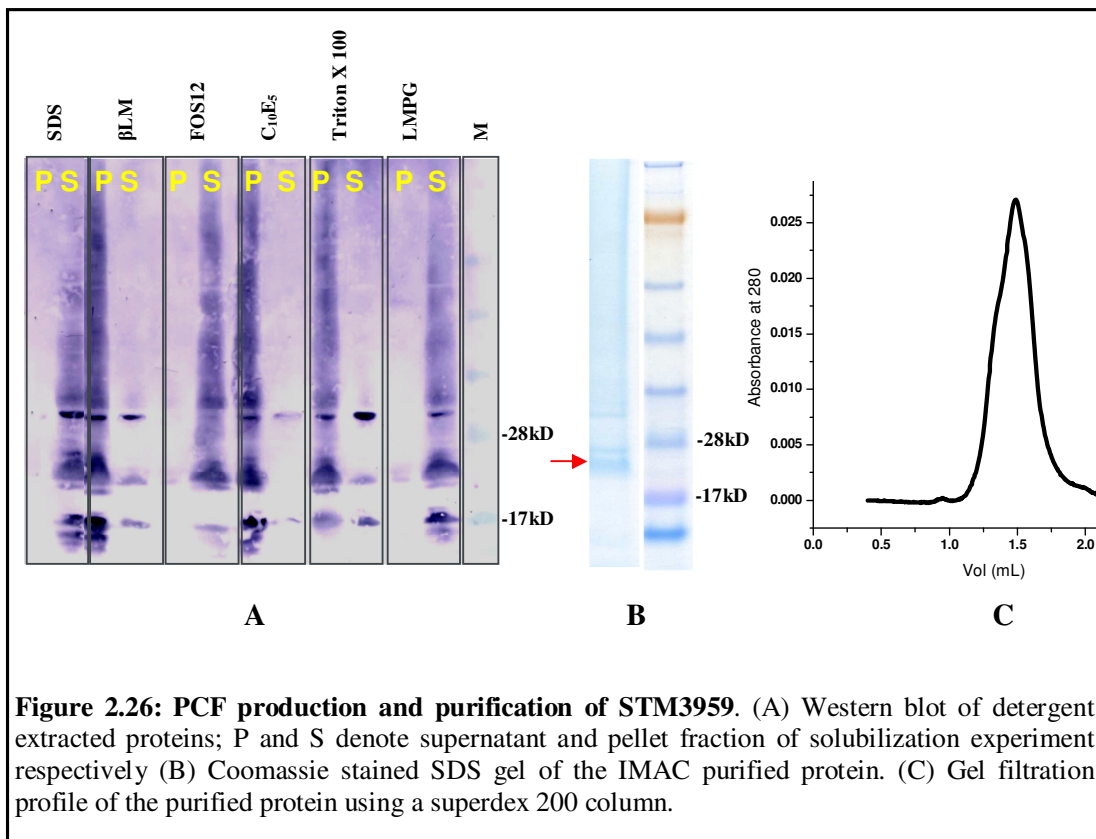




### 2.3.4 STM3959 of the RhtB family

Resistance to homoserin/threonin (RhtB) family proteins are small with 6 predicted transmembrane helices and molecular weights ranging between 22 and 23 kD. When checked for expression screening *in vivo*, they failed to express within the detection limit in our set of conditions (see appendices). I selected three members, namely, STM1270, STM3959 and STM3960 for cell-free expression. After initial screening in the continuous exchange mode, STM3959 was selected for further trials in PCF mode.

After detergent screening for solubilization (Fig. 2.26A), large scale PCF production in 2 ml reaction mix was carried out. The pellet was solubilized with LMPG and purified with  $\beta$ LM using a Ni-NTA affinity matrix. The purified protein could be observed in the Coomassie stained SDS gel migrating between 28 and 17 kD (Fig. 2.26B). In analytical gel filtration runs, the protein showed a broad peak (Fig. 2.26C).

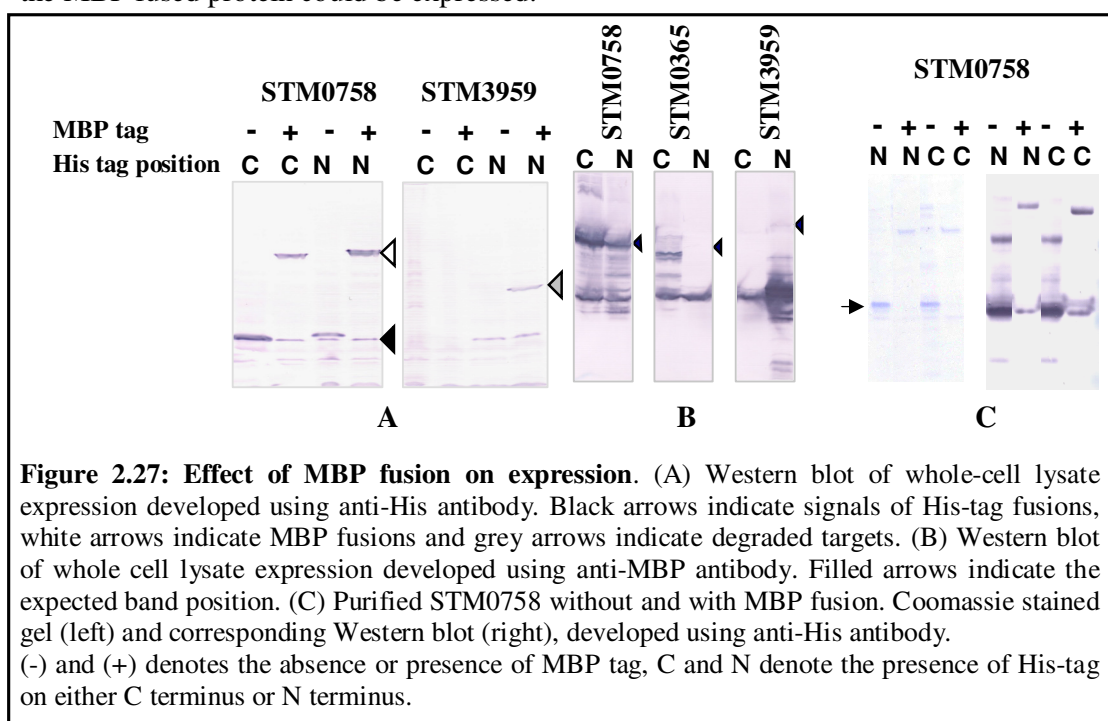


**Figure 2.26: PCF production and purification of STM3959.** (A) Western blot of detergent extracted proteins; P and S denote supernatant and pellet fraction of solubilization experiment respectively (B) Coomassie stained SDS gel of the IMAC purified protein. (C) Gel filtration profile of the purified protein using a superdex 200 column.

## 2.4 MBP fusion and effect on production

To evaluate the effect of full length MBP on the heterologous production of secondary active transporters, I fused MBP to the N-terminus of three prokaryotic membrane transporters belonging to 2 different families. In parallel, to distinguish the effect of MBP from His-tag on the production level of the target protein, both N- and C-terminal His-tag fusions were tested.

Table 2.8 provides information about the selected targets of respective families, the number of predicted transmembrane helices and the topology evaluated with several prediction programmes. The reason behind choosing the CDF family target was to check how MBP fusion affects the overall yield. On the other hand, the RhtB family transporters failed to express with the conventional strategy. So, the aim was to check if the MBP fused protein could be expressed.



The overall result of the whole cell lysate expression screening is summarized in Table 2.9 and the corresponding Western blots, developed using either anti-His antibody or anti-MBP antibody are depicted in Fig. 2.27 A and B.

Table 2.8: Selected transporter for MBP fusion, their predicted topology and experimental evidence

S N	Gene	Origin	Family	HMMTOP (TMH/N terminus)	TMHMM TMH/N terminus	TopPred (TMH/N terminus)	TMpred (TMH/N terminus)	SOSUI (TMH/N terminus)	Experimental topology (TMH/N terminus)	Reference
1	STM0758	ST	CDF	5/ in	5/ in	5/ in	5/ in	6/ -	YiiP ( <i>E. coli</i> ) 6/ In	Lu and Fu (2007)
2	STM0365	ST	RhtB	6/ out	6/ out	6/ out	6/ in	5/ -	-	-
3	STM3959	ST	RhtB	6/ out	6/ out	5/ out	5/ out	6/ -	-	-

SN: Serial number; ST: *S. typhimurium*; TMH: transmembrane helix

Table 2.9: Expression screening of whole cell lysate, determined by both anti-His and anti-MBP antibody

Gene	Family	Length (AAs)	Vector	N-term His <sub>10</sub> (anti-His)	MBP N-term His <sub>10</sub>		C-term His <sub>10</sub> (anti-His)	MBP C-term His <sub>10</sub>	
					(anti-His)	(anti-MBP)		(anti-His)	(anti-MBP)
STM0758	CDF	312	pTTQ18	+	++	+b	++	+	++b
STM0365	RhtB	210	pTTQ18	0	0	0	0	0	±b
STM3959	RhtB	206	pTTQ18	0	+ a	++ a b	0	0	0

(0) represents signal below detection limit, (+) represents 0.05 to 0.2 µg of protein, (++) represents 0.3 to 0.6 µg of protein. Protein concentrations were estimated with known amount of His-tagged protein (Positope: Invitrogen)

(a) protein mobility in SDS-PAGE was different from that of predicted molecular mass, (b) degradation of the recombinant protein was observed

Table 2.10: Summary of purification and yield of STM3959 with and without MBP fusion

Gene	Construct		Yield (mg/ 1 Lit culture)	% of purity	Ratio of decrease
	N- Term tag	C- Term tag			
STM0758	His <sub>10</sub>	Strep tag II	0.057 mg	80	
	His <sub>10</sub> - MBP	Strep tag II	0.033 mg	95	2 ( 3.4)
	-	His <sub>10</sub>	0.087 mg	70	
	MBP	His <sub>10</sub>	0.045 mg	95	1.5 (5.9)

Protein yield was measured by the Bradford assay; estimated purity (%) was judged according to the Coomassie stained SDS-PAGE gels; ratio of the increase (+) or decrease (-) of the yield has been calculated comparing the intensities of protein bands on Coomassie stained gels (in brackets: comparing Western blot signals)

## **RESULTS**

The anti-His antibody developed Western blot showed that STM0758 is slightly better produced with MBP when the His-tag was present at the N-terminus, but the effect was opposite when His-tag was present at the C-terminus. However, the production levels seem to be equal when the Western blot was developed using anti-MBP antibody. No production of STM0365 was detected when developed with an anti-His antibody, but the MBP fused STM0365 with the His-tag at the C-terminus showed signals of degradation when developed with an anti-MBP antibody. MBP fused STM3959 with N-terminal His-tag showed a signal, migrating at an altered position in anti-His developed Western blot. However, the anti-MBP developed blot showed several degraded bands with no clear signal at the desired position.

Only STM0758 was subjected to membrane preparation and protein purification from 1 liter culture volume (Fig. 2.27C). The comparative yield and purity, judged by Bradford assay and Coomassie stained gels are summarized in Table 2.10. This revealed that the actual production or integration of the protein into the membrane upon MBP fusion is lowered by 3.4 times and 5.9 times when a His-tag is present at the N-terminus and at the C-terminus, respectively.

## *Discussion*

---

### 3.1 Cation diffusion facilitator (CDF) family

#### 3.1.1 The proteins

The CDF proteins form a phylogenetically ubiquitous family of membrane transporters generally believed to play a role in the homeostasis of a wide range of divalent metal cations. CDFs are found in various membranes, including the bacterial cell membrane, the vacuolar membrane of both plant and yeast, and the Golgi apparatus of animals. All CDF proteins characterized in bacteria so far are involved in resistance to zinc and other heavy metal cations. They function typically as homodimers and use proton antiport to drive substrate translocation. The best characterized CDF proteins are YiiP and ZitB from *E. coli*; CzcD from *R. metallidurans* and ZnT3 and ZnT8 from human. Among my selected transporters, STM0758 and STM4061 are closely related to the *E. coli* homologues ZitB and YiiP respectively. STM0758 also shares significant identity (30-31%) with the human homologues. On the other hand, CDF proteins from *A. aeolicus* share significant identity with CzcD from *R. metallidurans* (Table 2.1). Phylogenetic analysis of the selected CDF proteins for this work and their well characterized homologues revealed three distinct clusters. STM0758 and Aq\_1073 are more closely related to ZitB of *E. coli*, while STM4061 and Aq\_2073 are closely related with YiiP of *E. coli*. Znt3, ZnT8 and CzcD have their own distinct locations in the tree (Fig. 2.1).

#### 3.1.2 Expression screening

There are several groups who have attempted to compare the use of different vectors, tags and expression strains in the production of integral membrane proteins. Eshaghi et al. (2005) analyzed the production of 49 *E. coli* proteins in *E. coli* using different vectors and strains and reported a success rate of 71%. However, for any single vector tested the number of targets produced was approximately 50%. Psakis et al. (2007) carried out a medium-throughput cloning and expression screening of integral membrane proteins from *Helicobacter pylori* using *E. coli* as the expression host. One-hundred sixteen *H. pylori* targets were cloned into two different vector systems and heterologously expressed in *E. coli*. Eighty-four percent of these proteins displayed medium to high expression. They could not observe any clear-cut correlation between expression levels and number

## DISCUSSION

of putative transmembrane spans, predicted functionality, and molecular mass. However, transporters of  $\leq 40$  kD and with an average of four to six transmembrane helices showed overall adequate expression, irrespective of the tag position. They also advocated that neither the location nor the length of the His-tag could solely define expression efficiency. However, targets with an N terminal His-tag displayed overall higher levels of expression. In another study, Lewinson et al. (2008) cloned 36 P-type ATPases from 11 different genomes into four different expression vectors (with T7 and arabinose promoter) having varying tag positions and analyzed the expression screening in 5 different *E. coli* expression hosts. They observed a significant influence of the affinity tag location. Some proteins expressed better with an N-terminal His-tag while others with C-terminal His-tag though at lower frequency. They also observed a better correlation between the expression level in whole-cell lysates and the expression level in the membrane fraction for proteins expressed under the control of arabinose promoter compared to proteins expressed under the control of the T7 promoter.

A similar study was also carried out in our lab by Surade et al. (2006). They presented comparative expression data for 37 secondary active transporters from *S. typhimurium*, *A. aeolicus* and *P. furiosus* in *E. coli* and *L. lactis* host cells. Four different vectors with two different tag positions were used for this study. An expression rate of approximately 50-60% was observed for any individual vector while 76% total expression was observed for all the vectors combined. The overall observations were: (i) *E. coli* is a robust production host for heterologous prokaryotic membrane proteins, (ii) there is no single *E. coli* expression host / vector system that shows overwhelming advantages over others, (iii) a combination of two or more different vectors and host strains should be preferred when feasible, (iv) screening of additional vectors is more advantageous than screening additional tag positions, (v) screening of orthologues, ie., same target from different organisms, increases the expression space significantly (20% in this study).

Therefore, in order to express my selected CDF transporters from *A. aeolicus* and *S. typhimurium*, I cloned them into pBAD, pTTQ18 and pQE vectors. In structural genomics projects, working in a highthroughput mode is a vital part of the overall strategy. In order to make the cloning step highthroughput, each of the three selected vectors (pBAD, pTTQ8 and pQE) were modified at multiple cloning sites (MCS) to have the same pair of restriction enzyme sites. Moreover, each vector has two different version based on the tag positions, A2 and C3. The A2 version has a deca His-tag at the C-

terminus whereas the C3 version has a deca His-tag at the N-terminus and a Strep tag II at the C-terminus. All 24 constructs were expressed quite well as judged by whole cell lysate screening. Aq\_2073 and STM0758 showed a slightly better expression with the His-tag at the N-terminus (Table 2.2). This is in agreement with other reports (Lewinson et al. 2008). A well expressing tag, such as a His-tag at the beginning of the ORF (N-terminus) often enhances the expression.

### 3.1.3 Functional complementation

The functional production of CDF transporters was primarily investigated *in vivo* by means of complementation. In the *E. coli* strain GG48, the zinc exporting primary pump ZntA and the CDF transporter ZitB were deleted. This manipulation renders the strain highly zinc sensitive. Grass and colleagues have used this system to demonstrate that the ZitB protein functions as a zinc ion transporter (Grass et al. 2001).

The complementation system of GG48 by additional transporters offers a great system to determine the functional production of the protein. Our results, the functional production of the STM0758 protein, demonstrate that this was properly targeted to the membrane and correctly folded. However, other proteins like Aq\_1073, Aq\_2073 and STM4061 could not complement the stressed cells. Upon multiple alignments of these four proteins (Fig. 1.4), and by drawing their phylogenetic tree (Fig. 2.1), one can see that proteins STM4061 and Aq\_2073 are closely related, which in turn are closely placed to YiiP of *E. coli*. Protein Aq\_1073 and STM0758 are closely placed to ZitB of *E. coli*. It is worth mentioning here that YiiP has been reported to transport ferrous ion *in vivo* (Grass et al. 2005). This relation indicates that the primary function of the STM4061 and Aq\_2073 proteins could be to relieve iron stress *in vivo* and not zinc stress. Such a function could explain the observed pattern. The inability of the *A. aeolicus* transporter Aq\_1073 to complement zinc transport function in *E. coli* strain GG48, could be attributed mainly to its hyperthermophilic origin or a need of specific lipid. *A. aeolicus* is known to grow at temperatures between 70 °C to 90 °C and proteins from this organism are reported to function optimally at or above 60 °C (Chintalapati et al. 2008). The different lipid environment in the *E. coli* membrane could be the other possible reason for its non-functionality (Opekarova and Tanner 2003, Hunte and Richers 2008).



### **3.1.4 Purification and stability of Aq\_2073**

The pBAD3 construct of Aq\_2073 was preferred for large scale production of the protein for crystallization purpose. The C3 version of the construct offers an optional second step of affinity purification using the strep tag, if necessary. Moreover, with dual tags the integrity of the purified protein can also be checked by Western blot.

An important step in purifying the membrane protein is the efficient extraction of the protein from the host membrane. This is achieved by solubilizing the membrane with a suitable detergent. Solubilization involves a number of intermediary states. It usually starts by destabilization of the lipid component of the membranes, a process that is accompanied by a transition of detergent binding by the membrane from a noncooperative to a cooperative interaction already below the critical micellar concentration (CMC). This leads to the formation of membrane fragments of proteins and lipids with detergent-shielded edges. In the final stage of solubilization membrane proteins are present as protomers, with the membrane inserted sectors covered by detergent. le Maire et al. (2000) have concluded that in general detergent binding as a monolayer ring, rather than a micelle, is the most probable mechanism.

Since, Aq\_2073 is from a hyperthermophilic source organism, the solubilization experiment was carried out both at 4 °C and 50 °C with a range of detergents reported to be useful for membrane protein crystallization. Detergents of the maltoside group ( $\beta$ LM and DM), LDAO and FOS12 could solubilize the protein efficiently with no significant effect of temperature (Fig. 2.4).

For purification of recombinant proteins, immobilized metal affinity chromatography (IMAC) is probably the method of choice for all structural genomics style of projects. A comparative analysis of different tags used for this purpose is documented by Lichty et al. (2005) and Waugh (2005). A single step IMAC using Ni-NTA matrix resulted into more than 90% purity of Aq\_2073. The protein showed a homogeneous peak in analytical gel filtration which is a reflection of its monodispersity (Fig. 2.5).

Due to their amphiphilic nature, membrane proteins tend to aggregate fast in unfavorable conditions, precluding crystallization. This is especially critical for membrane transporters because of their multiple membrane spans and large hydrophobic areas. Individual membrane transporter proteins demonstrate unique sensitivities for their environment, like, detergent, pH and temperature. Crystallization requires the protein to

be stable and monodisperse in solution. Previous work on membrane transporters has shown that the types of detergent and pH used are the most important parameters influencing the monodispersity of the protein (Boulter and Wang 2001; Li et al. 2001; Engel et al. 2002).

In a relevant study, Lemieux et al. (2003) observed a clear relationship between the ability of a detergent to preserve the monodispersity of glycerol-3-phosphate transporter (GlpT) and the possibility of yielding protein crystals. In this report, nine detergents that retained GlpT monodispersity produced protein crystals of various qualities. However, three detergents that were unable to preserve the monodispersity of the protein failed to yield any crystal. This clearly showed a strong correlation between the monodispersity and the crystallizability of a membrane protein in a particular detergent. Additionally, specific lattice contacts in any membrane protein crystal are made exclusively via protein-protein interactions, and a too large detergent micelle can be an obstacle for protein crystallization. Therefore, optimizing the detergent micelle size and shape by carefully choosing the right detergent or a mixture of detergents, the area available for the formation of lattice contacts can be essentially increased, thereby improving the protein crystallizability.

Aq\_2073 showed stability and monodispersity in pH 4 to pH 9. The  $\beta$ LM purified protein remained monodispersed for 15 days when kept at 4 °C at pH 8 and showed little sign of heterogeneity when kept at room temperature (Fig. 2.7). When subjected to detergent screening, Aq\_2073 exhibited monodispersity in a wide range of detergents and detergent mixtures (Table 2.3). One important observation was that the stability decreased in the order LM>UM>DM>NM and the protein completely aggregated in OG. This could be explained when we consider the detergent micelle size. Detergents with small carbon chain length have a smaller micelle size which probably is not sufficient to cover the hydrophobic surface area of the protein to keep it soluble or monodispersed. However, the protein was substantially stable in relatively harsh detergents like LDAO and FOS12. So, apart from the head group specificity, carbon chain lengths of the detergents also play an important role in membrane protein stability.

When the stability of Aq\_2073 was checked by DSC, it showed a  $T_m$  around 96 °C at physiological pH range. DSC profile also revealed that the  $T_m$  decreased with the decrease in pH (Fig. 2.8).

The stability of Aq\_2073 in such a wide range of conditions and an apparently higher  $T_m$  value in DSC could be attributed to its hyperthermophilic origin and this observation is in agreement with other reports in literature (Pfeil et al. 1997; Wassenberg et al. 1999; Razvi and Scholtz 2006).

### 3.1.5 Oligomeric state of Aq\_2073

Their existence as oligomer is a general feature of secondary transport proteins. The experimental determination of the physiologically relevant oligomeric state of a protein is technically demanding. However, several complementary techniques can be used to determine the oligomeric structure of proteins either in detergent solution or embedded in membranes. An overview of the different techniques, with their prerequisites, advantages and shortcomings are discussed in details by Veenhoff et al. 2002. Table 3.1 briefly describes the techniques I employed to determine the oligomeric states of Aq\_2073.

The retention volume of Aq\_2073 in calibrated gel filtration column was 1.33 ml which corresponds to an approximate mass of 280 kD. The molecular weight of Aq\_2073 calculated from the amino acid sequence corresponds to 43 kD and a Aq\_2073 dimer would have a molecular weight of 86 kD. However, this technique is applied to membrane protein in detergent solution. So, the contribution of bound detergent needs to be accounted for. According to Heuberger et al. (2002), for secondary transporters with multiple membrane spanning helices, the amount of detergent ( $\beta$ LM) bound is in the range of 150-180 molecules per monomer, corresponding to 70-90 kD. The amount of bound detergent is thus critical to interpret the gel filtration profile. Therefore, taking this fact into account, the observed elution volume of Aq\_2073 (Fig. 2.5B and Fig. 2.9C) would correspond to a dimer (~266 kD), rather a monomer (~133 kD) or a trimer (~400 kD). This is in agreement with the reported oligomerization state of YiiP, the CDF protein from *E. coli* (Lu and Fu 2007). In order to determine accurately the oligomeric state of a membrane protein by gel filtration chromatography, one has to run the protein in a series of detergents differing in the length of the alkyl chain (eg. 14M, 13M, LM, UM, DM, NM, OM) and eventually extrapolate the exact hydrodynamic radius of the protein molecule (Bamber et al. 2006). However, this is only possible when the protein is stable, eluting as a single narrow peak and assuming the same oligomeric state in the whole series of detergents.

**Table: 3.1 Overview of different oligomerization techniques**

Techniques	Principles and unique features	Prerequisites and drawbacks
<b>Size-exclusion chromatography</b>	<ul style="list-style-type: none"> <li>◆ In combination with a sedimentation coefficient, a reasonable estimate of the molecular mass can be obtained</li> </ul>	<ul style="list-style-type: none"> <li>◆ Amount of bound detergent/lipids needs to be determined</li> <li>◆ Dependent on protein shape</li> <li>◆ Calibration is problematic for membrane proteins</li> <li>◆ kD of the complex needs to be relatively low</li> </ul>
<b>Blue native gel electrophoresis</b>	<ul style="list-style-type: none"> <li>◆ Protein stops migrating when trapped at region of appropriate pore size</li> <li>◆ Many different samples can be analyzed simultaneously</li> <li>◆ Calibration on soluble proteins is possible when the additional mass of bound CBB is accounted for</li> </ul>	<ul style="list-style-type: none"> <li>◆ Protein solubility and stability in the presence of CBB</li> <li>◆ Non-equilibrium method</li> <li>◆ kD of the complex needs to be relatively low</li> </ul>
<b>Chemical cross-linking</b>	<ul style="list-style-type: none"> <li>◆ Simple method to demonstrate close proximity of proteins in their native state</li> </ul>	<ul style="list-style-type: none"> <li>◆ Accessibility of target sites</li> <li>◆ Modification of the protein is needed</li> <li>◆ Difficult to exclude nonspecific cross-links</li> </ul>
<b>Freeze-fracture EM</b>	<ul style="list-style-type: none"> <li>◆ Determination of the oligomeric state in the membrane</li> <li>◆ Empirically determined <math>1.4 \text{ nm}^2/\alpha</math> helix in combination with the predicted number of TM <math>\alpha</math> helices is used to estimate the oligomeric state</li> </ul>	<ul style="list-style-type: none"> <li>◆ Large excess of the protein of interest in the native membrane or incorporation of purified protein into liposomes</li> <li>◆ Secondary structure information is needed</li> <li>◆ Thickness of the metal film needs to be determined</li> </ul>

In BN-PAGE, Aq\_2073 migrated as two distinct bands, one just above 146 kD and the other between 242 kD and 480 kD (Fig. 2.9A). During blue native electrophoresis, the detergent bound to the proteins is replaced by the amphipathic Coomassie Brilliant Blue (CBB) dye. The mass of the bound CBB dye needs to be quantified as it influences the migration of the protein band. Heuberger et al. (2002) carried out a systematic study and concluded that bound CBB dye increases the molecular mass by 1.8 times for a protein with 12 transmembrane helices. Taking this conversion factor of  $\sim 1.8$ , the band migrating around 146 kD of Aq\_2073 corresponds to a dimer (154.8 kD). However there is another prominent band between 242 kD and 480 kD and a very faint band above that. These are probably higher oligomers (tetramer and above) of Aq\_2073. In glutaraldehyde crosslinking experiment, one could also observe bands corresponding to higher oligomers in addition to a clear band corresponding to a dimer (Fig. 2.9B). Whether these higher oligomeric states are a property of Aq\_2073 or due to aggregation, needs to be verified. Many different membrane proteins have been observed to exist in multiple oligomeric states. The MFS family sugar transporter LacS from *Streptococcus lactis* has been

reported to be in equilibrium between monomer and dimer (Heuberger et al. 2002), while GLUT1 and Band 3 anion exchanger have reported to exist both as dimers and tetramers (Veenhof et al. 2002).

Incorporation of Aq\_2073 into artificial liposome failed to generate prominent crystalline arrangements. Thus, the possibility of determining the oligomeric state in a more native like condition is excluded.

### 3.1.6 Substrate(s) of Aq\_2073

Members of CDF family proteins are reported to transport a range of heavy metal substrates. The *E. coli* CDF protein YiiP appears to be responsible for iron detoxification *in vivo* (Grass et al. 2005) but is capable of zinc and cadmium transport *in vitro* (Wei and Fu 2006). ZitB, the other CDF protein of *E. coli* has been reported to transport zinc and cadmium (Chao and Fu, 2004). Recently, Rahman et al. (2008) showed both by solid state NMR and metal ion uptake studies that ZitB is potentially capable of transporting not only zinc and cadmium but also nickel and copper. Various methods, like, calorimetry, equilibrium dialysis, fluorescence and NMR have all previously been used to examine substrate binding to transporters. I tried differential scanning calorimetry to check if any of the putative substrates could actually bind to Aq\_2073. If a substrate binds to the protein in question and stabilizes the native structure, it will increase the melting temperature ( $T_m$ ). On the other hand, if substrate binding destabilizes the native structure, the  $T_m$  will be lowered. Previously, Epanand et al. (1999 and 2001) found that the  $T_m$  of the glucose transporter GLUT 1 from human erythrocytes increased in the presence of glucose while it decreased in the presence of ATP. I used a His-tag free protein (removed by TEV digestion) for binding studies. The result showed that out of six metal ions (zinc, cadmium, ferrous, nickel and magnesium as negative control) tested only cadmium can significantly increase the melting temperature (by ~ 10 °C). Among the characterized CDF transporters so far, none has been reported to solely transport cadmium. For all the CDF transporters of prokaryotic origin, zinc remains the major transporting substrate. In complementation assay, Aq\_2073 failed to complement the zinc deficiency of *E. coli* strain GG48 (Fig.2.3). However, Aq\_2073, being a protein from the hyperthermophilic bacteria *A. aeolicus*, may not function optimally at 37 °C. For a similar reason, the substrate uptake assay in reconstituted liposomes with Aq\_2073

and a trapped fluorescent dye at room temperature, showed no significant transportation. So, further experiments, like isothermal titration calorimetry or reconstitution for substrate uptake assays in different lipids are needed to determine the exact substrates of Aq\_2073.

### 3.1.7 Construct design and crystallization of Aq\_2073

The protein itself is the most vital factor in protein crystallization. Flexible termini and loops in a protein often hinder the formation of high-quality crystals. After the removal of such flexible termini, the protein core tends to crystallize more readily. As a result, it is now a standard procedure for improving crystal quality to identify the protein core and redesign the protein construct accordingly. Doyle et al. (1998) had solved the structure of a truncated version of KcsA K<sup>+</sup> channel (amino acid residue 23 to 119) at 3.2 Å. Lemieux et al. (2003) reported that the crystallization of full length Glycerol-3-Phosphate transporter (GlpT) from *Escherichia coli* resulted into a shower of microcrystals. They subsequently generated several constructs and finally solved the structure of GlpT448 (amino acid residues 2-448) at 3.2 Å resolution (Huang et al. 2003). N-terminally truncated versions of FocA from *E. coli* and BetP from *Corynebacterium glutamicum* have been solved in recent years (Ressl et al. 2009; Wang et al. 2009).

Initial crystallization trials with Aq\_2073 resulted in small needle like crystals (Fig. 2.15). When the secondary structure of Aq\_2073 was analyzed using PredictProtein server (Fig. 2.11), the results indicated that C-terminal domain possesses several unstructured regions. Moreover, the C terminal domain has several lysine amino acid residues, which are least abundant in crystal contacts (Dasgupta et al. 1997). So, I made several constructs of Aq\_2073 based on homology alignment. In addition to making several C-terminally truncated versions, I also made two constructs only of the soluble C-terminal domain, CTD1 and CTD2 (Table 2.4). The atomic structure of the cytoplasmic domain of CDF proteins from *Thermus thermophilus* and *Thermotoga maritima* were solved (Cherezov et al. 2008; Higuchi et al. 2009).

Out of the 6 constructs generated, 4 were tried for purification and 3 out of them could be purified to homogeneity. The construct Aq\_2073 C1 was extensively characterized for detergent compatibility and stability.

Both Aq\_2073 and Aq\_2073 C1 were tried for 2D crystallization. 2D crystals are formed by reconstitution of purified protein into lipid bilayers and subsequent removal of detergents. So choosing the right lipid and a LPR (Lipid Protein Ratio) screen are the first step towards 2D crystallization. I tried two different lipids, *E. coli* polar extract and *E. coli* total extract. The incorporation of Aq\_2073 into liposome was only achieved with *E. coli* total lipid extract (Fig. 2.14). Other important parameters, like, ionic strength (pH), salt concentration, additives, temperature and detergents were screened. Most of the conditions resulted into protein incorporation in the liposome. Reconstitution at pH 7, LPR 0.5, 30 °C incubation for 10 days, minimal salt concentration (10-20 mM NaCl) with protein solubilized in  $\beta$ LM and lipid in OG resulted into optimum sized vesicles with crystalline patches. However, none of the condition tested resulted into a well ordered 2D crystal.

3D trials with full length and truncated Aq\_2073 resulted into some initial crystal forms (Fig. 2.15). However, extensive optimization of these conditions and trials with new conditions did not generate any optimistic results. Constructs of the C-terminal domain, CTD1 and CTD2 were also tried for 3D crystallization. But both constructs lack long term stability and in crystallization trials, 70%-80% of the drops showed amorphous precipitate. This observation indicates an inherent instability of the C-terminal domain itself.

Probably this large soluble C-terminal domain (more than 50% of the amino acids belong to the C-terminal domain according to TMHMM prediction) with inherent flexibility and abundance of lysine amino acid residues (least present in crystal contacts) made Aq\_2073 a challenging target for crystallization. Making more constructs and crystallization trials with additives that somehow make the C-terminal domain less flexible are the option one could try to obtain a well diffracting crystal of Aq\_2073.

## **3.2 Potassium uptake permease (KUP) family**

### **3.2.1 STM3880: expression, purification and stability**

In prokaryotes, accumulation of potassium takes place by multiple uptake systems. Among these, TrkA and KUP are constitutive low affinity systems while Kdp is known to be a high affinity system.

STM3880 is the only KUP protein present in *S. typhimurim*. It has 12 predicted transmembrane helices and a calculated mass of 69.3 kD. The C-terminal domain, predicted to be 182 amino acid long (TMHMM), forms a hydrophilic domain in the cytoplasm. The closest homologue of STM3880, the *E. coli* KUP protein was characterized *in vivo* by Zakharyan and Trchounian (2001). They reported that KUP is the major potassium uptake system in hyperosmotic stress at low pH (5.5). They also advocated that KUP is a proton-potassium symporter. So far, there is no report on the isolation of the protein from the membrane and subsequent characterization.

I cloned STM3880 into 3 different vectors with 2 different versions (A2 and C3) and found only pBADA2 version to express the protein (Section 2.2.1). The large scale production and IMAC purification resulted in pure protein in miligram amount from 1 L cell culture (Fig. 2.17). However, STM3880 lacked long term stability. As described in Table 2.6, several parameters were varied extensively to make the protein stable. However, only a low pH elution (Tris-phosphate, pH 6) could make the protein relatively stable. This suggests that the protein is extremely unstable when extracted from the native membrane and *in vitro* functional characterization and structural studies with STM3880 are a challenging task.

### 3.2.2 Solid supported membrane based electrophysiological studies

Recently there are several reports of membrane transporters characterized by solid supported membrane (SSM) based electrophysiology (Garcia-Celma et al. 2008 and 2009). Interestingly, the method is well established in our institute. I tried to functionally characterize STM3880 protein by SSM. Purified STM3880 was reconstituted into *E. coli* total lipid extract by the rapid dilution method. The freeze fracture images showed the reconstituted proteoliposome in Fig. 2.18. The SSM experiment was carried out at two pHs, pH 7.5 and pH 6. The activating solution contained KCl while the non-activating solution contained NaCl. Only at pH 6, a substrate concentration jump of 1 mM produced a detectable current. This indicates that the protein is probably active only at low pH in agreement with an earlier report (Zakharyan and Trchounian 2001). However, the amount of current generated was low compared to the extent of protein incorporation into the liposome. This observation could be explained by the poor stability of the purified protein. Probably, only a small fraction of the protein reconstituted into the liposome was



stably folded. Either improving the stability of the purified protein or an *in vivo* functional characterization could be alternative means to characterize STM3880.

### 3.3 Cell-free production

Initially, cell-free (CF) reactions were carried out in batch mode where all necessary components are present in a single compartment. This mode was accompanied by short lifespan (0.5-2 hrs) and consequently low yield. The main reason was rapid depletion of phosphate pool which took place even in the absence of protein synthesis. In turn, this leads to the accumulation of free phosphates, which can apparently complex with magnesium, and further inhibit protein synthesis. Later, this problem was overcome by Spirin and coworkers (1988) with the introduction of the continuous exchange cell-free reaction (CECF) system. In CECF, a compartment (RM) with high molecular weight compounds such as protein and nucleic acids is separated from the other compartment (FM) with low molecular weight precursors such as nucleotides and amino acids, by a semipermeable membrane. Thus, ensuring efficient exchange of low molecular weight products and precursors extends the reaction time to 14-16 hrs. Miligram amounts of protein production from a 1 ml reaction mixture are reported through the CECF reaction set up (Spirin et al. 1988; Shirokov et al. 2007). I used the CECF mode for all my cell-free reactions.

#### 3.3.1 Cell-free production and characterization of PF0780

PF0780 is a MATE family transporter from *P. furiosus* with 12 transmembrane helices and a calculated mass of 49.2 kD. *In vivo* produced PF0780 generated well diffracting crystals in our lab. My objective was to establish a purification protocol for cell-free produced PF0780 and if found stable, homogeneous and pure, then crystallizing the protein. By doing so, one can check the crystallizability of cell-free produced protein in addition to compare the folding and oligomerization behavior of *in vitro* and *in vivo* produced proteins. For production of PF0780, I followed two approaches: (i) producing the protein in absence of detergents as precipitates (PCF), (ii) in the presence of detergent micelles in soluble form (DCF).

### 3.3.1.1 PCF mode production and characterization

The precipitated protein obtained in the PCF mode is different from the protein isolated from inclusion bodies of cell based protein production. PCF produced proteins are readily soluble in relatively mild detergents yielding functional proteins unlike proteins in inclusion bodies which need strong denaturing detergents or chaotropic agents for solubilization. There are several reports in the literature which show functional solubilization of PCF produced proteins (Klammt et al. 2005 and 2007). PF0780 could be efficiently solubilized by LMPG, LMPC and FOS12. Relatively mild detergents like  $\beta$ LM and C<sub>10</sub>E<sub>5</sub> were less efficient in solubilizing the protein (Fig.2.20B). The LMPG solubilized protein was purified by IMAC and the detergent was exchanged to C<sub>10</sub>E<sub>5</sub> during the purification. SDS-PAGE revealed the protein to be more than 90% pure and migrated just below 38 kD which is normal for a membrane protein with a calculated mass of 49.2 kD.

Structural genomics projects deal with many IMPs which have either unknown function or putative substrates. So, establishing a functional assay is often difficult. In this situation one must employ other methods to benchmark the quality of CF produced proteins. I employed analytical gel filtration as an indicator for homogeneity, purity and stability for most of my target proteins. Typically a monodisperse protein which elutes as a single peak and stable over time is mostly correctly folded and functional.

PF0780 eluted as a single monodisperse peak and the gel filtration profile remained the same for at least 10 days when stored at 4 °C. I carried out several crystallization trials with PF0780 after gel filtration. When the protein was subjected to BN-PAGE, it migrated as two distinct bands, one at around 240 kD and the other around 480 kD (Fig. 2.2D). This indicates that the PF0780 is definitely oligomeric in nature and may exist as two different oligomeric forms. But whether this observed higher oligomeric form is a feature of PF0780 or originated due to concentration and storage, needs to be verified by other complementary experiments, like crosslinking studies or analytical ultracentrifugation. The yield of the PCF produced protein was ~ 200  $\mu$ g per ml of cell-free reaction. However, the final yield went down with every added step of purification like gel filtration and protein concentration by centrifugation.

### 3.3.1.2 Soluble production in DCF mode

Production of PF0708 as soluble form in the presence of detergent was carried out with a set of six detergents (Brij-35, Brij-58, Brij-78, Brij-98,  $\beta$ LM and digitonin) as described by Klammt et al. (2005). After choosing the suitable detergent, large scale production and purification was performed with Brij-35. The SDS-PAGE showed that the protein is less pure compared to the PCF produced protein and the final yield also went down by at least 20%-30% (Fig. 2.21).

So, I found that PCF mode production is a better option for structural and functional characterization of PF0780 considering its purity, homogeneity and stability. The final protein yield remains low for regular crystallization trials as this is still a major bottleneck for cell-free production of large membrane proteins.

### 3.3.2 Cell-free production and characterization of STM3476

STM3476 is a FNT family protein from *S. typhimurium* with 6 predicted transmembrane helices. Recent structures of FocA, a FNT family protein, showed that they are pentameric in nature and function as a channel rather than as a transporter. *In vivo* produced STM3476 initially showed functional transport of nitrite by SSM experiment in our lab (unpublished result). I established a cell-free production protocol for STM3476. Upon successful purification, I planned to carry out functional studies with reconstituted protein and crystallization if the yield is sufficient.

#### 3.3.2.1 Detergent screening in PCF and DCF mode

The precipitate mode production was followed by detergent screening to choose the right detergent for solubilization. The most efficient detergents turned out to be LMPC, LMPG and FOS12. These detergents are relatively harsh compared to mild detergents like maltosides and are known to efficiently solubilize the protein precipitates of PCF production (Klammt et al. 2005; Keller et al. 2008).

The DCF mode of production was carried out with a set of detergents (Brij-35, Brij-58, Brij-78, Brij-98,  $\beta$ LM) reported earlier to efficiently produce soluble membrane proteins. Additionally, one reaction was performed with 0.1 mg/ml (final concentration) of *E. coli* total lipid extracts solubilized in  $\beta$ LM. Specific lipid molecules are often tightly associated with membrane proteins and are known to play a significant role in functional

folding (Lee 2003). My screening results showed that Brij detergents (Brij 35, Brij58 and Brij78) were equally good in producing the detergent solubilized protein. Addition of lipids to  $\beta$ LM did not have a positive effect on yield compared to a reaction with only  $\beta$ LM. However, addition of lipids might increase the functional folding which remains to be proved once a functional assay has been established for cell-free produced STM3476.

### 3.3.2.2 Effect of temperature

Temperature is another important factor not only for protein production but also for functional folding in a coupled transcription translational cell-free system. In a relevant study, Iskakova et al. (2006) were able to increase the functional production of Green fluorescent protein (GFP) by either using a slower T7 transcriptase (mutant version) or reducing the reaction incubation temperature to 20 °C. I checked the effect of temperature on PCF and DCF mode production of STM3476 and also its effect on solubilizing the PCF precipitated protein. Both in PCF and DCF mode production, the yield went down linearly with the decreasing temperature (Fig. 2.23 A and C). On the other hand, solubilization of precipitates (PCF mode) at different temperatures (4 °C, 20 °C and 30 °C) with two different detergents revealed that the detergent head groups played a significant role rather than the temperature. FOS12 could solubilize the protein equally well at both the temperatures while  $\beta$ LM failed to do so. However, I could not comment about the amount of functionally produced protein at different temperatures because of a lack of established functional assay for STM3476.

### 3.3.2.3 Large scale PCF production, purification and characterization

Large scale production was carried out in PCF mode only. After solubilizing the protein precipitate by FOS12, the detergent was exchanged to  $\beta$ LM over the course of the purification. Coomassie stained SDS-PAGE showed the protein to be pure (Fig. 2.24A). The yield was between 400-500  $\mu$ g of protein from 1 ml reaction mix in different batches of purification. The cell-free produced and IMAC purified protein migrated as a major band between 17 kD and 28 kD. The calculated mass of the protein is 28 kD. However, careful observation revealed several other bands migrating just above 38 kD and 49 kD. These bands were also identified as STM3476 by peptide mass fingerprinting carried out in our in-house facility. As we know that FNT proteins assume higher oligomeric states,

these extra bands could be incompletely denatured or SDS resistant oligomeric states of STM3476.

The analytical gel filtration using superdex 200 showed that the protein is relatively unstable at pH 8 in  $\beta$ LM detergent (Fig. 2.24B). It eluted as a sharp peak near the void volume and another peak at around 1.32 ml. Both the peak fractions showed identical band pattern on a Coomassie stained SDS gel. Changing the pH to either higher (9) or lower (5, 6, 7) and the detergent from  $\beta$ LM to OG did not improve the elution profile markedly. A further characterization with different detergents, additives and pHs is necessary to improve the stability of cell-free produced STM3476. Several reconstitution trials by the rapid dilution methods were also carried out with purified STM3476. The success of the reconstitution attempts needs to be verified.

### 3.3.3 Production and characterization of SulP and RhtB proteins

STM1781 is a SulP family protein which showed extensive degradation when tried to produce *in vivo*. This is a relatively large protein with 11 predicted transmembrane helices and a calculated mass of 58.6 kD. On the other hand, RhtB family proteins are small with 6 predicted transmembrane helices and a calculated mass between 22-23 kD. Five members of the RhtB family from *S. typhimurium* were cloned and screened for *in vivo* production but without any success (see Appendix 1). So, proteins of these two families were ideal candidates for cell-free expression trials.

I was able to express STM1781 of SulP family in the PCF mode and could resolubilize the protein using FOS12 detergent. A large scale production and subsequent IMAC purification resulted in purified STM1781 without any degradation as evident from the Coomassie stained SDS-PAGE (Fig. 2.25B). The protein showed a better elution profile in gel filtration chromatography when purified in  $\beta$ LM rather FOS12. However, the yield of the final purified protein remained below 100  $\mu$ g per 1 mL reaction mix, making further characterization difficult and expensive. The low yield could be because of its large size and the high number of transmembrane helices in agreement with another report (Savage et al. 2007).

STM3959 of the RhtB family could also be produced by cell-free system in PCF mode. The precipitate was solubilized by FOS12 and purified by IMAC while changing the detergent from FOS12 to  $\beta$ LM. At pH 8, the protein eluted as a broad peak in gel

filtration chromatography (Fig.2.26). The purified STM3959 lacked long term stability and needs to be characterized further to improve the stability in order to perform functional and structural studies.

### **3.4 MBP fusion and its effect on production**

Several factors, such as protein length, number of transmembrane helices, source organism, function, presence and position of fusion partners (His-tag) and membrane topology of transporters have been suggested to play a role in the successful production of integral membrane proteins (Sachdev et al. 1998; Surade et al. 2006; Gordon et al. 2008).

I examined the effect of full length MBP and His-tag as fusion partners on the expression of three different secondary active transporters from two different families (Table 2.8). Out of these three, two transporters from RhtB family were not expressed as mentioned before. So, my aim was to check whether usage of a MBP fusion could result in full length proteins. But unfortunately, the RhtB family proteins STM0365 and STM3959 could not produced as MBP fusion proteins.

On the other hand, STM0758, predicted to have both N- and C- terminus inside the cytoplasm, was produced well as MBP fusion proteins. The whole cell lysate expression screening and detection by anti-His antibody (Western blot) showed that the production of STM0758 was slightly better with MBP when His-tag was present at the N-terminus, but the effect was opposite when the His-tag was present at the C-terminus. However, heterologously produced proteins are often not efficiently inserted into the membrane in *E. coli*, despite the presence of a signal peptide. Instead, they readily fold into a translocation incompetent conformation and accumulate as aggregates in the cytoplasm (Lee and Bernstein 2001; Korepanova et al. 2007). Since regular expression screening is performed on whole cell lysate, it does not distinguish between signals contributed by aggregated and properly membrane integrated proteins. So, I checked the actual expression level of membrane integrated protein by preparing membranes from 1 liter culture volume and subsequent protein purification. It was found that the actual production or integration of STM0758 into the membrane upon MBP fusion was lowered by 3.4 times and 5.9 times when a His-tag was present at N-terminus and C-terminus, respectively.

Probably the placement of MBP at the N-terminus of a protein having its ends in the cytoplasm interferes with correct insertion into the membrane and results in reduced production. This observation also supports the predicted topology of STM0758 with both ends in cytoplasm. It is interesting that other transporters with the N-terminus predicted outside showed an opposite effect (enhanced production) upon MBP fusion at the N-terminus (unpublished observations).

### **3.5 Overall discussion**

It is well recognized that structural studies of membrane proteins is very challenging. Not many proteins are well abundant in their native membrane. So, the success of membrane protein structural endeavor lies in getting sufficient amounts of heterologously produced protein. I tried several approaches, both *in vivo* and *in vitro* to address this core problem associated with membrane protein. In the following paragraphs, I want to discuss briefly several important factors to be considered when working with membrane proteins.

The most important criterion is to choose the suitable protein. But this is hardly predictable. However, one can select a target protein family and chose multiple proteins of that family from different organism. In addition, selecting multiple vectors and corresponding host cells for initial expression screening are beneficial. Together these factors are reported to increase the expression space coverage and in turn the chance of success (Surade et al. 2006). My approach to select 4 CDF targets from two different organisms proved advantageous as I found that though STM0758 was functionally active *in vivo*, Aq\_2073 of *A. aeolicus* was more suitable as a crystallization target because of its unique stability in a range of buffer conditions.

After a successful purification it is necessary to perform a stability check of the protein over time in addition to a detergent compatibility test. These initial screenings reveal the potential of that particular protein as crystallization target. Moreover, these also help to design conditions for crystallization trials.

Lipid plays an important role in membrane protein folding, stability and function. So, often it is necessary to add lipid during purification steps or to reduce the volume of washing buffer as every added washing step depletes the lipid attached to the protein. If the protein is pure and homogeneous after a single IMAC purification step, it can be directly subjected to crystallization trials after a dialysis step to remove imidazole.

## DISCUSSION

Sometimes, an extra gel filtration step to polish the protein may be detrimental as it removes lipid molecules necessary for crystallization.

Detergent is another important factor for membrane protein stability. Besides choosing the right detergent, choosing the optimum concentration is also important. While concentrating the protein, free detergent micelles also get concentrated. These concentrated micelles often produce detergent crystals in crystallization drops. So, while concentrating protein, one should consider the molecular mass of free detergent micelles and probably select the highest possible molecular weight cutoff.

Getting a well diffracting membrane protein crystal demands a lot of optimizations with all possible factors until one reaches the optimum condition. One attractive approach is to generate several truncated versions of the protein to find out the crystallization core construct and performing crystallization trials simultaneously. Many recent reports in literature on successful crystallization of membrane proteins indicate the success of this approach (Huang et al. 2003; Ressler et al. 2009; Wang et al. 2009).

I tried cell-free expression with targets from four different families and found that this method aptly complements the cell based production approach. Targets from RhtB family were not expressed *in vivo* but could be expressed and purified using CECF. STM1781 of SulP family was expressed and purified while the cell based production resulted in extensive degradation. PF0780 of MOP family was also purified to homogeneity and its stability was comparable to *in vivo* produced protein. However, I found the final yield of the cell-free produced protein as one of the major concern for structural studies. All my targets were large membrane proteins with 6-12 transmembrane helices. The observed yields of the purified proteins went down with an increase of the number of transmembrane helices. Savage et al. (2007) also reported a similar observation in a large scale study carried out with 120 membrane proteins. They also suggested that CF is a robust expression system and capable of producing more proteins than *in vivo* systems.

So, both *in vitro* and *in vivo* expression strategies that complement each other expands the overall expression space coverage and hence increases the chance of success of a structural genomics styled project.



## *Materials and Methods*

## 4.1 Materials

### 4.1.1 Chemicals

1,4-Dithiothreitol	Serva Elektrophoresis GmbH, Heidelberg
5-Bromo-4-chloro-3-indolylphosphate 4-toluidine salt	Biomol Feinchemikalien GmbH, Hamburg
Acrylamide 30% for PAGE	Carl Roth GmbH & Co. KG, Karlsruhe
Agarose, electrophoresis grade	Bethesda Research Laboratories GmbH, Neu Isenburg
Ammonium persulfate	Serva Elektrophoresis GmbH, Heidelberg
Ampicillin, sodium salt	Carl Roth GmbH & Co. KG, Karlsruhe
L-Arabinose	Sigma-Aldrich Chemie GmbH, Deisenhofen
Bacto Agar	Difco Laboratories, Detroit, Michigan, USA
Bacto Tryptone	Difco Laboratories, Detroit, Michigan, USA
Bacto Yeast Extract	Difco Laboratories, Detroit, Michigan, USA
Bovine serum albumin	Sigma-Aldrich Chemie GmbH, Deisenhofen
Bromophenol blue	Sigma-Aldrich Chemie GmbH, Deisenhofen
Charcoal	Carl Roth GmbH & Co. KG, Karlsruhe
Coomassie Brilliant Blue R-250	Serva Elektrophoresis GmbH, Heidelberg
Deoxynucleotide-5'-triphosphates, (dATP, dGTP, dTTP, dCTP)	Pharmacia Biotech, USA
Ethidium bromide	Bio-Rad Laboratories GmbH, Munich
Ethylenediamine tetraacetic acid	GERBU Biotechnik GmbH, Gaiberg
Glutaraldehyde	Sigma-Aldrich Chemie GmbH, Deisenhofen
Glycine	Serva Elektrophoresis GmbH, Heidelberg
Kanamycin	Carl Roth GmbH & Co. KG, Karlsruhe
N,N,N,'N,'-tetramethylethylenediamine	Sigma-Aldrich Chemie GmbH, Deisenhofen
Nitrobluetetrazolium, toluidine salt	Biomol Feinchemikalien GmbH, Hamburg
Other polyethylene glycols	Fluka chemie GmbH, Steinheim
Peptone	Sigma-Aldrich Chemie GmbH, Deisenhofen
Phenol	Carl Roth GmbH & Co. KG, Karlsruhe
N-(2-Hydroxyethyl)-piperazine-N'-2-ethanesulfonic acid	Carl Roth GmbH & Co. KG, Karlsruhe
Polyethyleneglycol 6000	Serva Elektrophoresis GmbH, Heidelberg
Silver nitrate	Sigma-Aldrich Chemie GmbH, Deisenhofen
Streptavidine-AP conjugate	Amersham Buchler GmbH & Co. KG, Braunschweig
Tetracycline-hydrochloride	Sigma-Aldrich Chemie GmbH, Deisenhofen
Tris-(hydroxymethyl)-aminomethane	Carl Roth GmbH & Co. KG, Karlsruhe
Polyethyleneimine	Sigma-Aldrich Chemie GmbH, Deisenhofen

### 4.1.2 Detergents

Sodium dodecylsulfate	Carl Roth GmbH & Co. KG, Karlsruhe
n-Dodecyl- $\alpha$ -D-maltoside	Glycon Biochemicals, Luckenwalde
n-Dodecyl- $\beta$ -D-maltoside	Glycon Biochemicals, Luckenwalde
n-Decyl- $\beta$ -D-maltoside	Glycon Biochemicals, Luckenwalde
n-Octyl- $\beta$ -D-glucopyranoside	Glycon Biochemicals, Luckenwalde

## **MATERIALS AND METHODS**

n-dodecylphosphocholine	Anatrace, Maumee, OH, USA
Tween® 20	Koch-Light Ltd., Haverhill, England
Dodecyl nonaoxyethylene ether	Sigma-Aldrich Chemie GmbH, Deisenhofen
1-myristoyl-2-hydroxy- <i>sn</i> -glycero-3-[phospho- <i>rac</i> -(1-glycerol)].	Avanti Polar Lipids Inc.
1-palmitoyl-2-hydroxy- <i>sn</i> -glycero-3-[phospho- <i>rac</i> -(1-glycerol)]	Avanti Polar Lipids Inc.
1-myristoyl-2-hydroxy- <i>sn</i> -glycero-3-Phosphocholine	Avanti Polar Lipids Inc.
Polyoxyethylene-(23)-lauryl-ether	Sigma-Aldrich Chemie GmbH, Deisenhofen
Polyoxyethylene-(10)-cetyl-ether	Sigma-Aldrich Chemie GmbH, Deisenhofen
Polyoxyethylene-(20)-cetyl-ether	Sigma-Aldrich Chemie GmbH, Deisenhofen
Polyoxyethylene-(2)-stearyl-ether	Sigma-Aldrich Chemie GmbH, Deisenhofen
Polyoxyethylene-(20)-stearyl-ether	Sigma-Aldrich Chemie GmbH, Deisenhofen
Polyoxyethylene-(20)-oleyl-ether	Sigma-Aldrich Chemie GmbH, Deisenhofen
Digitonin	Sigma-Aldrich Chemie GmbH, Deisenhofen

### **4.1.3 Lipids**

<i>E. coli</i> total lipid extract	Avanti Polar Lipids Inc.
<i>E. coli</i> polar lipid extract	Avanti Polar Lipids Inc.

### **4.1.4 Protease inhibitors**

Benzamidine	Sigma-Aldrich Chemie GmbH, Deisenhofen
Phenylmethylsulfonylfluoride	Carl Roth GmbH & Co. KG, Karlsruhe

### **4.1.5 Chromatographic matrices, pre-packed columns and instruments**

Hi-Trap column (1ml)	GE Healthcare, Munich
Ni-NTA matrix	Qiagen GmbH, Hilden
SP sepharose FF	GE Healthcare, Munich
Superdex 200 HR 10/30	GE Healthcare, Munich
Superdex 200 PC 3.2/30	Amersham Pharmacia Biotech, Freiburg
SMART System	Amersham Pharmacia Biotech, Freiburg
ÄKTA Prime	GE Healthcare, Munich
ÄKTA Purifier	GE Healthcare, Munich

### **4.1.6 Enzymes**

Lysozyme	AppliChem GmbH
Benzonase	Merck KGaA, Darmstadt
Restriction endonucleases	MBI Fermentas GmbH, St. Leon-Rot and New England Biolabs GmbH, Schwalbach
T4 DNA ligase	New England Biolabs GmbH, Schwalbach
T4 DNA polymerase	MBI Fermentas GmbH, St. Leon-Rot

**4.1.7 Antibodies**

Monoclonal anti-polyHistidine antibody	Sigma-Aldrich Chemie GmbH, Deisenhofen
Anti-MBP antibody	New England Biolabs GmbH, Schwalbach

**4.1.8 Kits**

Qiagen Plasmid Miniprep kit	Qiagen GmbH, Hilden
QIAquick Gel Extraction kit	Qiagen GmbH, Hilden
QIAquick PCR Purification kit	Qiagen GmbH, Hilden
BCA Protein Assay Kit	Pierce, Rockford, USA
LMW and HMW calibration kit	GE healthcare, Munich
Crystallization screening kits	Jena Biosciences GmbH, Jena, Germany, Hampton Research, Aliso Viejo, CA, USA Qiagen GmbH, Hilden

**4.1.9 Marker probes**

Prestained protein marker (broad range)	New England Biolabs GmbH, Schwalbach
SeeBlue <sup>®</sup> Plus2 Pre-Stained Standard	Invitrogen, Karlsruhe
DNA ladder	New England Biolabs GmbH, Schwalbach
Positope	Invitrogen, Karlsruhe

**4.1.10 Buffers, solutions and culture- media composition**

Agarose gel (1%)	
Agarose	1 g
TAE buffer 1X	100 ml
Ethidium bromide (10 mg/ml)	2.5 µL
AP buffer	
Tris buffer pH 9.5	100 mM
Sodium chloride	100 mM
Magnesium chloride	5 mM
DNA Loading buffer (4X)	
Tris buffer pH 7.4	50 mM
EDTA	5 mM
Glycerol	50% (v/v)
Bromophenol blue	0.05% (w/v)
LB medium	
Bacto tryptone	10 g/L
Bacto yeast extract	5 g/L
Sodium chloride	10 g/L
Transformation buffer for <i>E. coli</i>	
Calcium chloride	150 mM

## MATERIALS AND METHODS

Manganese chloride	55 mM
Potassium chloride	250 mM
PIPES Buffer pH 6.7	10 mM
Sterilised by filtration through 0.2 µm filter	
5X SDS gel running buffer	
Tris buffer pH 7.5	50 mM
Glycine	190 mM
Sodium dodecyl sulphate	0.5%
5X SDS-PAGE sample buffer	
Sodium dodecyl sulphate	10% w/v
Tris buffer pH 6.8	250 mM
Dithiothritol	500 mM
Glycerol	50% w/v
Bromophenol Blue	1 mg/ml
Transfer buffer	
Tris buffer pH 8.0	25 mM
Glycine	100 mM
Methanol	10%
10X TBST buffer	
Tris buffer pH 8.0	100 mM
Sodium chloride	1.5 M
Tween 20	0.5%
1X AP buffer	
Tris buffer pH 9.5	100 mM
Sodium chloride	100 mM
Magnesium chloride	5 mM
4X Lower Tris buffer	
Tris buffer pH 8.8	3 M
Sodium dodecyl sulphate	0.4% (w/v)
4X Upper Tris buffer	
Tris buffer pH 6.	8.3 M
Sodium dodecyl sulphate	0.4% (w/v)
Fixing solution	
Methanol	50%
Glacial acetic acid	10%
Water	q.s.
Staining solution	
Coomassie Brilliant Blue R-250	0.1%
Methanol	50%
Glacial acetic acid	10%

## MATERIALS AND METHODS

Water	q.s.
Destaining solution	
Methanol	40%
Glacial acetic acid	10%
Water	q.s.
Gel storage solution	
Glacial acetic acid	5%
Water	q.s.
Cell lysis buffer	
HEPES buffer pH 8.0	50 mM
Sodium chloride	150 mM
Benzonase	1 U/ml
Magnesium chloride	1 mM
Lysozyme	200 µg/ml
PMSF	1 mM
EDTA	1 mM
Membrane re-suspension buffer	
HEPES buffer pH 8.0	20 mM
Sodium chloride	20 mM
Other stock solutions	
Ampicillin (1000X)	100 mg/ml in water
Chloramphenicol (1000X)	34 mg/ml in ethanol
Carbenicillin (1000X)	50 mg/ml in water
APS solution	10 % (w/v) in water
Benzamidine (100x)	100 mM in water
PMSF stock solution (100X)	200 mM in isopropanol
Protogel	30% as purchased
TEMED solution	As purchased
APS solution	30% Ammonium persulphate in water
BCIP-solution	50 mg BCIP/ml in dimethylformamide
NBT solution	50 mg NBT/ml in 70% dimethylformamide

### **4.1.11 Apparatus and consumables**

#### General

Electroporation device	Biorad, Munich
I-blot	Invitrogen, Karlsruhe
Snap ID	Millipore
Shaker	Infors AG, Bottmingen, Switzerland
Spectrophotometer	Thermo Spectronic, Cambridge, UK
Vortex machine	Bender & Hobein AG, Zürich, Switzerland
Thermomixer 5436	Eppendorf GmbH, Hamburg
LiposoFast	Avestin Europe GmbH, Mannheim

**Centrifuges**

Sorvall	RC-5B Sorvall, Bad Homburg
Avanti J-20 XPI	Beckman-Coulter, Fullerton, CA, USA
Optima LE-80K	Beckman-Coulter, Fullerton, CA, USA
Tabletop ultracentrifuge	Beckman Instruments Inc., Palo Alto, USA
Sigma 3K 12	Sigma-Aldrich Chemie GmbH, Deisenhofen

**Consumables**

15 ml/ 50 ml culture tubes	Greiner bio-one
Disposable pipets	Sarstedt, Nuembrecht
Glass bottles	Schott AG
Microfuge tubes	Eppendorf GmbH, Hamburg
Pipette tips	Sarstedt, Nuembrecht
Syringe filters	Sarstedt, Nuembrecht

**4.1.12 Microorganisms**

<i>E. coli</i> DH5 $\alpha$	Invitrogen, Karlsruhe
<i>E. coli</i> NM554	Stratagene, La Jolla, USA
<i>E. coli</i> TOP10	Invitrogen, Karlsruhe
<i>E. coli</i> BL21-CodonPlus(DE3)-RP	Stratagene, La Jolla, USA

**4.1.13 Plasmids**

pTTQ18	Kind gift from Prof. Henderson, University of Leeds
pBAD	Invitrogen, Karlsruhe
pQE	Kind gift from Prof. Essen, University of Marburg
pMAL-p2X	New England Biolabs GmbH, Schwalbach
pJET1	Fermentas

**4.1.14 Components for cell-free productions**

pET 21a	Kind gift from Prof. Volker Dötsch, University of Frankfurt
<i>E. coli</i> K12 A19 strain	Kind gift from Prof. Volker Dötsch, University of Frankfurt
D tube dialyzer	Novagen

**Chemicals**

HEPES	AppliChem, Darmstadt
DTT	AppliChem, Darmstadt
Potassium acetate	AppliChem, Darmstadt
Magnesium acetate	AppliChem, Darmstadt
PMSF	AppliChem, Darmstadt
Tris-Acetate	AppliChem, Darmstadt
Polyethylene Glycol 8000	AppliChem, Darmstadt
Sodium azide	AppliChem, Darmstadt

## MATERIALS AND METHODS

Sodium chloride	AppliChem, Darmstadt
Potassium hydroxide	AppliChem, Darmstadt
Sodium hydroxide	AppliChem, Darmstadt
ATP, CTP, GTP, UTP	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Alanine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Arginine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Aspartate	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Asparagine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Cysteine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Glutamine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Glutamic acid	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Histidine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Isoleucine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Leucine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Lysine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Methionine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Proline	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Phenylalanine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Serine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Threonine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Tryptophan	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Tyrosine	Sigma-Aldrich Chemie GmbH, Deisenhofen
L-Valine	Sigma-Aldrich Chemie GmbH, Deisenhofen
Phosphoenolpyruvate	Sigma-Aldrich Chemie GmbH, Deisenhofen
Acetylphosphate	Fluka GmbH, Deisenhofen
Glycine	Fluka GmbH, Deisenhofen
5-Formyl-5,6,7,8-tetrahydrofolic acid	Fluka GmbH, Deisenhofen
Complete (Protease inhibitor)	Roche, Mannheim
E.coli tRNA	Roche, Mannheim
Pyruvate Kinase	Roche, Mannheim
RiboLock™	Fermentas



### **4.2 Methods**

#### **4.2.1 Selection of families and targets**

Target families and target proteins were selected based on the bioinformatic analysis presented at the TransportDB website ([www.membranetransport.org](http://www.membranetransport.org)). Priorities were given to families without an atomic structure

#### **4.2.2 General molecular biological techniques**

##### **4.2.2.1 DNA isolation**

Isolation of plasmid DNA from *E. coli* cells and extraction of DNA fragments from agarose gel were carried out by using suitable Qiagen kits. Protocols and recommendations provided by the manufacturer were followed.

##### **4.2.2.2 DNA restriction digestion**

Restriction digestion, incubation temperature and time were followed as recommended for the individual restriction enzyme by the manufacturers.

##### **4.2.2.3 Ligation**

Digested DNA fragments were purified by PCR purification kit by Qiagen in order to remove the protein and nucleotide impurities and to improve the ligation efficiency. Ligation of the DNA fragments was carried out in the presence of T4 DNA ligase either overnight at 16 °C or for 4 hrs at room temperature.

##### **4.2.2.4 Vector modification**

In order to be applicable for the high-throughput cloning, all of the vectors were modified. The existent multiple cloning site (MCS) was modified in such a way that all of the vectors had a comparable multiple cloning site. To allow parallel cloning of a single PCR product digested with appropriate restriction enzymes, into all vectors.

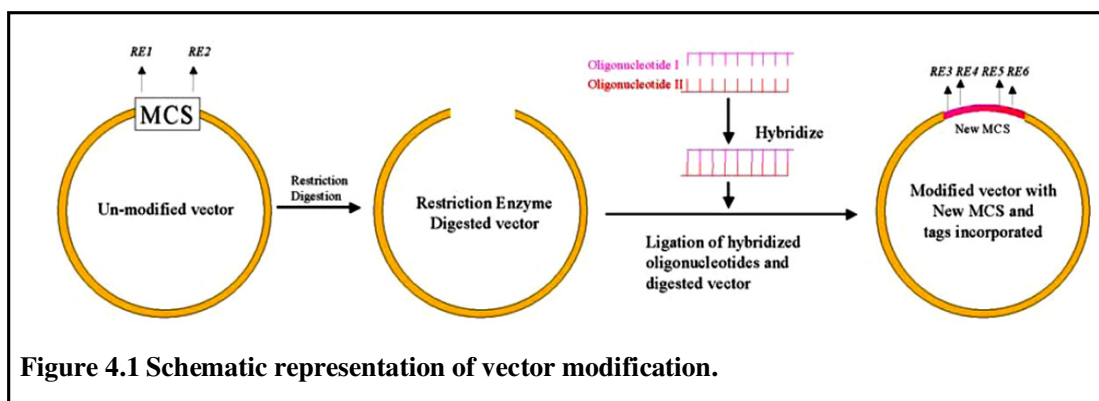
## MATERIALS AND METHODS

**Table 4.1** Details of restriction enzymes and oligonucleotides used for vector modification

Vector Version	Parent Vector	Restriction enzyme		Oligonucleotide for hybridization <sup>a</sup>		Oligonucleotide mix used for colony PCR
		I	II			
pTTQ18A2	pTTQ18	EcoRI	HindIII	pTTQ18-32	pTTQ18-33	1: pTTQ18-21, 2: pTTQ18-22
pTTQ18C3	pTTQ18	EcoRI	HindIII	pTTQ18-38	pTTQ18-39	1: pTTQ18-21, 2: pTTQ18-22
pQEA2	pQE	EcoRI	HindIII	pQ-1	pQ-2	1: pQ-5, 2: pQ-6
pQEC3	pQE	EcoRI	HindIII	pQ-3	pQ-4	1: pQ-5, 2: pQ-6
pBADA2	pBAD	NcoI	HindIII	pBAD-1	pBAD-2	1: pBAD-7, 2: pBAD-8
pBADC3	pBAD	NcoI	HindIII	pBAD-3	pBAD-4	1: pBAD-7, 2: pBAD-8

a: Please refer to the appendices for oligonucleotide sequence

Generation of pTTQ18A2 and C3 from parental pTTQ18 vector involves an intermediate pTTQ18A1 stage.



**Figure 4.1** Schematic representation of vector modification.

Generally, 5 µg vector (as obtained from the supplier or as a kind gift) was double digested with restriction enzymes I and II of table 4.1 and purified by agarose gel extraction. This digestion was performed in the absence of any alkaline phosphatase so that the 5' phosphate group remains intact which could be exploited for ligation. This step destroyed the existent multiple cloning site from the vector. In order to introduce a new cloning site, “Oligonucleotide I” and “Oligonucleotide II” were hybridised by heating a 10 µM mixture of them at 92 °C for 2 minutes. This reaction mixture was then allowed to cool down to room temperature for the purpose of annealing. This was

## **MATERIALS AND METHODS**

achieved by putting off the heating block and let it cool to room temperature. Different dilutions of hybridised oligonucleotides were made by diluting it with sterile water from 1:1 to 1:1000 parts in water. Of this diluted oligo, 1  $\mu$ l was then ligated to the purified and double digested vector (3  $\mu$ l) with T4 DNA ligase following the manufactures recommendations with proper controls. The total ligation reaction was performed in a 15  $\mu$ l volume and 3  $\mu$ l of this reaction mix was used for the transformation into chemically competent DH5 $\alpha$  cells. Positive clones were identified by isolating plasmid DNA and subsequent restriction digestion followed by DNA sequencing of the modified part of the plasmid.

### **4.2.2.5 Cloning for MBP fusion work**

The coding sequence for MBP with signal peptide (*malE* gene) was amplified by PCR from the commercially available pMAL-p2X vector (New England BioLabs), with both primers harboring *Bam*HI sites allowing 5' and 3' in-frame fusion with promoter and target proteins. The PCR product coding for MBP was first cloned into the pJET1 vector and then subcloned via *Bam*HI sites into the particular expression vectors (pTTQ18) carrying the target genes. Correct orientation and fusion with the target partner was verified by restriction analysis and sequencing.

### **4.2.3 General cell culture techniques for *E. coli***

#### **4.2.3.1 Preparation of chemically competent *E. coli* cells**

Appropriate *E. coli* strains like DH5 $\alpha$ , BL21(DE3), NM554 or TOP10 were streaked on a LB-agar plate without antibiotic and incubated at 37 °C overnight. From this plate a single colony was used for inoculating 2 ml LB medium and incubated with shaking at 200 RPM at 37 °C overnight. On next day, 100  $\mu$ l of this stationary phased *E. coli* culture was used to inoculate 100 ml LB medium containing 10 mM magnesium chloride. The flask was kept at 18 °C with shaking at 100 RPM until an OD<sub>600</sub> of 0.1 reached. Typically, it took 14-18 hrs to reach this growth phase. The culture was cooled on ice and centrifuged in a Sigma centrifuge at 5000 RPM at 4 °C for 10 minutes. The cell pellet was suspended in 33 ml of chilled transformation buffer and kept on ice for 15 minutes. This suspension was again centrifuged as before. The cell pellet was then finally suspended in 8 ml of transformation buffer and 600  $\mu$ l of DMSO was added as

cryoprotectant. The cell suspension was aliquoted and flash-frozen in liquid nitrogen before storing at  $-80\text{ }^{\circ}\text{C}$ .

### **4.2.3.2 Transformation of competent *E. coli* cells**

An aliquot of competent *E. coli* cells was thawed on ice. Typically for a single transformation 50  $\mu\text{L}$  of the competent cell suspension was transferred into a sterile and cooled microcentrifuge tube. To this fraction about 50 ng of plasmid DNA was added, the suspension was mixed and kept on ice for 10 minutes. Competent cells were heat shocked for 90 s at  $42\text{ }^{\circ}\text{C}$ . Immediately afterwards, the microcentrifuge tube was transferred to ice and kept on ice for additional 10 minutes. This suspension was directly plated onto a LB Agar plate with a suitable antibiotic to screen transformed *E. coli* cells.

### **4.2.4 Detection of protein production and protein visualization**

#### **4.2.4.1 Protein production screening in *E. coli***

For screening of gene expression in *E. coli*, only freshly transformed *E. coli* cells were used. LB medium was used as a culture medium for *E. coli*. The expression constructs were transformed into *E. coli* strains C43 (DE3) for pTTQ18 and pQE vectors, NM554 for pTTQ18 and pQE vectors, BL21 (DE3) for MBP fusion work and TOP10 for pBAD vector. Transformants were selected on LB agar plates containing 50  $\mu\text{g/ml}$  carbenicillin. For protein production, *E. coli* cells from overnight cultures of single colonies were transferred into 2 ml fresh LB medium in 24-well plates using a 1:40 dilution and were grown at  $37\text{ }^{\circ}\text{C}$  until the cultures reached an  $\text{OD}_{600}$  of 0.6. Then the cultures were induced with 0.5 mM isopropyl- $\beta$ -1-thiogalactoside (IPTG; for pTTQ18 and pQE) or 0.02% L-arabinose (for pBAD) and incubated for 3 hours. The cells from 1 ml fractions were harvested by centrifugation and finally stored at  $-20\text{ }^{\circ}\text{C}$ . This cell pellet was resuspended in 70  $\mu\text{l}$  of lysis Buffer (50 mM Tris-Cl, pH 7.9, 100 mM NaCl, 10 mM  $\text{MgSO}_4$ , 1x Complete protease inhibitor cocktail, 1 mg/ml lysozyme, 1U benzonase. After 10 minutes incubation at room temperature, 5  $\mu\text{l}$  of 10% SDS and 25  $\mu\text{l}$  of 4x SDS-PAGE loading buffer were added to the mixture to make it ready for analysis by SDS-PAGE.

### **4.2.4.2 Western blot procedure**

In order to detect recombinant proteins, produced with fusion tags, protein samples were separated by SDS-PAGE. Either home made 10 or 12% gels or commercially available pre-cast gels (Invitrogen) were used. After a proper run, protein transfer from the SDS gel to a PVDF membrane was achieved by semi-dry blotting. PVDF membranes were activated by soaking the appropriately sized membrane in methanol (1 minute) followed by thorough washing with de-ionized water (5 minutes) and soaking in transfer buffer. The transfer was achieved by applying a current of 1 mA/cm<sup>2</sup> for 1 hr. A broad range molecular weight ladder (NEB, P7708L or Invitrogen LC5925) was used as a reference. Upon successful transfer the PVDF membrane was blocked for about an hour with 1% BSA solution in TBST. All washing and incubation steps were done on a rocker at room temperature. The membrane was washed with TBST buffer 3 times each with duration of 5 minutes. The membrane was further incubated with anti-His antibody coupled to an alkaline phosphatase (1:2000 dilution with 1% BSA in TBST). For the detection of Strep tag-II, the PVDF membrane was incubated with 2 µg/ml of avidin in TBST buffer to block all the biotin, for about 15-30 minutes. Upon blocking biotin, the PVDF membrane was incubated with streptavidin, a Strep tag binding protein, coupled to alkaline phosphatase enzyme (1:2000 dilution in TBST) for about 1-2 hr. After incubation, the excess antibody solution was removed, and the PVDF membrane was washed 3 times for 5 minutes each with TBST buffer. It was followed by a wash with AP buffer 3 times for 5 minutes. The enzymatic activity of alkaline phosphatase was used to detect the recombinant protein indirectly. Alkaline phosphatase activity was detected by addition of a colour-developing reagent, 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) as recommended by the manufacturer.

Occasionally, “I blot” (Invitrogen) and “Snap ID” (miliopore) was used for transferring and blotting respectively, according to manufacturer’s protocol.

### **4.2.4.3 Coomassie staining procedure**

For the purpose of direct protein visualization after SDS-PAGE, the gels were transferred to Coomassie staining solution for about 20 minutes. Upon staining, the gel background was de-stained by putting the gels in Coomassie destaining solution until the backgrounds become clear and protein bands become clearly visible.

### **4.2.5 Protein purification**

#### **4.2.5.1 Protein production in large-scale cultures**

For isolating membranes for protein purification trials, *E. coli* cells were grown in large quantities (12-24 L). A small 5 ml LB medium was inoculated with a single colony with appropriate antibiotics and incubated in a shaker at 30 °C for about 8 hrs. This 5 ml culture was then used to inoculate 500 ml of LB medium for the preculture with proper antibiotics, and kept at 37 °C and shaken at 200 RPM overnight. This overnight culture was used as inoculum for the main expression culture. An inoculation ratio of 1:40 was used to inoculate LB medium with antibiotic for proper selection. *E. coli* cells were grown at 37 °C with shaking at 190 RPM until an OD<sub>600</sub> of 0.5-0.7 was reached. At this stage, cultures were induced with an appropriate concentration of inducer (final concentration of 0.2 to 1 mM IPTG and 0.0002 to 0.02 % L-arabinose).

#### **4.2.5.2 Membrane isolation in large scale**

For large-scale membrane isolation, cells were harvested by centrifuging the culture at 6000 RPM (with Avanti centrifuge, JLA-8.1000 rotor, Beckman) for 10 minutes. The cells were washed with a buffer containing 10 mM Tris or HEPES buffer, pH 8.0, with 150 mM sodium chloride. Harvested cells were stored at -20 °C or -80 °C for a maximum of one month. For membrane isolation, cells were suspended in the cell lysis buffer in a ratio of 1g of cells (wet weight of cells) per 10 ml of cell lysis buffer. The cell suspension was made homogeneous and filtered through a porous membrane so as to separate cell clumps from the uniformly suspended cells. The cell suspension was passed through a microfluidizer (Microfluidics Corp.) operated at 12000 PSI for 3-5 times. The lysis chamber was cooled by keeping it in ice. After lysis the cell lysate was centrifuged for 8000 RPM (Sorvall centrifuge, Sorval SLA-3000 rotor) at 4 °C for 30 minutes. The supernatant was then collected and ultracentrifuged at 43000 RPM (Beckman Coulter ultracentrifuge with 45 Ti rotor) for 1hr 30 minutes to spin down the membrane fraction. The membrane fraction was collected and suspended in the membrane resuspension buffer to a final protein concentration of 10 mg/ml.

### **4.2.5.3 Solubilization screen of membrane proteins with various detergents**

For the solubilization trials, 50  $\mu$ L of the membrane suspension buffer additionally containing 2% of different detergents ( $\beta$ -LM, DM, OG, NG, LDAO, C<sub>12</sub>E<sub>5</sub>, FOS12) were added to 50  $\mu$ L of the membrane suspension. The extraction was carried out at 4 °C for 1 hr. It was followed by another ultracentrifugation to separate solubilize from unsolubilized material. 10  $\mu$ l of the supernatant were analysed by Western blot. Solubilization of the membrane proteins for purification was carried out in the same way except that the scale of solubilization was higher. Mostly for purification  $\beta$ -LM was used as the detergent of choice unless found unsuitable.

### **4.2.5.4 Protein purification with affinity chromatography**

In order to purify proteins, the solubilize was loaded onto an Amersham HP Hi-trap 1 ml column. Usually 10 ml of solubilize were loaded onto a 1 ml column. In cases where protein production was low, higher amounts of solubilize were loaded with the help of a peristaltic pump. The unbound protein was washed off with the buffer in which solubilization had been carried out. Non-specifically bound protein was then removed by washing the column with washing buffer by applying a gradient of imidazole from 10 mM to 600 mM in a volume of 35-40 ml. The tagged protein of interest eluted as a last peak. In the batch method purification, the solubilize was allowed to bind to the Ni-NTA beads (Qiagen) manually. Bound beads were poured into an empty column and purification was done under gravity flow. After usual washing steps with 20-60 mM imidazole, proteins were finally eluted with 350 mM of imidazole. When detergent exchange was necessary during purification, the batch method was preferred over the automated method using Aekta Prime.

### **4.2.5.5 TEV protease digestion**

All our C3 version constructs have a TEV cleavage site preceding the His-tag. TEV cleavage was used to remove the His-tag. For this purpose imidazole was removed from the affinity purified protein by an additional buffer exchange step with a PD10 column. The protein was concentrated to 3 - 10 mg/ml using an Amicon ultra-15 centrifugal filter device (Millipore). The S219V mutant of TEV protease fused with a polyHis-tag was used for digestion. This mutated protease was purified and stored according to the supplier's instructions (Addgene; see web references). A protein to protease ratio of 5:1

## **MATERIALS AND METHODS**

to 50:1 was used. The digestion reaction was carried out in a buffer containing 50 mM HEPES buffer, pH 8.0, 150 mM sodium chloride or potassium chloride, 1 mM EDTA, 4 mM 2-mercapto-ethanol. The mixture was incubated overnight at 16 °C. After digestion, the reaction mixture was passed through a Ni-NTA matrix and the flow through and the wash fraction equivalent to two column volumes was collected. The bound fraction contained undigested protein, small peptides with polyHis-tag and a TEV protease, which could then be eluted with higher concentrations of imidazole. Completion of the reaction was monitored by a Western blot developed with anti-His antibody coupled to alkaline phosphatase.

### **4.2.5.6 Gel filtration chromatography**

Gel filtration chromatography was routinely used for protein preparation quality assessment as well as to “polish” the final preparations just before crystallization set-ups. Superdex-200 matrix was used for gel filtration. For analytical purposes a Superdex-200 PC 3.2/30 was used in a SMART system. For preparative purposes Superdex-200 HR 10/30 was used with an Aekta purifier or Aekta prime system. Gel filtration was carried out in a suitable buffer (HEPES pH 8.0; Tris pH 8.0; acetate buffer pH 4.0), salts (sodium chloride, potassium chloride), with detergent (1.5 – 4 times of CMC). The columns were calibrated using marker mixtures of the LMW and HMW column calibration kits (GE healthcare). Only portions representing a non-void volume fraction of the peaks were used for crystallization purposes.

### **4.2.6 Cell-free production of membrane proteins**

#### **4.2.6.1 S30 extract preparation**

The following buffers and components were prepared prior to the actual extract preparation.

#### **YPTG medium**

Yeast extract (60gm), trypton (96 gm) and NaCl (30 gm) were mixed in 4.5 L Millipore water and autoclaved in the fermenter. Later,  $\text{KH}_2\text{PO}_4$  (17.94 gm) and  $\text{K}_2\text{HPO}_4$  (54.73 gm) were dissolved in 1 L water and glucose (118.8 gm) in 500 ml water and autoclaved separately. These were then poured into the autoclaved fermenter with rest of the media, to make a final volume of 6 L.



## MATERIALS AND METHODS

### **Buffers**

S30 buffer A-----	10 mM	Tris-acetate
	14 mM	Mg(OAc) <sub>2</sub>
	60 mM	KCl
	6 mM	β-Mercaptoethanol (added freshly)
S30 bufferB-----	10 mM	Tris-Acetate
	14 mM	Mg(OAc) <sub>2</sub>
	60 mM	KCl
	1 mM	DTT (added freshly)
	0.1 mM	PMSF (added freshly)
S30 buffer C-----	10 mM	Tris-Acetate
	14 mM	Mg(OAc) <sub>2</sub>
	60 mM	KOAc
	0.5 mM	DTT (added freshly)

4M NaCl stock solution

**Strain:** *Escherichia coli* K12 A19, a kind gift from Volker Dötsch, Frankfurt University.

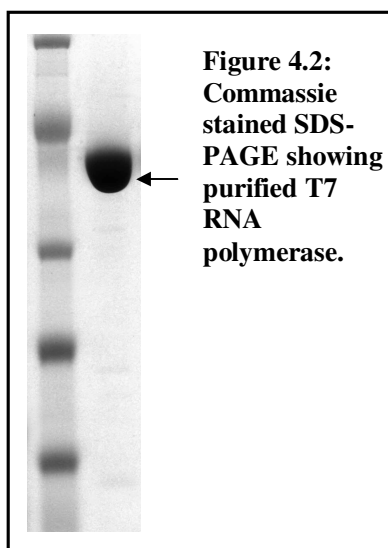
*E. coli* A19 strain was freshly streaked onto an antibiotic free plate from glycerol stock and incubated overnight at 37 °C. A single colony was inoculated into 5 ml LB media, grown for 10-12 hrs at 37 °C with shaking at 200 RPM and used as an inoculum for 60 ml LB culture. This 60 ml culture was grown overnight at 37 °C with shaking and used to inoculate the 6 L YPTG media in the fermenter (1:100). The culture in the fermenter was grown until an OD<sub>595</sub> of 4-4.5 was reached. The fermenter was then cooled down to 10 °C using ice cold water passing through a rubber tube wrapped around the fermenter for 40 minutes. All centrifugation bottles and buffers were pre-chilled to 4 °C. Cells were harvested by centrifugation for 20 minutes at 6000 RPM (Avanti centrifuge, JLA-8.1000 rotor, Beckman). The pellet was resuspended into 300 ml S30 buffer A and centrifuged twice for 10 minutes at 8000 RPM (Sorvall centrifuge, Sorvall SLA-3000 rotor). After another centrifugation at 8000 RPM using same rotor for 30 minutes, the cell pellet was weighed, resuspended in an appropriate volume of S30 buffer B (110% weight by volume) and subjected to cell breakage by French press at 20000 psi. The broken cells were centrifuged at 16000 RPM (~30,000g using a 45Ti rotor in a Beckman Coulter

## **MATERIALS AND METHODS**

ultracentrifuge) for 30 minutes and the upper 2/3<sup>rds</sup> of the supernatant was transferred to a fresh centrifuge tube and the centrifugation step was repeated. Again the upper 2/3<sup>rds</sup> of the supernatant was collected. After adjusting the final concentration of NaCl to 400 mM using a 4 M stock solution, the supernatant was incubated at 42 °C for 45 minutes. The solution turned turbid. This turbid extract was dialyzed (12-14 kD cutoff membrane) in at least 60x volume of S30 buffer C for 2-4 hrs and then overnight after a buffer change. This dialyzed extract was centrifuged at 16000 RPM (~30,000g using a 45Ti rotor in a Beckman Coulter ultracentrifuge) for 30 minutes and the upper 2/3<sup>rds</sup> of non-turbid supernatant was collected. This extract was then aliquoted, snap-frozen in liquid N<sub>2</sub> and stored at -80 °C for future use.

### **4.2.6.2 T7 RNA polymerase preparation**

pAR1219, the plasmid encoding T7 RNA polymerase, was a kind gift of Prof. Volker Dötsch, University of Frankfurt. The plasmid was freshly transformed into BL21(DE3) cells and a single colony was inoculated into 50 ml LB with ampicilin (50 µg/ml final concentration) and grown at 37 °C overnight. This culture was used to inoculate 2L LB in 1:100 dilution and grown at 37 °C with shaking. The induction was done at OD<sub>600</sub> 0.5-0.6 with 0.5 mM IPTG. The culture was grown 3 more hours after induction and the cells were harvested by centrifugation at 6000 RPM (Avanti centrifuge, JLA-8.1000 rotor, Beckman) for 15 minutes. The pellet was washed with washing buffer (50 mM Tris buffer, pH 8, 2 mM EDTA, 20 mM NaCl), centrifuged at 4000 RPM using same rotor for 20 minutes and resuspended into 24 ml of resuspension buffer (composition same as



wash buffer but with 0.25 mM PMSF and 1x Complete protease inhibitor cocktail). To this 24 ml volume, 6 ml of lysozyme (final conc. 0.3 gm/L) solution was added and incubated for 20 minutes on ice. To this, 2.5 ml of 0.8% deoxycholate was added and incubated further 20 minutes on ice. This mixture was then subjected to sonication (3x 30 seconds, 5 mm tip, 50% duty cycle and output 5) while cooling on ice. 5 ml of 2 M ammonium sulfate was added and the final volume was made up to 50 ml by resuspension buffer. Next, 5 ml of 10% Polyethyleneimine (sigma) was slowly added and

## **MATERIALS AND METHODS**

the mixture was stirred on ice for 20 minutes. The whole mixture was then centrifuged at 22000 RPM (Sorvall centrifuge, Sorval SLA-3000 rotor) for 20 minutes and the supernatant was recovered. To this supernatant, 0.82x volume of saturated ammonium sulfate solution (4.1 M) was added and the mixture was stirred on ice for 20 minutes. Afterwards, centrifugation was performed at 13400 RPM for 20 minutes using the same rotor and the pellet recovered was resuspended into 15 ml of buffer C-50 (20 mM sodium phosphate, pH 7.7, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, 1x complete protease inhibitor, 0.25 mM PMSF). This suspension was dialyzed for 3 hrs followed by 12 hr in 1 L buffer C-50 each. After dialysis, the suspension was centrifuged at 10000 RPM (Beckman Coulter ultracentrifuge with a 45 Ti rotor) for 15 minutes, the supernatant was recovered and loaded onto a SP sepharose FF column. This column of 50 ml bed volume had been manually prepared, fixed to a Aekta purifier system and pre-equilibrated with 1 column volume (CV) of C-50 buffer. After washing with 2CV of C-50 buffer, the protein was eluted with a linear gradient of 50-500 mM NaCl over 5CV. The flow rate was kept constant at 2 ml/minute. The fractions containing T7 RNA Polymerase were collected and pooled together. The pooled fractions (~25 ml) were dialyzed against 2 L of buffer C-0 (composition same as C-50 except no NaCl present) overnight with a buffer exchange after 4-6 hrs. The dialysate was centrifuged at 13000 RPM for 20 minutes, the pellet was recovered and resuspended in 4 ml of C-100 buffer (100 mM NaCl). 4 ml of 100% glycerol was added to this, rendering the final purified T7 RNA polymerase concentration to be 3.8 g/L. Aliquots of appropriate volumes were stored at -80 °C. The purity of the purified T7 RNA polymerase was analyzed by SDS-PAGE (Fig. 4.2).

### **4.2.6.3 Cell-free expression reaction mixture**

Cell-free expressions were performed mostly in a 70 µl reaction mix volume for analytical purposes and either 1 ml or its multiple (2, 4, 10 ml) for scaling up. The feeding mixture (FM) volumes were kept 17-19 times higher than the reaction mixture (RM) volume. The Mg<sup>2+</sup> and K<sup>+</sup> concentration always varied with the particular S30 extract batch and a prior titration for determination of the optimum concentration was necessary for every new S30 extract preparation. Table 4.2 describes the components and the concentration used for a typical analytical or scaled-up reaction.

## MATERIALS AND METHODS

**Table 4.2: Components and their typical concentration in FM and RM of cell-free reaction**

Components	Concentration in feeding mixture (FM)	Concentration in reaction mixture (RM)	Components only in reaction mixture (RM)	Concentration in reaction mixture (RM)
NaN <sub>3</sub>	0.10 %(W/V)	0.10 %(W/V)	Pyruvat Kinase	0.04 mg/ml
PEG 8000	3.99 %(W/V)	3.99 %(W/V)	tRNA	0.56 mg/ml
KOAc	320 mM	320 mM	RNAsin	0.4 U/μl
Mg(OAc) <sub>2</sub>	27 mM	27 mM	S30 Extrakt	0.3 to 0.4 % of total volume
HEPES pH 8	175 mM	175 mM	T7 RNA Polymerase	0.01 mg/ml
Complete	1x	1x	Plasmid (target gene)	0.015 μg/μl
Folinic acid	0.4 mM	0.4 mM		
DTT	2.5 mM	2.5 mM		
NTP mix	1x	1x		
PEP	40 mM	40 mM		
AcP	40 mM	40 mM		
Amino Acid mix	1.76 mM	1 mM		
RCWMDE mix <sup>a</sup>	1.8 mM	1.8 mM		

<sup>a</sup> RCWMDE is a mixture of Arginine, Cysteine, Tryptophan, Methionine, Aspartic acid and Glutamic acid.

In case of DCF reactions, detergents were typically used in the following concentrations: βLM (0.2-0.4%), Digitonin (0.4%), Brij-35 (0.1%), Brij-58 (1.5%), Brij-78 (1%), Brij-98 (0.2%), Triton X-100 (0.1-0.2%). However, the detergent concentration was also varied to find out the optimal expression condition.

### 4.2.6.4 Purification of cell-free produced protein

In case of DCF production, the reaction mixture was centrifuged at 14000 RPM (Eppendorf centrifuge 5417R) for 15 minutes and the supernatant was collected carefully. This supernatant was diluted 5 to 10 times with wash buffer (20 mM HEPES, pH 8, NaCl 300 mM) for purification, preferably containing purification detergent at higher CMC (typically βLM with 0.1% final concentration) and loaded onto pre-equilibrated Ni-NTA column/matrix. After an initial 10 column volume wash with this higher detergent concentration (10x CMC), further washing (20-30 CV) and elution (5 to 10 CV) were performed at 2x-3x CMC (0.02% or 0.03%). This procedure was followed to ensure maximum detergent replacement as almost always the purification detergent was different than the detergent used for DCF production.

## **MATERIALS AND METHODS**

For PCF production, the reaction mixture was centrifuged at 14000 RPM for 15 minutes and the recovered pellet was solubilized in a suitable buffer (typically 20 mM HEPES, pH 8, 150 mM NaCl, 2 mM  $\beta$ -mercaptoethanol and 1% LMPG or FOS12) for 1-2 hrs at 30 °C. Before scaling up, solubilization was always optimized for the most efficient detergent. The solubilizate was centrifuged at 14000 RPM for 15 minutes and the supernatant was collected. This supernatant was subjected to purification similarly as DCF products.

### **4.2.7 Determination of the oligomerization state**

#### **4.2.7.1 Cross-linking of proteins with glutaraldehyde**

For cross-linking experiment, glutaraldehyde was used as a chemical cross-linking reagent. Protein Aq\_2073 was purified with IMAC as discussed before. However, the buffer was chosen to be 20 mM HEPES, pH 8.0. Glutaraldehyde was added to the protein sample in concentrations of 20, 50, and 100 mM with a final protein concentrations of 0.1 mg/ml. A sample with 1% final concentration of SDS and 50 mM glutaraldehyde was also tested. Treated samples were kept at room temperature for 1 hr. At the end of the incubation period, Tris buffer was added to the samples at a final concentration of 100 mM to quench the un-reacted glutaraldehyde. Proteins samples were then run over a SDS-PAGE and proteins bands were visualized by Coomassie staining.

#### **4.2.7.2 Blue native (BN) gel electrophoresis**

BN gel electrophoresis was performed using the Invitrogen NativePAGE™ Novex Bis-Tris gel system. The NativePAGE™ Novex® Bis-Tris Gel System consists of the followings:

<b>Components</b>	<b>Details Composition</b>
NativePAGE™ Novex® Bis-Tris Mini Gels	(4-16%) gels, 10x10 cm
NativePAGE™ Sample Buffer (4X)	For 1x, BisTris buffer (50 mM), NaCl (50 mM), glycerol (10% w/v) and Ponceau S (0.001%), adjusted to pH7.2 with HCL

## **MATERIALS AND METHODS**

<b>Components</b>	<b>Details Composition</b>
NativePAGE™ 5% G-250 Sample additive	5% concentrated stock solution of Coomassie® G-250
NativePAGE™ Running Buffer (20X)	50 mM Bis Tris, 50 mM Tricin for 1x
NativePAGE™ Cathode Buffer additive (20x)	0.4% Coomassie® G-250

Using the above mentioned components, 1x running buffers were made as below

	<b>Reagents</b>	<b>Amount</b>
<b>Cathode buffer</b>	NativePAGE™ Running Buffer (20X)	10 ml
	NativePAGE™ Cathode Additive (20X)	10 ml
	Deionized Water	180 ml
	<b>Total Volume</b>	<b>200 ml</b>
<b>Anode buffer</b>	NativePAGE™ Running Buffer (20X)	50 m
	Deionized Water	950 ml
		<b>Total Volume</b>

After buffer preparation, loading samples were prepared on ice, as below.

<b>Reagent</b>	<b>Amount</b>
Protein sample	x µl
NativePAGE™ sample buffer (4X)	2.5 µl
NativePAGE™ 5% G-250 sample additive	0.25-1 µl
Deionized water to	10 µl

The gel cassette was marked outside with a marker pen to denote the wells. After setting up the gel cassettes into the running chambers, cathode and anode buffers were poured

## **MATERIALS AND METHODS**

carefully. Sample was loaded and the run was carried out at 4° C for 1hr at 150 volt and then 250 volt for 1-2 hrs. After the completion of the run, the gel was immersed into fixing solution (40% methanol and 10% acetic acid), microwaved for 45 seconds and shaken at room temperature for 15 minutes. The same procedure was repeated with destaining solution (40% methanol and 8% acetic acid) until the desired band became clearly visible.

### **4.2.8 Reconstitution of membrane protein into liposome**

#### **4.2.8.1 Preparation of the lipid stock**

Lipids were obtained from Avanti Polar Lipids as stocks in chloroform. Lipid was dried under constant flow of nitrogen for 2 hours. For 2D crystallization, dried lipids were dissolved in buffers containing 1-1.5% detergents ( $\beta$ LM, DM or OG) to a final concentration of 4 mg/ml, sonicated till homogeneity and stored in aliquots at -80° C.

For SSM studies lipid was resuspended in reconstitution buffer A (typically, 50 mM HEPES-Tris, pH 7.5, 100 mM NaCl) to a final concentration of 10 mg/ml and sonicated till a clear solution was obtained. This was then extruded through a 400nm filter using the LiposoFast liposome extruder (Avestin). The liposome diameter was checked by using dynamic light scattering. Aliquots of liposomes were stored at -80° C.

#### **4.2.8.2 Reconstitution procedure**

An aliquot of the liposome suspension (200  $\mu$ l, 10 mg/ml) was mixed with 22.5  $\mu$ l of 15% OG and the mixture was left on ice for 5-10 minutes. 0.2 mg of protein (for LPR 10) was added to the mixture and incubated on ice for 10 minutes. This lipid-detergent-protein mixture was diluted into 100 ml of pre-chilled reconstitution buffer B (typically 50 mM HEPES-Tris, pH 7.5, 300 mM NaCl) with moderate stirring and left for 10 minutes at 4 °C. The reconstituted proteoliposomes were precipitated with a Ti 60 rotor at 60000 RPM (Beckman Coulter ultracentrifuge) for 1 hr at 4 °C. The pellet was resuspended in reconstitution buffer B and centrifuged again. The obtained pellet was finally resuspended in storage buffer C (typically 50 mM HEPES-Tris, pH 7.5, 1 mM DTT or 2 mM  $\beta$  mercaptoethanol).

## **4.2.9 Functional and biophysical characterization**

### **4.2.9.1 Functional complementation of GG48 strain with CDF transporters**

GG48 is an *E. coli* strain devoid of ZntA and ZitB. ZntA is a zinc ATPase and ZitB is a secondary transporter responsible for efflux of zinc out of the cells. GG51 is a strain, which is devoid of ZitB only. It has been shown that the GG48 strain is sensitive to moderate concentrations of zinc. Chemically competent GG48 cells were transformed with the pTTQ18-C plasmid carrying genes of the CDF family. Transformation was followed by an expression screen to detect the expression of these genes under new experimental conditions (GG48 as host cells, 0.2 mM IPTG at 30 °C). First, a standard curve of zinc sensitivity to GG48 was determined by incubating GG48 with varying concentrations of zinc (0 to 200  $\mu$ M of zinc chloride in LB medium) at 30 °C with shaking at 200 RPM for several hours. In the case of transformed cells, a pre-culture was made inoculating a single colony in 2 ml of LB medium and incubating at 30 °C overnight. The next day, this culture was used to inoculate expression medium containing a varying amount of zinc chloride and 0.2 mM IPTG. Care was taken in this experiment to make an initial OD<sub>600</sub> of 0.05 and allowed the cell culture to grow at 30 °C at 200 RPM for several hours. At the end of the experiment, the sensitivity to zinc ion was determined by measuring the OD<sub>600</sub>.

### **4.2.9.2 Solid supported membrane (SSM) experiments**

The SSM measurement was performed according to Garcia-Celma et al. (2009). Briefly, 40  $\mu$ l of proteoliposomes at a protein concentration of 1 mg/ml were allowed to adsorb for 1 hr to an octadecanethiol/phosphatidylcholine hybrid bilayer on a gold surface (the sensor). The solution exchange protocol consisted in 3 phases each lasting 0.5 seconds(s). The nonactivating solution flows through the cuvette during the first and third phase (from 0 to 0.5 s, and from 1 to 1.5 s), whereas the activating solution flows during the second phase (from 0.5 to 1 s). The non-activating solution always contained 1 mM NaCl, and the activating solution contained 1 mM KCl. All solutions were buffered in 50 mM HEPES-TRIS at given pH values plus 1 mM DTT (Table 2.7 of result section). Currents were recorded throughout the entire time, and amplified with a current amplifier set to a gain of  $10^9$ - $10^{10}$  V/A and low pass filtering set to 300–1,000 Hz.



### **4.2.9.3 Differential scanning calorimetry**

The DSC measurements were performed using a VP-Capillary DSC system (Microcal, Inc., Northampton, MA) equipped with tantalum 61 cells, each with an active volume of 137  $\mu$ l. Protein samples were diluted to 0.5-1 mg/ml while the corresponding buffer was used as a reference. In most cases, the protein samples were dialyzed in the buffer prior to the experiment and this same buffer was used as a reference buffer. Both the sample and buffer were filtered routinely using a 20  $\mu$ m filter before the injection. The samples were scanned at a scan rate of 90  $^{\circ}$ C/hr unless otherwise stated, with an initial 15 minutes of equilibration at the starting temperature. Data were analyzed using Origin 7.0 software (OriginLab<sup>®</sup> Corporation, Northampton, MA). Thermograms were corrected by subtraction of buffer-only blank scans. The corrected thermograms were normalized for protein concentration.

### **4.2.10 2D crystallization, negative staining and screening by EM**

Protein typically at 1 mg/ml final concentration was mixed at the desired lipid to protein ratio (LPR) with lipid, incubated at room temperature for 1hr before being transferred to dialysis bags for two dimensional crystallization. Protein concentration or the pre incubation time before dialysis was varied. The temperature and time period for dialysis were 30  $^{\circ}$ C and 10 days but these criteria also varied with different detergents.

2  $\mu$ l of the dialyzed sample was placed on a carbon coated copper grid (400 mesh size). The sample was allowed to adsorb for a minute. The grid was stained by drops of 1.5% uranyl acetate. After soaking the excess solution, another drop of uranyl acetate was added on the grid. This ensured that any ingredients (phosphate, glycerol) in the sample buffer were removed in the first drops, resulting in better staining. The grids were screened in a Philips CM120 microscope equipped with a LaB<sub>6</sub> filament and operated at 120KV. Images were recorded in low dose mode typically at a magnification of 45,000 x with an exposure resulting in 20 electrons per  $\text{Å}^2$ .

### **4.2.11 3D crystallization**

Three dimensional crystallization trials were done with IMAC and gel filtration purified proteins. Drops of varying sizes (100 nl, 200 nl, 300 nl, 400 nl, 1  $\mu$ l) were set either by

## **MATERIALS AND METHODS**

using Cartesian, Honey bee (Zinsser Analytic), Mosquito (TTP Labtech), Crystalmat (Rigaku) or manually. Different commercial crystallization screens (Hampton, Jena Bioscience, Qiagen etc.) were used along with lab-made buffers to generate or improve crystals. Hanging drops or sitting drops were mostly used for crystallization. Protein to buffer ratio was mostly 1:1 though other ratios were also tried. Protein concentrations were varied from 2 mg/ml to 10 mg/ml.

Crystallization drops were inspected using a Leica M165 C stereomicroscope and image capture system.

---

## *References*

---

**References**

- Ahn, J.H., Chu, H.S., Kim, T.W., Oh, I.S., Choi, C.Y., Hahn, G.H., Park, C.G., and Kim, D.M. 2005. Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochem Biophys Res Commun* **338**: 1346-1352.
- Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J.D. 1994. *Molecular biology of the cell*. (Third edition). Garland publishing Inc. New York and London.
- Aleshin, V.V., Zakataeva, N.P., and Livshits, V.A. 1999. A new family of amino-acid-efflux proteins. *Trends Biochem Sci* **24**: 133-135.
- Andre, N., Cherouati, N., Prual, C., Steffan, T., Zeder-Lutz, G., Magnin, T., Pattus, F., Michel, H., Wagner, R., and Reinhart, C. 2006. Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. *Protein Sci* **15**: 1115-1126.
- Anton, A., Grosse, C., Reissmann, J., Pribyl, T., and Nies, D.H. 1999. CzcD is a heavy metal ion transporter involved in regulation of heavy metal resistance in *Ralstonia* sp. strain CH34. *J Bacteriol* **181**: 6876-6881.
- Anton, A., Weltrowski, A., Haney, C.J., Franke, S., Grass, G., Rensing, C., and Nies, D.H. 2004. Characteristics of zinc transport by two bacterial cation diffusion facilitators from *Ralstonia metallidurans* CH34 and *Escherichia coli*. *J Bacteriol* **186**: 7499-7507.
- Arechaga, I., Miroux, B., Runswick, M.J., and Walker, J.E. 2003. Over-expression of *Escherichia coli* F1F(o)-ATPase subunit a is inhibited by instability of the uncB gene transcript. *FEBS Lett* **547**: 97-100.
- Auer, M., Kim, M.J., Lemieux, M.J., Villa, A., Song, J., Li, X.D., and Wang, D.N. 2001. High-yield expression and functional analysis of *Escherichia coli* glycerol-3-phosphate transporter. *Biochemistry* **40**: 6628-6635.
- Bamber, L., Harding, M., Butler, P.J., and Kunji, E.R. 2006. Yeast mitochondrial ADP/ATP carriers are monomeric in detergents. *Proc Natl Acad Sci U S A* **103**: 16224-16229.
- Beckham, K.S., Potter, J.A., and Unkles, S.E. 2009. Formate-nitrite transporters: Optimisation of expression, purification and analysis of prokaryotic and eukaryotic representatives. *Protein Expr Purif* **71**: 184-189
- Berrier, C., Park, K.H., Abes, S., Bibonne, A., Betton, J.M., and Ghazi, A. 2004. Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent. *Biochemistry* **43**: 12585-12591.

## REFERENCES

- Bogdanov, M., Sun, J., Kaback, H.R., and Dowhan, W. 1996. A phospholipid acts as a chaperone in assembly of a membrane transport protein. *J Biol Chem* **271**: 11615-11618.
- Bogdanov, M., Umeda, M., and Dowhan, W. 1999. Phospholipid-assisted refolding of an integral membrane protein. Minimum structural features for phosphatidylethanolamine to act as a molecular chaperone. *J Biol Chem* **274**: 12339-12345.
- Bonander, N., Hedfalk, K., Larsson, C., Mostad, P., Chang, C., Gustafsson, L., and Bill, R.M. 2005. Design of improved membrane protein production experiments: quantitation of the host response. *Protein Sci* **14**: 1729-1740.
- Boulter, J.M., and Wang, D.N. 2001. Purification and characterization of human erythrocyte glucose transporter in decylmaltoside detergent solution. *Protein Expr Purif* **22**: 337-348.
- Bowie, J.U. 2005. Solving the membrane protein folding problem. *Nature* **438**: 581-589.
- Broeer, S., and Wagner, C.A. 2003. Introduction to membrane transport. In *Membrane Transporter Diseases*. (eds. B. Stefan, and W.C. A.). Kluwer Academic/Plenum Publishers, New York.
- Bruylants, G., Wouters, J., and Michaux, C. 2005. Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design. *Curr Med Chem* **12**: 2011-2020.
- Byrne, R., Levin, J.G., Bladen, H.A., and Nirenberg, M.W. 1964. The in Vitro Formation of a DNA-Ribosome Complex. *Proc Natl Acad Sci U S A* **52**: 140-148.
- Cabanne, C., Pezzini, J., Joucla, G., Hocquellet, A., Barbot, C., Garbay, B., and Santarelli, X. 2009. Efficient purification of recombinant proteins fused to maltose-binding protein by mixed-mode chromatography. *J Chromatogr A* **1216**: 4451-4456.
- Calhoun, K.A., and Swartz, J.R. 2005. An economical method for cell-free protein synthesis using glucose and nucleoside monophosphates. *Biotechnol Prog* **21**: 1146-1153.
- Chang, G., and Roth, C.B. 2001. Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293**: 1793-1800.
- Chao, Y., and Fu, D. 2004. Kinetic study of the antiport mechanism of an Escherichia coli zinc transporter, ZitB. *J Biol Chem* **279**: 12043-12050.
- Chen, Y., Song, J., Sui, S.F., and Wang, D.N. 2003. DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in Escherichia coli. *Protein Expr Purif* **32**: 221-231.

## REFERENCES

- Cherezov, V., Hofer, N., Szebenyi, D.M., Kolaj, O., Wall, J.G., Gillilan, R., Srinivasan, V., Jaroniec, C.P., and Caffrey, M. 2008. Insights into the mode of action of a putative zinc transporter CzcB in *Thermus thermophilus*. *Structure* **16**: 1378-1388.
- Chintalapati, S., Al Kurdi, R., van Scheltinga, A.C., and Kuhlbrandt, W. 2008. Membrane structure of CtrA3, a copper-transporting P-type-ATPase from *Aquifex aeolicus*. *J Mol Biol* **378**: 581-595.
- Chumpolkulwong, N., Sakamoto, K., Hayashi, A., Iraha, F., Shinya, N., Matsuda, N., Kiga, D., Urushibata, A., Shirouzu, M., Oki, K., et al. 2006. Translation of 'rare' codons in a cell-free protein synthesis system from *Escherichia coli*. *J Struct Funct Genomics* **7**: 31-36.
- Clegg, S., Yu, F., Griffiths, L., and Cole, J.A. 2002. The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli* K-12: two nitrate and three nitrite transporters. *Mol Microbiol* **44**: 143-155.
- Daley, D.O., Rapp, M., Granseth, E., Melen, K., Drew, D., and von Heijne, G. 2005. Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308**: 1321-1323.
- Dasgupta, S., Iyer, G.H., Bryant, S.H., Lawrence, C.E., and Bell, J.A. 1997. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins* **28**: 494-514.
- Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. 1985. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Å resolution. *Nature* **318**: 618-624.
- DeVries, J.K., and Zubay, G. 1967. DNA-directed peptide synthesis. II. The synthesis of the alpha-fragment of the enzyme beta-galactosidase. *Proc Natl Acad Sci U S A* **57**: 1010-1012.
- Dobrovetsky, E., Lu, M.L., Andorn-Broza, R., Khutoreskaya, G., Bray, J.E., Savchenko, A., Arrowsmith, C.H., Edwards, A.M., and Koth, C.M. 2005. High-throughput production of prokaryotic membrane proteins. *J Struct Funct Genomics* **6**: 33-50.
- Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., and MacKinnon, R. 1998. The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **280**: 69-77.
- Drew, D., Lerch, M., Kunji, E., Slotboom, D.J., and de Gier, J.W. 2006. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* **3**: 303-313.
- Dutzler, R., Campbell, E.B., and MacKinnon, R. 2003. Gating the selectivity filter in Cl<sup>-</sup> chloride channels. *Science* **300**: 108-112.

## REFERENCES

- Engel, C.K., Chen, L., and Prive, G.G. 2002. Stability of the lactose permease in detergent solutions. *Biochim Biophys Acta* **1564**: 47-56.
- Eband, R.F., Eband, R.M., and Jung, C.Y. 1999. Glucose-induced thermal stabilization of the native conformation of GLUT 1. *Biochemistry* **38**: 454-458.
- Eband, R.F., Eband, R.M., and Jung, C.Y. 2001. Ligand-modulation of the stability of the glucose transporter GLUT 1. *Protein Sci* **10**: 1363-1369.
- Eroglu, C., Cronet, P., Panneels, V., Beaufils, P., and Sinning, I. 2002. Functional reconstitution of purified metabotropic glutamate receptor expressed in the fly eye. *EMBO Rep* **3**: 491-496.
- Eshaghi, S., Hedren, M., Nasser, M.I., Hammarberg, T., Thornell, A., and Nordlund, P. 2005. An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. *Protein Sci* **14**: 676-683.
- Esposito, D., and Chatterjee, D.K. 2006. Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol* **17**: 353-358.
- Falke, D., Schulz, K., Doberenz, C., Beyer, L., Lilie, H., Thiemer, B., and Sawers, R.G. 2009. Unexpected oligomeric structure of the FocA formate channel of *Escherichia coli*: a paradigm for the formate-nitrite transporter family of integral membrane proteins. *FEMS Microbiol Lett* **303**: 69-75
- Franke, I., Resch, A., Dassler, T., Maier, T., and Bock, A. 2003. YfiK from *Escherichia coli* promotes export of O-acetylserine and cysteine. *J Bacteriol* **185**: 1161-1166.
- Garcia-Celma, J.J., Dueck, B., Stein, M., Schlueter, M., Meyer-Lipp, K., Leblanc, G., and Fendler, K. 2008. Rapid activation of the melibiose permease MelB immobilized on a solid-supported membrane. *Langmuir* **24**: 8119-8126.
- Garcia-Celma, J.J., Smirnova, I.N., Kaback, H.R., and Fendler, K. 2009. Electrophysiological characterization of LacY. *Proc Natl Acad Sci U S A* **106**: 7373-7378.
- Goder, V., Junne, T., and Spiess, M. 2004. Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Mol Biol Cell* **15**: 1470-1478.
- Goeddel, D.V. 1990. Systems for heterologous gene expression. *Methods Enzymol* **185**: 3-7.
- 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.
- Gordon, E., Horsefield, R., Swarts, H.G., de Pont, J.J., Neutze, R., and Snijder, A. 2008. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. *Protein Expr Purif* **62**: 1-8.

## REFERENCES

- Grass, G., Fan, B., Rosen, B.P., Franke, S., Nies, D.H., and Rensing, C. 2001. ZitB (YbgR), a member of the cation diffusion facilitator family, is an additional zinc transporter in *Escherichia coli*. *J Bacteriol* **183**: 4664-4667.
- Grass, G., Otto, M., Fricke, B., Haney, C.J., Rensing, C., Nies, D.H., and Munkelt, D. 2005. FieF (YiiP) from *Escherichia coli* mediates decreased cellular accumulation of iron and relieves iron stress. *Arch Microbiol* **183**: 9-18.
- Griffith, D.A., Delipala, C., Leadsham, J., Jarvis, S.M., and Oesterhelt, D. 2003. A novel yeast expression system for the overproduction of quality-controlled membrane proteins. *FEBS Lett* **553**: 45-50.
- Grisshammer, R., Grunwald, T., and Sohal, A.K. 2002. Characterization of an antibody Fv fragment that binds to the human, but not to the rat neurotensin receptor NTS-1. *Protein Expr Purif* **24**: 505-512.
- Grisshammer, R., Little, J., and Aharony, D. 1994. Expression of rat NK-2 (neurokinin A) receptor in *E. coli*. *Receptors Channels* **2**: 295-302.
- Guan, L., Smirnova, I.N., Verner, G., Nagamori, S., and Kaback, H.R. 2006. Manipulating phospholipids for crystallization of a membrane transport protein. *Proc Natl Acad Sci U S A* **103**: 1723-1726.
- Guffanti, A.A., Wei, Y., Rood, S.V., and Krulwich, T.A. 2002. An antiport mechanism for a member of the cation diffusion facilitator family: divalent cations efflux in exchange for K<sup>+</sup> and H<sup>+</sup>. *Mol Microbiol* **45**: 145-153.
- Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**: 4121-4130.
- Halsey, J.F., Mountcastle, D.B., Takeguchi, C.A., Biltonen, R.L., and Lindenmayer, G.E. 1977. Detection of a ouabain-induced structural change in the sodium, potassium-adenosine triphosphatase. *Biochemistry* **16**: 432-435.
- He, G.X., Kuroda, T., Mima, T., Morita, Y., Mizushima, T., and Tsuchiya, T. 2004. An H(+)-coupled multidrug efflux pump, PmpM, a member of the MATE family of transporters, from *Pseudomonas aeruginosa*. *J Bacteriol* **186**: 262-265.
- Heuberger, E.H., Veenhoff, L.M., Duurkens, R.H., Friesen, R.H., and Poolman, B. 2002. Oligomeric state of membrane transport proteins analyzed with blue native electrophoresis and analytical ultracentrifugation. *J Mol Biol* **317**: 591-600.
- Higuchi, T., Hattori, M., Tanaka, Y., Ishitani, R., and Nureki, O. 2009. Crystal structure of the cytosolic domain of the cation diffusion facilitator family protein. *Proteins* **76**: 768-771.



## REFERENCES

- Huang, Y., Lemieux, M.J., Song, J., Auer, M., and Wang, D.N. 2003. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301**: 616-620.
- Hunte, C., and Michel, H. 2002. Crystallisation of membrane proteins mediated by antibody fragments. *Curr Opin Struct Biol* **12**: 503-508.
- Hunte, C., and Richers, S. 2008. Lipids and membrane protein structures. *Curr Opin Struct Biol* **18**: 406-411.
- Ishihara, G., Goto, M., Saeki, M., Ito, K., Hori, T., Kigawa, T., Shirouzu, M., and Yokoyama, S. 2005. Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. *Protein Expr Purif* **41**: 27-37.
- Iskakova, M.B., Szaflarski, W., Dreyfus, M., Remme, J., and Nierhaus, K.H. 2006. Troubleshooting coupled in vitro transcription-translation system derived from *Escherichia coli* cells: synthesis of high-yield fully active proteins. *Nucleic Acids Res* **34**: e135.
- Jewett, M.C., and Swartz, J.R. 2004. Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol Bioeng* **86**: 19-26.
- Johansson, L.C., Wohri, A.B., Katona, G., Engstrom, S., and Neutze, R. 2009. Membrane protein crystallization from lipidic phases. *Curr Opin Struct Biol* **19**: 372-378.
- Kanaya, S., Yamada, Y., Kudo, Y., and Ikemura, T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* **238**: 143-155.
- Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. 2001. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**: 993-1000.
- Keller, T., Schwarz, D., Bernhard, F., Dotsch, V., Hunte, C., Gorboulev, V., and Koepsell, H. 2008. Cell free expression and functional reconstitution of eukaryotic drug transporters. *Biochemistry* **47**: 4552-4564.
- Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. 1999. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* **442**: 15-19.
- Kihara, A., Akiyama, Y., and Ito, K. 1995. FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proc Natl Acad Sci U S A* **92**: 4532-4536.

## REFERENCES

- Kim, D.M., Kigawa, T., Choi, C.Y., and Yokoyama, S. 1996. A highly efficient cell-free protein synthesis system from *Escherichia coli*. *Eur J Biochem* **239**: 881-886.
- Kim, D.M., and Swartz, J.R. 2000. Prolonging cell-free protein synthesis by selective reagent additions. *Biotechnol Prog* **16**: 385-390.
- Kim, R.G., and Choi, C.Y. 2001. Expression-independent consumption of substrates in cell-free expression system from *Escherichia coli*. *J Biotechnol* **84**: 27-32.
- Klammt, C., Schwarz, D., Fendler, K., Haase, W., Dotsch, V., and Bernhard, F. 2005. Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J* **272**: 6024-6038.
- Klammt, C., Srivastava, A., Eifler, N., Junge, F., Beyermann, M., Schwarz, D., Michel, H., Doetsch, V., and Bernhard, F. 2007. Functional analysis of cell-free-produced human endothelin B receptor reveals transmembrane segment 1 as an essential area for ET-1 binding and homodimer formation. *FEBS J* **274**: 3257-3269.
- Kommer, A.A., Dashkova, I.G., Esipov, R.S., Miroshnikov, A.I., and Spirin, A.S. 2005. Synthesis of functionally active human proinsulin in a cell-free translation system. *Dokl Biochem Biophys* **401**: 154-158.
- Korepanova, A., Gao, F.P., Hua, Y., Qin, H., Nakamoto, R.K., and Cross, T.A. 2005. Cloning and expression of multiple integral membrane proteins from *Mycobacterium tuberculosis* in *Escherichia coli*. *Protein Sci* **14**: 148-158.
- Korepanova, A., Moore, J.D., Nguyen, H.B., Hua, Y., Cross, T.A., and Gao, F. 2007. Expression of membrane proteins from *Mycobacterium tuberculosis* in *Escherichia coli* as fusions with maltose binding protein. *Protein Expr Purif* **53**: 24-30.
- Kresheck, G.C., Adade, A.B., and Vanderkooi, G. 1985. Differential scanning calorimetry study of local anesthetic effects on F1ATPase and submitochondrial particles. *Biochemistry* **24**: 1715-1719.
- Kuo, A., Bowler, M.W., Zimmer, J., Antcliff, J.F., and Doyle, D.A. 2003. Increasing the diffraction limit and internal order of a membrane protein crystal by dehydration. *J Struct Biol* **141**: 97-102.
- Kuroda, T., and Tsuchiya, T. 2009. Multidrug efflux transporters in the MATE family. *Biochim Biophys Acta* **1794**: 763-768.
- Kutukova, E.A., Livshits, V.A., Altman, I.P., Ptitsyn, L.R., Ziyatdinov, M.H., Tokmakova, I.L., and Zakataeva, N.P. 2005. The *yeaS* (*leuE*) gene of *Escherichia coli* encodes an exporter of leucine, and the Lrp protein regulates its expression. *FEBS Lett* **579**: 4629-4634.

## REFERENCES

- le Maire, M., Champeil, P., and Moller, J.V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta* **1508**: 86-111.
- Lederman, M., and Zubay, G. 1967. DNA-directed peptide synthesis. 1. A comparison of T2 and Escherichia coli DNA-directed peptide synthesis in two cell-free systems. *Biochim Biophys Acta* **149**: 253-258.
- Lee, A.G. 2003. Lipid-protein interactions in biological membranes: a structural perspective. *Biochim Biophys Acta* **1612**: 1-40.
- Lee, H.C., and Bernstein, H.D. 2001. The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc Natl Acad Sci U S A* **98**: 3471-3476.
- Lee, S.M., Grass, G., Haney, C.J., Fan, B., Rosen, B.P., Anton, A., Nies, D.H., and Rensing, C. 2002. Functional analysis of the Escherichia coli zinc transporter ZitB. *FEMS Microbiol Lett* **215**: 273-278.
- Lemieux, M.J., Song, J., Kim, M.J., Huang, Y., Villa, A., Auer, M., Li, X.D., and Wang, D.N. 2003. Three-dimensional crystallization of the Escherichia coli glycerol-3-phosphate transporter: a member of the major facilitator superfamily. *Protein Sci* **12**: 2748-2756.
- Lewinson, O., Lee, A.T., and Rees, D.C. 2008. The funnel approach to the precrystallization production of membrane proteins. *J Mol Biol* **377**: 62-73.
- Li, X.D., Villa, A., Gownley, C., Kim, M.J., Song, J., Auer, M., and Wang, D.N. 2001. Monomeric state and ligand binding of recombinant GABA transporter from Escherichia coli. *FEBS Lett* **494**: 165-169.
- Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J., and Tan, S. 2005. Comparison of affinity tags for protein purification. *Protein Expr Purif* **41**: 98-105.
- Liuzzi, J.P., and Cousins, R.J. 2004. Mammalian zinc transporters. *Annu Rev Nutr* **24**: 151-172.
- Locher, K.P., Lee, A.T., and Rees, D.C. 2002. The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**: 1091-1098.
- Lu, M., Chai, J., and Fu, D. 2009. Structural basis for autoregulation of the zinc transporter YiiP. *Nat Struct Mol Biol* **16**: 1063-1067.
- Lu, M., and Fu, D. 2007. Structure of the zinc transporter YiiP. *Science* **317**: 1746-1748.
- Luirink, J., von Heijne, G., Houben, E., and de Gier, J.W. 2005. Biogenesis of inner membrane proteins in Escherichia coli. *Annu Rev Microbiol* **59**: 329-355.

## REFERENCES

- Lundstrom, K.H. 2006. in *Structural Genomics on Membrane Proteins*. CRC Press Taylor & Francis Group, Boca Raton.
- Matthaei, J.H., and Nirenberg, M.W. 1961. Characteristics and stabilization of DNAase-sensitive protein synthesis in E. coli extracts. *Proc Natl Acad Sci U S A* **47**: 1580-1588.
- Michel, H 1983. Crystallization of membrane proteins. *TIBS* **8**: 56-59
- 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.
- Monne, M., Hessa, T., Thissen, L., and von Heijne, G. 2005. Competition between neighboring topogenic signals during membrane protein insertion into the ER. *FEBS J* **272**: 28-36.
- Morita, Y., Kodama, K., Shiota, S., Mine, T., Kataoka, A., Mizushima, T., and Tsuchiya, T. 1998. NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrob Agents Chemother* **42**: 1778-1782.
- Nelson D.L., Cox M.M. 2000. *Lehninger Principles of Biochemistry*. (Third edition). Worth Publishers
- Nies, D.H. 2003. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol Rev* **27**: 313-339.
- Opekarova, M., and Tanner, W. 2003. Specific lipid requirements of membrane proteins—a putative bottleneck in heterologous expression. *Biochim Biophys Acta* **1610**: 11-22.
- Ostermeier, C., Iwata, S., Ludwig, B., and Michel, H. 1995. Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat Struct Biol* **2**: 842-846.
- Paulsen, I.T., and Saier, M.H., Jr. 1997. A novel family of ubiquitous heavy metal ion transport proteins. *J Membr Biol* **156**: 99-103.
- Pfeil, W., Gesierich, U., Kleemann, G.R., and Sterner, R. 1997. Ferredoxin from the hyperthermophile *Thermotoga maritima* is stable beyond the boiling point of water. *J Mol Biol* **272**: 591-596.
- Pratt, J. M. 1984. Coupled transcription-translation in prokaryotic cell-free system. In *Coupled Transcription-Translation in Prokaryotic Cell-Free System*. (eds. Hames, B. D. and Higgins, S. J.) Oxford: IRL Press, pp. 179–209.

## REFERENCES

- Psakis, G., Nitschkowski, S., Holz, C., Kress, D., Maestre-Reyna, M., Polaczek, J., Illing, G., and Essen, L.O. 2007. Expression screening of integral membrane proteins from *Helicobacter pylori* 26695. *Protein Sci* **16**: 2667-2676.
- Rahman, M., Patching, S.G., Ismat, F., Henderson, P.J., Herbert, R.B., Baldwin, S.A., and McPherson, M.J. 2008. Probing metal ion substrate-binding to the *E. coli* ZitB exporter in native membranes by solid state NMR. *Mol Membr Biol* **25**: 683-690.
- Rauhut, R., and Klug, G. 1999. mRNA degradation in bacteria. *FEMS Microbiol Rev* **23**: 353-370.
- Raunser, S., and Walz, T. 2009. Electron crystallography as a technique to study the structure on membrane proteins in a lipidic environment. *Annu Rev Biophys* **38**: 89-105.
- Razvi, A., and Scholtz, J.M. 2006. Lessons in stability from thermophilic proteins. *Protein Sci* **15**: 1569-1578.
- Regnier, P., and Arraiano, C.M. 2000. Degradation of mRNA in bacteria: emergence of ubiquitous features. *Bioessays* **22**: 235-244.
- Ressl, S., Terwisscha van Scheltinga, A.C., Vonrhein, C., Ott, V., and Ziegler, C. 2009. Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. *Nature* **458**: 47-52.
- Rigell, C.W., de Saussure, C., and Freire, E. 1985. Protein and lipid structural transitions in cytochrome c oxidase-dimyristoylphosphatidylcholine reconstitutions. *Biochemistry* **24**: 5638-5646.
- Roy, A., Shukla, A.K., Haase, W., and Michel, H. 2008. Employing *Rhodobacter sphaeroides* to functionally express and purify human G protein-coupled receptors. *Biol Chem* **389**: 69-78.
- Ryabova, L.A., Vinokurov, L.M., Shekhovtsova, E.A., Alakhov, Y.B., and Spirin, A.S. 1995. Acetyl phosphate as an energy source for bacterial cell-free translation systems. *Anal Biochem* **226**: 184-186.
- Sachdev, D., and Chirgwin, J.M. 1998. Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*. *Biochem Biophys Res Commun* **244**: 933-937.
- Saier, M.H., Jr. 2000. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* **64**: 354-411.
- Saier, M.H., Jr., Eng, B.H., Fard, S., Garg, J., Haggerty, D.A., Hutchinson, W.J., Jack, D.L., Lai, E.C., Liu, H.J., Nusinew, D.P., et al. 1999. Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim Biophys Acta* **1422**: 1-56.

## REFERENCES

- Sánchez-Ruiz, J.M., and Mateo, P.L. 1987. Differential scanning calorimetry of membrane proteins. *Revis Biol Celular* **11**: 15-45.
- Savage, D.F., Anderson, C.L., Robles-Colmenares, Y., Newby, Z.E., and Stroud, R.M. 2007. Cell-free complements in vivo expression of the E. coli membrane proteome. *Protein Sci* **16**: 966-976.
- Schmidt-Krey, I. 2007. Electron crystallography of membrane proteins: two-dimensional crystallization and screening by electron microscopy. *Methods* **41**: 417-426.
- Schneider, B., Junge, F., Shirokov, V.A., Durst, F., Schwarz, D., Dotsch, V., and Bernhard, F. 2010. Membrane protein expression in cell-free systems. *Methods Mol Biol* **601**: 165-186.
- Schulz, P., Garcia-Celma, J.J., and Fendler, K. 2008. SSM-based electrophysiology. *Methods* **46**: 97-103.
- Shirokov, V.A., Kommer, A., Kolb, V.A. & Spirin, A.S. 2007. Continuous-exchange protein-synthesizing systems. In *Methods of Molecular Biology*. (eds. Grandi, G.). Humana Press, Totowa, NJ, 19–57.
- Sobczak, I., and Lolkema, J.S. 2005. Structural and mechanistic diversity of secondary transporters. *Curr Opin Microbiol* **8**: 161-167.
- Spangfort, M.D., Ipsen, H., Sparholt, S.H., Aasmul-Olsen, S., Larsen, M.R., Mortz, E., Roepstorff, P., and Larsen, J.N. 1996. Characterization of purified recombinant Bet v 1 with authentic N-terminus, cloned in fusion with maltose-binding protein. *Protein Expr Purif* **8**: 365-373.
- Spirin, A.S., Baranov, V.I., Ryabova, L.A., Ovodov, S.Y., and Alakhov, Y.B. 1988. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242**: 1162-1164.
- Spirin A. S., Swartz R. 2008. *Cell-free Protein Synthesis Methods and Protocols*. Wiley-VCH Verlag GmbH & Co. KgaA.
- Su, X.Z., Chen, J., Mizushima, T., Kuroda, T., and Tsuchiya, T. 2005. AbeM, an H<sup>+</sup>-coupled *Acinetobacter baumannii* multidrug efflux pump belonging to the MATE family of transporters. *Antimicrob Agents Chemother* **49**: 4362-4364.
- Suppmann, B., and Sawers, G. 1994. Isolation and characterization of hypophosphite-resistant mutants of *Escherichia coli*: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. *Mol Microbiol* **11**: 965-982.
- 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.

---

## REFERENCES

- Tate, C.G. 2001. Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett* **504**: 94-98.
- Valent, Q.A., de Gier, J.W., von Heijne, G., Kendall, D.A., ten Hagen-Jongman, C.M., Oudega, B., and Luirink, J. 1997. Nascent membrane and presecretory proteins synthesized in *Escherichia coli* associate with signal recognition particle and trigger factor. *Mol Microbiol* **25**: 53-64.
- Veenhoff, L.M., Heuberger, E.H., and Poolman, B. 2002. Quaternary structure and function of transport proteins. *Trends Biochem Sci* **27**: 242-249.
- Vrljic, M., Garg, J., Bellmann, A., Wachi, S., Freudl, R., Malecki, M.J., Sahm, H., Kozina, V.J., Eggeling, L., Saier, M.H., Jr., et al. 1999. The LysE superfamily: topology of the lysine exporter LysE of *Corynebacterium glutamicum*, a paradigm for a novel superfamily of transmembrane solute translocators. *J Mol Microbiol Biotechnol* **1**: 327-336.
- Waight, A.B., Love, J., and Wang, D.N. 2010. Structure and mechanism of a pentameric formate channel. *Nat Struct Mol Biol* **17**: 31-37.
- Wakagi, T., Oshima, T., Imamura, H., and Matsuzawa, H. 1998. Cloning of the gene for inorganic pyrophosphatase from a thermoacidophilic archaeon, *Sulfolobus* sp. strain 7, and overproduction of the enzyme by coexpression of tRNA for arginine rare codon. *Biosci Biotechnol Biochem* **62**: 2408-2414.
- Wallin, E., and von Heijne, G. 1995. Properties of N-terminal tails in G-protein coupled receptors: a statistical study. *Protein Eng* **8**: 693-698.
- Wang, Y., Huang, Y., Wang, J., Cheng, C., Huang, W., Lu, P., Xu, Y.N., Wang, P., Yan, N., and Shi, Y. 2009. Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* **462**: 467-472.
- Ward, A., Hoyle, C., Palmer, S., O'Reilly, J., Griffith, J., Pos, M., Morrison, S., Poolman, B., Gwynne, M., and Henderson, P. 2001. Prokaryote multidrug efflux proteins of the major facilitator superfamily: amplified expression, purification and characterisation. *J Mol Microbiol Biotechnol* **3**: 193-200.
- Ward, A., O'Reilly, J., Rutherford, N.G., Ferguson, S.M., Hoyle, C.K., Palmer, S.L., Clough, J.L., Venter, H., Xie, H., Litherland, G.J., et al. 1999. Expression of prokaryotic membrane transport proteins in *Escherichia coli*. *Biochem Soc Trans* **27**: 893-899.
- Ward, A., Sanderson, N.M., O'Reilly, J., Rutherford, N.G., Poolman, B. and Henderson, P.J.F. 2000. *Membrane Transport*. (Editor Baldwin, S.A.). Oxford Univ. Press, Oxford, pp. 141-166.
- Wassenberg, D., Welker, C., and Jaenicke, R. 1999. Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J Mol Biol* **289**: 187-193.

## REFERENCES

- Waugh, D.S. 2005. Making the most of affinity tags. *Trends Biotechnol* **23**: 316-320.
- Wei, Y., and Fu, D. 2006. Binding and transport of metal ions at the dimer interface of the Escherichia coli metal transporter YiiP. *J Biol Chem* **281**: 23492-23502.
- Weiss, H.M., and Grisshammer, R. 2002. Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli. *Eur J Biochem* **269**: 82-92.
- Weiss, H.M., Haase, W., and Reilander, H. 1998. Expression of an integral membrane protein, the 5HT5A receptor. *Methods Mol Biol* **103**: 227-239.
- White, J.F., Trinh, L.B., Shiloach, J., and Grisshammer, R. 2004. Automated large-scale purification of a G protein-coupled receptor for neurotensin. *FEBS Lett* **564**: 289-293.
- White, S.H., and von Heijne, G. 2004. The machinery of membrane protein assembly. *Curr Opin Struct Biol* **14**: 397-404.
- Wright, D.J., Leach, I.B., and Wilding, P. 1977. Differential scanning calorimetric studies of muscle and its constituent proteins. *J Sci Food Agric* **28**: 557-564.
- Yeliseev, A., Zoubak, L., and Gawrisch, K. 2007. Use of dual affinity tags for expression and purification of functional peripheral cannabinoid receptor. *Protein Expr Purif* **53**: 153-163.
- Zakataeva, N.P., Aleshin, V.V., Tokmakova, I.L., Troshin, P.V., and Livshits, V.A. 1999. The novel transmembrane Escherichia coli proteins involved in the amino acid efflux. *FEBS Lett* **452**: 228-232.
- Zakharyan, E., and Trchounian, A. 2001. K<sup>+</sup> influx by Kup in Escherichia coli is accompanied by a decrease in H<sup>+</sup> efflux. *FEMS Microbiol Lett* **204**: 61-64.
- Zhang, H., and Cramer, W.A. 2004. Purification and crystallization of the cytochrome b6f complex in oxygenic photosynthesis. *Methods Mol Biol* **274**: 67-78.
- Zubay, G. 1973. In vitro synthesis of protein in microbial systems. *Annu Rev Genet* **7**: 267-287.



**World wide webs**

[http://blanco.biomol.uci.edu/Membrane\\_Proteins\\_xtal.html](http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html)

<http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>

<http://www.addgene.org/pgvec1>

[http://www.apsu.edu/thompsOnj/Anatomy & Physiology/2010/2010 Exam Reviews/Exam 1 Review/Ch03 The Cell and Membrane Structure.htm](http://www.apsu.edu/thompsOnj/Anatomy%20&%20Physiology/2010/2010%20Exam%20Reviews/Exam%201%20Review/Ch03%20The%20Cell%20and%20Membrane%20Structure.htm)

<http://www.cbs.dtu.dk/services/TMHMM/>

<http://www.membranetransport.org>

<http://www.tcdb.org>

*Appendix*

---

## Appendix 1

### Cloning and expression screening of RhtB family members

Resistance to homoserine/threonine(RhtB) family proteins are involved in exporting different amino acids and conferring resistance to these amino acids. Very few proteins of this family are functionally characterized. I selected five RhtB family proteins from *S. thphimurium*, cloned them in to various vectors and tried to heterologously overproduce them in *E. coli* host cells. However, none of the constructs showed detectable expression when the expression screening and subsequent Western blots were developed as described in Materials and Methods. Table A1 shows the cloning and expression results in details.

**Table A1 : Details of RhtB family members expression screening**

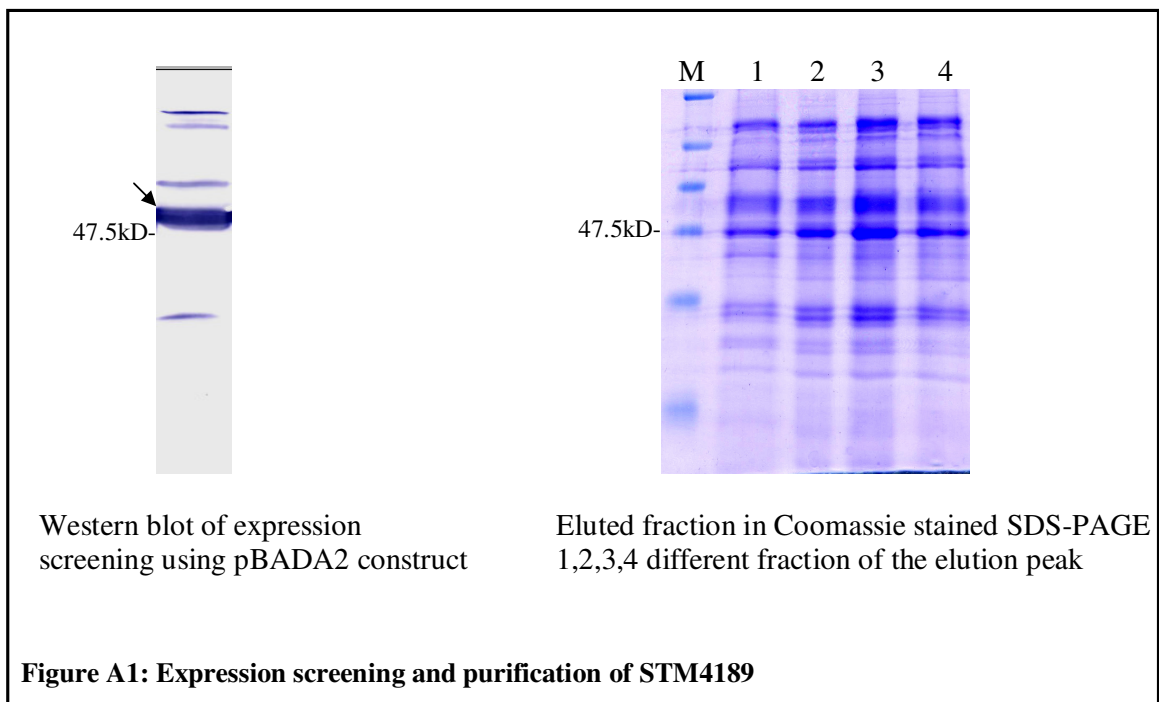
Protein	AA	MW (kD)	TM	pI	PTTQ18A2	PTTQ18C3	pQEA2	pQEC3	pBADA2	pBAD3
<u>STM0365</u>	210	23.13	6	11.28	C, NE	C, NE	C, NE	C, NE	C, NE	C, NE
<u>STM1270</u>	212	22.94	6	9.75	C, NE	C, NE	C, NE	C, NE	C, NE	CF
<u>STM2645</u>	195	21.29	6	9.78	C, NE	C, NE	C, NE	C, NE	C, NE	C, NE
<u>STM3959</u>	206	22.49	6	10.01	C, NE	C, NE	C, NE	C, NE	C, NE	C, NE
<u>STM3960</u>	206	22.32	6	10.62	C, NE	C, NE	C, NE	C, NE	C, NE	C, NE

AA: Number of amino acids, MW(kD): calculated molecular weight in kilo Dalton, TM: predicted transmembrane helices, pI: theoretical Isoelectric point  
C: cloned, CF: cloning failed, NE: not expressed

## Appendix 2

### Cloning, expression and purification of PNaS family protein from *S. typhimurium*

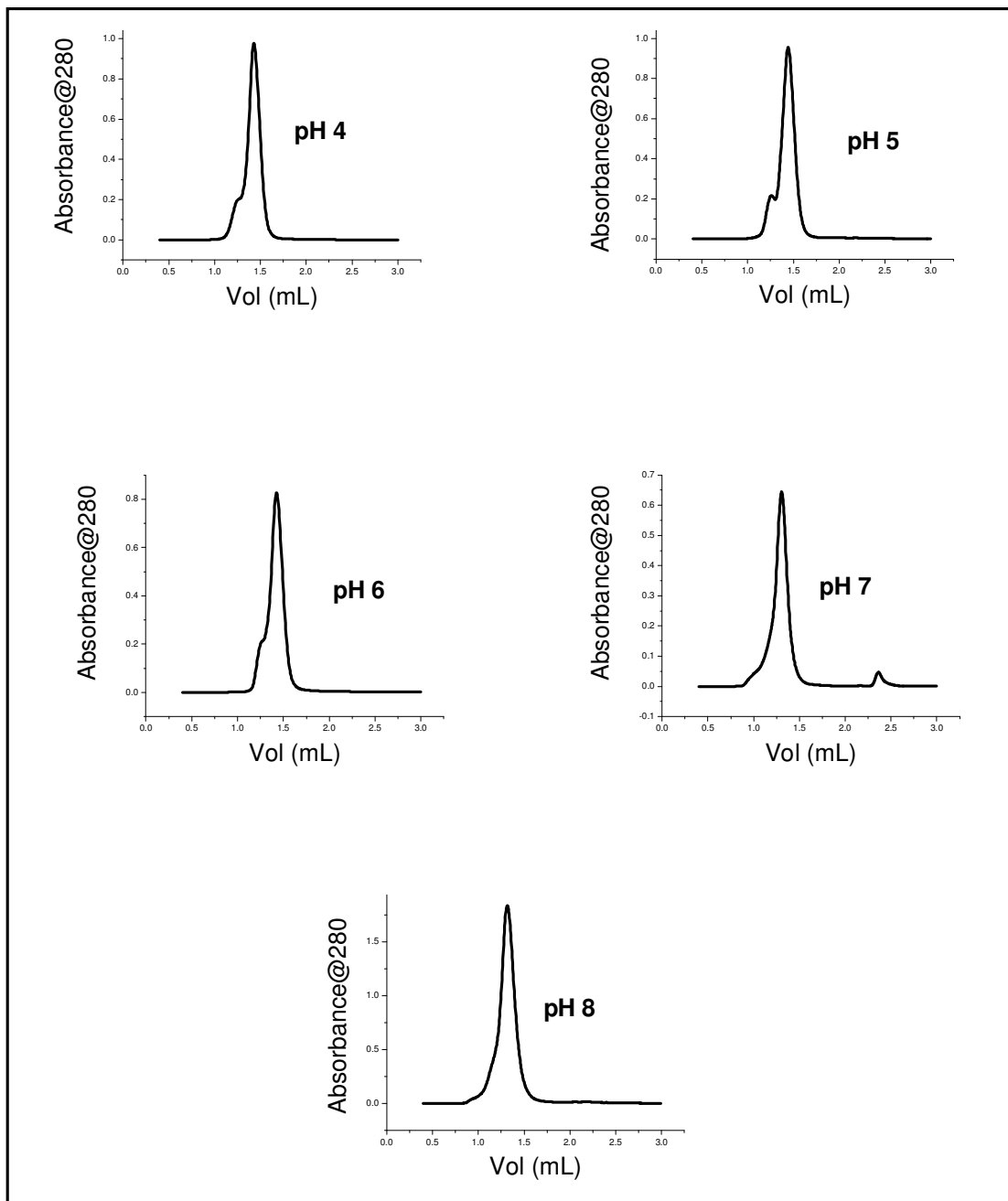
STM4189, the only PNaS family member of *S. typhimurium* was cloned into pBAD, pTTQ18 and pQE vectors, each having two different constructs (A2 and C3) based on tag position. These 6 constructs were subjected to expression screening and only the pBADA2 construct was found to be expressing when checked in Western blot. In order to purify the protein, membrane was prepared from 6 L LB culture. Solubilization screening revealed FOS12 to be the preferred detergent while  $\beta$ LM could solubilize the protein moderately. However, the purification of the protein using IMAC was not successful. The elution fraction always contained a lot of impurities as evident from Coomassie stained SDS- PAGE. It became hard to distinguish the protein band from other impurities in SDS- PAGE. Several optimizations with buffer pHs and other additives did not improve the elution profile. One probable reason could be that the protein was folded in such a way, that the His-tag remained inaccessible by the affinity column matrices. Hence, the protein failed to bind properly with the column and eluted with impurities (Fig A1).



### Appendix 3

#### Analytical gel filtration profiles of Aq\_2073 at different pHs

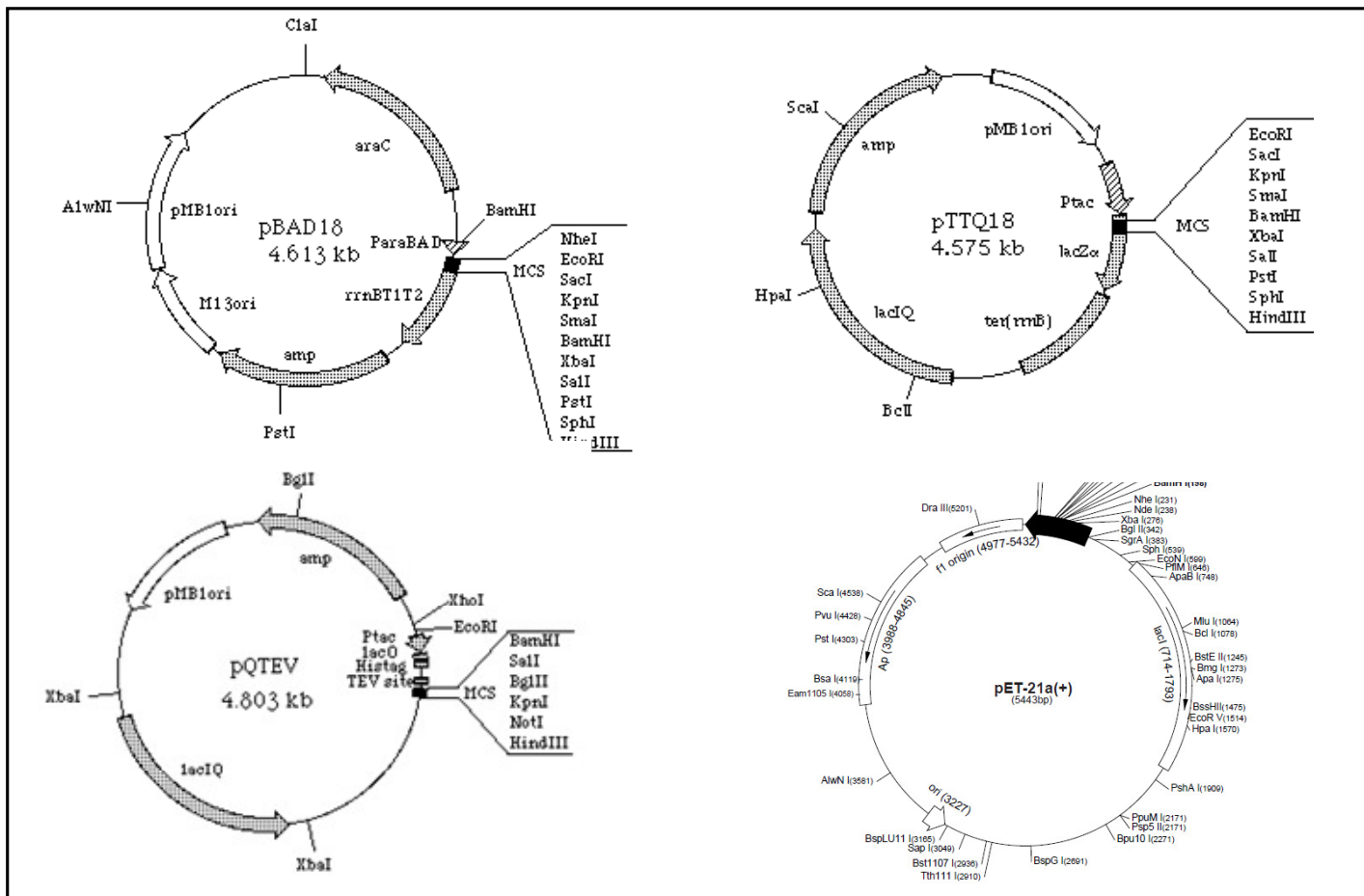
A superdex 200 column in SMART system was used for gel filtration analysis.



Appendix 4

Schematic representation of the vectors used for this thesis work

Basic vector system; vectors were than modified based on purpose.





## **Acknowledgements**

To begin with, I would like to thank my supervisor Prof. Hartmut Michel, for giving me this opportunity to work under him as a PhD student. I am grateful to him for his generous support, guidance and the freedom he provided me to bring my PhD to fruition. In spite of his busy schedule, he was always available for discussion and providing new ideas. I consider my stay under his guidance as a privileged and valuable experience in my research career.

I like to thank Prof. Bernd Ludwig for being my university supervisor and his pleasant behavior whenever I approached him.

I am grateful to Dr. Janet Vonck for helping me in two dimensional crystallization trials, Dr. Winfried Haase for freeze fracture studies and Prof. Klaus Fendler for SSM studies. I am thankful to Prof. Dr. Volker Doetsch from Frankfurt University for providing me the opportunity to carry out my initial cell-free expression work in his lab.

I am highly indebted to Dr. Tonie Baars for reading my thesis with great patience and providing helpful suggestions. I am thankful to Mr. Michael Jähme for the German translation of my thesis summary and Ms. Suzan Rührer for its meticulous reading. Together, we formed the cell-free group in our department which was later joined by Ms. Lakshmi Srinivasan. It has been a learning experience to work with them as a team.

My special thanks to Dr. Sachin Surade for introducing me to the different techniques of membrane protein biochemistry at the beginning of my PhD career. I have learned a lot from him and always cherished the friendship we had. I am thankful to Ms. Cornelia Münke, Ms. Hanne Mueller and Ms. Sofia Hollschwandner for running the lab smoothly and providing necessary help whenever required. Many of my jogging endeavors were possible because of Ms. Hanne Mueller. I feel fortunate to work with other members of our structural genomics group, Dr. Adriana Ryčovska and Dr. Tsuyoshi Nonaka. They have always been very friendly and providing helpful suggestions throughout my PhD. Our structural genomics group was later joined by Ms. Jagdeep Kaur and Mr. Vivekanand Malviya. I thank them for their company in and off the lab and wish them



good luck for their PhD career. Thanks to all members of the structural genomics and the cell-free group for their continuous help and support. I am also thankful to Dr. Julian Langer for mass spectroscopic analyses of my protein samples and Dr. Yvonne Thielmann for setting up crystallization drops with my protein.

Like the lab, my office was also a convivial place. I am thankful to Mr. Florian Hilbers for innumerable English translations of my official letters and also to my previous office mate Mr. Philipp Schleidt for being a warm friend.

I would like to thank all scientific and non scientific staffs of Max Planck institute for making my research life a smooth one. I am thankful to Frau Hildegard Weber for uninterrupted media supply, Frau Rosemarie Schmidell for numerous article requests, Mr. Paolo Lastrico for printing posters and Frau Solveigh McCormack for her help and cordial behavior. I would also like to thank Max Planck society, International Max Planck Research School and SFB807 for providing financial support and creating a vibrant scientific atmosphere in the campus.

I was fortunate enough to make many friends in and around Frankfurt. There are too many to be mentioned here. I am thankful to Mr. Rajsekhar Paul and Mr. Shaik Syed Ali for being good friends and for the great time we spend together. I am thankful to Mr. Anirban Roy, Mr. Rudra Sekhar Manna, Ms. Ankita Roy, Mr. Chandramouli Reddy, Ms. Ankita Srivastava, Mr. Yogesh Bhargava and Ms. Ajeeta Nyola for their help and support.

I want to thank Mr. Soumya Chatterjee, who is like my elder brother, always supporting and encouraging me to all my endeavors.

Last but not the least, my sincere gratitude to my parents and my dear sister Rajashree whose unconditional support and encouragement bring me at this stage of my life. They have always been my greatest motivation to follow my dreams.

**Devrishi Goswami**

## **Devrishi Goswami**

Dept. of Molecular Membrane Biology  
Max Planck Institute of Biophysics  
Max Von Laue str. 3  
Frankfurt am Main, D-60438  
Germany  
Devrishi.Goswami@biophys.mpg.de  
Phone: 0049-6963031050



## **CURRICULUM VITAE**

### **Education**

- 2006-2010      PhD, Max Planck Institute of Biophysics and Johann Wolfgang Goethe  
University, Frankfurt am Main, Germany
- 2002-2004      MSc (Biotechnology), Pondichery University, Pondicherry, India
- 1999-2002      BSc (Zoology), Burdwan University, India

### **Research experience**

- 2006-2010                      Max Planck Institute of Biophysics, Frankfurt am Main, Germany  
*Heterologous production and characterization of selected secondary  
active transporters from the CDF, KUP, MOP, FNT, RhtB and SulP  
families*  
Supervisor: Prof. Dr. Hartmut Michel
- 2004-2006                      Indian Institute of Science, Bangalore, India  
*Understanding the relationship between protein stability, expression  
level and temperature sensitive phenotype (Ts)*  
Supervisor: Prof. Raghavan Varadarajan
- 2003-2004                      Pondichery University, Pondicherry, India  
*Biochemical and DNA fingerprinting analysis of pseudomonads bacteria  
from banana rhizosphere (root soil)*  
Supervisor: Prof. Natarajan Sakthivel
- 2003 (May-July)                      Indian Institute of Chemical Biology, Kolkata, India  
*Genotypic and haplotypic analysis of tumor suppressor protein p53 gene  
in squamous cell carcinoma of the head and neck*  
Supervisor: Prof. Susanta Roychoudhury

### **Manuscripts accepted/under preparation**

- **Goswami D**, Surade S, Michel H. (2010). Heterologous production and functional characterization of four CDF transporters from hyperthermophilic and mesophilic origin. Manuscript ready to be submitted.
- Rycovska A, **Goswami D**, Kaur J, Michel H. (2010). The production level of secondary active transporters affected by polyhistidine tag and MBP depends on topology. Manuscript ready to be submitted.
- Bajaj K, Dewan PC, Chakrabarti P, **Goswami D**, Barua B, Baliga C, Varadarajan R. (2008). *Biochemistry* 47(49):12964-73. PMID: 19006334.
- Naik PR, Sahoo N, **Goswami D**, Ayyadurai N, Sakthivel N.(2008). *Microb Ecol* 56(3):492-504. PMID: 18347847.

### **Conferences, workshops, meetings**

- Poster presentation at E-MeP General Assembly meeting. 26th–27th October 2009, Aston University, Aston Triangle, Birmingham, UK.
- Advanced Training Workshop in Bioinformatics of Membrane Proteins, 22-24 February 2008, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK.
- Oral presentation at EMBN Workshop: Membrane Proteins and Proteomics, 28-31 January 2008, Groningen, Netherlands.

### **Awards and Fellowships**

- IMPRS (International Max Planck Research School) PhD fellowship, 2006-2010, Frankfurt, Germany.
- Council for Scientific & Industrial Research (<http://csirhrdg.res.in/>) “Junior Research Fellowship” by Govt. of India, to pursue PhD in India (2004). Ranked among top 20% in the qualifying CSIR-NET exam and received call for SPM fellowship.
- University gold medalist in Masters (2006). College topper in Bachelors (2002). Recipient of National scholarship for postgraduate studies (2002).

### **Personal details**

Gender: Male

Nationality: Indian

Date and place of birth: 1<sup>st</sup> September, 1980 (West Bengal, India)