

# **Structure-function studies of ABC transporters**

## **HlyB from *Escherichia coli* and ChoX from *Sinorhizobium meliloti***

Inaugural-Dissertation

zur

Erlangung des Doktorgrades der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

**Christine Oswald**

aus Memmingen

Januar 2008

Aus dem Institut für Biochemie  
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der  
Mathematisch-Naturwissenschaftlichen Fakultät der  
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. L. Schmitt  
Koreferent: Prof. Dr. G. Groth

Tag der mündlichen Prüfung: 23.01.2008





## Summary

ATP binding cassette (ABC) transporters represent a superfamily of ubiquitous membrane proteins which utilize energy derived from binding and hydrolysis of ATP to fuel the translocation of substrates. The transport substrates are highly diverse in their chemical characteristics and their size.

Generally, this superfamily can be divided in importers and exporters. In contrast to export systems, archaeal and bacterial importers exhibit an additional substrate binding protein (SBP), which captures the substrate and delivers it to the ABC transporter. Binding of the SBP to the transporter triggers ATP hydrolysis and substrate transport. Notably, the substrate specificity of importers is solely determined by the SBP, whereas for exporters the transmembrane domains account for substrate specificity.

Haemolysin B (HlyB), an ABC transporter from *Escherichia coli*, fuels the transport of the 107 kDa toxin HlyA. Together with the membrane fusion protein HlyD and the outer membrane protein TolC this ABC transporter constitutes a type I secretion system, which allows for substrate transport across two membranes in a single step.

Central questions for a mechanistic understanding of ABC transporters are how substrate specificity is determined and how substrate binding is triggering ATP binding and hydrolysis. As HlyA is a rather large substrate with respect to the transporter, it is extremely intriguing to unravel the transport mechanism of HlyB.

To give an insight into ABC transporter function, a structural determination of HlyB was anticipated. HlyB was successfully overexpressed in *Lactococcus lactis* and purified to homogeneity by affinity chromatography and gelfiltration. Crystallization studies allowed for the determination of first conditions for a successful HlyB crystallization.

The choline transporter ChoXWV from *Sinorhizobium meliloti* represents an ABC importer. Interestingly, it could be shown *in vitro* that the SBP binds not only choline but also acetylcholine. Determination of the X-ray structure of the SBP ChoX provided a detailed picture of the molecular determinants of choline and acetylcholine binding. ChoX is composed of two lobes, which together form the substrate binding site consisting of an aromatic box. Comparison of acetylcholine binding with carbamylcholine binding in the nicotinic acetylcholine receptor revealed similarities, which provide first evidence for a common ancestor.

Most interestingly, different unliganded structures of ChoX allowed for the identification of a rotational subdomain. Inward rotation of this subdomain upon substrate binding may present a “sensor-transmitter switch”. Only if this “sensor-transmitter switch” is making contacts with the substrate, binding of the SBP to the transporter triggers a series of events leading to ATP hydrolysis and the import of substrate. The proposed “transport triggering mechanism” guarantees that ATP is hydrolyzed only if the binding protein carries substrate.

## Zusammenfassung

ATP binding cassette (ABC) Transporter bilden eine Klasse von Membranproteinen, die in allen Bereichen des Lebens vorkommen. Die Gewinnung von Energie aus ATP-Bindung und Hydrolyse ermöglicht es diesen Membranproteinen Substrate zu transportieren. ABC Transporter sind in Exporter und Importer unterteilt. Im Gegensatz zu Exportern besitzen Importer eine zusätzliche Untereinheit, das Substratbindeprotein (SBP), welches das Substrat erfasst und es an den Transporter liefert. Die Bindung des SBP an den Transporter selbst induziert ATP-Hydrolyse und Substrattransport. Während bei Exportern die Substratspezifität über die Transmembrandomänen vermittelt wird, geschieht dies bei Importern interessanterweise nur über das SBP.

Haemolysin B (HlyB), ein ABC Transporter aus *Escherichia coli*, energetisiert den Transport des 107 kDa großen Toxins HlyA. Dieser ABC Transporter bildet zusammen mit dem Membranfusionsprotein HlyD und dem äußeren Membranprotein TolC ein Typ I Sekretionssystem, welches das Substrat in einem Schritt über zwei Membranen transportiert.

Für ein molekulares Verständnis des Mechanismus von ABC Transportern ist es von besonderem Interesse, die Vermittlung der Substratspezifität zu verstehen. Weiterhin ist der Zusammenhang zwischen Substratbindung und ATP-Hydrolyse ungeklärt. Die Aufklärung des Transportmechanismus von HlyB erscheint besonders interessant, da HlyA im Vergleich zu seinem Transporter, ein sehr großes Substrat darstellt. Um einen Einblick in den ABC Transportmechanismus zu erhalten, sollte HlyB strukturell charakterisiert werden. Dazu wurde es in *Lactococcus lactis* überexprimiert und mittels Affinitätschromatographie und Gelfiltration zur Homogenität aufgereinigt. Kristallisationsstudien ermöglichten die erfolgreiche Etablierung von Kristallisationsbedingungen für HlyB.

Der Cholin Transporter ChoXWV aus *Sinorhizobium meliloti* gehört der Gruppe der ABC Importer an. *In vitro* Studien zeigten, dass das SBP dieses Transporters nicht nur Cholin sondern auch Acetylcholin binden kann. Um ein molekulares Verständnis der Bindung von Cholin und Acetylcholin zu erhalten, wurde die Kristallstruktur von ChoX in substratgebundenen Zuständen aufgeklärt. ChoX besitzt zwei Domänen, die gemeinsam die Substratbindestelle bilden und eine aromatische Box generieren. Der Vergleich der Acetylcholinbindung in ChoX mit der Bindung von Carbamylcholin im nikotinischen Acetylcholinrezeptor, zeigt Ähnlichkeiten auf, die erstmals auf einen gemeinsamen Vorläufer hindeuten. Die Strukturbestimmung von ChoX in unterschiedlichen ligandfreien Konformationen erlaubte die Identifizierung einer noch nicht beschriebenen, rotierenden Subdomäne. Da die Subdomäne bei Bindung des Liganden nach Innen rotiert, kann diese als Ligandsensor dienen. Die Interaktion des SBP in ligandgebundenem Zustand mit dem Transporter induziert Veränderungen im Transporter, die letztendlich zur ATP-Hydrolyse führen. Liegt das SBP in ligandfreiem Zustand vor, so verhindert die Subdomäne eine effektive Interaktion mit dem Transporter. Damit kann ATP-Verbrauch reguliert und verschwenderische ATP-Hydrolyse vermieden werden.

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	Membrane transport	1
1.2	ABC transporters	2
1.2.1	The general blueprint of ABC transporters	3
1.2.2	Structural information on ABC transporters	5
1.2.3	Transport mechanism	13
1.3	Protein translocation in bacteria	15
1.3.1	Overview	15
1.3.2	Type I secretion	15
1.3.3	The haemolysin system - a paradigm for type I secretion	17
1.4	The role of choline for <i>Sinorhizobium meliloti</i>	19
1.4.1	Choline as phosphatidylcholine precursor	19
1.4.2	The role of choline in osmoprotection	19
1.4.3	Choline as carbon and nitrogen source	20
1.4.4	Choline uptake	20
1.5	Aims and objectives	21
<b>2</b>	<b>MATERIALS AND METHODS</b>	<b>23</b>
2.1	Materials	23
2.1.1	Chemicals	23
2.1.2	Enzymes and standards and antibodies	25
2.1.3	Microorganisms, plasmids and media	25
2.1.4	Buffers and solutions	27
2.1.5	Chromatographic materials and others	28
2.1.6	Instruments	28
2.2	Microbiological methods	29
2.2.1	The expression system	29
2.2.2	Overnight culture of <i>L. lactis</i>	30
2.2.3	Preparation of cryostocks	30
2.2.4	Nisin A production	30
2.2.5	Expression of HlyB in <i>L. lactis</i>	30
2.3	Biochemical methods	31
2.3.1	Determination of protein concentration	31
2.3.2	Concentrating a protein sample	31
2.3.3	SDS-polyacrylamide gel electrophoresis (SDS-PAGE)	31
2.3.4	Immuno-blotting	32
2.3.5	Preparation of membranes	32
2.3.6	Solubilization screen	33
2.3.7	Purification of HlyB	34
2.3.8	Gelfiltration chromatography	34
2.3.9	Detergent exchange	35
2.4	Crystallographic methods	35
2.4.1	Fundamentals of crystal growth	35
2.4.2	Crystallization of HlyB	41
2.4.3	Crystallization of ChoX	42
2.4.4	Crystallization by seeding	43
2.4.5	Basic principles of X-ray crystallography	45
2.4.6	Twinning - a crystal growth disorder	58
2.4.7	Cryocrystallography	63
2.4.8	Data collection and processing	63

2.4.9	Determination of the structure of ChoX in complex with choline-----	64 -
2.4.10	Determination of the structure of ChoX in complex with acetylcholine and in the unliganded closed state-----	64 -
2.4.11	Determination of the structure of ChoX in unliganded semi-closed state -----	65 -
2.4.12	Analysis of the obtained models and graphical visualization -----	65 -
2.4.13	Structure deposition-----	65 -

### 3 RESULTS ----- 66 -

<b>3.1</b>	<b>The ABC transporter HlyB-----</b>	<b>66 -</b>
3.1.1	Heterologous overexpression of HlyB in <i>L. lactis</i> -----	66 -
3.1.2	Solubilization of HlyB -----	66 -
3.1.3	Purification of HlyB-----	68 -
3.1.4	Detergent exchange -----	71 -
3.1.5	Crystallization of HlyB -----	73 -
<b>3.2</b>	<b>The substrate binding protein ChoX-----</b>	<b>75 -</b>
3.2.1	Expression and purification of ChoX -----	75 -
3.2.2	Ligand binding assay-----	76 -
3.2.3	Crystallization-----	77 -
3.2.4	Data collection and processing-----	81 -
3.2.5	Obtaining the ChoX/choline model -----	82 -
3.2.6	Generating models for ChoX/acetylcholine and the unliganded states of ChoX -----	85 -
3.2.7	Structures of ChoX complexed with ligands-----	90 -
3.2.8	Structures of ChoX in various unliganded states -----	98 -

### 4 DISCUSSION ----- 103 -

<b>4.1</b>	<b>The ABC transporter HlyB from <i>Escherichia coli</i>-----</b>	<b>103 -</b>
4.1.1	The solubilization and purification of HlyB -----	103 -
4.1.2	Crystallization of HlyB -----	106 -
4.1.3	Outlook -----	107 -
<b>4.2</b>	<b>The substrate binding protein ChoX from <i>Sinorhizobium meliloti</i>-----</b>	<b>108 -</b>
4.2.1	Binding affinities for choline, acetylcholine and glycine-betaine -----	108 -
4.2.2	Seeding and twinning – is there a correlation?-----	109 -
4.2.3	Structural aspects of choline and acetylcholine binding-----	110 -
4.2.4	Binding mode of acetylcholine -----	111 -
4.2.5	Binding of acetylcholine in ChoX compared to carbamylcholine binding in the AChBP of the nicotinic acetylcholine receptor of <i>Lymnaea stagnalis</i> -----	113 -
4.2.6	The rotational subdomain Ia -----	114 -
4.2.7	Outlook -----	123 -

# 1 Introduction

## 1.1 Membrane transport

The biological membrane presents an essential component of the cell. By establishing boundaries, the plasma membrane guarantees not only for the individuality of a cell, it also maintains different properties of the inner compartment, the cytosol, in comparison to the extracellular space. Furthermore, intracellular membranes form organelles thereby enabling compartmentalization within the cell. Each of these compartments has its own characteristic ensemble of enzymes and differs in its chemical composition<sup>1</sup>. A biological membrane can also be seen as a compartment itself, which, due to restricted two-dimensional mobility, enables faster recognition of reaction partners within the membrane as compared to the three-dimensional diffusion within organelles<sup>1</sup>.

A simplistic view of a biological membrane describes it as a fluid lipid bilayer which is impermeable to ions and only partially permeable for polar molecules<sup>2</sup>. Hence, a lot of hydrophilic nutrients essential for metabolism cannot enter the cell by diffusion. Nevertheless, these components are important for cellular function. For example ions, unable to pass this barrier freely, are involved not only in the maintenance of the electrochemical gradient but also act as cofactors in enzymatic reactions. Furthermore, the flux over this selective barrier needs to be maintained for signaling molecules, toxins, antibiotics and genetic information. Therefore, the cell has developed different means for distinctive and highly coordinated transport.

The transport of specific molecules is accomplished by membrane proteins. Based on functional and phylogenic aspects these proteins are classified according to the transporter classification system<sup>3</sup> (<http://www-biology.ucsd.edu/~msaier/transport/>). Generally, membrane proteins can be divided into two main classes, channels and carriers. Channels facilitate a diffusion-limited transport through an aqueous pore. Thereby, the substrate interacts only weakly with the membrane protein. In contrast, carriers form tighter interactions with the substrate and undergo conformational changes in order to translocate the substrate.

One group of carriers is formed by the electrochemical potential-driven transporters, which according to their name make use of the electrochemical gradient to fuel substrate transport. Members of this family work in one of three different modi<sup>3</sup>, uniport, symport or antiport.

Uniporters transport a single substrate along a gradient. In cases where substrate needs to be transported against a gradient this process is coupled to a translocation of a second substrate along an electrochemical gradient. Symporters transport both substrates in the same directions whereas antiporters translocate to opposite directions. The major facilitator superfamily belongs to this group of carrier proteins with the symporter lactose permease representing one of the best studied membrane proteins<sup>4</sup>.

Primary active transporters form another group of carriers. These utilize a primary energy source to transport substrates. The energy can be derived in different ways, for example by hydrolysis of a phosphoric acid anhydride bond, by light absorption or by oxidoreduction. Thus, within this group direct usage of mechanical, electrical, chemical or electromagnetic energy is energizing the transport. The structurally well characterized photosynthetic reaction centre<sup>5</sup> and also ATP binding cassette (ABC) transporters<sup>6</sup> are prominent members of this category.

In contrast to the before mentioned transporters, group translocators modify their substrate during transport. This can be achieved by phosphorylation as it occurs in the bacterial phosphotransferase system<sup>7</sup>.

## **1.2 ABC transporters**

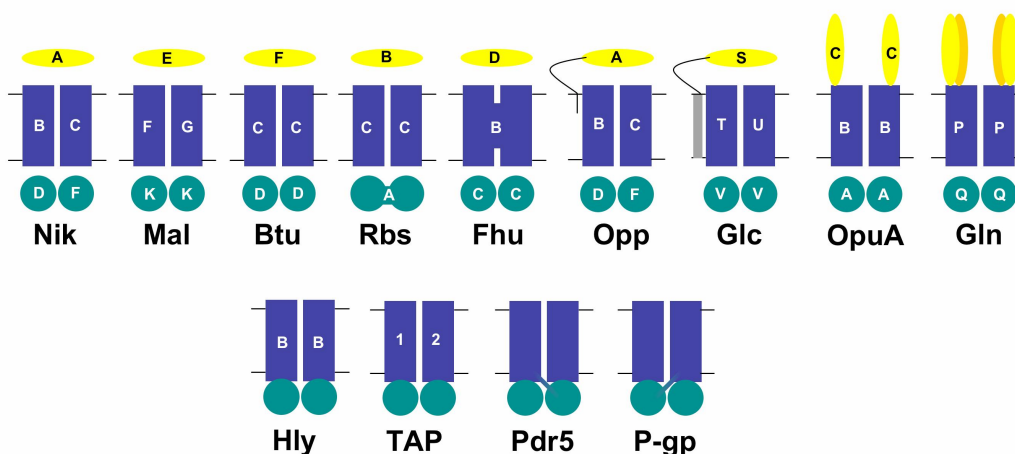
ABC transporters<sup>6</sup> form one of the largest transmembrane protein superfamilies and are found in all phyla<sup>8</sup>. They use the energy generated by ATP binding and hydrolysis to transport a substrate (also called allocrite<sup>9</sup>) across a cellular membrane. ABC transporters are divided into two groups, exporters and importers. For exporters the substrate is situated in the same compartment as the ATP binding cassette, whereas for importers the substrate and the ATP binding cassette are positioned on opposite sides of the membrane. Exporters are found in eukaryotes, whereas in bacteria and archaea, exporters as well as importers were discovered. The allocrites display different chemical characteristics and a huge difference in their size<sup>8</sup>. One class of transport substrates are small ions like molybdate, the substrate of the molybdate importer ModABC<sup>10</sup>. The cystic fibrosis transmembrane conductance regulator (CFTR), a ligand gated ion channel<sup>11,12</sup> allowing for chloride ion fluxes, represents another example for this substrate class. But also nutrients like sugars<sup>13,14</sup> or vitamins<sup>15</sup> are translocated via ABC transporters. Another prominent member of the ABC transporter family is P-glycoprotein<sup>16</sup>, which transports hydrophobic compounds. This transporter is highly expressed in cancer cells

and presents a major obstacle in cancer chemotherapy<sup>16</sup>. Furthermore, small peptides are transported from the cytosol to the ER lumen by another member of this protein superfamily. This process is accomplished by the transporter associated with antigen processing (TAP or ABCB2/ABCB3)<sup>17</sup>. The largest substrate revealed up to now is the 900 kDa S-layer protein<sup>18</sup>. The human genome comprises 49 ABC transporter genes. Many of these have been associated with genetic diseases. A prominent example is cystic fibrosis, the most common genetically inherited disease among Caucasians, which is caused by mutations within the *CFTR* (or *ABCC7*) gene<sup>11</sup>. The Dubin-Johnson syndrome<sup>19</sup> or Stargardt disease<sup>20</sup> are conferred by mutations in *ABCC2* and *ABCA4*. Although not being directly linked to a disease, the overexpression of multidrug resistance protein 1 (MDR1, ABCB1 or P-glycoprotein) and multidrug resistance-related proteins (MRPs), involved in the efflux process of xenobiotic compounds, establish a major obstacle in cancer chemotherapy<sup>16</sup>.

### 1.2.1 The general blueprint of ABC transporters

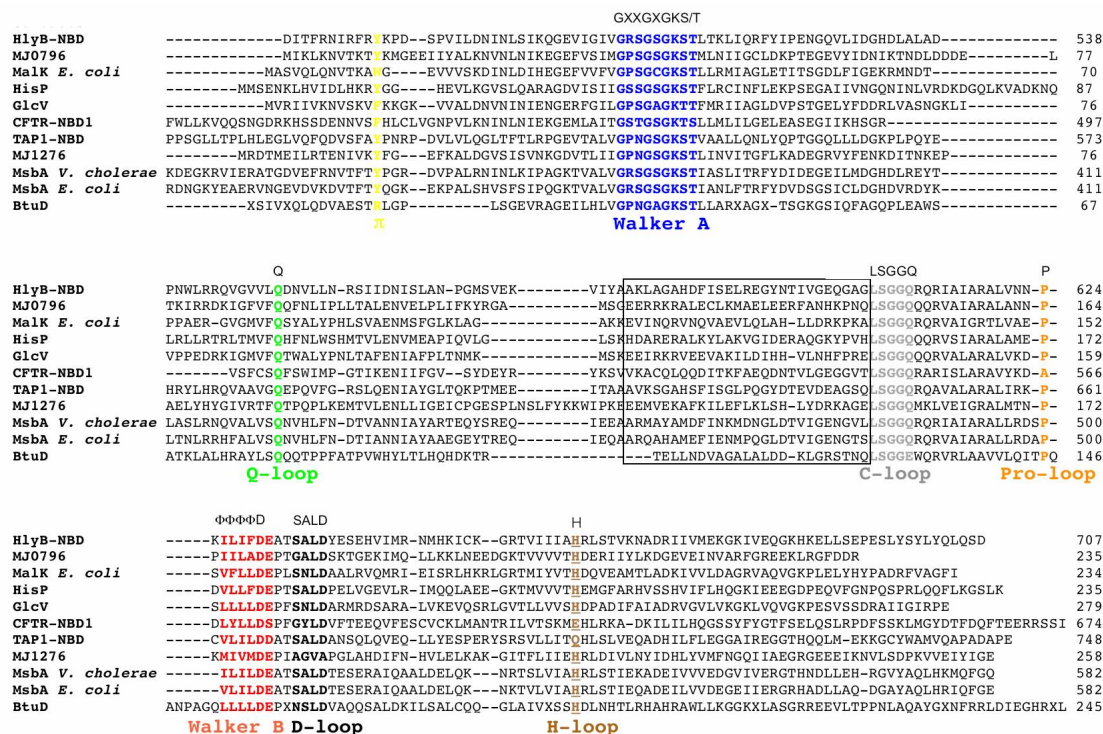
Independent of the direction of transport or the transported substrate, all ABC transporters share a common minimal core architecture composed of two hydrophilic nucleotide binding domains (NBDs) and two transmembrane domains (TMDs). As implicated in the name, the NBDs, situated adjacent to the membrane, bind and hydrolyze ATP. The TMDs form a translocation pathway and mediate substrate binding. TMDs are  $\alpha$ -helical hydrophobic subunits which typically expose between 4 and 10 transmembrane spanning helices<sup>15,21</sup>. Furthermore, some ABC transporters consist of additional accessory domains that often show regulatory functions and are not a part of the core translocator<sup>22</sup>. Importers exhibit an additional subunit, the substrate binding protein (SBP). In Gram-negative bacteria this subunit is located in the periplasmic space, whereas in Gram-positive bacteria and archaea it is anchored to the membrane via a lipid or a transmembrane peptide<sup>22</sup>. In some systems, one or two substrate binding subunits are fused to one TMD (e.g. OpuA from *Lactococcus lactis*), resulting in either two or four substrate binding subunits per functional transporter<sup>23</sup>.

All kinds of different possibilities for the fusion of the TMDs and NBDs can be found in ABC transporters. In some transporters each domain is encoded by a separate open reading frame. One NBD can be fused with one TMD to form a so-called half-size transporter, which need to be fully assembled to homodimers (TMD-NBD)<sub>2</sub> or heterodimers (TMD-NBD)(TMD\*-NBD\*) to constitute a functional ABC transporter. Full-size transporters have all four core domains fused on a single polypeptide (TMD-NBD-TMD\*-NBD\*). Figure 1.1 shows examples for different domain architectures found in ABC transporters.



**Figure 1.1** Domain architectures found in ABC transporters. The nucleotide binding domain and the transmembrane domain are depicted in green and blue, respectively. The substrate binding protein of imports is shown in yellow. Adapted from Biemans-Oldehinkel *et al.*<sup>22</sup>

The remarkable substrate diversity of ABC transporters is reflected in sequence variance of the TMDs. The high sequence conservation of the NBDs implies that there may be a common mechanism for the use of energy to drive substrate transport. The NBDs comprise the sequences which classify a protein as ABC-ATPase<sup>24</sup>. A sequence alignment of different NBDs is depicted in Figure 1.2.



**Figure 1.2** Sequence alignment of different NBDs. Conserved motifs are highlighted in color. The Walker A and B motif are shown in blue and red, respectively. The C-loop, or signature motif, is shown in grey. The Q-loop and the Pro-loop are depicted in green and orange, respectively. The D-loop is shown in black and the H-loop in brown. The residue mediating  $\pi$ -stacking interactions with ATP is colored yellow. The structurally diverse region is highlighted by a box.<sup>25</sup>

Three characteristic sequence motifs can be found within this domain. The Walker A<sup>26</sup> (or P-loop; colored blue) and the Walker B (red) motif form a nucleotide binding fold common to all P-loop ATPases<sup>27</sup>. The Walker A motif comprises the sequence GXXGXGKS/T, where X represents any amino acid. The Walker B motif is built up of four hydrophobic residues followed by an aspartate (ΦΦΦΦD). The hallmark of ABC transporters is the C-loop (grey) or signature motif (LSGGQ) situated between Walker A and Walker B<sup>28</sup>. Furthermore, sequence analysis shows an invariable glutamine and a proline which form the Q- and the Pro-loop, respectively<sup>29</sup> (green and orange). Additionally, the D-loop (SALD; colored black) is situated a few residues after the Walker B motif. An important histidine residue, giving the H-loop (brown) its name, is placed downstream of the D-loop. Another conserved sequence is revealed in the NBD of exporters. The X-loop (TEVGERG) is located N-terminally of the C-loop and possibly transmits ATP binding to the TMD<sup>30</sup>. As this sequence is described for ABC transporters with an export function only, it is not highlighted in Figure 1.2.

In contrast to the high sequence conservation of NBDs the TMDs show only little sequence similarities. One exception is the EAA motif present in transporters with import function<sup>31</sup>. This motif was described first for the TMD of the maltose importer and is located in the cytoplasmic loop N-terminally of the penultimate transmembrane helix<sup>32</sup>. Biochemical evidence suggested a role for this motif in the recognition of the NBD<sup>31</sup>.

The identification of different conserved sequences does not explain their biochemical roles. To answer the question of functionality three-dimensional structures of fully assembled ABC transporters or their NBDs are extremely beneficial.

### **1.2.2 Structural information on ABC transporters**

#### *Isolated nucleotide binding domains*

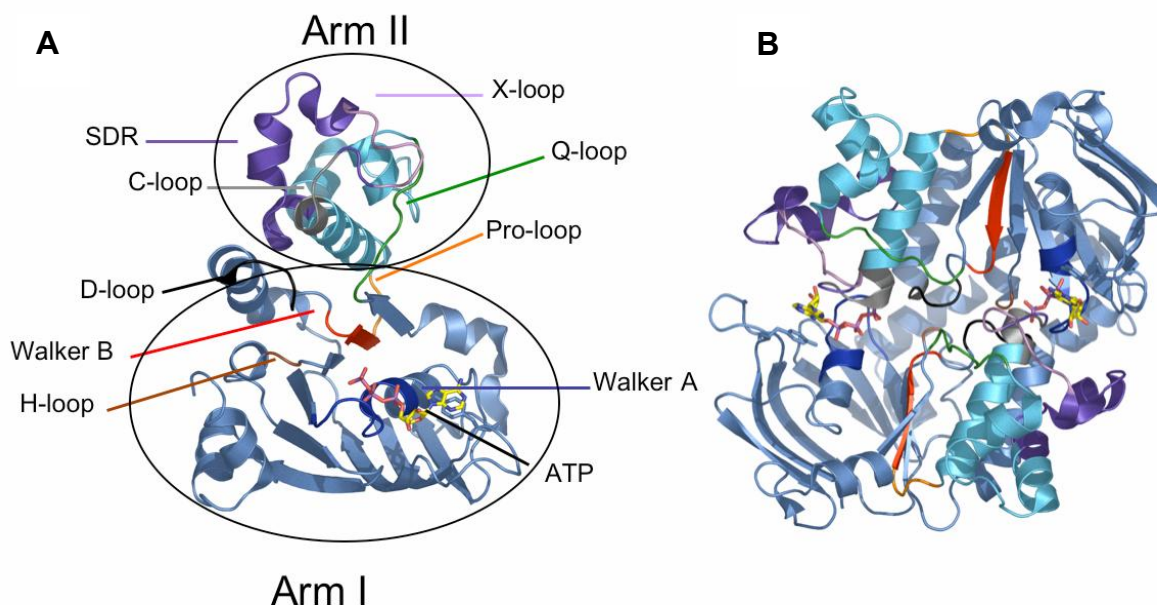
Until today NBD structures of many ABC transporters have been solved. The structures were analyzed in various states, with ATP, ADP and in nucleotide free form as well as with pyrophosphate or the unhydrolyzable AMP-PNP. Furthermore, some of the structures even expose Mg<sup>2+</sup>, an essential cofactor for ATP hydrolysis. Table 1.1 comprises a list of isolated NBDs structures published so far.

Name	Organism	Substrate	Resolution	PDB-code	Year
HisP	<i>Salmonella typhimurium</i>	ATP	1.5 Å	1B0U	1998 <sup>33</sup>
MalK	<i>Thermococcus litoralis</i>	Pyrophosphate	1.9 Å	1G29	2000 <sup>34</sup>
MalK	<i>Escherichia coli</i>	Nucleotide free open Nucleotide free semi-open ATP Mg*ADP	2.9 Å 2.8 Å 2.6 Å 2.3/2.8 Å	1Q1E 1Q1B 1Q12 2AWN/ 2AWO	2003 <sup>35</sup> 2003 <sup>35</sup> 2003 <sup>35</sup> 2005 <sup>36</sup>
MalK	<i>Pyrococcus horikoshii</i>	Nucleotide free ATP	2.3 Å	1V43 1VCI	2004 <sup>37</sup> 2004 <sup>37</sup>
MJ0796	<i>Methanococcus jannaschii</i>	Mg*ADP ATP (E171Q)	2.7 Å 1.9 Å	1F3O 1L2T	2001 <sup>38</sup> 2002 <sup>39</sup>
MJ1267	<i>Methanococcus jannaschii</i>	Mg*ADP Nucleotide free	1.6 Å 2.5 Å	1G6H 1GAJ	2001 <sup>40</sup> 2001 <sup>40</sup>
TAP1	Human	Mg*ADP	2.4 Å	1JJ7	2001 <sup>41</sup>
GlcV	<i>Sulfolobus solfataricus</i>	Nucleotide free (A) Nucleotide free (B) Mg*ADP Mg*AMP-PNP Mg*ATP (G144A)	1.65 Å 2.1 Å 2.1 Å 1.95 Å 1.45 Å	1OXS 1OXT 1OXU 1OXV 1OXX	2003 <sup>42</sup> 2003 <sup>43</sup> 2003 <sup>43</sup> 2003 <sup>43</sup> 2003 <sup>43</sup>
HlyB	<i>Escherichia coli</i>	Nucleotide free Mg*ATP (H662A) ADP (wt) ADP (E631Q) ADP (H662A) ATP (H662A) ATP (E631Q)	2.6 Å 2.5 Å 1.6 Å 1.9 Å 1.7 Å 2.6 Å 2.7 Å	1MT0 1XEF 2FF7 2FFB 2FFA 2FGJ 2FGK	2003 <sup>29</sup> 2005 <sup>44</sup> 2006 <sup>45</sup> 2006 <sup>45</sup> 2006 <sup>45</sup> 2006 <sup>45</sup> 2006 <sup>45</sup>
CFTR	Mouse	AMP-PNP ATP Nucleotide free ATP ADP ATP	2.5 Å 2.35 Å 2.2 Å 2.2 Å 2.55 Å 3.0 Å	1Q3H 1R0Z 1R0W 1R0X 1R0Y 1R10	2004 <sup>46</sup> 2004 <sup>46</sup> 2004 <sup>46</sup> 2004 <sup>46</sup> 2004 <sup>46</sup> 2004 <sup>46</sup>
CFTR	Human	ATP (F508A) ATP (ΔF508)	2.25 Å 2.3 Å	1XMI 1XMJ	2005 <sup>47</sup> 2005 <sup>47</sup>
CysA	<i>Alicyclobacillus acidocaldarius</i>	Nucleotide free	2.0 Å	1Z47	2005 <sup>48</sup>

**Table 1.1** List of structures of isolated NBDs.

The overall structure of the ATP binding domain shows a stubby L-shaped form which can be divided into two arms. To get an impression of the overall structure, the HlyB NBD at 2.5 Å (PDB accession code: 1XEF) is depicted in Figure 1.3. Arm I (residues 467-549 and residues 624-707, HlyB numbering) is shown in pale blue and arm II (residues 550-623, HlyB numbering) in cyan<sup>44</sup>. Arm I holds the binding site for ATP and is built up by  $\alpha$ -helices as well as  $\beta$ -sheets. As this catalytic domain shows similarities to the structures of RecA<sup>49</sup> and the F<sub>1</sub>-ATPase<sup>50</sup>, this part of the NBD is often called the “Rec-A-like” domain or the “F<sub>1</sub>-like” domain. Arm II is composed solely by  $\alpha$ -helices and, therefore, it is named the helical

domain. The Pro- and the Q-loop present the linkers between the helical domain and the catalytic domain. Arm II exposes the so-called structurally diverse region (SDR), which is shown in violet in Figure 1.3. The name is derived from the observation that structures of different NBDs reveal a region of roughly 30-40 amino acids that has only little structural similarity<sup>29</sup>. Within this region the lately identified X-loop is positioned<sup>30</sup>, which is important in recognition of the TMD. The X-loop is colored light violet in Figure 1.3.



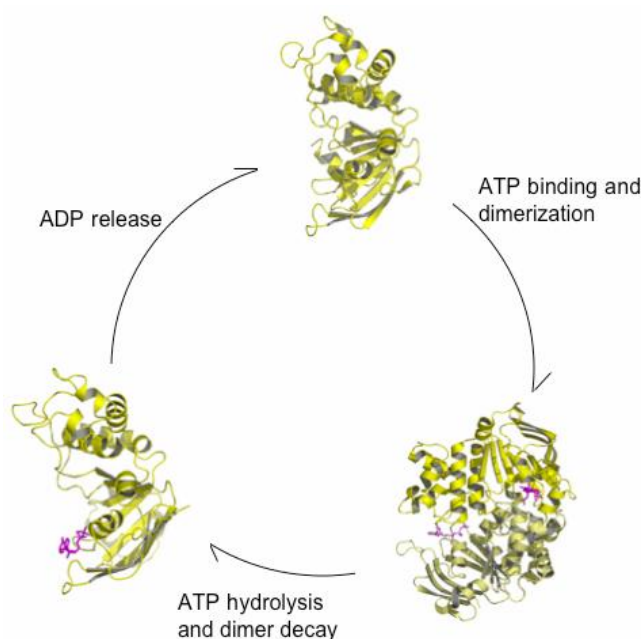
**Figure 1.3** Sequence motifs of NBDs exemplified in HlyB NBD (PDB accession code: 1XEF). Color coding is the same as in figure 1.2. The X-loop is colored light violet. **A** represents an “artificial” ATP-bound monomer. **B** The composite ATP-bound dimer is depicted.

The binding of the nucleotide is accomplished by different molecular contacts. An aromatic residue (Tyr477 in HlyB), recently named the A-loop<sup>51,52</sup>, forms  $\pi$ -stacking interactions with the adenine moiety of the ATP. The phosphate groups of the nucleotide contact the Walker A motif (depicted in blue). The last residues of the Walker B motif interact with the  $\gamma$ -phosphate via a water molecule. To understand the function of the C-loop/signature motif (depicted in grey) the dimeric arrangement of the NBDs needs to be considered (Figure 1.3b). The binding of ATP is fully accomplished only in a dimer formation with the C-loop of the trans-monomer completing the cis-binding site.

Biochemical and structural analyses revealed that the H-loop histidine in HlyB, and probably in many other ABC transporter NBDs, is important for the hydrolysis event<sup>44</sup>. Together with the last residue of the Walker B motif the H-loop histidine forms a catalytic dyad<sup>53</sup>, stabilizing

the transition state of the hydrolysis event. This residue interacts not only with ATP but also with the trans-monomer. Therefore, the histidine was also named “linchpin”<sup>44</sup>.

Structures of NBDs in different liganded states allow for the analysis of the catalytic cycle of isolated NBDs. Figure 1.4 shows the catalytic cycle exemplified with the available structures of the HlyB NBD<sup>45</sup>. In an unliganded state the NBD exposes a monomeric arrangement. The binding of ATP results in an induced fit and in a reduction of flexibility of the NBD<sup>40</sup>. Upon ATP binding the catalytic domain and the helical domain rotate towards each other by  $\sim 18^\circ$  (for HlyB NBD). The bound ATP acts as molecular glue and induces dimer formation. After hydrolysis the phosphate ion possibly exits the dimer via a phosphate tunnel<sup>45</sup> and the dimer decays due to the lack of  $\gamma$ -phosphate. This is accompanied by an outward movement of  $\sim 17^\circ$ . ADP release finally sets the NBD back to the unliganded state, where it awaits a second round of ATP binding and hydrolysis.

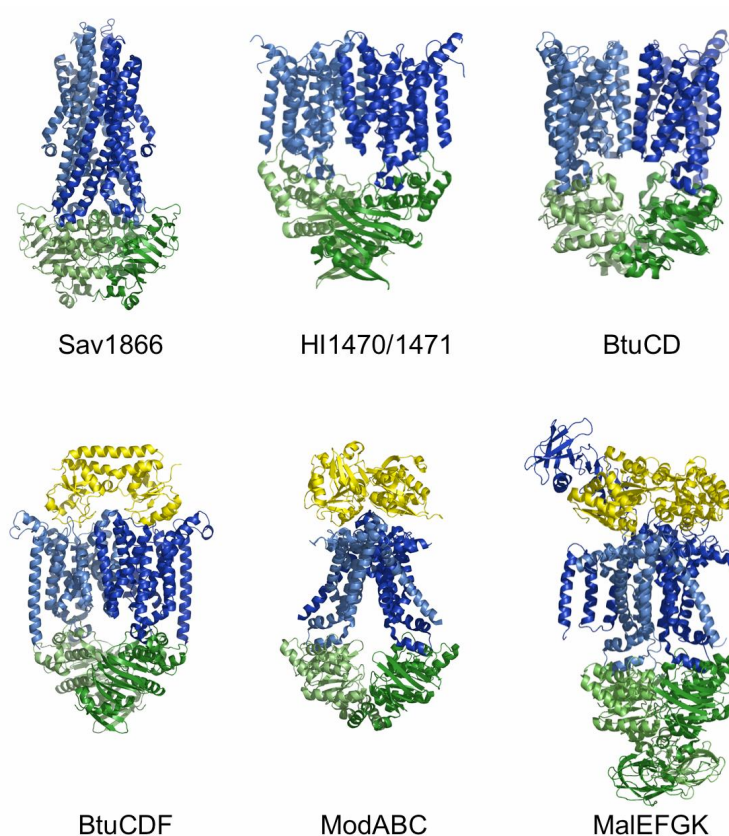


**Figure 1.4** The catalytic cycle of ABC transporter NBDs exemplified on the HlyB-NBD.

### *Assembled ABC transporters*

In addition to the structure of BtuCD, presenting the first successfully solved assembled ABC transporter structure<sup>15</sup>, in the last two years more fully assembled ABC transporter structures have been elucidated<sup>10,30,54-57</sup> (presented in Figure 1.5). Several structures of MsbA from different organisms have been solved before, however they all have been retracted<sup>58</sup> due to erroneous software used for the data conversion. Recently, these structures have been published again<sup>59</sup>, but due to the low resolution they will not be further considered here.

The obtained structural information of five different ABC transporters represents a major step in ABC transporter research and improves the understanding of mechanistic processes developed within this membrane protein superfamily. An overview of the available fully assembled ABC transporter structures is given in Table 1.2.



**Figure 1.5** Structures of full length ABC-transporters. NBDs are colored green the TMDs blue and the substrate binding proteins yellow.

Name	Organism	Substrate	TM-segments	TMD conformation	Resolution	PDB-code	Years
BtuCD BtuCDF	<i>Escherichia coli</i>	Cycloortho-vanadate nucleotide free	20	Outward facing  Closed pathway	3.2 Å  2.6 Å	1L7V  2QI9	2002 <sup>15</sup>  2007 <sup>54</sup>
Sav1866	<i>Staphylococcus aureus</i>	ADP AMP-PNP	12	Outward facing Outward facing	3.0 Å 3.4 Å	2HYD 2ONJ	2006 <sup>30</sup> 2007 <sup>55</sup>
HI1470/ HI1471	<i>Haemophilus influenzae</i>	Nucleotide free	20	Inward facing	2.4 Å	2NQ2	2007 <sup>56</sup>
ModABC	<i>Archaeoglobus fulgidus</i>	Nucleotide free	12	Inward facing	3.1 Å	2ONK	2007 <sup>10</sup>
MalEFGK	<i>Escherichia coli</i>	ATP and maltose	8+6	Outward facing	2.8 Å	2R6G	2007 <sup>57</sup>
MsbA	<i>Escherichia coli</i>	Nucleotide free open	12	Outward facing	5.3 Å	3B5W	2007 <sup>59</sup>
	<i>Vibrio cholerae</i>	Nucleotide free closed		Outward facing	5.5 Å	3B5X	2007 <sup>59</sup>
	<i>Salmonella thyphimurium</i>	AMP-PNP		Inward facing	3.7 Å	3B5Y	2007 <sup>59</sup>
	<i>Salmonella thyphimurium</i>	ADP*Vi		Inward facing	5.5 Å	3B5Z	2007 <sup>59</sup>

**Table 1.2** List of fully assembled ABC transporter structures.

The structure of the multidrug transporter Sav1866<sup>30</sup> constitutes the only three-dimensional representation of an exporter in a mediate (~3.0 Å) resolution. The structures of a putative metal-chelate-type transporter HI1470/1471<sup>56</sup> as well as the vitamin B<sub>12</sub> transporter BtuCD<sup>15</sup> represent ABC importers. Especially interesting for the function of importers are the very recently solved structures of BtuCDF<sup>54</sup>, the molybdate transporter ModABC<sup>10</sup> and the maltose transporter MalEFGK<sup>57</sup> which show the transporters in presence of their binding protein. The SBPs of MalFGK and BtuCDF exhibit an open state, whereas ModABC displays a closed state.

Sav1866 as well as ModABC has 12 transmembrane segments, whereas BtuCD and HI1470/1471 are composed of 20 transmembrane segments. In contrast, the maltose importer forms a heterodimer constituted of MalF and MalG, which display eight and six transmembrane segments, respectively.

Although crystallized in the presence of ADP, Sav1866 shows an outward facing conformation probably reflecting the ATP bound state<sup>30</sup>. The structure of the Sav1866 with the unhydrolyzable AMP-PNP reveals the same conformation. Unique to the exporter are the rather extensive intracellular loops protruding into the cytoplasmic space.

Within all five structures so-called “coupling helices” form part of a cytoplasmic loop. The TMD-NBD transmission interface is composed by these “coupling helices” (within the TMD)

and by the Q-loop of the NBD. Notably, the coupling helix from the Sav1866 monomer *a* stands in contact with the NBD of monomer *b*, thereby generating tight cross interactions between TMD*a*-NBD*b* and vice versa. The importer structures do not expose such cross interactions. Instead a larger gap at the centre of the transporter is revealed.

The importer structures in complex with their respective binding protein show interactions between periplasmic loops and both lobes of the binding protein<sup>10,54,57</sup>. Salt bridges as well as charged residues are involved in the SBP-TMD interface.

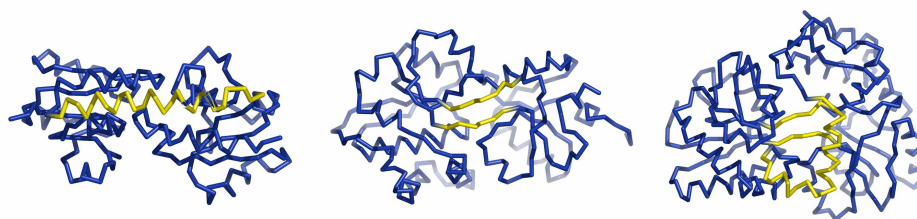
The maltose importer shows a large periplasmic loop P2, which folds in an Ig-like domain<sup>57</sup>. This loop interacts with the N-terminal lobe of the maltose binding protein.

Interestingly, BtuCDF and MalEFGK show a periplasmic loop which is suggested to displace the substrate from the binding site and therefore is named “scoop”<sup>57</sup>. However, a corresponding loop cannot be identified in the ModABC structure, suggesting different displacement mechanisms<sup>60</sup>.

The maltose transporter is the first structure of a fully assembled ABC transporter in presence of its substrate. The substrate is situated in the cavity formed by the outward facing TMDs, approximately halfway into the lipid bilayer.

### *Substrate binding proteins*

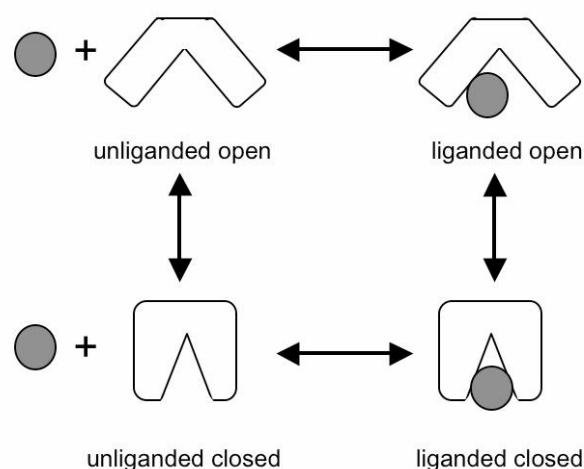
The first structure of a substrate binding protein, the arabinose binding protein, was published in 1977<sup>61</sup>. Since then many more structures of SBPs have been solved. Although substrate binding proteins display high diversity in their recognized substrates and only low sequence homologies, the determined structures exhibit similar overall folds. All structures reveal a bilobal fold consisting of two globular domains. Each domain shows a central  $\beta$ -sheet with flanking  $\alpha$ -helices. The two lobes are connected via a flexible linker. Dependent on the number of segments contributing to this linker, binding proteins can be classified into three different classes. Many sugar binding proteins, e.g. maltose binding protein (MalE), belong to the type I binding protein (Figure 1.6, right), which comprises three linker segments. Prominent examples for type II binding proteins, exposing two linker segments, are the amino acid binding proteins HisJ<sup>62,63</sup> and LAO (Lys, Arg and ornithine)<sup>64</sup> (Figure 1.6, middle). For a third class only one linker segment is characteristic (Figure 1.6, left). BtuF, the vitamin B<sub>12</sub> binding protein of the ABC transporter BtuCD, belongs to this class<sup>65,66</sup>.



**Figure 1.6** Substrate binding proteins from different groups showing one (left), two (middle) and three (right) linker segments.

For all classes of SBPs the substrate binding site lies in a gorge between the two globular domains. There, the substrate is captured and often shielded from the surrounding solution. In addition to liganded structures many unliganded structures reveal a more open conformation. Theoretic conformational energy calculations<sup>67</sup> as well as small angle X-ray scattering experiments<sup>68</sup> indicate a closure of the protein induced by ligand binding. This finding leads to the proposal of a “Venus-Fly trap mechanism”, that describes the hinge-bending motion occurring in binding proteins upon ligand binding<sup>67</sup>.

In addition to the liganded closed and unliganded open conformations, a few structures have been solved that show the binding protein in a ligand-free closed state<sup>69</sup> or in a liganded open state<sup>70,71</sup>. These findings in conjunction with computer simulations calculating the population distribution of ribose binding protein<sup>72</sup> in different states suggest an equilibrium between two different unliganded states, the open and the closed. The binding of substrate shifts the equilibrium towards the closed state. The liganded open state probably reflects an intermediate conformation encountered before the substrate is liberated into the translocation pathway formed by the transporter. Different states of the SBP and possible interactions with the ligand are presented in Figure 1.7.



**Figure 1.7** Different states of substrate binding proteins. If no substrate is bound, the SBP can undergo opening and closing movement. In presence of ligand an interaction in open as well as in closed form can be mediated.

Several studies on isolated transport systems in complex with their SBP indicated a role for the SBP apart from its task of capturing the substrate and delivering it to the transporter. Studies on the glycine-betaine/proline-betaine<sup>73</sup>, maltose<sup>74</sup> and histidine<sup>75-77</sup> transporters, either in detergent solution or reconstituted in liposomes, all showed similar results regarding the ATPase activity. By adding unliganded SBP to the transporter, ATP hydrolysis is induced. But only when liganded SBP was added ATPase activity could be stimulated fully. These findings revealed a new assignment for the SBP, the transmission of a signal through the TMD to the NBD and subsequent stimulation of ATP hydrolysis.

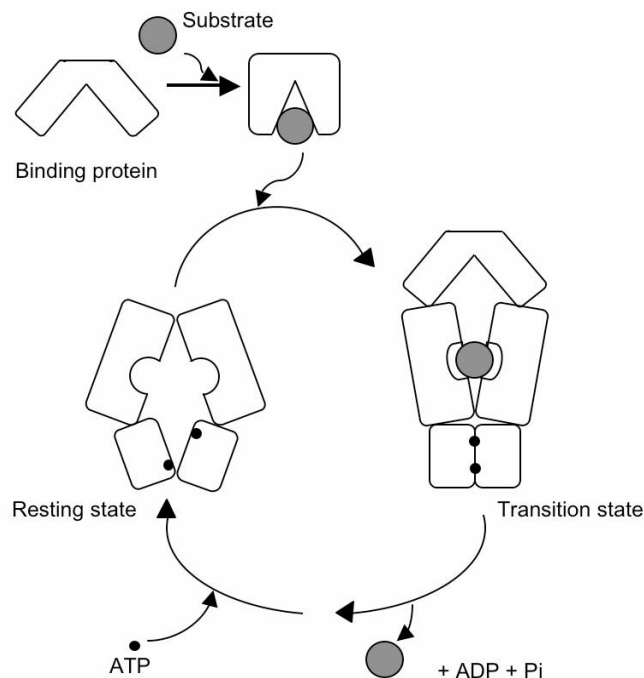
### 1.2.3 Transport mechanism

The determined structures of ABC transporters expose different states. The molybdate importer<sup>10</sup> and the HI1470/1471<sup>56</sup> show an inward facing conformation of the TMD in the nucleotide free state. In contrast, the structure of BtuCDF<sup>54</sup> exposes a nucleotide free state but rather a closed pathway is formed by the TMDs. Thus, this structure possibly reflects an intermediate state. Although crystallized with ATP only ADP is present in the NBD of the structure of Sav1866. However, the NBDs display a conformation which reflects the composite dimer identified in isolated NBDs<sup>30</sup>. As the overall conformation of Sav1866 with the unhydrolyzable ATP analogue AMP-PNP exposes a similar conformation as the ADP bound structure, it was anticipated that ABC transporters display an outward facing conformation in the ATP bound state<sup>55</sup>. This view was further manifested by the structure of the maltose importer in presence of ATP revealing an outward facing conformation<sup>57</sup>.

With the help of the available structures a very basic general mechanism was proposed by Dawson and colleagues<sup>60</sup>. In principal this model reflects ideas already put forward in 2001 by Davidson and coworkers<sup>78</sup> but elaborates on TMD and NBD communication. In a nucleotide free state the TMDs expose an opening to the inside, as was uncovered for the molybdate transporter. As the composite dimer, formed by ATP loaded NBDs, is not formed yet, the NBDs exhibit a state that can be compared to the monomeric state of isolated NBD structures. With the help of the NBD-TMD transmission interface, consisting of the Q-loop of the NBDs and the coupling helices of the TMDs, the ATP driven dimer formation of the NBDs can effectively be transmitted to the TMDs. As response, the TMDs close on the cytoplasmic side and open at the outside (outward facing conformation). For importers the substrate binding protein may interact with the inward facing conformation of the transporter<sup>78</sup>. After ATP binding and NBD dimer formation the transporter is converted into the outward facing conformation. In this outward facing conformation importers can accept

the substrate from the substrate binding protein. After the release of ADP the transporter flips back into the inward facing conformation and the substrate can be released.

However, Oldham and colleagues<sup>57</sup> proposed a model taking the ATP concentration in the cell into further consideration (Figure 1.8). In the resting state, the NBDs are probably loaded with ATP but the dimer interface is kept open by the TMDs. The loaded SBP can interact with the transporter in the resting state and initiates the NBD closure with a concomitant opening of the TMDs to the periplasmic side. This outward facing conformation is called the transition state, where the affinity of the SBP to the substrate is reduced<sup>78</sup>. Here, the SBP is bound in an open conformation to the transporter. ATP hydrolysis and transport of the substrate reset the transporter into the resting state.



**Figure 1.8** Mechanism of substrate import. ATP is bound to the transporter in the inward facing resting state. After the binding of substrate, the SBP can interact with the transporter in the resting state and induce a dimer formation of the NBDs. This dimer formation generates an outward facing conformation of the transporter and the SBP is opened to deliver the substrate to the transporter (transition state). After ATP hydrolysis the dimer decays, fresh ATP is bound and the transporter is transferred back to the resting state. Adapted from Oldham *et al.*<sup>57</sup>

Dawson and colleagues stated that for exporters the described process may be reversed<sup>79</sup>. The substrate is bound in an inward facing conformation and released after ATP driven conversion of the transporter to the outward facing conformation.

## **1.3 Protein translocation in bacteria**

### **1.3.1 Overview**

Even though the synthesis of proteins is accomplished within the cell, many proteins are found on the cell surface, in the cell envelope or even in the extracellular space. To bring these proteins to their designated location, bacteria have developed different transport systems.

As Gram-negative bacteria possess two membranes, the inner and the outer membrane, three fundamentally different transport systems can be distinguished. These are divided with respect to which membrane has to be crossed during the transport. Transport via the inner membrane delivers proteins to the periplasm, which either reside there or are further transported via the outer membrane to the extracellular space. Another possibility is the transport across two membranes in a single step without periplasmic intermediates. Gram-positive bacteria, however, only exhibit one membrane. Thus, transport is similar to the transport via the inner membrane of Gram-negative bacteria<sup>80</sup>.

Translocation via the inner membrane can be accomplished by two different systems. The well-studied Sec-system is involved in translocation of unfolded proteins. Fully folded proteins constitute transport substrates for the twin arginine targeting<sup>81</sup> (TAT)-system. Proteins translocated with the help of the Sec-system<sup>82</sup> to the periplasm, can be further transferred across the outer membrane. This process is mainly maintained by the type II secretion pathway.

Undoubtedly, the translocation of proteins via the transenvelope of the cell in one step presents a rather complex process. In Gram-negative bacteria this task is fulfilled by three different systems. The most “simple” of these processes is type I secretion<sup>83</sup>, which necessitates three different membrane proteins. A more complex process is the type III secretion system, which is utilized for the translocation of flagellar proteins or effector proteins. Type III systems exhibit at least nine conserved proteins<sup>84</sup>.

### **1.3.2 Type I secretion**

The type I secretion system represents a rather simple arrangement of three membrane protein components. Two of these proteins are residing in the inner membrane and the third is situated in the outer membrane of Gram-negative bacteria. Although rather simple in

composition, the type I system accomplishes a complex task, the translocation of a substrate across two membranes in a single step. In this process no stable periplasmic intermediates are generated<sup>85</sup>.

Despite the difference in size, ranging from 10 kDa to ~900kDa<sup>86</sup>, the substrates of type I secretion systems show some common characteristics<sup>86</sup>. In contrast to the N-terminal signal sequence targeting proteins to the Sec-system, the substrates of type I secretion systems normally expose a C-terminal secretion sequence which in general remains intact after transport. The translocated substrates are proteins with a rather acidic pI and a low amount of cysteines. Most of them also exhibit glycine rich repeats, which are composed of the sequence GGXGXDXXX, with X being any amino acid. These repeats, occurring up to 80 times in a single substrate, form  $\beta$ -roll structures upon binding of  $\text{Ca}^{2+}$ -ions<sup>87</sup>. Due to the repeating glycine motif this type of substrate is classified as a RTX toxin (repeats in toxin).

The proteins secreted by type I secretion systems have been identified to function as toxins, like *E. coli* Haemolysin A<sup>88</sup>, or hydrolases, like the metalloprotease PrtG from *Erwinia chrysanthemi*<sup>89</sup>. However, for some substrates no function could be identified so far.

As mentioned before the translocator complex is composed of three proteins. An ABC transporter, residing in the inner membrane, is utilized to energize the transport. Furthermore, the second component, a so-called membrane fusion protein (MFP), is also situated in the inner membrane. MFPs are commonly built up by a membrane spanning  $\alpha$ -helix, which is preceded by a relatively small cytoplasmic stretch, and a rather large periplasmic part that forms the C-terminus. The third component is the outer membrane protein (OMP) which exposes a  $\beta$ -barrel structure within the membrane. With the help of the MFP the ABC transporter and the OMP are linked and a continuous channel is formed, which builds the translocation pathway for the substrate.

In a first step of transport initiation the C-terminal secretion sequence of the substrate interacts with the ABC transporter. This contact signals the recruitment of the other transport components<sup>90,91</sup>. In some cases the substrate also interacts with the cytosolic part of the MFP, an interaction essential for OMP recruitment<sup>92</sup>. The binding of ATP to the ABC transporter destabilizes the previously established interaction with the substrate, suggesting that the substrate is released during the hydrolysis cycle<sup>93</sup>. The actual transport of the substrate through the tunnel still remains unclear and might be rather different depending on the substrate size. Noteworthy is the discovery that mutations in the outer membrane protein and in the MFP can affect the activity of the secreted protein, which is probably due to misfolding

or aggregation of the protein<sup>94,95</sup>. Thus, the generated pathway might not only function as transport tunnel but also as a folding cage.

### 1.3.3 The haemolysin system - a paradigm for type I secretion

Certain uropathogenic *E. coli* strains produce haemolysin A<sup>96</sup>, a 107 kDa toxin which exhibits haemolytic activity<sup>97,98</sup>. This toxin is secreted by the bacteria with the help of a type I secretion system encoded by the *hly* operon (*hlyCABD*). The *hly* operon was first identified in the mid 80ies<sup>99</sup> and comprises the genes for the toxin HlyA, two integral membrane proteins, HlyB and HlyD, as well as the acyltransferase HlyC. HlyC, which is catalyzing the activation of proHlyA by acylation of two lysine residues, is dispensable for secretion<sup>100</sup>.

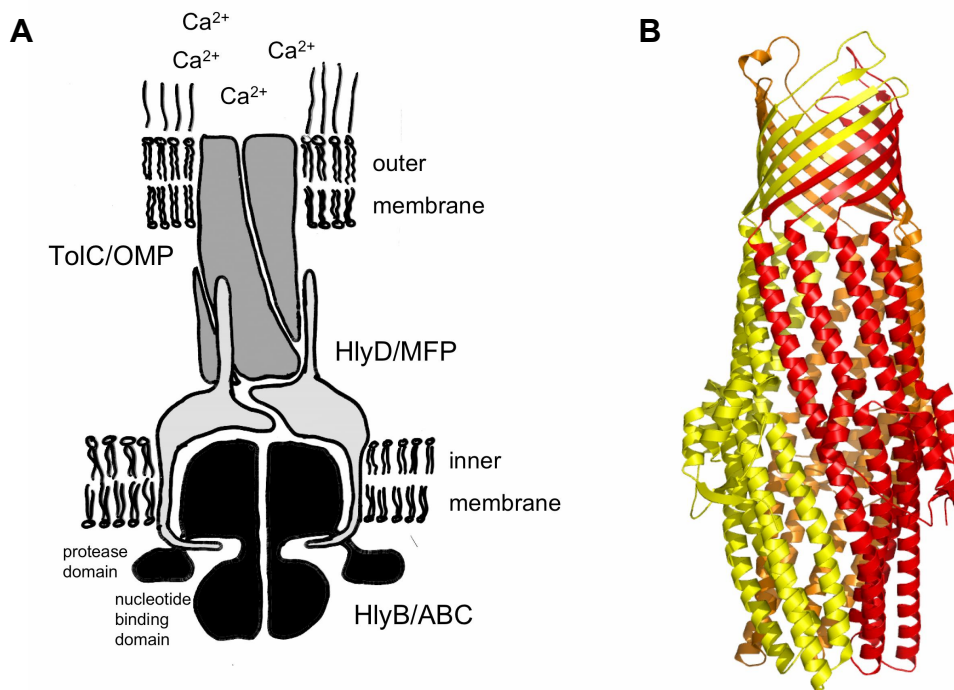
HlyB and HlyD form the inner membrane components of the secretion system, the ABC transporter and the membrane fusion protein, respectively. Together with the outer membrane porin TolC a tunnel is formed, through which HlyA is translocated.

To date structural information on this transport system is limited to the OMP TolC and the NBDs of HlyB<sup>29,44,101</sup> (see chapter 1.2.2). In contrast to the available structures of ABC transporters, an additional N-terminal domain (~150 aa), preceding the first transmembrane helix, is revealed by topology analysis of HlyB. This domain presents a degenerated C39 peptidase with a replacement of the catalytically active Cys by a Tyr. The role of this domain for the secretion process is still unknown.

Only little “structural” information is available on the second component of the inner membrane complex, the membrane fusion protein HlyD. The N-terminal cytosolic part of HlyD forms an amphipathic helix followed by charged residues<sup>92</sup>. Adjacent to a single transmembrane helix, a rather large helical domain and a C-terminal domain composed of  $\beta$ -strands are revealed<sup>102</sup>.

The structure of TolC<sup>103</sup> reveals a trimeric composition. It forms a channel tunnel of 140 Å, which spans the outer membrane as  $\beta$ -barrel and forms a helical barrel within the periplasmic space. The tapered periplasmic end of this tunnel is suggested to open by a rotational untwisting movement during substrate translocation<sup>103</sup>.

Whereas structural analysis of TolC<sup>103</sup> and crosslinking studies of HlyD<sup>90</sup> suggest a trimeric arrangement, the ABC transporter is supposed to be a dimer. This symmetry break presents one of the puzzling issues concerning the haemolysin secretion system. A simple schematic representation of the haemolysin type I secretion system as well as the determined OMP TolC is depicted in Figure 1.9.



**Figure 1.9** The type I secretion system. **A** A schematic presentation of the haemolysin A secretion system. The ABC transporter HlyB and the membrane fusion protein HlyD are residing in the inner membrane. TolC is situated in the outer membrane. **B** The structure of the TolC at 2.1 Å resolution (PDB accession code: 1EK9)<sup>103</sup>.

Crosslinking studies on the haemolysin system revealed that HlyB and HlyD form a precomplex<sup>90</sup>. Interaction of the substrate with this inner membrane complex results in recruitment of the outer membrane protein TolC<sup>90</sup>. The charged residues following the amphipathic helix of the MFP have been shown to be involved in this process<sup>92</sup>. Furthermore, surface plasmon resonance experiments revealed an interaction between the HlyB NBD and a HlyA fragment, containing the secretion sequence<sup>93</sup>. In this study it is suggested that secretion is triggered by the displacement of HlyA from the NBD due to ATP binding. These results give an insight on how the translocation complex might be initiated but they lack to provide information about the mechanism of substrate transport. One remaining central question is the stoichiometry of hydrolyzed ATP per substrate. The translocation of a 107 kDa substrate seems to challenge the before mentioned ABC transport mechanism, deprived from ABC transporters which transport rather small substrates. Therefore, further biochemical and structural information on the individual components as well as interacting parts involved in the haemolysin system would definitely be of major benefit to understand the complex mechanism of type I secretion systems.

## **1.4 The role of choline for *Sinorhizobium meliloti***

The Gram-negative bacterium *Sinorhizobium meliloti* (*S. meliloti*) is commonly found in soil and the rhizosphere where it lives in symbiosis with plants of the *Medicago*, *Melilotus* and *Trigonella* species. *S. meliloti* induces the formation of nodules at the root of these plants. Within these nodules differentiated bacteria, so-called bacteroids, reduce the atmospheric nitrogen to ammonia. Ammonia can further be taken up by the plant in form of ammonium. In return the bacteria are provided with nutrients by the plants.

### **1.4.1 Choline as phosphatidylcholine precursor**

Phosphatidylcholine (PC) is a phospholipid commonly found in the membrane of eukaryotes but rarely distributed in prokaryotes. In *E. coli* and *Bacillus subtilis* (*B. subtilis*), the Gram-negative and Gram-positive model systems, PC does not constitute a big part of the membrane. However, it has been estimated that ~10% of all bacteria contain PC<sup>104</sup> as a major phospholipid. *S. meliloti* belongs to this small group and it was speculated that the microsymbiont needs this specific lipid to form a successful interaction with the host root<sup>105</sup>.

Two different pathways for the biosynthesis of PC have been discovered in eukaryotes, the CDP-choline and the methylation pathway<sup>106</sup>. The methylation pathway, based on the conversion of phosphatidylethanolamine to PC via threefold methylation, had been long identified in bacteria<sup>107</sup>. Recently, a second pathway for PC biosynthesis in *S. meliloti* was described<sup>108</sup>. Here, phospholipid is formed directly from choline, which is extruded by the plant, and CDP-diacylglycerol with the help of the enzyme phosphatidylcholine synthase. This process enables *S. meliloti* to efficiently produce PC, a major phospholipid component of its membrane.

### **1.4.2 The role of choline in osmoprotection**

Soil bacteria are subjected to varying weather conditions. Therefore, the cell has to react quickly to drastic changes in osmolarity by keeping the cell turgor at a constant level. If the environment exhibits a high osmolarity, at dry weather conditions, water efflux has to be avoided. In a first step potassium ions are imported before compatible solutes are accumulated. The compatible solutes are either imported or synthesized within the cell.

Compatible solutes are organic osmolytes that can be accumulated to high concentrations without interacting with cellular functions. According to the preferential exclusion model<sup>109</sup>, these molecules are excluded from the hydration shell surrounding the protein within the cell. The unfavorable interactions of these osmolytes with the protein surface result in a non-homogenous distribution of the compatible solutes and the protein hydration shell can be maintained.

In rhizobial species trehalose, ectoine, glycine-betaine and proline-betaine have been identified as effective compatible solutes. Although structurally similar to glycine-betaine choline shows no osmoprotective effect. However, with the help of the *bet* operon, comprising the four genes *betI*, *betC*, *betB* and *betA*, the oxidation of choline to glycine-betaine in *S. meliloti* can be guaranteed. Thus, it is possible to utilize choline for the stabilization of the cell turgor.

#### **1.4.3 Choline as carbon and nitrogen source**

Uptake of choline is maintained at high and low osmolarities implying another function apart from osmoprotection. It has been shown that *S. meliloti* can catabolize choline as well as glycine-betaine<sup>110,111</sup>. Surprisingly, for *S. meliloti* it is possible to survive on choline or glycine-betaine as sole carbon and nitrogen source. Before choline can be successfully catabolized it needs to be oxidized to glycine-betaine first, as described before. Glycine-betaine is further demethylated to glycine which then is converted to serine. Thus, *S. meliloti* has developed a possibility to survive on choline, a compound readily found in the rhizosphere<sup>110</sup>.

#### **1.4.4 Choline uptake**

As choline is employed in various processes this compound plays a central role in *S. meliloti*. Dependent on the usage of choline, different transport systems equipped with adequate biochemical characteristics have been developed. For osmoprotection, choline import has to be inducible by high ionic strengths. Also it should be fast to successfully counteract the change in osmolarity. In contrast, the uptake of choline for catabolism and the synthesis of PC do not necessitate an induction by high ionic strengths.

Transport assays in different cultural conditions revealed three distinct systems in *S. meliloti*<sup>112</sup>. Two of the transport systems are constitutively expressed and the third is induced by choline. The choline inducible transporter shows a high affinity for choline and is

insensitive to the osmolarity of the medium. Therefore, it is possibly not playing a role in osmoprotection. The second transporter which shows no effect on salt addition conveys a low affinity for choline and is constitutively expressed. The third transport system is also constitutively expressed and exhibits a high affinity for choline. It also displays sensitivity to the environmental salt concentration.

In a second survey three genes, *choXWV*, encoding for a substrate binding protein dependent ABC transporter were identified<sup>113</sup>. This transport system encompasses a NBD ChoV, being composed of ~350 amino acids, and a hydrophobic domain ChoW (~280 residues) exposing six putative transmembrane helices. The substrate binding protein ChoX, a 309 amino acid protein, shows specificity for choline. Dupont *et al.* also analyzed the affinities of the binding protein to different substrates. Thereby, a high affinity for choline (2.7  $\mu$ M) and a reduced affinity for acetylcholine (145  $\mu$ M) were revealed. Surprisingly, no affinity for glycine-betaine was detected<sup>113</sup>.

## 1.5 Aims and objectives

Due to the ubiquitous occurrence of ABC transporters and their role in diverse processes from nutrient import in bacteria to the development of multidrug resistance in humans, this superfamily of membrane proteins gains continuous interest in the membrane transport field. Independent of the origin or the directionality of transport, i.e. import or export, it appears that they all share the same principal mechanism that drives substrate translocation. For this process the energy derived by ATP binding and hydrolysis is utilized. The continuous effort to characterize ABC transporters biochemically and structurally showed some fruitful results in the case of isolated NBDs. However, studies on the full length membrane protein with all its components are essential to the understanding of the mechanism underlying ABC transporter function. To unravel the mechanism of ABC transporters more structural information has to be generated that can further complement biochemical data and also generate a solid basis for more biochemical analyses.

In membrane transport research, biochemical and structural investigations often address the central question of substrate recognition. Dependent on their function, ABC transporters expose two fundamentally different ways of conveying substrate specificity. Importers expose an additional subunit, the SBP, which possesses all molecular determinants for specific substrate recognition. In contrast the substrate specificity of exporters is mediated by the

TMDs. It is fascinating how the diverse substrates found for ABC transporters can be translocated using a similar molecular assembly. And even more astonishing is the translocation of large substrates like the 107 kDa HlyA with this assembly.

Even though the structural and biochemical information on ABC transporters is growing, it still remains to be uncovered how the substrate can trigger the actual transport event. For importers it had been shown that the binding of the SBP to the transporter triggers ATP hydrolysis and subsequent import. However, due to the nature of the SBP to undergo an opening and closing motion in the absence of substrate, it remains a mystery how the ABC transporter can possibly sense the presence of substrate for effective triggering of the transport cycle.

To understand the substrate specificity of exporters, a structural characterization of HlyB, the ABC transporter fueling HlyA translocation in *E. coli*, was aimed for. As the toxin HlyA is rather large it was further anticipated that the three-dimensional structure might give an insight on how it is possible to translocate these substrates. In a first step, an expression and purification of the protein was to be established that would allow a successful isolation of purified HlyB yielding amounts suitable for protein crystallization trials. Finding conditions for the crystallization of a protein is a trial and error process. Therefore, in a second step, the purified protein is to be subjected to crystallization trials to evaluate the best conditions for crystallization and eventually obtain crystals that allow structure calculations.

To approach the answer of substrate specificity of ABC importers this work contains studies on ChoX, a choline binding protein from *S. meliloti*. It has been shown that choline as well as acetylcholine interacts with this binding protein. The structural determination of ChoX in complex with these two different substrates would allow an evaluation of the molecular basis of choline and acetylcholine binding. A structure of ChoX in complex with acetylcholine could be compared to other acetylcholine binding proteins, thereby possibly revealing interesting similarities to acetylcholine receptors. Besides, it was anticipated that structures of the SBP in absence and presence of substrate might reveal interesting differences on the surface of the binding protein that could help to shed light on the question how the presence of substrate can be effectively transmitted to the transporter and the adjacent NBD.

## 2 Materials and methods

### 2.1 Materials

#### 2.1.1 Chemicals

Acrylamide/N,N'-Methylenbisacrylamide solution 30% (w/v) / 0.8% (w/v)	Carl Roth
Ammoniumperoxodisulfate (APS)	Carl Roth
Ascorbate	Sigma
Acidic acid	Carl Roth
Bromphenolblue	Carl Roth
Chloramphenicol	Sigma
Coomassie Brilliant Blue R-250	Carl Roth
Coomassie Plus	Pierce
Enhanced chemiluminescence System	GE Healthcare
Ethanol	Carl Roth
Ethylenediamine-tetra-acidic acid (EDTA)	Carl Roth
Glucose	Carl Roth
Glycerol, 98%	Carl Roth
Glycine	Carl Roth
HEPES	Carl Roth
Hydrochloric acid, 37%	Carl Roth
Imidazole, BioChemical Ultra	Fluka
Magnesium chloride	Fluka
Magnesium sulfate	Fluka
Methanol	Carl Roth
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Carl Roth
Paraffin oil	Sigma
Peptone from beef	Carl Roth
Peptone from casein	Carl Roth
Peptone from soy	Carl Roth
Sodium chloride	Fluka
Sodium dodecyl sulfate (SDS)	Carl Roth
Sodium hydroxid	Fluka
Tris	Carl Roth
Tween 20	Carl Roth
Yeast extract	Carl Roth
$\beta$ -Glycerol phosphate	Sigma

### *Detergents*

The detergents utilized for the initial detergent screen were part of the Solution Master Detergent KIT purchased from Anatrace.

Detergents used for purification were purchased choosing the highest purity available.

Decyl- $\beta$ -D-maltopyranoside (DM)	Glycon
Dodecyl-N,N-dimethylglycine	Anatrace
Dodecyl- $\beta$ -D-maltopyranoside (DDM)	Glycon
FOS-CHOLINE14	SynphaBase
Tetradecyl-N,N-dimethylamine-N-oxide (TDAO)	Anatrace

### *Chemicals for protein crystallization*

Ammonium acetate	Fluka
Ammonium sulfate	Fluka
Benzamidine	Fluka
BICINE	Fluka
Cacodylate	Fluka
Calcium chloride	Hampton Research
Citric acid anhydrous	Fluka
Ethylenglycol	Sigma
Glycerol	Carl Roth
HEPES	Fluka
Imidazol	Fluka
Lithium sulfate	Fluka
Magnesium acetate	Sigma
Magnesium chloride	Fluka
Magnesium sulfate	Fluka
MES	Fluka
MPD	Hampton Research
Nickel chloride	Sigma
PEG 400	Hampton Research
PEG 550 MME	Fluka
PEG 1000	Fluka
PEG 1500	Fluka
PEG 2000 MME	Fluka
PEG 3350	Hampton Research
PEG 4000	Fluka
PEG 6000	Fluka
Sodium acetate	Fluka
Sodium chloride	Fluka
Sodium dihydrogen citrate anhydrous	Fluka
Tris	Fluka
Zinc sulfate	Riedel de Haen
1,2,3-Heptantriol	Fluka
2-Methylpentandiol	Fluka
2-tert-Butanol	Fluka

### *Kits for protein crystallization*

Crystallization Basic Kit for Proteins	Sigma
Crystallization Extension Kit for Proteins	Sigma
MembFac	Hampton Research
The Mb Class	Qiagen
The Mb Class 2	Qiagen

### *Substrates used in protein crystallization*

Acetylcholine	Sigma
Adenosine-5'-diphosphate (ADP)	Sigma
Adenosine-5'-triphosphate (ATP)	Sigma
Choline	Sigma

## **2.1.2 Enzymes and standards and antibodies**

Benzonase	Merck
Bovine serum albumine solution	Pierce
Desoxyribonuclease I (DNase I)	Fluka
Lysozyme	Fluka
$\alpha$ -HlyB from rabbit	provided by Prof. K. Kuchler
$\alpha$ -rabbit-HRP from mouse	Dianova

### *Molecular weight marker for SDS-PAGE*

Prestained Molecular Weight Marker	MBI Fermentas
------------------------------------	---------------

### *Molecular weight marker for gel filtration chromatography*

Molecular Weight Marker Kit	Sigma
-----------------------------	-------

## **2.1.3 Microorganisms, plasmids and media**

### *Microorganisms*

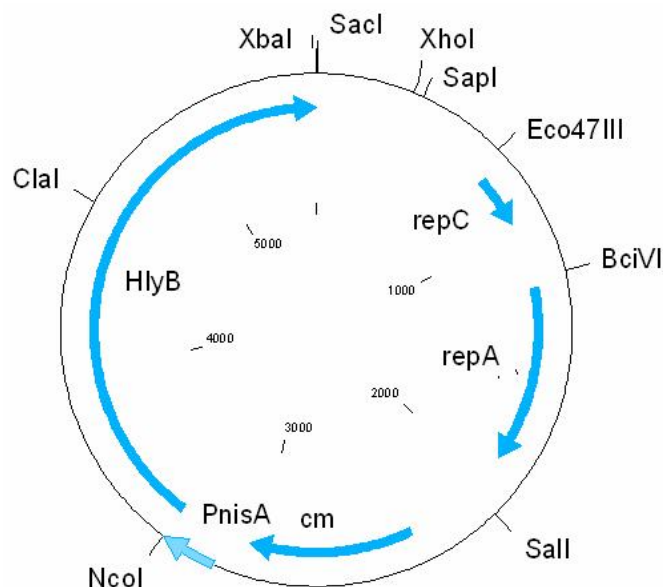
#### *L. lactis:*

<u>Strain</u>	<u>Reference</u>
NZ9000	114
NZ9700	114

#### *Plasmids*

<u>Plasmid</u>	<u>Reference</u>	<u>Resistance marker</u>
pNZ8048	115	Cm <sup>R</sup>

The gene inserted into the plasmid *pNZ8048* was either wild type HlyB with an N-terminal decahistidine tag or a mutant of HlyB (further called H662A mutant) comprising an N-terminal octahistidine tag and the mutations C4S, C653S and H662A. A vector card of the plasmid is shown in Figure 2.1



**Figure 2.1** Vector card of pNZ8048 with the inserted HlyB gene.

#### *Medium used for growth of L. lactis*

Medium and solutions for bacterial growth were prepared using deionized water and were either heat sterilized or sterile filtered. A 20% glucose solution was sterilized separately and added to a final concentration of 0.5% to the heat sterilized components of the M17-medium. When necessary, chloramphenicol was added shortly before inoculation of the media with bacteria.

<u>M17-Medium per liter</u>		<u>Chloramphenicol solution</u>		<u>Glucose solution</u>	
β-Glycerophosphate	19.0 g	Choramphenicol	30 mg/ml	Glucose	200 g/l
Peptone from casein	5.0 g	Ethanol	70 % (v/v)		
Peptone from soy	5.0 g				
Peptone from beef	5.0 g				
Yeast extract	2.5 g				
Ascorbate	0.5 g				
Magnesium sulfate	0.25 g				

### 2.1.4 Buffers and solutions

Buffers were prepared with MilliQ water, stored and supplemented with detergent prior to use.

<u>Buffer 1</u>		<u>Buffer 2</u>	
HEPES pH 7.4	50 mM	Tris/HCl pH 7.8	50 mM
NaCl	150 mM	NaCl	150 mM
		Glycerol	10% (v/v)
<u>Buffer 3</u>		<u>Buffer 4</u>	
Tris/HCl pH 7.8	50 mM	Tris/HCl pH 7.8	50 mM
NaCl	150 mM	NaCl	150 mM
Glycerol	10% (v/v)	Glycerol	10% (v/v)
Imidazole pH 8.0	10 mM	Imidazole pH 8.0	200 mM
<u>Buffer 5</u>			
HEPES pH 7.4	20 mM		
NaCl	100 mM		
Glycerol	10% (v/v)		

#### *Solutions for SDS-PAGE*

<u>Separation-gel buffer</u>		<u>Stacking-gel buffer</u>	
Tris/HCl pH 8.85	1.5 M	Tris/HCl pH 6.8	0.5 M
Sodium dodecyl sulfate	0.4% (w/v)	Sodium dodecyl sulfate	0.4% (w/v)
<u>10x Running buffer</u>		<u>Sample buffer</u>	
Tris	0.25 M	Tris/HCl pH 6.8	0.1M
Glycine	1.9 M	SDS	4%
Sodium dodecyl sulfate	1% (w/v)	Bromphenol blue	0.02%
pH adjusted to 8.3		Glycerol	20%
<u>Coomassie stain</u>		<u>Destain</u>	
Acidic acid	10% (v/v)	Acidic acid	10% (v/v)
Methanol	40% (v/v)	Methanol	40% (v/v)
Coomassie Brilliant	0.25% (w/v)		
Blue -250			

## *Solutions for Western blotting*

<u>Transfer buffer</u>		<u>TBS</u>	
10x Running buffer	10% (v/v)	NaCl	250 mM
Methanol	20% (v/v)	Tris/HCl pH 8.0	20 mM
H <sub>2</sub> O	70% (w/v)		
<u>TBS-T</u>		<u>Blocking solution</u>	
NaCl	250 mM	3% BSA in TBS-T	
Tris/HCl pH 8.0	20 mM		
Tween 20	0.1% (v/v)		

## **2.1.5 Chromatographic materials and others**

### *Affinity chromatography*

IDA Hi Trap Chelating column, 1ml	GE Healthcare
-----------------------------------	---------------

### *Gelfiltration chromatography*

Superdex 200 10/300 GL	GE Healthcare
------------------------	---------------

### *Others*

Amicon Ultra 15 (MWCO 100 kDa)	Millipore
CryoLoop	Hampton Research
CrystalDrop pregreased lid	Greiner Bio-One
CrystalQuick plate	Greiner Bio-One
Crystal clear sealing tape	Jena Bioscience
24-well crystallization plate	Greiner Bio-One
Dispo Dialyzer, 1ml (MWCO 50kDa)	Spectrum
PVDF membrane	Pall life Science
Whatman-paper	Schleicher & Schüll
96-well micro plate	Greiner Bio-One
Spin-X, 0.22 µm	Corning

## **2.1.6 Instruments**

### *Chromatographic instruments*

FPLC System	GE Healthcare
Äkta Basic	GE Healthcare

### *Optical instruments*

ELISA plate reader Polarstar Galaxy	BMG Labortechnologie
CHEMI GENIUS <sup>2</sup>	Syngene

### *Centrifuges*

Sorvall Evolution RC	Thermo
Eppendorf 5417R	Eppendorf
Sorvall Discovery M120 SE	Thermo
Sorvall Discovery 90 SE	Thermo

### *Electrophoretic instruments*

Power supply EPS 200	GE Healthcare
Electrophoresis unit	University Düsseldorf
Semidry Blot system	BIO-RAD

### *Incubator*

Incubator	Heraeus
Shaking incubator	Infors

### *Others*

Basic Z System cell disruptor	Constant Systems
Balances	Satorius
Milli-Q <sup>50</sup> Plus	Millipore
SDS-PAGE gel casting unit	University Düsseldorf
Dewar	Taylor Wharton

## **2.2 Microbiological methods**

### **2.2.1 The expression system**

HlyB is heterologously overexpressed in the *L. lactis* strain NZ9000<sup>115</sup>. The plasmid *pNZ8048* comprising a chloramphenicol resistance gene and a *nisA* promoter<sup>116</sup> is used for the overexpression of the ABC transporter. The *HlyB* gene is located directly after the *nisA* promoter, and induced by addition of nisin A. The wild type *HlyB* gene encoded for a decahistidine tag, whereas the mutant H662A *HlyB* gene encoded for an octahistidine tag.

### **2.2.2 Overnight culture of *L. lactis***

An overnight culture of *L. lactis* NZ9000 is prepared by inoculation of 200 ml sterile M17-medium supplemented with 0.5 % glucose and 5 µg/ml chloramphenicol. Either *L. lactis* from cryostocks or freshly transformed cells grown on agar plates were used. The culture was incubated at 30 °C for 14-16 hours. Overnight cultures of NZ9700 were prepared in the same way without addition of chloramphenicol.

### **2.2.3 Preparation of cryostocks**

Cryostocks were prepared by mixing 800 µl of an overnight culture with 200 µl of sterile glycerol in a 1.5 ml cryo-tube. The mixture was frozen in liquid nitrogen and stored at -80 °C.

### **2.2.4 Nisin A production**

Nisin A was produced utilizing the *L. lactis* strain NZ9700. 20 ml of an overnight culture were used to inoculate one liter of M17-medium. The culture was grown at 30 °C for 8-10 hours. The cells were pelleted for 15 min at 6000 g and 4 °C and the supernatant, containing the secreted nisin A, was stored at -20 °C.

### **2.2.5 Expression of HlyB in *L. lactis***

HlyB was expressed in the *L. lactis* strain NZ9000 transformed with plasmid *pNZ8048*. Two liters of sterile M17-medium supplemented with 5 µg/ml chloramphenicol and 0.5% glucose was inoculated with 40 ml of an overnight culture. Cells were incubated at 30 °C and shaken slightly every hour to allow for aeration. The growth of the cells was monitored by determination of the optical density at 600 nm (OD<sub>600</sub>). When the cells reached an OD<sub>600</sub> of 1.0 the expression was induced by adding the nisin A containing supernatant to a final concentration of 0.4%. The culture was further incubated for another 1.5 hours to allow for protein expression. Then, the cultures were cooled on ice. Subsequently, the cells were harvested by centrifugation for 15 min at 6000 g and 4 °C. The supernatant was discarded. The obtained cell pellet was washed by resuspending the cells in 100 ml (per cells of 2 liter culture) of buffer 1 supplemented with 1 mM PMSF and 1 mM EDTA. The suspension was centrifuged (15 min, 6000 g at 4 °C) and the obtained cell pellet was washed a second time. After centrifugation, the obtained cell pellet was stored at -20 °C.

## **2.3 Biochemical methods**

### **2.3.1 Determination of protein concentration**

A modified Bradford assay was utilized to assess the concentration of a protein solution. 300 µl of a commercially available Coomassie Plus solution were mixed with 10 µl of protein solution in a 96-well micro plate. As reference a bovine serum albumin (BSA) standard containing BSA at concentrations from 0-2 mg/ml was also placed in the plate and mixed with the Coomassie Plus solution. After 10 min of incubation at room temperature the protein-Coomassie mixture was spectroscopically analyzed by measuring the absorption at 584 nm of the solution with an ELISA plate reader. The absorption of the standard containing no BSA was used as background and subtracted from the other absorption values. The absorption values of the BSA standard were plotted against the protein concentrations and utilized to determine a standard calibration curve. The standard curve was utilized to calculate the concentration of the protein solutions.

If the protein concentration was too high and the absorption was saturated, the protein solution was diluted prior to the concentration determination.

### **2.3.2 Concentrating a protein sample**

Protein solutions were concentrated with an Amicon Ultra concentrator containing a low binding regenerated cellulose membrane. The concentrator was pretreated with buffer to minimize protein loss. Up to 15 ml of protein solution were placed in the concentrator and centrifuged at 3000 g until the desired concentration was reached.

### **2.3.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)**

The SDS-PAGE was utilized to assess the purity of a protein sample. The protein sample is denatured by the sample buffer and concentrated in the stacking gel to form a sharp band before it enters the 10% separation gel, where the proteins are separated according to their molecular weight.

Glass plates for nine gels were placed in a SDS-PAGE casting unit. 18.75 ml of separation buffer was mixed with 25 ml of 30% acrylamide solution, 31.25 ml of deionized water, 300 µl 10% (w/v) APS solution and 70 µl TEMED. The solution was immediately poured into the prepared casting unit. When the separation gel was polymerized the stacking gel was prepared

by mixing 5 ml of stacking buffer with 5.25 ml of 30% acrylamide solution, 28 ml of deionized water, 210  $\mu$ l of 10% (w/v) APS and 70  $\mu$ l TEMED. This solution was poured on top of the separation gel and teflon spacers were fitted between two glass plates to form the sample application wells. Fully polymerized gels were stored at 4 °C.

For the separation of the protein sample 20  $\mu$ l of protein solution was mixed with 5  $\mu$ l 5x sample buffer. One polyacrylamide gel was placed into the electrophoresis unit and the chambers were filled with running buffer. The teflon spacer was removed and 20  $\mu$ l of the prepared protein sample was filled into the sample application wells. The electrophoresis was performed at 160 V for 1.5 hours.

When the electrophoresis process was finished the gel was stained with Coomassie stain for ~1 hour. Unbound Coomassie dye was removed by adding destain solution until the background of the gel became clear.

#### **2.3.4 Immuno-blotting**

In a first step the proteins were separated by SDS-PAGE. Whatman filter paper was soaked in transfer buffer and placed on a Semidry Blot system (BIO-RAD). The PVDF membrane was treated with pure methanol and placed on layers of the soaked filter paper. Subsequently, the polyacrylamide gel was put on the membrane and topped with soaked filter paper. By applying a constant current of 100 mA per gel the proteins were transferred from the gel to the membrane in 1 hour.

After the immobilization of the proteins on the PVDF membrane the membrane was placed in blocking solution to diminish unspecific binding of the antibody to the membrane. Then, the membrane was shaken for 1 hour in a solution of the first antibody (anti-HlyB) diluted 1:200 in TBS-T. Thereafter, to remove all unbound antibody, the membrane was washed three times for 10 min in TBS-T before it was placed for another hour in a 1: 20000 dilution of the second antibody (anti-rabbit) in TBS-T. The second antibody was conjugated to a horseradish peroxidase. After washing the membrane twice in TBS-T and once in TBS for 10 min the peroxidase reaction was induced with the enhanced chemiluminescence System (GE Healthcare). The light emission was detected with a CHEMI GENIUS<sup>2</sup>.

#### **2.3.5 Preparation of membranes**

Usually, cells of 16 liters of culture were used to prepare membranes, which were stored at

-80 °C. The stored cell pellet was defrosted and thereafter resuspended in 100 ml of buffer 1 supplemented with 1 mM PMSF and 1 mM EDTA. 10 mg/ml lysozyme was added and the resuspension was incubated at 30 °C for 30 min. To break down DNA benzonase and 6 mM MgSO<sub>4</sub> was added and left to react at room temperature for 5 min. The cells were cooled on ice for 10 min. Cell rupture was performed by four times passage of the cell suspension through a cooled cell disruptor (Constant Systems) using a pressure of 2.7 kbar. The obtained solution was centrifuged for 40 min at 13000 g and 4 °C to remove cell debris and unbroken cells. A second centrifugation of the supernatant at 28700 g for 45 min at 4 °C guaranteed for removal of cell debris. The obtained supernatant was subjected to ultracentrifugation for 1 hour at 200000 g and 4 °C to pellet the cell membranes. 10 ml of buffer 2 was added and the membranes were homogenized with a potter on ice. The concentration of the membranes was adjusted to 15-20 mg/ml and the membranes were stored at -80 °C.

### 2.3.6 Solubilization screen

Membranes with a concentration of 15-20 mg/ml were defrosted. 90 µl were mixed with 10 µl of a 10 % detergent solution. The utilized detergents, listed in Table 2.1, were part of the Solution Master Detergent KIT (Anatrace). The protein/detergent mixture was incubated for 30 min on ice to allow for solubilization of the protein. Cloudy solutions were warmed to 30 °C for a few minutes to bring the detergent back to solution. Unsolubilized membranes were removed by centrifugation for 1 hour at 200000 g and 4 °C. Samples of the protein/detergent mixture before and after centrifugation were taken to compare on a western blot. The samples were diluted 1:10 with buffer 2 and 10 µl of each sample was used for the SDS-PAGE and subsequent western blotting.

Detergent	Detergent	Detergent
ANAPOE®-20	CYMAL®-7	n-Decyl-β-D-thiomaltopyranoside
ANAPOE®-35	FOS-CHOLINE®-12	n-Undecyl-β-D-thiomaltopyranoside
ANAPOE®-C12E9	FOS-CHOLINE®-14	n-Dodecyl-β-D-thiomaltopyranoside
ANAPOE®-C12E10	FOS-CHOLINE®-16	n-Dodecyl-N,N-dimethylglycine
ANAPOE®-C13E8	FOS-MEA®-10	Octaethylene glycol monododecyl ether (C12E8)
ANAPOE®-X-100	Dodecyltrimethylammoniumchloride	Pentaethylene glycol monododecyl ether (C10E5)
ANAPOE®-X-114	Hexadecyltrimethylammoniumchloride	Sucrose monododecanoate
ANZERGENT® 3-14	Octadecyltrimethylammoniumchloride	Dimethyldodecylphosphine oxide
CYCLOFOS™-6	n-Undecyl-α-D-maltopyranoside	n-Tetradecyl-N,N-dimethylamine-N-oxide (TDAO)
CYCLOFOS™-7	n-Undecyl-β-D-maltopyranoside	n-Dodecyl-N,N-dimethylamine-N-oxide (DDAO)
CYMAL®-6	n-Dodecyl-α-D-maltopyranoside	

**Table 2.1** List of detergents employed in the solubilization screen.

### 2.3.7 Purification of HlyB

The membrane suspension prepared from cells of 8 liters of bacteria culture was defrosted. FOS-CHOLINE14 was added to a final concentration of 1% (w/v) and the mixture was stirred at 8 °C for 30 min. The unsolublized membranes were removed by sedimentation for 1 hour at 200000 g and 4 °C. The supernatant containing the HlyB/detergent micelles was further purified in an immobilized metal ion affinity chromatography (IMAC) step using IDA material loaded with  $\text{Zn}^{2+}$ . The  $\text{Zn}^{2+}$ /IDA complex makes interactions with the N-terminal polyhistidine tag. Three high trap chelating columns (GE Healthcare) were loaded with  $\text{Zn}^{2+}$  and equilibrated with buffer 3 supplemented with 0.0115% (w/v) FOS-CHOLINE14. The protein/detergent solution was applied on a series of three columns and non-specifically bound protein was washed off with three column volumes (CV) buffer 3 supplemented with 0.0115% FOS-CHOLINE14. A step gradient of three CV using 15% (or 20% for wild type HlyB) of buffer 3 and 85% (or 80% for wild type HlyB) of buffer 4, both buffers supplemented with 0.0115% FOS-CHOLINE14, was applied to remove untightly bound impurities. The protein was eluted with 10 CV of buffer 4 containing 0.0115% FOS-CHOLINE14. Fractions of 1 ml were collected, their concentration determined by Bradford assay and their purity assessed by SDS-PAGE.

The peak fractions were pooled and concentrated to 250-300  $\mu\text{l}$  with an Amicon Ultra concentrator containing a low binding regenerated cellulose membrane (Millipore). The concentrated sample was dialyzed against buffer 5 supplemented with 0.0115% FOS-CHOLINE14. The protein solution was placed in a Dispo Dialyzer and dialyzed for 16 hours against this buffer, then the buffer was changed and the protein solution was further dialyzed for another 2-4 hours at 8 °C. Finally the concentration of the sample was determined.

### 2.3.8 Gelfiltration chromatography

The dispersity of the prepared sample was analyzed by gelfiltration chromatography. A Superdex 200 10/300 GL column was used in an Äkta System. The column was equilibrated with 1 CV of buffer 5 supplemented with the desired detergent. 500  $\mu\text{l}$  of protein sample was applied on the column at a flow of 0.5 ml/min. The peak (maximum at 11 ml) was collected in 1 ml fractions.

The column was calibrated with a mixture of soluble proteins containing Carbonic anhydrase (29 kDa), Albumin (66 kDa), Alcohol dehydrogenase (150 kDa),  $\beta$ -Amylase (200 kDa), Apoferrin (443 kDa) and Tyroglobulin (669 kDa).

### 2.3.9 Detergent exchange

In an attempt to find a suitable detergent for protein purification, HlyB was solubilized in FOS-CHOLINE14. The protein was immobilized on a  $\text{Zn}^{2+}$ /IDA column and the FOS-CHOLINE14 detergent micelle was exchanged against a different detergent micelle by washing with buffer containing the new detergent. The detergents employed in the analysis were dodecyl- $\beta$ -D-maltopyranoside, decyl- $\beta$ -D-maltopyranoside, tetradecyl-N,N-dimethylamine-N-oxide and dodecyl-N,N-dimethylglycine.

Membranes (of 2 liter cell culture) were treated as described in 2.3.7. After ultracentrifugation the supernatant was applied on one high trap chelating column, which was loaded with  $\text{Zn}^{2+}$  and equilibrated with buffer 3 supplemented with 0.0115% (w/v) FOS-CHOLINE14. Subsequently, the column was washed with 20 CV of buffer 3 containing 0.0115% FOS-CHOLINE14 and afterwards with 14 CV of buffer 3 supplemented with the new detergent at 2.5 x CMC (critical micelle concentration). Thereafter, the protein was eluted with 10 CV of buffer 4 containing the new detergent at 2.5 x CMC.

The purity of the sample was analyzed by 10 % SDS-PAGE and the protein yield was assessed by a Bradford assay. The dispersity of the sample was analyzed by gel filtration chromatography using a Superdex 200 10/300 GL. For this step the protein was concentrated in an Amicon Ultra concentrator prior to application on the gel filtration column. The column was equilibrated and run with buffer 5 supplemented with the new detergent at 2.5 x CMC.

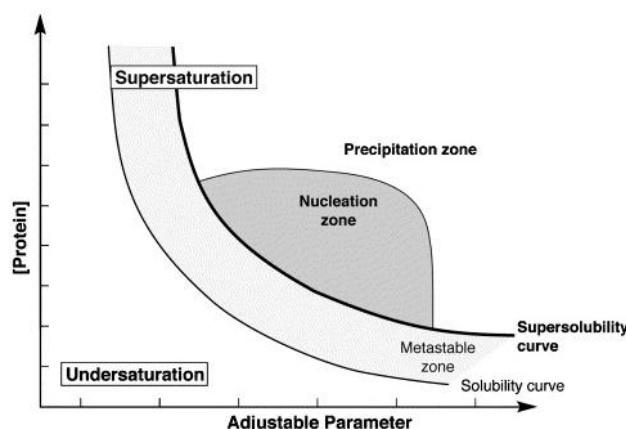
## 2.4 Crystallographic methods

### 2.4.1 Fundamentals of crystal growth

Although the use of robotics has greatly advanced in the last decade, obtaining protein crystals still represents the bottleneck in X-ray crystallography. Growing protein crystals almost always involves excessive screening for suitable crystallization conditions as the search represents an empirical process. Crystal growth can be separated into two steps, nucleation and the enlargement of the nuclei<sup>117</sup>, two processes occurring at certain regions in the phase diagram of the protein solution. Therefore, obtaining information on the phase behavior of a solution helps to find suitable crystallization conditions.

The solubility curve divides the phase diagram of a protein solution into two zones, the undersaturated and the supersaturated zone (see Figure 2.2). At the solubility curve the protein

solution is in equilibrium between solid and dissolved matter. The area under this curve represents the undersaturated zone, where neither nucleation nor crystals growth can emerge. Above the solubility curve, in the supersaturated zone, crystals can grow and nuclei form. However, nucleation and crystal growth appear at different regions within the supersaturated area. Nuclei arise in the nucleation zone and, further, may grow there as well. In contrast, the metastable zone enables crystal growth but not nucleation. If very high supersaturation is reached, the nucleation process, which can be described as ordered precipitation, might be overruled by unordered precipitation. Thus, amorphous precipitate, a phenomenon often encountered in protein crystallization experiments, is formed. Therefore, the ideal crystallization experiment establishes conditions at which the protein reaches a level of supersaturation without forming amorphous precipitant but at which formation of a few nuclei is maintained. Once nuclei formed, the concentration of the protein in solution is lowered and ideally drops back into the metastable zone, where further growth of the nuclei is supported.



**Figure 2.2** Example of a phase diagram of a protein solution. The solubility curve divides the phase diagram in an undersaturated and a supersaturated zone. The supersaturated zone can be further divided into the metastable, the nucleation and the precipitation zone. Taken from Chayen 2005<sup>117</sup>.

Even with the theoretical knowledge of nuclei formation and crystal growth, it is impossible to predict a priori which conditions are best to be used for protein crystallization. In theory, the phase diagram of the protein in every crystallization experiment would need to be established and evaluated to find the most suitable condition for nucleation and growth. However, knowing a few points in the phase diagram facilitates the empirical process of establishing crystallization conditions.

### *Means of reaching supersaturation*

For a successful crystallization experiment a protein solution of a certain concentration has to reach the supersaturated zone. This is achieved by manipulation of different variables which influence the solubility of the protein solution. For example:

- temperature
- salt concentration (higher: “salting out”; lower: “salting in”)
- pH-value
- addition of polymers

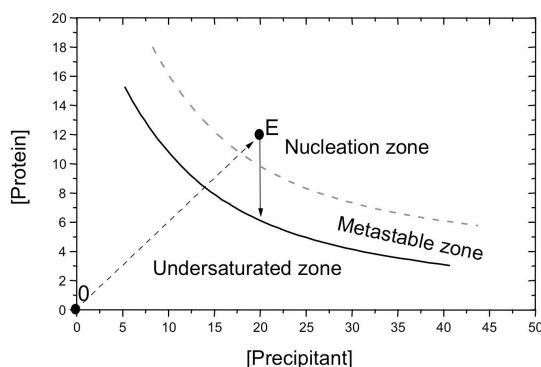
For a crystallization experiment, one or more of these variables are effected in the crystallization setup. The precipitant itself may be a salt at high concentration, a polymer or an organic solvent.

### *Batch/microbatch crystallization*

In earlier days the batch crystallization was applied to crystallize substances. Supersaturation is achieved by directly mixing the precipitant solution with the protein solution. The phase diagram in Figure 2.3 exemplifies the process of a successful batch experiment. One drawback of this method is that only one point, the endpoint E which is reached directly after mixing the components, is investigated in the experiment. If E lies within the nucleation zone, the protein concentration in solution is lowered and possibly the metastable zone is reached. Since the invention of vapor diffusion methods the batch crystallization has lost its importance.

Nevertheless, in the 90ies, the batch method has gained new interest as so-called microbatch experiments<sup>118,119</sup>. These experiments are conducted under oil. The protein and precipitant are mixed directly under a layer of paraffin oil. As the oil seals the experiment and circumvents evaporation, the setup is well suitable for small volumes. A way of modifying the experiment is to use mixtures of silicon and paraffin oil<sup>120</sup>. Silicon oil does not seal the drop completely and the drop can dry out, whereas paraffin oil allows for very little evaporation. Mixing the two components in numerous ratios generates the possibility to combine the batch experiment with evaporation (see also vapor diffusion), thus making it possible to vary the point of investigation (E) during the experiment. One disadvantage of the microbatch method is that no volatile organic solvents (e.g. dioxane or phenol) or organic compounds which interact with the oil can be applied in the experiment. Furthermore, it is not possible to use PEGs at very high concentrations (above 30%) as they partition into the oil. However, the setup is easily and quickly performed and allows for the use of robotics<sup>121</sup>. Though not anticipated, for

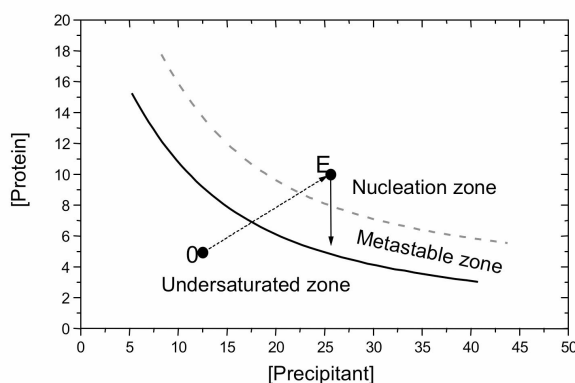
dodecylmaltoside it was shown that hardly any detergent is migrating into the oil<sup>122</sup>. Therefore, this method is ideal for membrane protein crystallization.



**Figure 2.3** The process of a successful batch experiment exemplified in the phase diagram.

### *Crystallization by vapor diffusion*

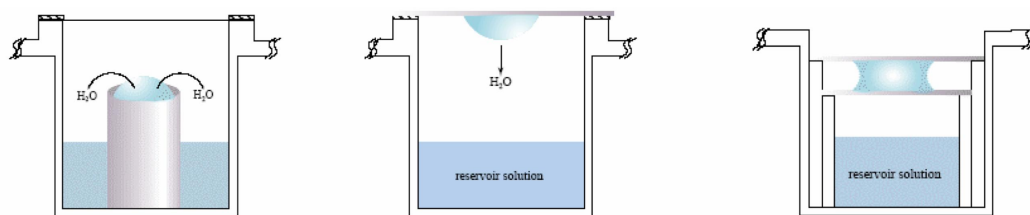
In contrast to the batch method, the vapor diffusion experiment explores a wider area of the phase diagram (Figure 2.4) thereby increasing the probability of reaching the nucleation zone. For this experiment the protein is mixed with a precipitant solution (point 0 in the phase diagram). Then a drop of this mixture is placed over a reservoir, containing the precipitant solution. The experiment is sealed and the drop is left to equilibrate with the reservoir solution, which exceeds the volume of the drop by 100-500 times. In an ideal vapor diffusion setup the equilibration process governs the protein solution through the phase diagram, from the undersaturated zone to the nucleation zone (endpoint E). During the equilibration nuclei can form, thus the protein concentration in solution is lowered and the crystals can grow.



**Figure 2.4** The process of a successful vapor diffusion experiment exemplified in the phase diagram.

The vapor diffusion method can be performed as sitting drop, hanging drop or sandwich drop as depicted in Figure 2.5 on the left, middle and right, respectively. The drop surface

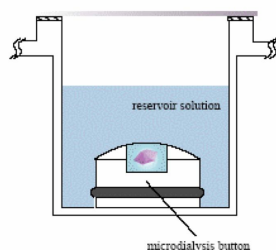
influences the kinetic of reaching the equilibrium between the drop and the reservoir, thereby often influencing the amount of crystals as well as their size.



**Figure 2.5** Different methods of setting up a vapor diffusion experiment. The sitting drop (left), the hanging drop (middle) and the sandwich drop (right) are exemplified. Adapted from Hampton Research.

### *Dialysis*

The dialysis experiment represents a rather tedious setup and necessitates high quantities of protein. This is probably the reason why in the last decade the application of the dialysis experiment became less important. However, if the protein shows an inverse phase diagram, the dialysis experiment represents a useful method for crystallization. In this technique the protein solution is placed in a microdialysis bottom (Figure 2.6) and the setup is covered with a semipermeable dialysis membrane. By placing the microdialysis button in water, the protein solution is slowly equilibrated against the surrounding solution, thus lowering the buffer and salt concentration of the setup (“salting in” effect). When the nucleation zone is reached crystals can start to form and, once reaching the metastable zone, grow to their final size. Furthermore, proteins with a “normal” phase diagram can also be crystallized with the dialysis method. In this case, a precipitant solution, containing precipitant, buffer and/or salt, is used as dialysis solution. However, this “salting out” application has been widely replaced by the vapor diffusion experiment.



**Figure 2.6** A dialysis setup. The protein solution is placed in the microdialysis button and dialyzed against a reservoir. Adapted from Hampton Research.

### *Choice of crystallization conditions*

Different proteins in various solutions show different behavior. In principle, it is impossible to predict a phase diagram for a protein solution a prior. Thus, to explore the best crystallization

conditions for a protein, an empirical approach based on trial and error has to be utilized. Often commercially available screens are employed in initial crystallization experiments.

A grid screen is a systematic approach to evaluate crystallization conditions. In grid screens two parameters are varied, while the others are kept constant. Hereby, the steps can be varied in their size allowing for a broader or a more narrow screening. However, a disadvantage of the grid screen is the necessity to explore many different conditions before a successful condition is found. With the emergence of sparse matrix screens<sup>123</sup> in the 90ies the application of grid screens was greatly reduced. Nowadays, mostly optimization of an initial successful condition is performed based on grid screening.

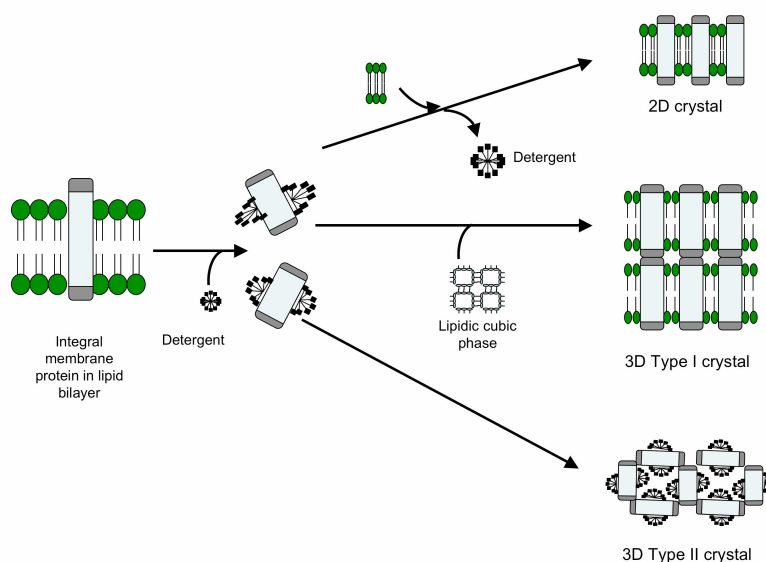
In 1991 Jancarik and Kim published a first set of 50 independent, “arbitrary” crystallization conditions<sup>123</sup>, thereby starting a new era in protein crystallization. These conditions were empirically derived based on the knowledge of so-far successful crystallization conditions. Sparse matrix screens are composed of solutions which are very different in their composition. Thus, a broad variety of different buffers, salts and precipitants can be evaluated in one screen. Sparse matrix screens are therefore often utilized in initial crystallization experiments.

#### *Membrane protein crystallization*

Crystals form an ordered array of a specific unit. In the case of soluble proteins this unit is composed of only protein and possibly substrate. However, when integral membrane protein crystals are formed, this unit does not only exhibit protein (and substrate) but also detergent and likely lipid molecules. Detergents generate micelles, which form a belt around the hydrophobic part of the membrane protein and represent an obstacle in the crystallization process, as they do not mediate crystal contacts. At certain concentrations detergent molecules form micelles by shielding the hydrophobic tails and exposing the hydrophilic head groups to the solvent. Furthermore, the complex phase behavior of detergents also has to be considered, when crystallizing integral membrane proteins.

Figure 2.7 gives a schematic impression, how membrane protein crystallization can be conducted. In a first step, the integral membrane protein has to be extracted from its natural environment, the lipid bilayer, and purified. This is achieved with the help of lipid like molecules, the detergents. Re-introducing lipids to the purified protein and removing detergent simultaneously facilitates 2D crystallization. Two different methods can be pursued to obtain 3D crystals. Type II crystals are built up of stacked micelles and are generally

formed when the protein is crystallized as a protein/detergent complex. A more tedious and less commonly used method is the crystallization in lipidic cubic phases<sup>124</sup>. Here a mixture of monoolein or monopalmitolein and aqueous buffers is used to form a cubic phase. By addition of the protein/detergent complex the integral membrane protein is inserted to the cubic phase. The use of precipitants induces the formation of a lamellar phase, which allows for a layered stacking of the native-like membranes. The crystals formed by this method are so-called type I crystals. Dependent on which method is used, it is possible to crystallize integral membrane proteins as type I as well as type II crystals. As the generation of cubic phases is rather time consuming and a large amount of protein is needed for this method, it is more common to grow type II crystals instead of type I crystals.



**Figure 2.7** Possible ways of crystallizing a membrane protein. In a first step the protein is solubilized in detergent and purified. By reintroducing lipids 2D crystals can be grown. 3D crystals are obtained by directly crystallizing the protein/detergent complex or by adding lipidic cubic phases. Adapted from Bergfors<sup>125</sup>.

## 2.4.2 Crystallization of HlyB

Crystallization of the integral membrane protein HlyB was performed as vapor diffusion experiments in 96-well hanging drop as well as in 96-well sitting drop setups. The concentration of the protein solution was adjusted to 7-15 mg/ml. The hanging drop setup was performed in CrystalQuick plates (Greiner), where the reservoir well was filled with 200  $\mu$ l motherliquid. 1  $\mu$ l of the protein solution was placed on a pregreased CrystalDrop lid and mixed in a 1:1 ratio with precipitant. Then, the lid with the crystallization setups was placed over the reservoirs of the CrysalQuick plate and sealed airtight. The plates were stored either

at 277 K or at 293 K. Sitting drop setups were performed with Corning CrystalEX microplates with conical flat bottom. As the reservoir well is smaller in these plates only 150  $\mu$ l precipitant was used in each setup. The plates were sealed with Crystal Quick Sealing tape.

Initial crystallization trials were performed with the commercial screens MbClass and MbClass II Suites from Nextal, Crystallization Basic and Extension Kit for Proteins from Sigma and a self-made kit based on JBScreen Classic 1, 2 and 7 from Jena Bioscience.

Prior to the crystallization setup the protein solution was filtered through a Spin-x cellulose acetate filter (7 min at 9000 g, 4 °C). CocrySTALLIZATION trials were performed with wild type HlyB and ADP as well as with the H662A mutant and ATP. The cocrySTALLIZATION trials were performed by mixing the protein solution with the ligand and leaving it to react on ice for 30 min prior to setting up the drops. The final concentration of the substrate was adjusted to 5 mM. A few experiments were carried out with 10 mM Na-azide in addition to 5 mM ATP as substrate.

To optimize promising conditions, grid screens around the initial conditions were constructed and crystallization setups were performed in 24-well plates (for details see crystallization of ChoX). As an alternative to the grid screening method, screening with commercially available screens as additives to the initial condition was performed<sup>126</sup>. Here, 75% of the initial precipitant was mixed with 25% of the additive precipitant.

### **2.4.3 Crystallization of ChoX**

Crystallization experiments with ChoX and various substrates as well as in unliganded states were performed as vapor diffusion experiments in a 24-well setup. The protein was provided in a 100 mM Tris/HCl pH 7.0 buffer by the group of Prof. Dr. E. Bremer (Marburg University). Setups were performed in this buffer at concentrations varying from 10-15 mg/ml. Protein was mixed with substrate and incubated on ice for 30 min. Considering the  $K_d$  values for the different substrates, the final ligand concentration was chosen to be 1 mM for choline and acetylcholine. The protein/substrate solution was filtered through a Spin-x cellulose acetate filter (7 min at 9000 g, 4 °C) prior to setup. 1 ml of precipitant solution was placed in the reservoir wells of 24-well crystallization plates, which were greased with high vacuum grease prior to use. 1  $\mu$ l of protein solution was mixed with 1  $\mu$ l of precipitant solution on a glass cover slide. This slide was placed over the reservoir and sealed tightly. Crystallization plates were either stored at 277 K or at 293 K. For initial screening sparse matrix screens from Sigma (Crystallization Basic and Extension Kit for Proteins) were used and crystal optimization was performed using grid screens around initial conditions.

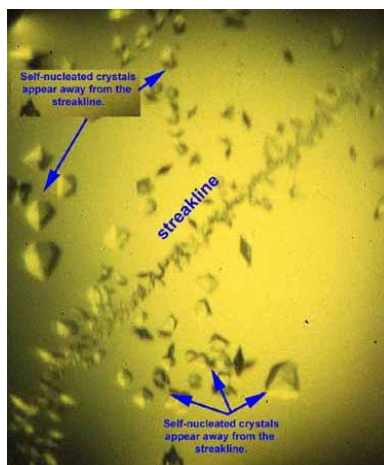
Crystallization of Se-Met substituted protein was carried out at a concentration of 9 mg/ml in a buffer containing 10 mM Tris/HCl pH 7.0. The Se-Met substituted protein was treated like native ChoX, i.e. protein was mixed with choline and acetylcholine, incubated on ice and filtered. The successful crystallization conditions of the native ChoX/choline provided a starting point for the grid screens designed to crystallize the Se-Met substituted ChoX. Vapor diffusion experiments were carried out in a 24-well crystallization plate as a hanging drop experiment.

#### **2.4.4 Crystallization by seeding**

##### *Fundamentals*

The formation of crystal nuclei is more likely to occur in a highly supersaturated phase. However, the ordered growth of crystals is supported best in the metastable zone. The method of seeding crystals has been established to separate the two events of nucleation and crystal growth<sup>127</sup>. This is accomplished by transferring a seed, a crystal or other solid material on which further growth can occur, from one condition, where the level of supersaturation is high, to a similar condition at a lower level of supersaturation. This state can be obtained by lowering the precipitant or protein concentration<sup>128</sup>.

Generally, it can be distinguished between macroseeding and microseeding. In macroseeding a single crystal is transferred which is washed in stabilizing solutions before it is placed in the equilibrated new setup. However, this method is very time consuming and tedious. Microseeding utilizes submicroscopic nuclei as seeds. These can be prepared by smashing a crystal in a stabilizing solution and preparing dilution stocks, which are added to the new crystallization setup. Smashing the crystal can be accomplished in different ways<sup>129</sup>, for example by vortexing, sonification or with seed beads. Another technique inoculating the new drop with microseeds is the so-called streak seeding method<sup>130</sup>. Here, a crystal is touched with a thick fiber, often a cat whisker or a horse tail hair that can trap seeds of the crystals in its overlapping cuticles<sup>127</sup>. Then, the hair is streaked through the new setup. If the concentration of the new setup is not chosen too low, where crystals tend to melt, or to high, which gives rise to spontaneous nucleation, crystals can grow along the streak line as seed induced nucleation. Self induced nucleation may also occur somewhere in the drop<sup>130</sup>. A typical, successful streak seeding experiment is shown in Figure 2.8.



**Figure 2.8** Crystals grown by streak seeding. Microcrystals, along the streakline, facilitate crystal growth. Self induced nucleation appears further from the streakline. From Bergfors<sup>131</sup>.

The advantage of the seeding method becomes obvious when nucleation is the most time consuming step of the crystallization experiment. The whole experiment can be accelerated when nucleation is uncoupled from crystal growth. Thus, it is possible to optimize crystallization hits in shorter time. Additionally, nuclei obtained from crystals of one protein species can be transferred to a setup composed of a closely related protein, such as a mutant, Se-Met substituted protein or the same protein with a different substrate. This heterogeneous seeding is also called cross-seeding. Lately, investigations in the field of nucleation focus on the development of the “universal nucleant”, a material which heterogeneously acts as a seed for any protein<sup>132</sup>, to overcome the bottleneck of protein crystallization. However, this process is hindered by the fact that proteins exhibit various shapes and diameters. Thus, a nucleant has to be found which meets all different requirements.

#### *Seeding with ChoX/choline*

One problem in the crystallization of ChoX was not to obtain crystals per se, but to produce crystals of ChoX in an unliganded state. Besides, the structures solved from crystals grown in a setup of ChoX with acetylcholine showed electron density for a choline molecule in the binding site but not for an acetylcholine molecule. This is probably due to acid catalyzed hydrolysis of acetylcholine to choline and acetate during crystal formation. To obtain crystals in unliganded and in acetylcholine bound form it was anticipated that microseeding might be successful.

Therefore, ChoX/choline crystals, grown in a usual hanging drop setup, were used as microseeds. Crystals serving as a seed were either touched with a seeding tool purchased from Hampton Research (HR8-133) or a home made variant of the same. Then the seeding tool was

streaked across two newly setup protein/precipitant drops before touching the original crystal again. The new hanging drop experiment was sealed as quickly as possible after inoculation with the microseeds. As quick crystal growth was desired, neither the protein concentration nor the precipitant concentration was changed drastically compared to the original setup. Seeding experiments were performed with ChoX in complex with acetylcholine and with unliganded ChoX. The seeding conditions were prepared as a grid screen around the conditions which yielded the original seed crystal.

## 2.4.5 Basic principles of X-ray crystallography

### *Crystal systems and space groups*

A crystal is build up by a periodic space-filling arrangement of a unit cell, the building block of a crystal. Hereby, the unit cells can be transfered into each other via translation in three dimensions. The unit cell is defined by its lattice parameters, the cell lengths  $a$ ,  $b$  and  $c$  and the cell angles  $\alpha$ ,  $\beta$  and  $\gamma$ . There are seven unique crystal systems, which are defined by the properties of the unit cell.

The unit cell itself may be composed in a symmetric fashion. Possible symmetry operations that a crystal can exhibit are mirror planes, inversion centers, rotation axis and screw axis. However, due to the chirality of biomolecules, mirror planes and inversion centers are not allowed in protein crystals leaving rotation axis or screw axis as the only possible symmetry applications. Combining the unit cell parameters with these symmetry operations establishes the space groups. Without inversion centers and mirror planes 65 space groups are possible for protein crystals. Depending on the space group, the unit cell is composed of different amounts of asymmetric units (ASUs), which are related to each other by the symmetry operations inherent to the space group of the crystal. Table 2.2 lists the seven possible crystal systems and their minimum symmetry elements.

Crystal system	Conditions imposed on cell geometry	Minimal symmetry element
Triclinic	None	None
Monoclinic	$\alpha = \gamma = 90^\circ$ , $\beta \neq 90^\circ$	One 2-fold axis
Orthorhombic	$\alpha = \beta = \gamma = 90^\circ$	Three perpendicular 2-fold axes
Tetragonal	$a = b$ ; $\alpha = \beta = \gamma = 90^\circ$	One 4-fold axis
Trigonal	$a = b$ ; $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$	One 3-fold axis
Hexagonal	$a = b$ ; $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$	One 6-fold axis
Cubic	$a = b = c$ ; $\alpha = \beta = \gamma = 90^\circ$	Four 3-fold axes

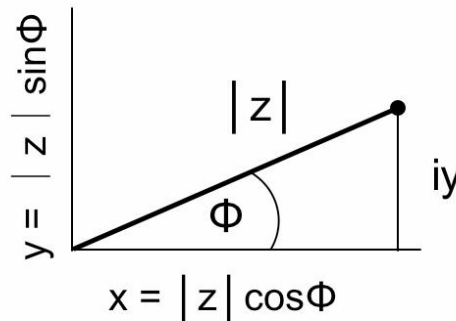
**Table 2.2** The seven crystal systems, their cell geometry and minimal symmetry elements<sup>133</sup>.

### Working with waves

The scattering of X-rays by substances can be described as an interaction of the electromagnetic incident X-ray beam with the electrons of the atoms. The incident beam forces the electrons to oscillate in the same frequency as the incident beam. Upon returning to the unexcited state radiation is emitted by the electrons, which has the same wavelength as the X-ray beam. Thus, the incident wave and the emitted wave exhibit the same wavelength. However, the phase is shifted. To understand the theory of X-ray diffraction and structure determination it is essential to be familiar with some basic concepts that make working with waves easier. Therefore, a few principles are explained here.

For describing waves the Argand diagram often is used. In principle an Argand diagram is a way of representing complex numbers. A complex number  $z$  can be written as  $x + iy$  where  $x$  is its real part and  $y$  is the imaginary part. Or it is expressed as  $|z|e^{i\Phi}$  where  $|z|$  is the amplitude and  $\Phi$  is the phase. By representing a complex number in an Argand diagram (Figure 2.9) with the real part written on the x-axis and the imaginary part displayed on the y-axis, the following equation can easily be derived.

$$z = x + iy = |z|(\cos\phi + i \sin\phi) = |z|e^{i\phi} \quad (2.3)$$



**Figure 2.9** Representation of a complex number in an Argand diagram.

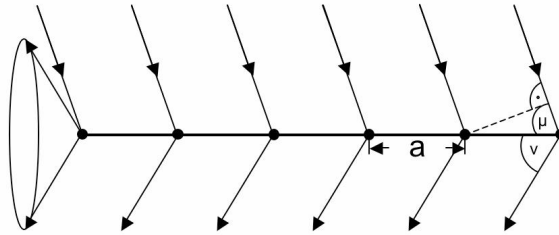
The complex number notation represents a convenient way to describe waves. With a given wavelength a wave can be represented as a single complex number, which is determined by its amplitude and its phase.

$$W = Ae^{i(2\pi x / \lambda + \phi)} \quad (2.4)$$

*Principles of diffraction - Laue equations, Bragg's law and the Ewald construction*

To explain the diffraction of a beam on a three-dimensional lattice, different approaches can be used. Max von Laue, who discovered X-ray diffraction in 1912, formulated the Laue equations. Rather straightforward and easy to understand is Bragg's law, introduced 1913 by Lawrence Bragg.

To understand the Laue equations we can first consider the interference of rays on a one dimensional lattice being represented by equally spaced scatterers on a line in Figure 2.10.



**Figure 2.10** Interference of rays in a one dimensional lattice with the spacing **a** of the scatterers. The angle of the incident beam is  $\mu$ . The outgoing rays are scattered at an angle  $\nu$ . Adapted from Massa<sup>134</sup>.

Positive interference of the scattered waves is only occurring if all outgoing waves are in phase with each other. Therefore, the path-difference of the incoming wave and the outgoing wave has to be  $n\lambda$ , with  $n = 0, 1, 2, 3, \dots$ . If the incident beam forms an angle  $\mu$  with the line of scatterers, diffraction only occurs if the beam is diffracted with an angle  $\nu$  relative to the line of scatterers. As the diffracted wave is spreading in all directions, a cone, called Laue cone, can be formed with the opening of  $2\nu$  that represents the directions of the possible diffracted beams.

The condition stated above can be formulated as follows:

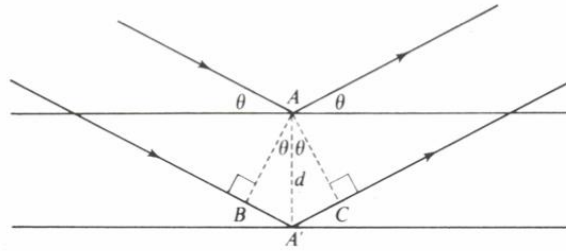
$$a \cos \mu_a + a \cos \nu_a = n\lambda \quad (2.5)$$

A crystal lattice is built up of lines of atoms, or scatterers, in all three dimensions. Therefore, two more Laue equations for the other two dimensions have to be formulated.

$$\begin{aligned} a \cos \mu_a + a \cos \nu_a &= n\lambda \\ b \cos \mu_b + b \cos \nu_b &= n\lambda \\ c \cos \mu_c + c \cos \nu_c &= n\lambda \end{aligned} \quad (2.6)$$

Only if all three equations are satisfied simultaneously or all three Laue cones intersect at one point, diffraction is possible.

Bragg showed that X-rays impinging on two parallel planes which exhibit an interplanar spacing of  $d_{hkl}$ , only show diffraction if the angle of the incoming beam is equal to the angle of the outgoing beam and their path-difference is  $n\lambda$ . This is represented in Figure 2.11.



**Figure 2.11** Reflection of parallel beams on parallel planes with interplanar spacing  $d$ . Adapted from Woolfson<sup>125</sup>.

Bragg's law is formulated in the following equation:

$$n\lambda = 2d_{hkl} \sin \theta \quad (2.7)$$

If the path difference is one wavelength  $n = 1$ , if it is two wavelength  $n = 2$  and so on.  $n$  determines the order of the reflection.

Braggs law can also be expressed as

$$\lambda = 2d_{hkl} \sin \theta \quad (2.8)$$

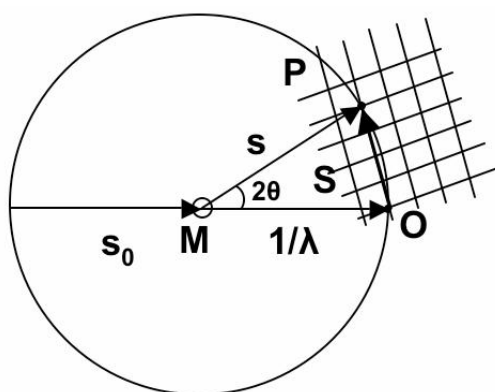
where each reflection is generated by a set of planes with the spacing  $d_{nk,nh,nl}$ .

In a unit cell many different sets of parallel planes can be inscribed. A set of planes intersects the unit cell at the points  $h$ ,  $k$  and  $l$  and is defined by these indices, also called Miller indices. Here,  $h$  describes the intersection of the plane with the cell edge  $a$ . If the edge  $a$  is intersected  $a/n$ ,  $h$  exposes the value  $n$ . The same principle applies for  $k$  and  $l$  intersecting the cell edges  $b$  and  $c$ , respectively.

With this background it is easily understood that higher order reflections are generated by a set of planes exhibiting smaller spacing and thus carry higher resolution information.

To construct the direction of scattering, the reciprocal lattice and the Ewald construction was introduced. A set of parallel planes, on which diffraction of a beam can occur, can be also expressed by a vector  $\mathbf{S}$  which has the length of the spacing between the planes and is oriented along their normal. For each set of hkl-planes such a vector can be constructed, thereby the origin of the reciprocal lattice is the starting point of each of these vectors. The end of the vector determines a point in the reciprocal lattice. Due to the reciprocal relation sets of planes with higher hkl-indices display a smaller vector.

By constructing Bragg's law in reciprocal space and taking the reciprocal lattice in consideration the diffraction direction can be determined. This construction, represented in two dimensions in Figure 2.12, is also called Ewald sphere. The incident beam  $\mathbf{s}_0$ , exhibiting a length of  $1/\lambda$ , is directed to the origin  $O$  of the reciprocal lattice. Then a circle with radius  $1/\lambda$  and its centre  $M$  being placed on the line  $\mathbf{s}_0$  is drawn. Diffraction occurs if the vector  $\mathbf{S}$ , with its start at the origin  $O$ , has its endpoint  $P$  on the Ewald sphere. Under these conditions the resulting scattered beam is  $\mathbf{s}$ .



**Figure 2.12** The Ewald construction relates the incident beam with the angle of reflection with the help of the reciprocal lattice. The construction of the Ewald circle is explained in the text.

#### *Structure factors, electron density and how they relate*

The scattered ray, which produces a reflection, can be described as a Fourier series of scattered rays generated by the individual atoms in a unit cell. The Fourier series that describes this term is called a structure factor equation. Thus, a structure factor is a wave composed of contributions of all atoms in a unit cell.

Diffraction occurs by interaction of the impinging beam with the electrons of a molecule. Therefore, the diffracted beam contains information of the electron clouds of a molecule. The goal of crystallography is to calculate the electron density function  $\rho(x,y,z)$  in dependence of the coordinates  $x$ ,  $y$  and  $z$ , which is a periodic function as it repeats itself due to the regular array of a crystal.

Imagining that the electron density in the unit cell is composed of small volume elements, each structure factor can be written as a Fourier summation over these volume elements. Each term of this summation represents the diffraction of the electrons constituting to this volume element, i.e. the diffraction of the averaged electron density within this volume element.

Thus, the structure factor  $F_{hkl}$  can be expressed by a Fourier series with the contribution of the averaged electron density of each of these volume elements or with infinitesimal small volume elements with the average values of  $\rho(x,y,z)$ .

$$F_{hkl} = \int_V \rho(x, y, z) e^{2\pi i(hx+ky+lz)} dV \quad (2.9)$$

As this equation is a Fourier series, its transform, also a Fourier series with the structure factor  $F_{hkl}$  as amplitude, can be written as

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad (2.10)$$

### *The phase problem*

The aim of the crystallographic process is to obtain an interpretable counter map of the electron density distribution in the unit cell. The electron density distribution is given by a Fourier summation (see equation 2.10) with structure factors presenting the amplitudes of the waves that compose the Fourier series.

The structure factor describes a diffracted X-ray beam and thereby also expresses a wavefunction with a frequency equal to the frequency of the incident X-ray beam and an amplitude  $|F(hkl)|$  which is proportional to the square root of the measured intensities  $I(hkl)$ . However, the phase of a structure factor can not be determined directly from the diffraction experiment. The calculation of the electron density map necessitates phase information of

each structure factor which has to be extracted indirectly. Different methods have been developed to gain phase information. Their basic concepts will be described in the following.

- Isomorphous replacements
- Anomalous diffraction
- Molecular replacement

These methods all utilize the Patterson function in the process of working back from the diffraction pattern to the crystal structure. Introductory to the description of the phasing methods the Patterson function shall be considered in detail.

### *The Patterson function*

The Patterson function  $P(uvw)$ , developed in 1934 by Lyndo Patterson, is also called self convolution of a structure. In principle it is a Fourier summation like the electron density function presented in equation 2.10 with the difference that in a Patterson function no structure factors are used as coefficients but the measured intensities of the reflections. Thus, no phase information is needed for the calculation of a Patterson function. The map can be calculated directly from the data obtained by the diffraction experiment.

$$P(uvw) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F_{hkl}|^2 \cos(2\pi(hu + kv + lw)) \quad (2.11)$$

Here  $u$ ,  $v$  and  $w$  represent vectors locating a point in the Patterson map.

Another way of expressing the Patterson function is given in equation 2.12 where  $u$  and  $r_1$  represent vectors.

$$P(\vec{u}) = \int_{\vec{r}_1} \rho(\vec{r}_1) \rho(\vec{r}_1 + \vec{u}) d\vec{r}_1 \quad (2.12)$$

A peak in the Patterson map at  $(uvw)$  means that two atoms are placed in the real cell which have the coordinates  $(x,y,z)$  and  $(x + u, y + v, z + w)$ , with the Patterson function relating these atoms.

### *Multiple Isomorphous Replacement (MIR)*

Heavy atoms, exposing many electrons concentrated in a small sphere, contribute strongly to the diffraction of a protein crystal. Thus, by comparing the diffraction pattern of a protein crystal containing heavy atoms (derivative crystal) with the diffraction pattern of a native crystal, i.e. without heavy atoms, differences in the intensities of the reflections are revealed. These differences can help to estimate the phases of the structure factors. However, a prerequisite of isomorphous replacement is the availability of at least two isomorphous crystals, of a native and a heavy atom incorporated crystal. Isomorphous crystals show the same unit cell dimensions and same symmetry. This is often achieved by cocrystallization with a heavy atom or soaking a native crystal in a heavy atom solution.

The first step of isomorphous replacement is to locate the heavy atom. This is achieved by employing a difference Patterson function. In this Patterson function the difference in the structure factor amplitudes of the derivative and the native crystal,  $|F_{PH}| - |F_P|$ , is used as coefficient.

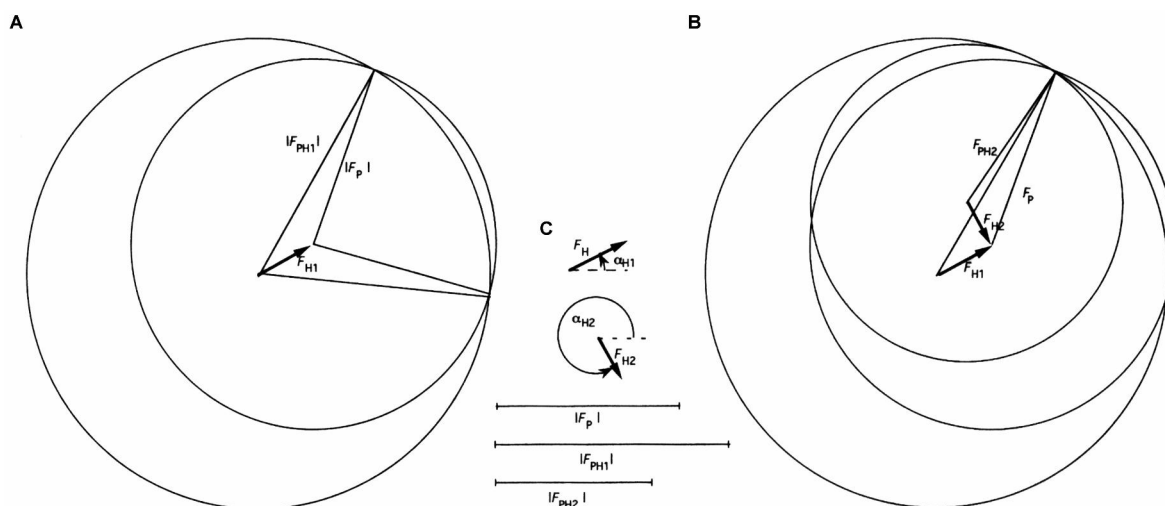
$$P(uvw) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \left( |F_{PH}(hkl)| - |F_P(hkl)| \right)^2 \cos(2\pi(hu + kv + lw)) \quad (2.13)$$

In this step for each set of heavy atoms two possible positions are determined,  $(x + u, y + v, z + w)$  and  $(x - u, y - v, z - w)$ . Structure factors produced solely by the heavy atoms are calculated and further employed in phase determination. As two possible positions for the heavy atoms are obtained, both have to be employed in the process and later it is evaluated which of these sets gives the right hand.

As all atoms in the unit cell contribute to each structure factor, the structure factors of the derivative crystal are composed of the contribution of the protein without heavy atoms and the contribution of the heavy atoms alone. For any structure factor this can be formulated as:

$$F_{PH} = F_P + F_H \quad (2.14)$$

As we know the position of the heavy atoms we can solve this vector equation for  $F_P$ . Figure 2.13a shows the graphic representation of this procedure, also known as Harker construction.



**Figure 2.13** Harker construction for the graphical determination of phases. **A** The determination of phases with one heavy atom derivative is shown. **B** Determination with two heavy atom derivatives. **C** Known variables used for the Harker construction. Adapted from Blow<sup>135</sup>.

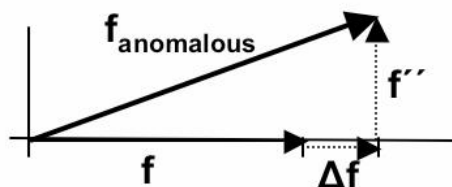
The length of vectors  $F_{PH}$  and  $F_P$  is defined by the measured intensities of the derivative and the native dataset. Besides, the vector  $F_H$  is known from the determination of the heavy atom positions by the difference Patterson function. First a circle with the radius of  $|F_P|$  is sketched which represents all possible phase angles of  $F_P$ . The vector  $F_H$  is drawn such that the vector ends at the centre of the circle. The beginning of the vector is taken as the center for a second circle with a radius of  $|F_{PH}|$ . This circle represents all possible phase angles for  $F_{PH}$ . The vector equation given in 2.14 holds true at the intersection of the two circles. Thus, two possible phases for  $F_P$  are determined. The determination of two more possible phases is achieved by a second Harker construction using a second derivative. The intersection point of a construct using both heavy atom derivatives determines the phase for the structure factor  $F_P$  (Figure 2.13b).

In this way structure factors for all values of  $(h,k,l)$  can be determined and an initial electron density map can be calculated using equation 2.10. Dependent on the decision made for the set of heavy atoms either an electron density map for a protein with L-amino acids and right handed  $\alpha$ -helices or its mirror image is displayed. The correct choice of heavy atoms has to be evaluated by the crystallographer with the help of the resulting electron density map.

### *Multiple Wavelength Anomalous Diffraction (MAD)*

A second method to determine phases exploits the characteristic of heavy atoms to change their scattering behavior at certain wavelengths<sup>136-138</sup>. If the wavelength of the incident X-ray beam approaches the characteristic absorption edge of a specific heavy atom, the assumption that the scattering electron is a free electron holds no longer true. The electron also absorbs

energy and thereby is promoted into a higher electronic state. Due to the absorption of photons the total coherent scattering is reduced and the phase of the coherently scattered rays is changed. This is represented in Figure 2.14.



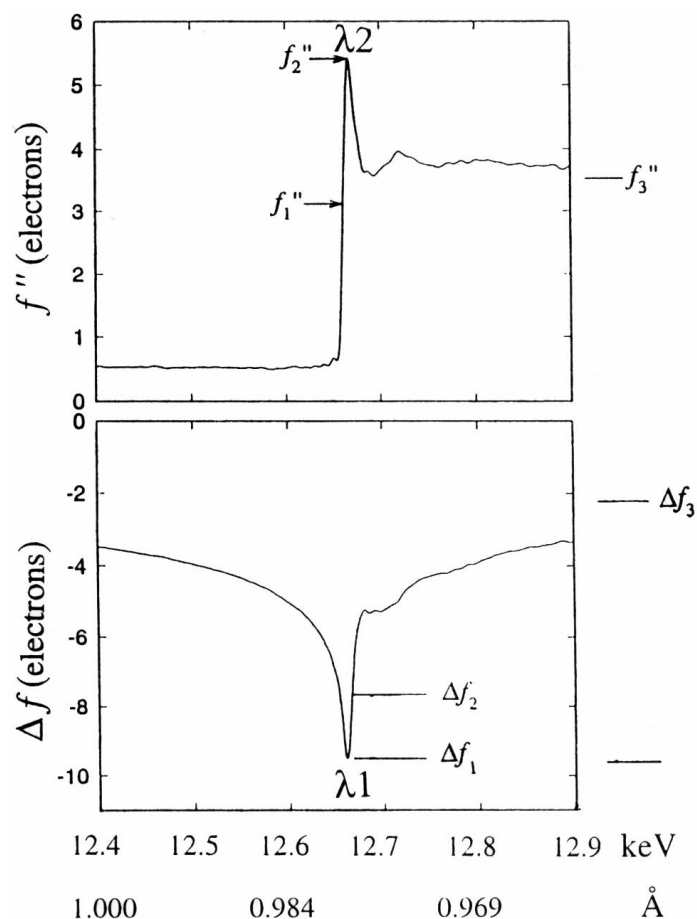
**Figure 2.14** The anomalous scattering of a heavy atom at its absorption edge. The anomalous scattering is composed of a real ( $\Delta f$ ) and an imaginary ( $f''$ ) component.

The scattering factor of an anomalous atom is composed of a real and an imaginary component and can be written as:

$$f_{\text{anomalous}} = f + \Delta f + if'' \quad (2.15)$$

Due to anomalous scattering the structure factors of a Friedel pair,  $F_{hkl}$  and  $F_{-h-k-l}$ , normally showing same amplitudes and a relation of their phase by  $180^\circ$ , do not exhibit the same amplitudes and their phases become independent.

This effect is exploited in multiple wavelength anomalous diffraction. To obtain phases with MAD, often the amino acid methionine in a protein molecule is substituted with Se-methionine, with Se exhibiting an absorption edge at  $0.9795 \text{ \AA}$ . The imaginary component  $f''$  and the real component  $\Delta f$  of the anomalous scattering of Se is presented in Figure 2.15. Also three wavelengths are shown at which an optimum MAD experiment can be conducted.



**Figure 2.15** The composition of the anomalous scattering of a Se-atom, when the absorption edge is approached.  $\Delta f$  and  $f''$  represent the real and the imaginary component, respectively. In an MAD experiment data is collected at different wavelengths. The maximum of  $\Delta f$  is reached at wavelength  $\lambda_1$ . At  $\lambda_2$  the imaginary component of the anomalous scattering shows its maximum. The third wavelength  $\lambda_3$  is not shown in the pane. Adapted from Drenth<sup>133</sup>.

For an MAD experiment data at three different wavelengths are collected on a single “heavy atom incorporated” crystal. These are the wavelengths  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ , also called the inflection, peak and remote wavelength.  $\lambda_1$  is chosen such that the real component  $\Delta f$  of the anomalous scattering is at its minimum. At  $\lambda_2$  the imaginary component  $f''$  of the anomalous scattering is at its maximum.  $\lambda_3$  represents a wavelength where the anomalous scattering is minimized.

To obtain phases with a MAD experiment two different processes, the anomalous MAD and the isomorphous MAD are distinguished. The latter process uses the large differences in the real component  $\Delta f$  of the anomalous scattering, whereas the relatively small contribution of the imaginary component  $f''$  is removed by averaging the Friedel mates of the inflection and the peak dataset. Again, the heavy atom positions can be determined with the help of a difference Patterson function for the remote and inflection datasets. The determination of two possible phases is analogous to the phase determination in a single isomorphous replacement by construction of intersecting Harker circles.

In a second step the difference between the Friedel mates, which reaches a maximum in the peak dataset, is utilized to calculate an anomalous difference Patterson function. Harker construction yields another set of phase angles. By a combined Harker construction with the isomorphous and the anomalous component one possible phase angle is obtained.

### *Molecular replacement*

In cases where a structure of a protein is available and the structure of a mutant of this protein or the protein in complex with a ligand has to be determined, the available phase information can be exploited to obtain initial phases for the new model. This method is called molecular replacement and in principle can be applied to all proteins showing similar folding, which is often assessed by the homology of their sequences.

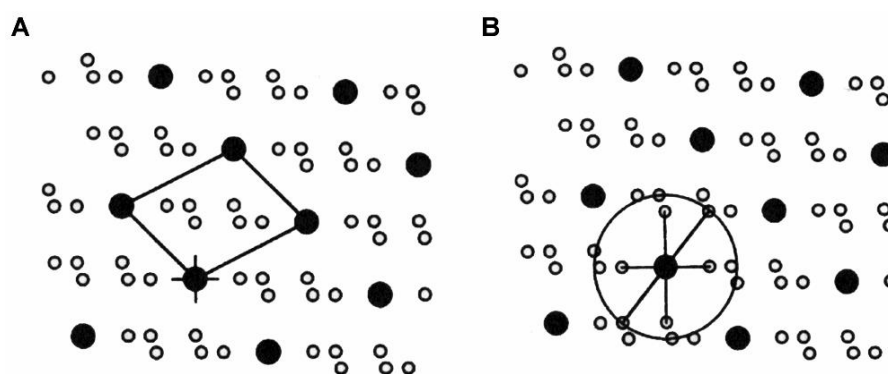
If the models are isomorphous, as often the case for a protein and a mutant, the phases can directly be calculated with equation 2.10 using the structure factor amplitudes derived of the new diffraction data and the phase angles of the existing model.

In nonisomorphous cases the existing model needs to be oriented in the new unit cell in such way that the resulting phases for the old model get as good as possible. Structure factors can be calculated from the newly positioned old model. The phases of these calculated structure factors are deduced for calculating an initial electron density utilizing the new diffraction data.

The process of orienting the model in the new unit cell is divided in two steps. First the orientation of the model, independent of its position in the unit cell, needs to be found by rotating the model. In a second step the correct position within the unit cell is assigned by translation of the model in three dimensions.

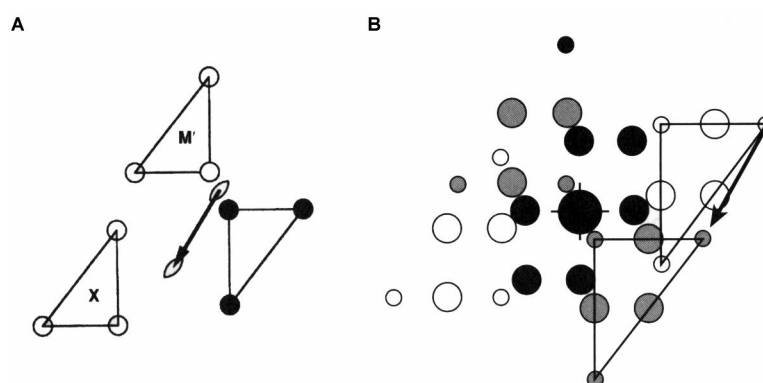
Both steps can be performed with the help of the Patterson function. Within the Patterson map different types of peaks are revealed, the cross- and the self-vector peaks. Vectors originating from two atoms of the same molecule in the unit cell are called self-vectors. If the two atoms belong to different molecules within the unit cell, the vectors are called cross-vectors. As the distance between two atoms of different molecules mostly is larger than the distance between atoms within the same molecule the peaks in the Patterson map originating from self-vectors are concentrated around the origin. Peaks from cross-vectors are located further away. The correct orientation of the model can be ascertained with the help of self-vectors whereas the translation, used to position the model correctly in the unit cell, is achieved with the help of cross-vectors.

The Patterson function describes the vector between two atoms independent of the position of the atoms in the unit cell. A rotation of a molecule in real space would result in the rotation of its Patterson map around the origin of the map. Thus, if the Patterson map of the model is rotated, a position is revealed where the model Patterson map overlays best with the Patterson map derived of the new reflection data. Thus, information about the rotation function can be extracted. The rotation function search is exemplified in two dimensions in Figure 2.16.



**Figure 2.16** The rotation function search. **A** The Patterson map of the unknown crystal structure is outlined with its unit cell. **B** The Patterson function of the search model is rotated such that the peaks of the model function overlay with the Patterson map of the unknown crystal. The rotation function can be determined. Adapted from Blow<sup>135</sup>.

After the determination of the correct orientation of the model the next step is to position the correctly oriented model in the unit cell. This step is only needed in cases where the crystal exhibits symmetry other than the lattice symmetry and the crystal origin needs to be defined relative to the rotation axes. Therefore, the position of the rotation axes in the crystal is determined relative to the orientation of the model. This can be achieved by calculating a Patterson function for the model with an arbitrary chosen rotation axis. The location of the axis influences the position of the Patterson peaks but not their constellation (see Figure 2.17). Comparing this Patterson function with the Patterson map of the reflections reveals the self-vectors on the same position and the cross-vectors being displaced. With the help of the translation function a vector can be determined which relates the cross-vector peaks, as shown in a two-dimensional example in Figure 2.17. By doubling this vector and applying it to the arbitrary chosen axis of the model the position of the real axis is revealed. Thus, the position of the model relative to the rotation axes can be determined.



**Figure 2.17** The translation function search to obtain a twofold axis of an unknown structure relative to the origin of the model structure. **A** The symmetry related model  $M'$  generated by an arbitrary chosen twofold axis. By translation  $M'$  can be superimposed with the unknown structure  $X$ . Therefore, the vector of this translation applied on the arbitrary chosen axis relates the model structure with the unknown structure  $X$ . **B** Using a common origin the Patterson functions of the search model and  $MM'$  and the unknown structure are overlaid. The peaks resulting from self-vectors are indicated in black, those of cross-vectors are indicated in open circles (X) or as grey circles ( $MM'$ ). The vector which relates the cross-vector peaks of the model  $MM'$  with the cross-vector peaks of  $X$  is double the vector that relates the twofold axes shown in **A**. Adapted from Blow<sup>135</sup>.

## 2.4.6 Twinning - a crystal growth disorder

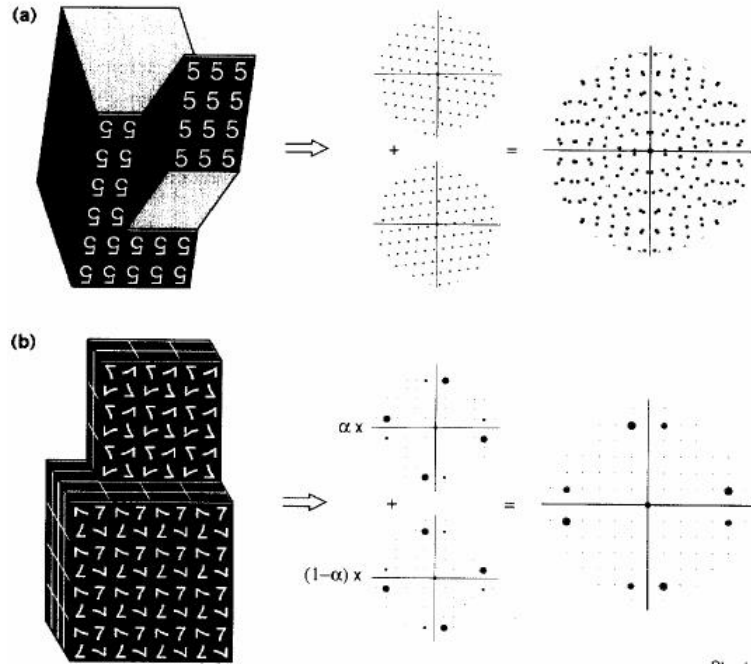
### *Basic concept*

In a perfect crystal all protein molecules are aligned highly ordered. However, during crystal growth a perfect crystal is hardly ever obtained, instead the crystal often exhibits several growth defects. The conformation of the molecule building up the crystal lattice can vary in space leading to the modeling of alternate conformations in the final structure. Or the crystal can be built up by mosaic microdomains, which, even though high local order is maintained, leads to the loss of global order. A twinned crystal exhibits multiple crystalline domains, which are related by a certain symmetry operation, the so-called “twin law”. Dependent on the lattice of the crystalline domains, the twinning can be classified as non-merohedral, when the lattices of the domains match in two dimensions, or merohedral, where all three dimensions of the lattices correspond (see Figure 2.18).

Whereas the before mentioned growth anomalies are rather often encountered in protein crystals, crystal twinning occurs more rarely and has long been considered a severe problem.

The diffraction of a non-merohedral twin shows a superposition of the diffractions of the separate domains. Therefore, it can be easily recognized by looking at the diffraction pattern. During the indexing procedure one of the lattices can be exclusively identified and processed. In contrast, in case of a merohedrally twinned crystal the diffraction patterns of the separate domains overlap perfectly and both lattices contribute to each reflection. The diffraction

pattern of a merohedrally twinned crystal often exhibits a higher symmetry than the actual crystal lattice. A special case of merohedral twinning is encountered when the crystal lattice by chance coincides with a higher symmetry system. In this case the twin law is an operation of the higher symmetry group. One example is a monoclinic crystal, which exhibits  $\beta \sim 90^\circ$ , showing an apparent orthorhombic symmetry.



**Figure 2.18** Twinned crystals and their effect on the diffraction pattern. **A** Non-merohedral crystal. **B** Merohedral crystal. Adapted from Yeates<sup>139</sup>.

A merohedrally twinned crystal shows a higher symmetry in his diffraction pattern than the actual symmetry of the crystal is implying. Hence, it is necessary to detwin the measured intensities, which represent the sum of inflections obtained by the separate twin domains. Hereby, it is important to know the twinning factor  $\alpha$ , which describes the ratio of the twinned domains according to the volume they occupy. In case of perfect twins,  $\alpha$  takes the value 0.5. The intensities of two twin-related, observed reflections  $I_1$  and  $I_2$  are composed of the individual reflections  $J_1$  and  $J_2$  according to the following equation.

$$\begin{aligned} I_1 &= \alpha J_1 + (1 - \alpha) J_2 \\ I_2 &= (1 - \alpha) J_1 + \alpha J_2 \end{aligned} \quad (2.1)$$

These equations can be solved if  $\alpha$  is determined and does not take the value 0.5. When  $\alpha=0.5$ , the equations become indeterminate. The true intensities of the reflections are calculated with

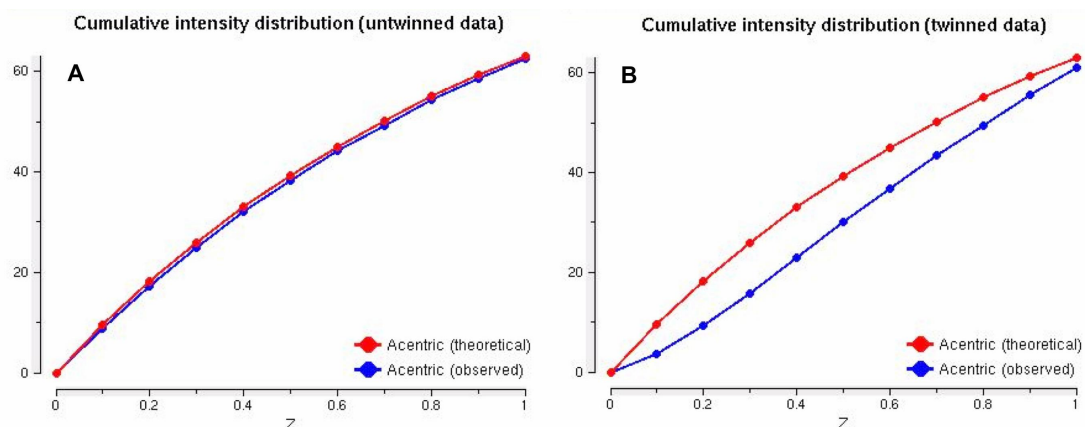
$$\begin{aligned}
J_1 &= [(1-\alpha)I_2 - \alpha I_1]/(1-2\alpha) \\
J_2 &= [(1-\alpha)I_1 - \alpha I_2]/(1-2\alpha)
\end{aligned}
\tag{2.2}$$

The twin operation combines two reflections. In general a reflection pattern exhibits much more intermediate reflection than strong and weak reflections. Therefore, when two reflections are combined due to twinning the probability that two very weak or two very strong reflections are added up is very low. The resulting distribution of reflection intensities is narrower around the mean value in case of twinned crystals. Thus, twinning can be detected by analyzing intensity distributions and often  $\alpha$  can be derived by these analyses. The refinement of data derived from twinned crystals can be eased by knowing the twinning fraction. But as  $\alpha$  only represents an estimate, the quality of structure of twinned data never is as good as from untwinned data.

#### *Detecting and working with twinning*

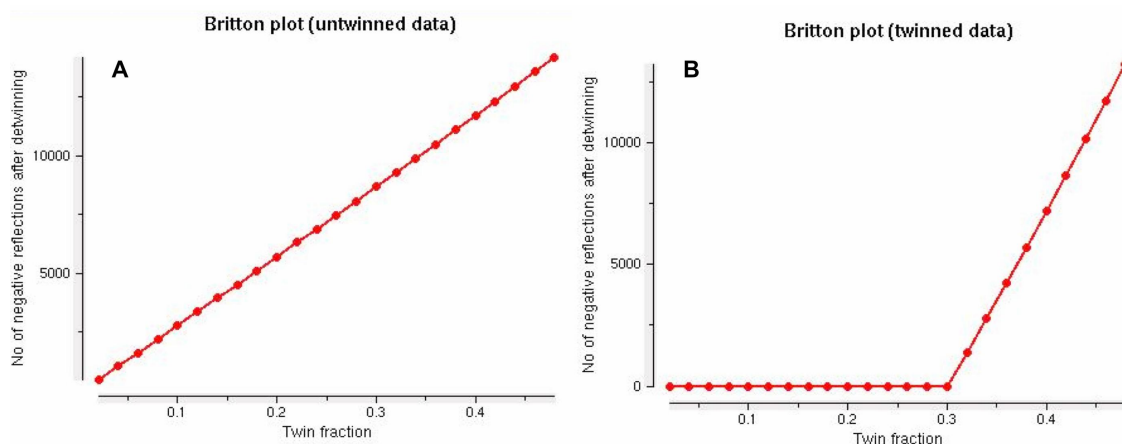
To judge if data is twinned, one first indication can be derived by the Wilson ratio  $\langle F \rangle^2 / \langle I \rangle^{140}$ , which should take a value of 0.785 and 0.885 for untwinned and twinned data, respectively. Also analyzing the mean average intensity ratio  $\langle I^2 \rangle / \langle I \rangle^2$  can give an idea if data is twinned or not. In case of perfect twinning it exhibits a value of 1.5 whereas untwinned data show a value of 2.0.

A more graphic representation to analyze twinning is given in the Rees plot (Figure 2.19), which regards the acentric reflection (i.e. phase angle not  $0^\circ$  or  $180^\circ$ ). The fraction of local average intensity  $z$  is plotted against the cumulative distribution of  $z$ . The fraction of weak intensities is lower in a twinned crystal than in an untwinned crystal. This is reflected in the cumulative distribution of  $z$ , which takes a more sigmoidal progression in the twinned case, whereas it is rather exponential for untwinned data<sup>141</sup>.



**Figure 2.19** The Rees plot gives the cumulative intensity distribution of  $z$  of acentric reflections. **A** Rees plot for untwinned data. **B** Rees plot for twinned data. Taken from CCP4<sup>142</sup>.

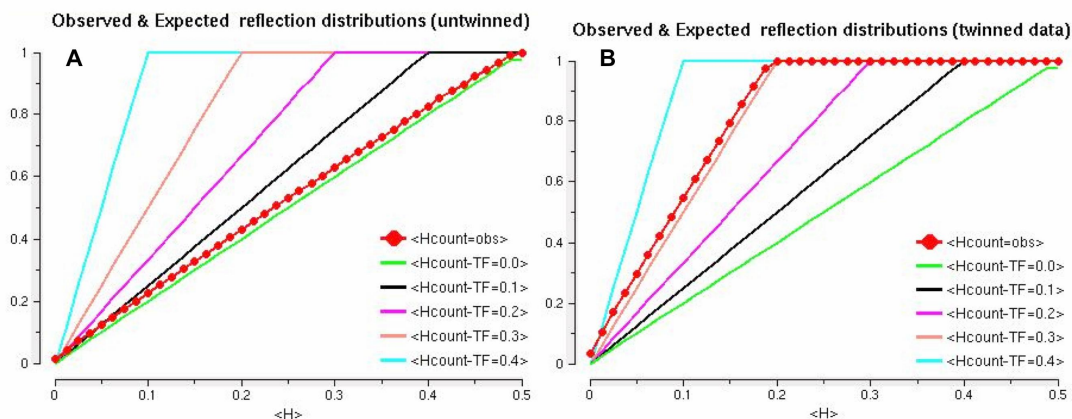
Another method to analyze for possible twinning is given by the Britton plot of negative intensities. It is based on the fact that during a detwinning procedure proposing a too high value for  $\alpha$  results in negative intensities<sup>143</sup>. Plotting the fraction of negative intensities against the assumed value of  $\alpha$  shows a curve that can be fitted by two linear asymptotes. The crossing point of these asymptotes gives an estimate for the twinning fraction  $\alpha$ <sup>144</sup>. A disadvantage of this method is the high sensitivity to accurately measured weak reflections. An example of a Britton Plot is shown in Figure 2.20, where this plot for untwinned and twinned data is shown.



**Figure 2.20** The Britton plot. **A** Example for untwinned data. **B** Britton plot for twinned data. Taken from CCP4<sup>142</sup>.

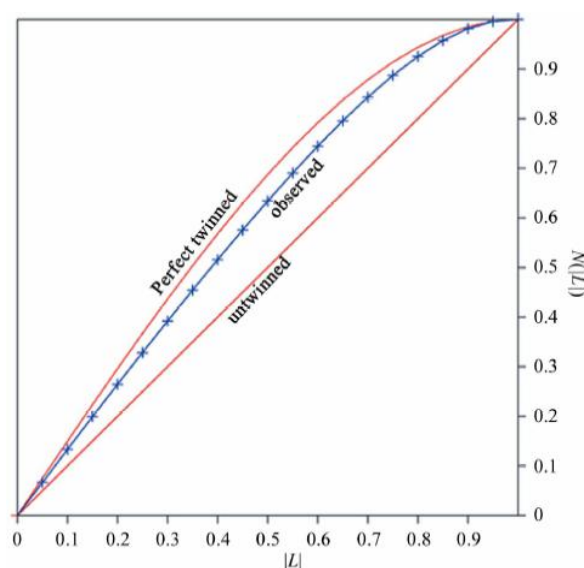
Yeates proposed yet another test of analyzing data for the twinning fraction<sup>145,146</sup>. With  $H$  being defined as  $H = |I_1 - I_2| / (I_1 + I_2)$ , the cumulative distribution of  $H$  indicates a curve with its slope being equal to  $1/(1-2\alpha)$ . This plot proves to be rather robust, especially as it is

independent of weak reflections, which often are prone to be more erroneous. An example of an Yeates  $S(H)$ -plot is shown in Figure 2.21.



**Figure 2.21** The Yeates plot gives the cumulative intensity distribution of  $H$ . **A** Yeates plot for untwinned data. **B** Yeates plot for twinned data. Taken from CCP4<sup>142</sup>.

Yeates and Padilla developed an even more robust test to analyze for twinning and for the twinning fraction<sup>147</sup>. As the  $S(H)$ -plot is regarding twin related reflections, abnormalities like anisotropy are obscuring the analysis as reflections are systematically different. The here presented more accurate test uses the same principle as the  $S(H)$ -plot, however, the intensities defining  $L$ , with  $L = |I_1 - I_2| / (I_1 + I_2)$  are not related by a twinning law, but are selected on the basis that they are in close vicinity in reciprocal space. An example for such a plot is shown in Figure 2.22.



**Figure 2.22** Examples of the L-function for untwinned (lower red curve) and perfectly twinned data (upper red curve). The observed data may be positioned between both curved (blue curve) and thus show partial twinning. Taken from Sultana<sup>148</sup>.

The twinning fraction  $\alpha$  has to be estimated as accurately as possible for working with twinned data. If the twinning fraction below 0.3 a detwinning procedure can be applied and refinement can be pursued with the detwinned data. If the twinning fraction approaches 0.5 programs like SHELXL<sup>149</sup> or CNS<sup>150</sup> can be used for refinement. In both the approach of Redinbo and Yeates<sup>151</sup> is implemented to account for twinned data.

#### **2.4.7 Cryocrystallography**

Protein crystals are sensitive to X-ray radiation and often suffer radiation damage. Especially high resolution data is affected by this phenomenon. Therefore, cryocrystallography has been developed to reduce radiation damage<sup>152</sup>. Additionally, the quality of the data is improved as thermal motion is reduced. In cryocrystallography the crystal is rapidly frozen, often in liquid nitrogen, and is kept at a low temperature during data collection. Usually, before freezing the crystal it is transferred to a cryoprotective buffer which hinders ice formation. Thus, the water in the crystal forms an amorphous glass. Often glycerol, PEG 400, ethylene glycol, or 2-methylpentandiol may be applied as cryoprotective solutions. Dependent on the constitution of the drop a certain amount of the Cryoprotectant is needed to successfully hinder ice formation. The concentration of the cryoprotectant necessary for certain conditions is often empirically analyzed, although some systematical work has been performed in this field<sup>153</sup>. In this thesis, all crystals analyzed for their diffraction quality or utilized for data collection were cryoprotected. With the help of a nylon loop they were transferred to the cryoprotective buffer, soaked for 10-30 seconds and subsequently frozen in liquid nitrogen. Afterwards, they were stored in a dewar containing liquid nitrogen until their application for diffraction analysis.

#### **2.4.8 Data collection and processing**

Diffraction analysis and data collection of the cryoprotected and frozen crystals was performed either at the Max-Planck beamline BW6 or at the EMBL beamline BW7a at DESY (Deutsches Elektronen Synchrotron), Hamburg. Datasets were collected, optimizing the collection strategy with BEST<sup>154</sup>. The wavelength used for collection was 1.05 Å and 0.98 Å for the choline and acetylcholine complexed ChoX, and 1.033 Å and 0.90 Å for the unliganded closed and semi-closed ChoX, respectively. The datasets were processed either with Denzo/Scalepack<sup>155</sup> or with XDS<sup>156</sup>.

#### **2.4.9 Determination of the structure of ChoX in complex with choline**

For the structural determination of ChoX in complex with choline, phases had to be determined by multiple anomalous diffraction using a Se-Met substituted ChoX crystal. The position of the Se-atoms and subsequent determination of the phase angles for the structure factors was performed with the program SOLVE<sup>157</sup>. The obtained electron density was further optimized by maximum likelihood modifications with the program RESOLVE<sup>158</sup>. Automated model building<sup>159</sup> with RESOLVE and iterative rounds of manual model building in COOT<sup>160</sup> and isotropic B-factor refinement with REFMAC5<sup>161</sup> (CCP4-suite<sup>142</sup>) resulted in the generation of one monomer of the two monomers present in the ASU.

Molecular replacement with MOLREP<sup>162</sup> using the high resolution dataset generated the second monomer of the ASU. The structure was further refined by iterative rounds of manual rebuilding in COOT utilizing  $1F_oF_c$  and  $2F_oF_c$  electron density maps and TLS<sup>163,164</sup> or B-factor refinement using REFMAC5. 5% of the reflections were omitted from the refinement, these generated the reflection set for calculation of the free R-factor<sup>165</sup>.

During refinement the electron density for the substrate became more pronounced. Therefore, choline was fitted in the structure. Initial coordinates and geometric parameter files were generated with PRODRG<sup>166</sup>. Automated water detection and building at a cutoff of  $3.0\sigma$  was performed with ARP/wARP<sup>167</sup>. Subsequently, the water positions were verified manually.

#### **2.4.10 Determination of the structure of ChoX in complex with acetylcholine and in the unliganded closed state**

The datasets were analyzed for possible twinning, as the  $\beta$  angle exhibited a value near to  $90^\circ$ . The twinning fraction was estimated using the Padilla Yeates server<sup>147</sup>. The model of ChoX in its choline complexed state was reduced to a monomer and depleted of its water and ligand molecules. This reduced model was utilized for molecular replacement in Phaser<sup>168-170</sup> and Molrep<sup>162</sup> for the acetylcholine bound and the unliganded closed structure. As the datasets showed high twinning fractions, the refinement was further carried out using SHELXL<sup>149</sup> with h,-k,-l as twinning operator. SHELXL allows for refinement of the twinning factor. Manual building was performed in COOT<sup>160</sup> and the water was automatically detected with Arp/wARP<sup>167</sup>. For acetylcholine the model coordinates and the geometric parameter file were generated in PRODRG<sup>166</sup> and the molecule was fitted in the  $2F_oF_c$  density map.

#### **2.4.11 Determination of the structure of ChoX in unliganded semi-closed state**

First, data used for refinement was cut at 2.6 Å as the highest resolution shell showed elevated  $R_{\text{meq-d-F}}$  values. The data was analyzed for possible twinning with the Padilla Yeates server<sup>147</sup>, which suggested a high twinning fraction. However, as the resolution is not sufficient for successful refinement with SHELXL<sup>149</sup>, the refinement was carried out as follows.

The reduced model, as described in 2.4.10 was used as input for molecular replacement with Molrep<sup>162</sup>. A clear solution was found for one monomer in the twinned spacegroup  $P2_12_12_1$ , which was further refined in COOT<sup>160</sup> and Refmac5<sup>161</sup>. The refined model was used as input for a second molecular replacement run in Molrep<sup>162</sup> using the data processed in spacegroup  $P2_1$ . The resulting monomer was refined as described before. The final R-factors were obtained with CNS<sup>150</sup> by using the refinement for twinned data assuming a perfect twinned crystal and a twinning operator of h,-k,-l.

#### **2.4.12 Analysis of the obtained models and graphical visualization**

The quality of the obtained models was assessed with the program PROCHECK<sup>171</sup>. Secondary structure elements were detected with Stride<sup>172</sup>. To detect possible hinges and rigid bodies the HingeProt server was utilized<sup>173</sup>. The rotational movement of the subdomain was analyzed DynDom<sup>174</sup>. Superimposition of structures was carried out with the program LSQMAN<sup>175</sup>. All representations of structures were generated with PyMOL<sup>176</sup>.

#### **2.4.13 Structure deposition**

The obtained structures were deposited at the protein databank (PDB). Their accession codes are 2REG, 2RIN, 2RF1 and 2REJ for the choline bound, acetylcholine bound, unliganded closed and unliganded semi-closed forms, respectively.

## 3 Results

### 3.1 The ABC transporter HlyB

#### 3.1.1 Heterologous overexpression of HlyB in *L. lactis*

N-terminally His-tagged HlyB was heterologously overexpressed in a *L. lactis* NZ9000 strain with the plasmid *pNZ8037* containing the *HlyB* gene. *L. lactis* was grown semi-anaerobically in M17-medium at 30 °C. The flasks were shaken every hour to allow for slight aeration. The expression level is decreased when too much air is available for *L. lactis*<sup>177</sup>. Therefore, it was ensured that the two liter culture flasks were filled with two liters of medium.

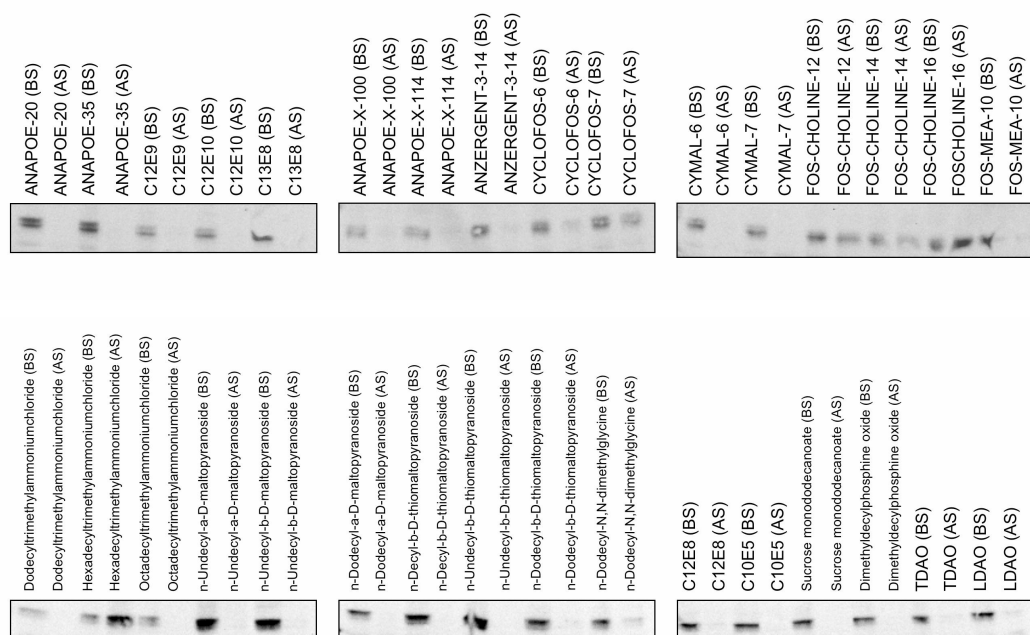
At an OD<sub>600</sub> of 1.0, the expression of HlyB was induced by adding 0.4% of the Nisin A solution. After induction the cells were further incubated for 1.5 hours to reach an optimal expression level. Cells were harvested by centrifugation (15 min at 6000 g, 4 °C) which yielded 2-3 ml of cell pellet per liter cell culture. The cells were washed twice in buffer 1 supplemented with 1 mM PMSF and 1 mM EDTA and stored at -80 °C.

#### 3.1.2 Solubilization of HlyB

32 different detergents were used to assess the ability of solubilizing HlyB. Table 3.1 gives an overview of the utilized detergents, their classification and the success in solubilization as characterized by the amount of solubilized HlyB.

To evaluate the amount of solubilized HlyB, the western blot technique was used. The solubilization was achieved by adding detergent solution to a final concentration of 1% (w/v) to membranes, mixing the solution and incubating the mixture on ice for 30 min. In some cases the detergent partially precipitated and formed a cloudy solution (*italics* in Table 1.1), which was warmed up (30 °C) to bring the detergent back in solution.

To remove unsolubilized membranes from detergent-protein complexes the solutions were pelleted (200000 g, 4 °C, 1h). The obtained supernatant (AS) was analyzed by western blotting and compared to the sample before solubilization (BS) as shown in Figure 3.1.



**Figure 3.1** Western blots using anti-HlyB antibody to assess the solubilization behavior of HlyB in different detergents. BS: before solubilization; AS: after solubilization.

Generally, HlyB is solubilized efficiently if zwitterionic detergents such as the FOS-CHOLINE group or ANZERGER 3-14 are used. Within this group, detergents exposing a phosphatidylcholine head group, like FOS-CHOLINE16 or CYCLOFOS7, extracted HlyB most efficiently. The best results were obtained with long alkyl-groups. This trend is revealed in the CYCLOFOS group and in the FOS-CHOLINE group, where FOS-CHOLINE16 solubilizes more HlyB than the shorter FOS-CHOLINE12 or 14.

Under the chosen conditions, hexadecyltrimethylammonium chloride is the best detergent to solubilize the membrane.

Detergent	Classification	Solubilization behavior
ANAPOE®-20	non-ionic	-
ANAPOE®-35	non-ionic	-
ANAPOE®-C12E9	non-ionic	-
ANAPOE®-C12E10	non-ionic	-
ANAPOE®-C13E8	non-ionic	-
ANAPOE®-X-100	non-ionic	-
ANAPOE®-X-114	non-ionic	-
ANZERGER® 3-14	zwitterionic	+
CYCLOFOS™-6	zwitterionic	+
CYCLOFOS™-7	zwitterionic	++
CYMAL®-6	non-ionic	-
CYMAL®-7	non-ionic	-
FOS-CHOLINE®-12	zwitterionic	++

Detergent	Classification	Solubilization behavior
FOS-CHOLINE®-14	zwitterionic	++
FOS-CHOLINE®-16	zwitterionic	+++
FOS-MEA®-10	ionic	-
Dodecyltrimethylammoniumchloride	ionic	-
Hexadecyltrimethylammoniumchloride	ionic	++++
Octadecyltrimethylammoniumchloride	ionic	-
n-Undecyl- $\alpha$ -D-maltopyranoside	non-ionic	-
n-Undecyl- $\beta$ -D-maltopyranoside	non-ionic	-
n-Dodecyl- $\alpha$ -D-maltopyranoside	non-ionic	-
n-Decyl- $\beta$ -D-thiomaltopyranoside	non-ionic	-
n-Undecyl- $\beta$ -D-thiomaltopyranoside	non-ionic	-
n-Dodecyl- $\beta$ -D-thiomaltopyranoside	non-ionic	-
n-Dodecyl-N,N-dimethylglycine	zwitterionic	+
Octaethylene glycol monododecyl ether (C12E8)	non-ionic	-
Pentaethylene glycol monodecyl ether (C10E5)	non-ionic	-
Sucrose monododecanoate	non-ionic	-
Dimethyldodecylphosphine oxide	partially zwitterionic	-
n-Tetradecyl-N,N-dimethylamine-N-oxide (TDAO)	partially zwitterionic	-
n-Dodecyl-N,N-dimethylamine-N-oxide (DDAO)	partially zwitterionic	+

**Table 3.1** List of detergents utilized in the solubilization screen. The detergents are classified and their ability to solubilize HlyB is characterized.

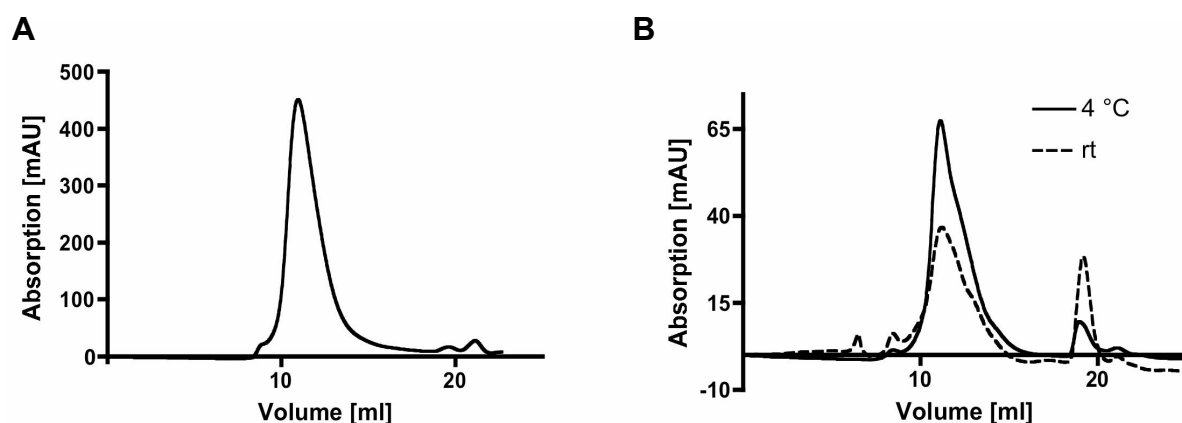
### 3.1.3 Purification of HlyB

The solubilization screen showed good results with detergents of the FOS-CHOLINE group. Thus, the solubilization of HlyB was performed employing one of the FOS-CHOLINE detergents. Generally, detergents with longer alkyl chains form larger micelles, which is undesired in membrane protein crystallization<sup>178</sup>. Therefore, purification of HlyB was performed with FOS-CHOLINE14.

After defrosting the membrane suspension 1% (w/v) FOS-CHOLINE14 was added and the solubilization was carried out by stirring the solution slowly at 8 °C. Unsolubilized membranes were removed by a second ultracentrifugation step (60 min at 200000 g, 4 °C). The purification of HlyB was achieved by immobilized metal ion affinity chromatography utilizing the N-terminal polyhistidine tag at 4 °C. The purified protein was concentrated to 10-15 mg/ml with an Amicon Ultra 15 concentrator of a molecular weight cutoff of 100 kDa. The solution was further dialyzed in a Dispo Dialyzer against buffer 5, supplemented with 0.0115% (w/v) FOS-CHOLINE14. This preparation yielded 0.7-1.0 mg pure protein per liter of cell culture.

To analyze the preparation for monodispersity and possible aggregation the protein solution was separated on a gelfiltration column (Superdex 200 10/300 GL) using buffer 5 supplemented with 0.0115% FOS-CHOLINE<sup>14</sup>. The calibration of the chromatogram with respect to the molecular weight was performed with a protein standard of soluble proteins. A typical example of a gelfiltration run is shown in Figure 3.2a.

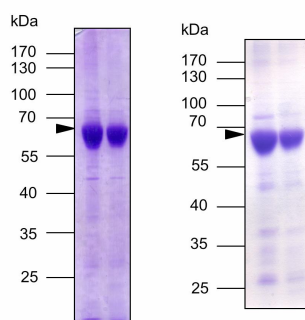
The stability of HlyB was also analyzed by gelfiltration chromatography. Two samples, one stored at 4 °C the other at room temperature for nine days, were applied on a Superdex 200 gelfiltration column. Whereas the sample stored at 4 °C showed no aggregation the sample stored at room temperature showed a very small aggregation peak (Figure 3.2b). However, both samples exhibited a slight shoulder formation to larger volumes. Besides, the samples show a peak at 19 ml, which is more pronounced in the sample stored at room temperature. This peak probably is due to protein degradation in the sample, which could also explain the shoulder formation seen at the main peak. Nevertheless, the protein purification results in a rather stable sample preparation, which is suitable for protein crystallization experiments.



**Figure 3.2** Gelfiltration chromatography. **A** Directly after purification. **B** After storage for nine days at 4 °C (full line) or at room temperature (rt; dashed line).

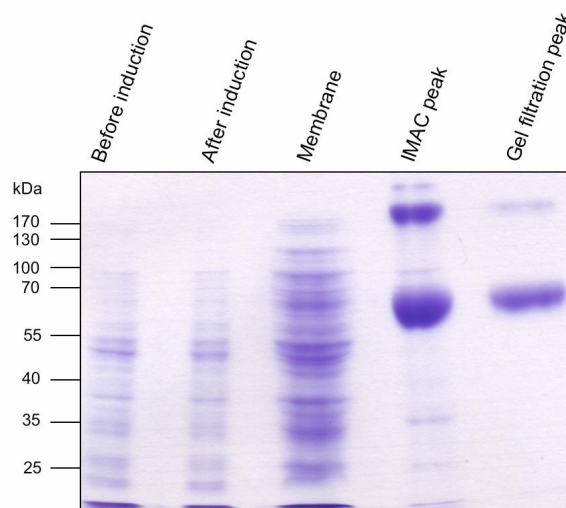
Crystallization experiments were conducted with wild type HlyB as well as with mutant (H662A) HlyB. For the isolated NBD the mutation of His662 to Ala showed a deficiency in ATP hydrolysis<sup>44</sup>. Therefore, it was anticipated that the full length protein would show a similar deficiency, which might help to stabilize the dimeric arrangement of the NBDs. The expressed proteins both comprised an N-terminal polyhistidine tag. The wild type protein exposed a tag of ten histidine residues, and the tag of the mutant protein consisted of eight histidine residues. Whereas the purification of wild type HlyB yielded a sample of 95% purity, the mutant HlyB showed slightly more impurities resulting in a 90% pure sample (see Figure 3.3). This discrepancy in the purity is due to the fact that the additional histidines in the

wild type protein enable a tighter binding of the histidine tag to the  $\text{Zn}^{2+}$ /IDA column. Thus, a washing step with 20% of buffer 4 can be employed. Only 15 % of buffer 4 could be used when an octahistidine tag was fused to the protein. With the higher amount of imidazole during washing interactions between impurities and column can be successfully decreased. In both cases the purification of HlyB yielded 0.7-1.0 mg per liter cell culture.



**Figure 3.3** 10% SDS-PAGE of purified wild type HlyB (left) and H662A HlyB (right) with 0.0115% FOS-CHOLINE14. The peak fractions are shown. HlyB is indicated by an arrow.

Further impurities of the sample were reduced by using high detergent concentrations during the purification steps (see Figure 3.4). Therefore, all buffers were supplemented with 0.46% FOS-CHOLINE14. As an example for the purification process, a 10% SDS-PAGE is depicted in Figure 3.4 showing a cell lysate before and after induction (lane 1 and 2 respectively), the isolated membrane (lane 3) and the purified protein after the IMAC (lane 4) step and after gelfiltration chromatography (lane 5). Comparing the cell lysate before and after induction reveals only a faint band for HlyB. This little amount of overexpressed protein can be successfully enriched. The purified sample in lane 4 exposes very little impurities. HlyB shows two bands, one at 70 kDa and a second band (~200 kDa) accounting for higher oligomers, which show resistance to SDS. In comparison to figure 3.3, sample impurities are greatly reduced; only two very weak bands (35 kDa and ~90 kDa) are revealed in the SDS-PAGE. As lane 5 reveals, an additional gelfiltration step improves purity even more.



**Figure 3.4** 10% SDS-PAGE of purified HlyB with 0.46% FOS-CHOLINE14. Lane 1 and 2 show cell lysate before and after induction with nisin A. Lane 3 displays the membrane preparation. Lane 4 and 5 show the purified protein after IMAC and after gelfiltration chromatography, respectively.

Purification of HlyB at high concentration of detergent resulted in a preparation of high purity. This phenomenon can be explained by the fact that at such detergent concentrations more empty detergent micelles are present during the purification. Thus, impurities are able to integrate in different detergent micelles. The protein detergent complex, which at low detergent amounts may also include impurities, is depleted of its impurities and HlyB can be purified as sole HlyB/detergent complex.

### 3.1.4 Detergent exchange

As the solubilization assay revealed, HlyB was solubilized mainly with detergents comprising a phosphatidylcholine headgroup. However, the success rate of HlyB crystallization with FOS-CHOLINE 14 was very low. Therefore, it was anticipated that purification of HlyB in different detergents and crystallization with the resultant protein samples may improve crystallization hits. To circumvent the problem of solubilization with non-phosphatidylcholine detergents, HlyB was solubilized with FOS-CHOLINE 14 and the detergent micelle surrounding HlyB was to be replaced in a subsequent purification step. A detergent exchange during gelfiltration chromatography often is non-quantitative<sup>179</sup>. Thus, detergent exchange during the IMAC step was aimed for.

In this approach the protein/detergent sample, obtained after solubilization with FOS-CHOLINE14 is applied on a  $\text{Zn}^{2+}$ -IDA column. The column was equilibrated with buffer 3 supplemented with 0.0115% FOS-CHOLINE14. After an initial wash step with buffer containing FOS-CHOLINE14, all subsequent steps were performed with buffers containing

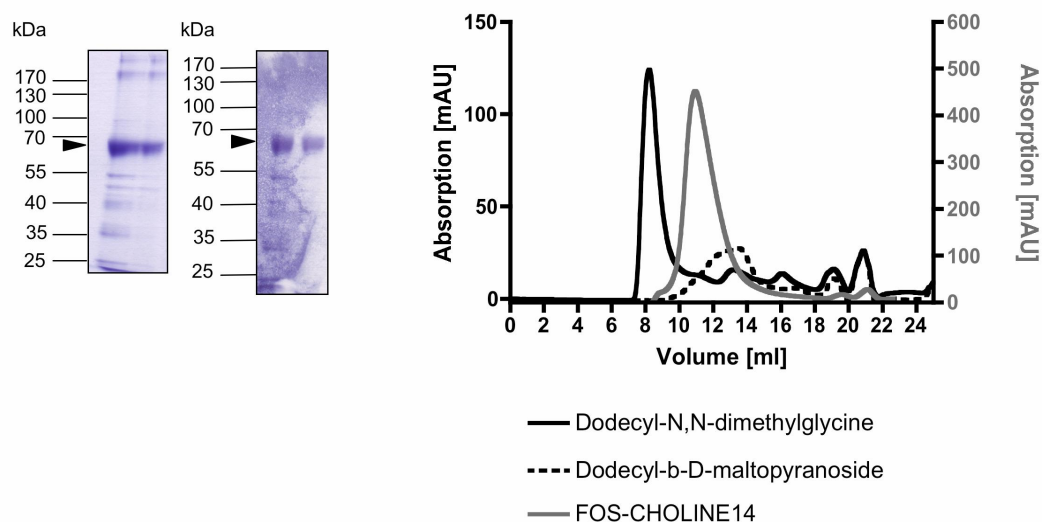
the new detergent. The fractions obtained by affinity purification were analyzed on a 10% SDS-PAGE to access the purity of the sample. The yield was determined with a Bradford assay. The protein sample was concentrated with an Amicon Ultra 15 concentrator (molecular weight cutoff: 100 kD) and applied on a Superdex 200 10/300 GL gelfiltration column to analyze for possible aggregation.

Yields, comparable to those of an established FOS-CHOLINE14 purification, were obtained with the zwitterionic detergents tetradecyl-N,N-dimethylamine-N-oxide and dodecyl-N,N-dimethylglycine, but the purity of the sample was poor and aggregates were formed (see Figure 3.5). The sugar based detergents dodecyl- $\beta$ -D-maltopyranoside and decyl- $\beta$ -D-maltopyranoside showed an acceptable yield. However, the preparation with decyl- $\beta$ -D-maltopyranoside exposed mainly aggregates. Although the size exclusion chromatography shows a rather broad peak (see Figure 3.5), the purification with dodecyl- $\beta$ -D-maltopyranoside appears successful and may form a base for further optimization. Table 3.2 summarizes the results of the detergent exchange screen.

Detergent	Concentration (w/v)	Yield [mg/l culture]	Purity after IMAC	Size exclusion chromatography
Dodecyl- $\beta$ -D-maltopyranoside	0.019%	0.5	+	broad peak
Decyl- $\beta$ -D-maltopyranoside	0.218%	0.5	++	mainly aggregates
Tetradecyl-N,N-dimethylamine-N-oxide	0.019%	0.9	-	aggregates on filter
Dodecyl-N,N-dimethylglycine	0.103%	0.9	-	Aggregates

**Table 3.2** The detergent utilized in the detergent exchange screen and the concentrations at which they were utilized are listed. The success of the exchange is judged by the yield, the purity of the sample after the IMAC step and the behavior on a gelfiltration column.

Most detergents utilized in the detergent exchange led to an aggregation of the protein. One detergent, dodecyl- $\beta$ -D-maltopyranoside, yielded low amounts of protein, which was not aggregated. However, it showed a rather broad peak in the gelfiltration chromatogram (see Figure 3.5). This is probably due to the impurities, as observed in an SDS-PAGE. The detergent exchange only yielded low amounts of protein with high impurities. Thus, this method is of little use for the preparation of HlyB for crystallization.



**Figure 3.5** SDS-PAGE and gel filtration chromatogram of the detergent exchange to dodecyl-N,N-dimethylglycine (left gel, black full line) and dodecyl- $\beta$ -D-maltopyranoside (right gel, black dashed line). Comparison to FOS-CHOLINE14 is shown (grey line). HlyB is indicated by an arrow.

### 3.1.5 Crystallization of HlyB

The established preparation of HlyB in FOS-CHOLINE14 was utilized to find suitable crystallization conditions. Vapor diffusion setups as well as crystallization in an unconventional batch method were used in the search for conditions that yielded crystals of HlyB.

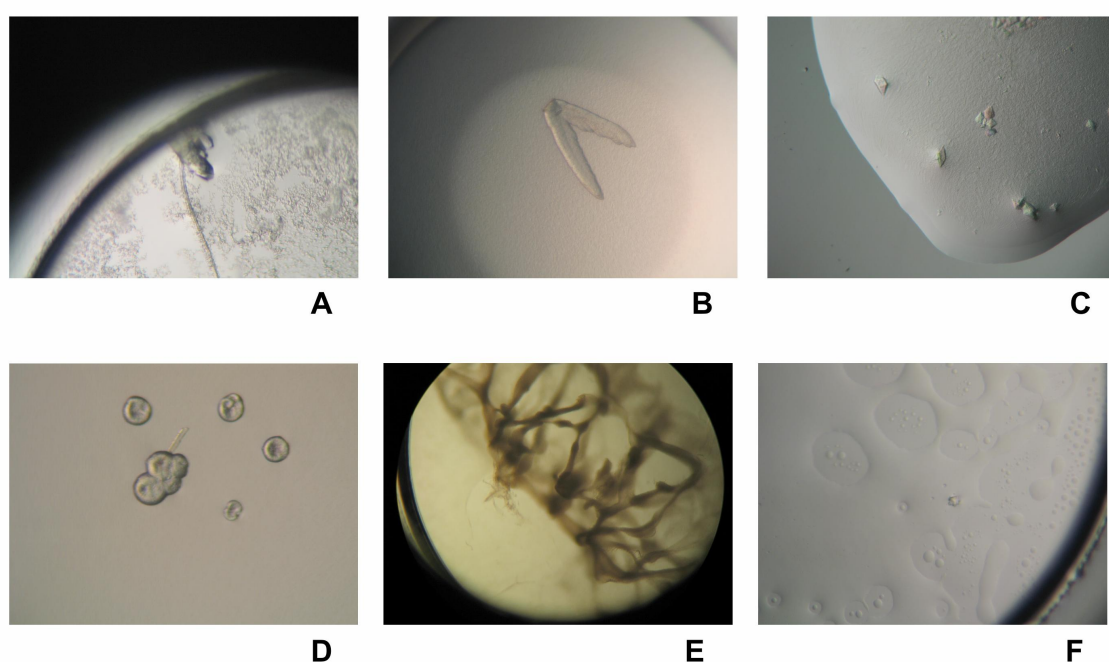
In the vapor diffusion setup, the MB ClassI and MB ClassII (Nextal/Qiagen) as well as the Crystallization Basic Kit for Proteins and the Crystallization Extension Kit for Proteins (Sigma) were applied. Furthermore, a self made screen based mainly on the Jena Bioscience Screens Classic 1, 2 and 7 was utilized as well. Crystallization of the protein sample was performed in 96-well plates either as a sitting or a hanging drop setup with a protein concentration of 10-20 mg/ml. Subsequently, the plates were stored at 4, 10 or 18 °C. The wild type protein was either supplemented with 5 mM ADP or used without ligand. As ATPase activity was anticipated the protein solution was not setup with ATP. However, the H662A HlyB mutant, which is deficient in ATP hydrolysis, was usually mixed with 5 mM ATP prior to setup. In some screens the protein solution was supplemented with an additive. The alcohols 1,2,3-heptantriol (2%) and 2-methylpentandiol (3%) as well as benzamidine (2%) were added to evaluate their effect on crystallization. Also crystallization with 0.46% FOS-CHOLINE14 was performed to analyze the effect of high detergent concentrations.

1,2,3-Heptantriol did not influence the setups greatly. However, setups performed with 2-methylpentandiol showed higher amounts of brown precipitate implicating a negative impact

on protein stability. Drops containing benzamidine exposed less precipitate in comparison to the setups without the additive. Increased phase separation was revealed when high amounts of FOS-CHOLINE14 were added.

Generally, at temperatures higher than 4 °C more precipitation was observed. Conditions containing a higher concentration of salt as precipitant, e.g. sodium chloride, potassium phosphate or ammonium sulfate at concentrations up to 2.5 M, showed no protein precipitation. Solutions containing PEG 4000 as precipitant often exposed phase separation at lower pH than 5.5. Higher molecular PEGs, e.g. PEG 6000 or PEG 8000, induced phase separation, especially when combined with buffers of acidic pH. MPD used in higher concentrations than 15%, results in brown protein precipitate. In contrast, the precipitant PEG 2000 often yields protein precipitation as well, but the drops show no brown color. Using PEG 400 as precipitant is mainly successful at concentrations of 30% and more. Other small PEGs, i.e PEG 550 MME and PEG 600 were only used at higher concentrations and showed similar results as PEG 400.

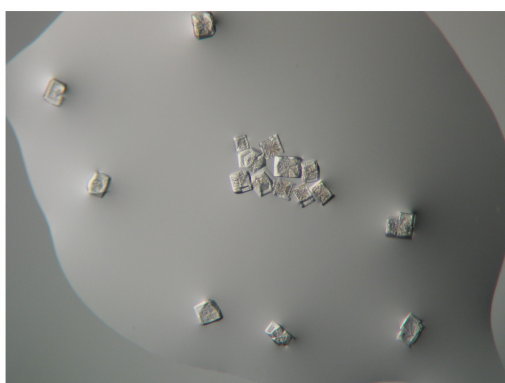
Drop phenomena encountered during the crystallization experiments of HlyB are depicted in Figure 3.6.



**Figure 3.6** Phenomena encountered at the crystallization experiments of HlyB. **A,B,C** Salt crystals **D** Spherulites **E** Protein precipitate **F** Phase separation.

An initial test was performed to assess the protein stability at different pH values. Therefore, protein was mixed with different buffers and placed over a reservoir. After a few months crystals appeared in the drop which contained cacodylate pH 6.6 as buffer. To reproduce these

crystals the concentration of the buffer and its pH was varied, different temperatures were analyzed, the FOS-CHOLINE14 concentration was increased and detergents were used as additives. The drops were placed over a reservoir, where the volume of the reservoir was changed by adding different amounts of paraffin oil. Crystals appeared in different conditions, but analysis at the beam-line BW6 (DESY, Hamburg) revealed that most of these were salt crystals. However, one of the crystals (see Figure 3.7) analyzed showed no salt or detergent diffraction but very weak protein diffraction. The crystal appeared at 25 °C in a setup comprising 0.46% FOS-CHOLINE14 and 1 M cacodylate pH 6.4. As cryoprotectant 5% ethylene glycol was chosen. Attempts to reproduce the crystal were unsuccessful which may be due the batch setup.



**Figure 3.7** Crystals appearing at 25 °C in a modified batch setup with HlyB in 0.46% FOS-CHOLINE and 1 M cacodylate at pH 6.4 used as precipitant.

## 3.2 The substrate binding protein ChoX

ChoX from *S. meliloti* was kindly provided and affinity determination was performed by the group of Prof. Erhard Bremer (University Marburg). Here, the expression and purification as well as the binding assay shall be described briefly.

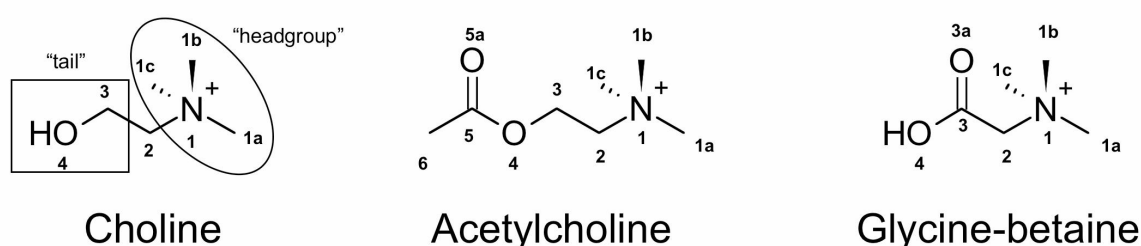
### 3.2.1 Expression and purification of ChoX

ChoX was expressed in *E. coli* BL21 (DE3) (pLysS) using a pET20b vector. The cells were grown on minimal media. Periplasmic cell rupture was performed by osmotic shock. The cells were resuspended in ice-cold buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole pH 8.0 and 500 mM sucrose). Thereafter, they were incubated on ice for 30 min. Cell debris and unbroken cells were removed by ultracentrifugation and the supernatant was applied on a Ni-NTA column for affinity purification using the C-terminal hexahistidine tag. The column was

washed with buffer A (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl and 10 mM imidazole pH 8.0). To elute the protein from the affinity column a gradient from 10-250 mM imidazole was used using buffer A and B (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl and 250 mM imidazole pH 8.0). The buffer was exchanged by dialysis against 10 mM Tris/HCl pH 7.0 for crystallization experiments and 10 mM Tris/HCl pH 7.0 and 100 mM NaCl for binding assays.

### 3.2.2 Ligand binding assay

The affinity of ChoX to its native ligand choline as well as to acetylcholine and glycine-betaine (see Figure 3.8) was monitored by binding assays using radioactively labeled ligand.

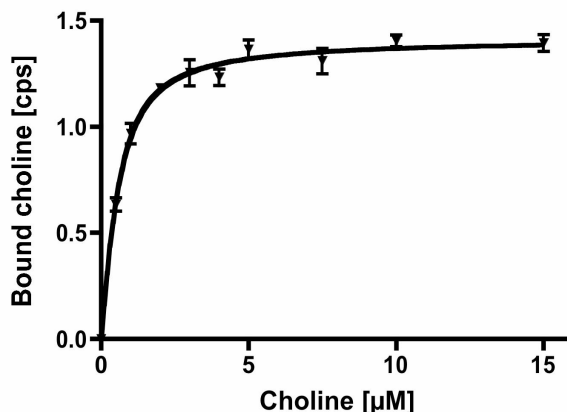


**Figure 3.8** Substrates used for analyses. For clarity the atoms, apart from H-atoms, are labeled. As an example, the "headgroup" and the "tail" of choline are indicated by an eclipse and a rectangle, respectively.

To allow binding of the substrate to the protein, ChoX (500 nM) was incubated with <sup>14</sup>C-labeled substrate. Substrate at various concentrations (0.1 to 25 μM for choline, 0.5 to 25 μM for glycine-betaine and 0.5 to 25 μM for acetylcholine) was added to the protein and incubated at room temperature. After 5 min the protein was precipitated with ice-cold, saturated ammonium sulfate and further incubated on ice for 10 min. The obtained suspension was passed through a filter and the precipitate was washed with ammonium sulfate to remove unspecifically bound substrate. Measuring the radioactivity of the bound substrate enables the determination of the K<sub>d</sub> values. The used substrate concentration is plotted against the counts of the bound substrate. K<sub>d</sub> values are calculated according to equation 3.1, which accounts for the high amount of protein concentration used in the substrate binding assay<sup>180</sup>. Here, T represents the protein concentration used, K<sub>d</sub> the dissociation constant, B<sub>max</sub> the maximum amount of bound substrate and x the substrate concentration. The determined K<sub>d</sub> values were 0.3 ± 0.1 μM, 3.8 ± 0.6 μM and 22.8 ± 6.9 μM for choline, glycine-betaine and acetylcholine, respectively.

$$y = (K_d + T + x) - \sqrt{(K_d + T + x)^2 - 4Tx} / (2/TB_{\max}) \quad (3.1)$$

Figure 3.9 exemplifies the determination of the  $K_d$  value for the binding of choline to ChoX.



**Figure 3.9** The determination of the  $K_d$  value for choline binding to ChoX. The used substrate concentration is plotted against the counts of the radioactively labeled, bound substrate.

### 3.2.3 Crystallization

To obtain initial crystallization conditions for ChoX with its substrates choline (Figure 3.8) and acetylcholine (see Figure 3.8), the vapor diffusion method in a hanging drop setup was utilized.

Commercially available sparse matrix screens (Crystallization Basic Kit for Proteins, Crystallization Extension Kit for Proteins, Sigma) were used to examine a broad variety of crystallization conditions. Setups were performed at 4 °C as well as at room temperature.

The protein (10-15 mg/ml in 100 mM Tris pH 7.0) was mixed with the respective substrate and incubated on ice for 30 min prior to crystallization. The final concentration of the ligand in the protein solution was 1 mM. A hanging drop setup with a protein to precipitant ratio of 1:1 and a final volume of 2  $\mu$ l proved most efficient for crystal growth. All setups were performed in 24-well crystallization plates.

An initial hit for the crystallization of ChoX in complex with its native ligands choline and acetylcholine was obtained with condition number 37 of the Crystallization Basic Kit for Proteins (100 mM Sodium acetate pH 4.6, 8% PEG 4000). Crystals appeared as small clusters of thin plates. This initial crystallization hit was further optimized in order to grow larger and more voluminous crystals. The optimization procedure was based on fine screening with grid screens. These grid screens were designed with 100 mM Sodium acetate at varying pH values

(pH 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.8, 5.0, 5.1). Different PEGs at varying concentrations, summarized in

Table 3.3, were used as precipitant.

Precipitant	Concentration
PEG 1000	12, 18% (w/v)
PEG 1500	12, 14, 16, 18% (w/v)
PEG 2000 MME	10, 12, 14, 18% (w/v)
PEG 3350	10, 12, 14, 16, 18, 20, 22% (w/v)
PEG 4000	6, 10, 16, 18, 20% (w/v)

**Table 3.3** Different PEGs and their concentrations used in the grid screen for optimization of the initial hit of ChoX.

With the help of the designed grid screens crystals could be grown in various conditions. The most suitable crystals were obtained at 20% PEG 3350, 100 mM Sodium acetate pH 4.8. However, it was notoriously hard to grow single crystals. Mostly, bunches of plates with a few outstanding extra large plates emerged in the crystallization drop. Figure 3.10 shows some examples of crystals grown with the hanging drop method as described before. All crystals appeared within three to four weeks and grew to their final size within another two weeks.



**Figure 3.10** Examples of crystals of ChoX grown with the hanging drop vapor diffusion method. Wild type ChoX cocrystallized with choline (left) and acetylcholine (middle). Se-Met substituted ChoX (right) cocrystallized with choline.

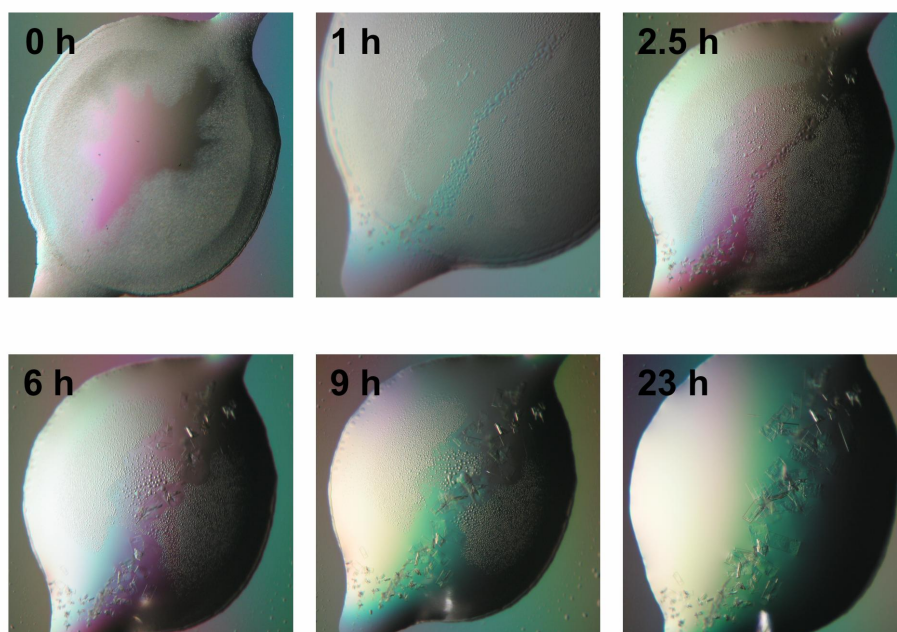
The grid screen yielded crystals from setups with ChoX mixed with choline as well as with acetylcholine, but it remained impossible to find suitable crystallization conditions for ChoX in an unliganded form. As analyzed during structure determination of the data obtained from crystals grown with acetylcholine, the electron density corresponded to choline but not

acetylcholine. Thus, it was assumed that during crystal growth the acetylcholine in the setup was subjected to hydrolysis. The resulting choline species probably bound to ChoX due to the higher affinity.

To circumvent this obstacle a method for quicker crystals growth was needed. Therefore, microseeding by the streak seeding method was established<sup>127</sup>.

Usually, for microseeding experiments, the concentration of the precipitant solution is lowered in comparison to the original condition. However, as quick crystal growth was favored, the precipitant concentration had to be high enough to allow the rapid formation of crystals. Therefore, a broader spectrum of precipitant concentration had to be screened. Precipitant solutions used in the microseeding setups contained 100 mM Sodium acetate pH 4.9/5.0/5.1 and 10-22% PEG 3350. With the help of a horse tail hair microcrystals, presenting nuclei for crystal growth, were transferred from a ChoX/choline crystal to a new setup by dragging the hair through the fresh setup. After two drops were inoculated with nuclei in this way, the ChoX/choline crystal was touched again and another two drops were inoculated with new nuclei. Subsequently, plates were either stored at room temperature (20 °C) or at 4 °C.

This method allowed for quick crystal growth yielding first small crystals within only 2 hours after inoculation. After ~24 hours, crystals, which were already large enough to be used for data collection, were mounted in a cryoloop and frozen in liquid nitrogen to prevent hydrolysis of acetylcholine. The speed of crystallization is exemplified in a time-course depicted in Figure 3.11. Furthermore, it was assumed that ligand free ChoX constantly undergoes an opening and closing movement. This rather dynamic species hinders nuclei formation. To overcome this nucleation problem, seeding into drops containing unliganded ChoX was performed.



**Figure 3.11** Time course of crystal growth utilizing the microseeding method by streak seeding. Crystals appear after 2 hours and grow to sizes suitable for data collection within 24 hours.

With this microseeding method crystals of ChoX/acetylcholine, ChoX/glycine-betaine and ChoX in unliganded states could be produced, which all proofed to be suitable for data collection.

Crystallization conditions of the crystals used for data collection are summarized in Table 3.4.

Parameter	Choline	Acetyl- choline	Glycine- betaine	Unliganded (closed)	Unliganded (semi-closed)
Reservoir solution	100 mM Na- acetate pH 4.6, 18% PEG 3350	100 mM Na- acetate pH 5.0, 16% PEG 3350	100 mM Na- acetate pH 5.1, 18% PEG 3350	100 mM Na- acetate pH 5.1, 22% PEG 3350	100 mM Na- acetate pH 5.0, 18% PEG 3350
temperature	277 K	293 K	277K	277 K	293 K
protein solution	9 mg/ml; 10 mM Tris/HCl pH 7.0, 1 mM choline	10 mg/ml ; 10 mM Tris/HCl pH 7.0, 130 mM NaCl, 1 mM acetylcholine	14 mg/ml ; 10mM Tris/HCl pH 7.1, 50 mM NaCl, 100 mM GB	14 mg/ml ; 10 mM Tris/HCl pH 7.0, 10 mM NaCl	14 mg/ml ; 10 mM Tris/HCl pH 7.0, 10 mM NaCl
protein:reservoir	1:1	1:1	1:1	1:1	1:1
microseeding	no	yes	yes	Yes	Yes

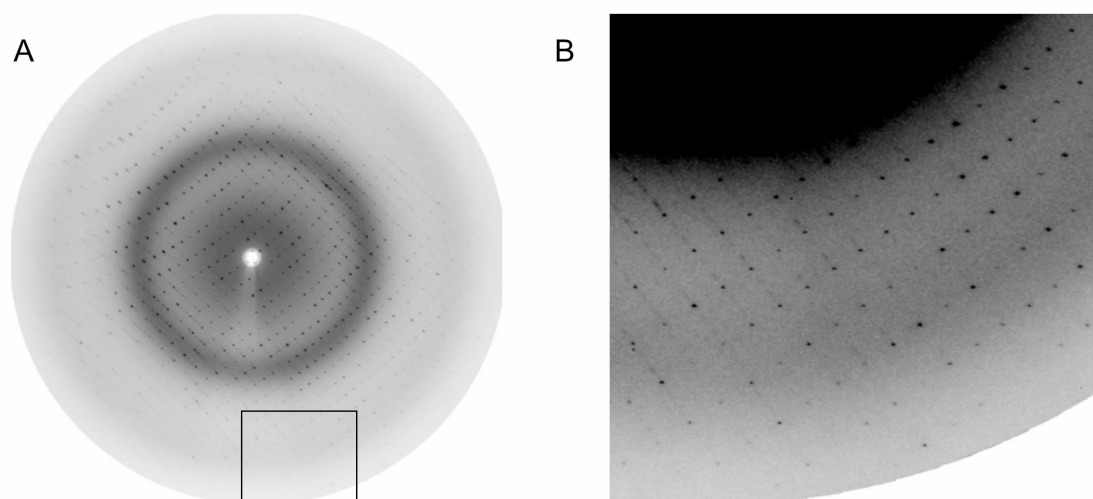
**Table 3.4** Crystallization conditions for crystals grown with the microseeding method.

### 3.2.4 Data collection and processing

To allow data collection at 100 K, which reduces radiation damage, all crystals of ChoX were treated with a cryoprotective buffer prior to freezing in liquid nitrogen. Therefore, the obtained crystal plates, as shown in Figure 3.10, were separated from each other using micro-tools. The crystals gained by this procedure were further fished out of the drop with the aid of a nylon loop. Before freezing the crystals in liquid nitrogen, they were transferred to the cryoprotective buffer, in which they were soaked for 1 min. Cryoprotectant solutions consisted of the corresponding reservoir solution supplemented with 25% (v/v) ethylene glycol.

Data collection was performed at the Max-Planck beamline BW6 (ChoX/choline) and the EMBL beamline BW7A (ChoX/acetylcholine, unliganded ChoX closed, unliganded ChoX semi-closed and ChoX/glycine-betaine) at the “Deutsches Elektronen Synchrotron” (DESY, Hamburg) at 100 K (for a summary see

Table 3.5). As a typical example for the diffraction quality of ChoX crystals a diffraction picture of a ChoX/acetylcholine crystal, diffracting to 1.8 Å, is shown in Figure 3.12.



**Figure 3.12** Typical diffraction image obtained from a ChoX/acetylcholine crystal. **A** Diffraction quality at low resolution. **B** The quality of the diffraction at high resolution. A close up is shown of the region marked by a rectangle in **A**.

Data processing, which involves the steps of indexing, integrating and scaling the reflections, was performed either using DENZO<sup>155</sup> (ChoX/choline) or XDS<sup>156</sup> (ChoX/acetylcholine, unliganded ChoX closed and unliganded ChoX semi-closed). All datasets belonged to the

space group P2<sub>1</sub>, but they exposed different cell parameters. Processing statistics are summarized in

Table 3.5.

	Choline (native)	Acetylcholine	Unliganded closed	Unliganded semi-closed
Space group	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>
Cell dimensions				
a, b, c (Å)	30.8, 185.4, 42.5	31.3, 213.1, 42.9	30.9, 196.2, 42.8	34.1, 232.5, 47.3
β (°)	91.8	90.1	90.1	90.2
Wavelength	1.05	0.98	1.033	0.90
Resolution (Å)*	93-1.9 (1.93-1.90)	20-1.8 (1.90-1.80)	20-1.6 (1.70-1.60)	20-2.5 (2.60-2.50)
<sup>a</sup> R <sub>sym</sub> / <sup>b</sup> R <sub>mrgd-F</sub> *	<sup>a</sup> 4.5 (31.3)	<sup>b</sup> 5.5 (19.7)	<sup>b</sup> 10.4 (32.2)	<sup>b</sup> 22.2 (46.0)
I / σI*	16.7 (3.9)	14.9 (4.4)	8.3 (3.4)	7.8 (3.5)
Completeness (%)*	90.3 (83.7)	93.9 (87.6)	99.1 (97.1)	91.2 (29.3)
Redundancy	14.1	1.87	3.5	3.4
No. unique reflections	37385	48201	66140	22282

**Table 3.5** Processing statistics of the ChoX datasets with bound choline and acetylcholine and in unliganded states. \* Values for the highest resolution shells are given in brackets.

$$R_{sym} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

$$R_{mrgd-F} = \frac{\sum |A_{I_{h,P}} - A_{I_{h,Q}}|}{0.5 * \sum A_{I_{h,P}} + A_{I_{h,Q}}}$$

As the completeness and the R<sub>mrgd-F</sub> in the highest resolution shell for the unliganded semi-closed dataset are not satisfying, the data was cut at 2.6 Å for structure determination.

### 3.2.5 Obtaining the ChoX/choline model

#### *Solving the phase problem*

During the diffraction experiment X-rays are scattered by the electrons of the protein forming the crystal. To obtain the information about the position of the electrons it is necessary to have all characteristics of the diffracted rays. The wavelength does not change during the experiment and therefore presents no problem. While the amplitude of a scattered ray can be related to the intensity detected during the diffraction experiment, the information of the phase can not be deduced. Therefore, different methods have been developed to obtain this phase information.

Multiple wavelengths anomalous diffraction (MAD)<sup>136,137,181</sup>, was utilized to solve the phase problem. Therefore, selenomethionine substituted protein<sup>138</sup> was crystallized (see Figure 3.10

on the right) using a grid screen around the conditions which were successful for the wild type protein.

Table 3.6 lists the parameters for successful crystallization of Se-Met substituted protein.

Parameter	Se-Met ChoX/Choline
reservoir solution	100 mM Na-acetate pH 4.2, 25 % PEG 3350
Temperature	277 K
protein solution	9 mg/ml; 10 mM Tris pH 7.0, 1 mM choline
protein:reservoir	1 $\mu$ l:1 $\mu$ l
Microseeding	No

**Table 3.6** Crystallization conditions for successful crystallization of Se-Met substituted ChoX.

As described before crystals of the Se-Met substituted ChoX were transferred to a cryoprotective buffer (100mM Na-acetate pH 4.2, 25% PEG 3350, 25% EG) and frozen in liquid nitrogen. Subsequently, a crystal was subjected to data collection at the Max-Planck beamline BW6 at the DESY, Hamburg. To obtain phase information via the MAD method it is necessary to use data at different wavelengths. Therefore, datasets were collected using wavelengths of 0.9791 Å (peak), 0.9794 Å (inflection) and 0.9500 Å (remote). Data processing was performed using Denzo<sup>155</sup>. Processing statistics of the Se-Met substituted ChoX/choline crystal are shown in

Table 3.7.

	Se-Met ChoX/choline (Peak)	Se-Met ChoX/choline (Inflection)	Se-Met ChoX/choline (Remote)
Space group	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	42.2, 210.3, 30.3	42.2, 210.3, 30.3	42.2, 210.3, 30.3
$\beta$ (°)	91.3	91.3	91.3
Wavelength (Å)	0.9791	0.9794	0.9500
Resolution (Å)*	105.4-2.9 (2.94-2.90)	105.4-2.9 (2.94-2.90)	105.4-2.9 (2.94-2.90)
<i>R</i> <sub>sym</sub> (%)	9.6 (25.8)	8.4 (28.5)	8.4 (28.1)
<i>I</i> / $\sigma$ *	20.5 (6.0)	13.5 (3.6)	13.1 (3.6)
Completeness (%)*	99.9 (98.7)	99.5 (97.6)	99.2 (95.5)
No. reflections	78302	38876	38427
No. unique reflections	11734	11799	11833
Redundancy	6.7	3.3	3.2

**Table 3.7** Processing statistics for the MAD dataset of the Se-Met substituted ChoX. \* Values for the highest resolution shells are given in brackets.

$$R_{sym} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

ChoX, depleted of its signal sequence, exhibits seven methionines. The Matthews coefficient<sup>182</sup> for Se-Met substituted ChoX (32200 Da) with the given cell constants shows a value of 2.1 Å<sup>3</sup>/Da, which correlates with 40% solvent content in the crystal for two molecules in the ASU. Therefore, fourteen Se-sites were expected in the ASU of the Se-Met substituted crystal. Using SOLVE<sup>157</sup> eleven of the fourteen sites could be located resulting in a mean figure of merit of 49%. With the help of the position of the Se atoms phases for the individual structure factors could be obtained and an initial electron density could be calculated. Further density modification and automated model building with RESOLVE<sup>158,159</sup> provided an initial model for ~35% of the ASU content. Finally, by iterative cycles of manual building in COOT<sup>160</sup> using 2F<sub>o</sub>-F<sub>c</sub> and 1F<sub>o</sub>-F<sub>c</sub> density maps, and restrained maximum likelihood refinement cycles utilizing isotropic B-factor refinement in Refmac5<sup>161</sup>, one monomer of the asymmetric unit could be obtained. To assess the quality of the model the R-factor and the free R-factor (R<sub>free</sub>)<sup>183</sup> were monitored. The figure of merit was controlled to ensure that the quality of the phase information is improving with the ongoing refinement. The obtained monomer is constituted of residues 31-314, lacking the N-terminal residues 28-31 (residues 1-27 belong to the signal sequence) and C-terminal (315-325) residues of the protein plus the hexahistidine tag.

#### *Refinement of wild type ChoX/choline*

As a high resolution dataset of ChoX/choline was available, it was anticipated that molecular replacement with the monomer obtained from the Se-Met substituted dataset on this dataset could help to derive the second monomer for the wild type ChoX/choline. Molecular replacement with MOLREP<sup>162</sup> indicated two clear solutions in the rotation function search yielding a model with two monomers in the asymmetric unit with a correlation coefficient of 47.9% and a R-factor of 45.5%. This model was further refined by iterative rounds of manual building in COOT<sup>160</sup> and restrained refinement in Refmac5<sup>161</sup>. In the first rounds non-crystallographic restraints were opposed, however, those were released at a later stage in the refinement process as the model quality improved. As the 2F<sub>o</sub>-F<sub>c</sub> and the 1F<sub>o</sub>-F<sub>c</sub> electron density map clearly showed substrate density, choline was fitted into the model. The necessary coordinate and geometric parameter files were generated with PRODRG<sup>166</sup>. Restrained refinement was combined with TLS refinement<sup>163,164</sup> using individual domains (residue 31-114 plus residue 235-314 for domain I and residue 120-230 for domain II) of each monomer as separate TLS group. Water molecules were detected and placed automatically

using ARP/wARP<sup>167</sup> at a threshold of 3.2 sigma, and subsequently analyzed manually for their correct position. The obtained model shows an R- and free R-factor of 18.0 and 22.2 %, respectively. It contains 269 waters and two protein monomers, monomer A being composed of residues 29-318 and monomer B, built up by residues 29-317. One choline molecule is bound per protein monomer. The Ramachandran plot showed 98.9% of the residues in the most favored and the additionally allowed region. The refinement statistics of the native ChoX/choline dataset are shown in

Table 3.8.

	Choline (native)
<b>Resolution (Å)</b>	20-1.9
<b><math>R_{\text{work}} / R_{\text{free}}</math></b>	18.0/22.2
<b>No. atoms</b>	4745
Protein	4462
Ligand/ion	14
Water	269
<b>B-factors</b>	
Protein	16.2/20.9
Ligand/ion	6.2/6.8
Water	25.3
<b>R.m.s. deviations</b>	
Bond lengths (Å)	0.007
Bond angles (°)	1.052
<b>Ramachandran Plot (%)</b>	
most favored	91.8 (453)
additionally allowed	7.1 (34)
generously allowed	0.8 (4)
disallowed regions	0.2 (1)

**Table 3.8** Refinement statistics obtained for the native ChoX/choline model at 1.9 Å. For the Ramachandran plot the values in brackets give the number of amino acids found in the respective region.

### 3.2.6 Generating models for ChoX/acetylcholine and the unliganded states of ChoX

#### *Obtaining phases for ChoX complexed with acetylcholine and unliganded ChoX*

Molecular replacement with the ChoX/choline model was utilized to obtain phases for the datasets of ChoX in complex with acetylcholine and ligand free ChoX. Therefore, the model generated in the refinement of ChoX/choline was modified by reducing it to one monomer (monomer A) and depleting it of water molecules as well as ligand molecules. The resulting monomer was utilized as input model in molecular replacement trials on the ChoX/acetylcholine, unliganded ChoX and unliganded semi-closed ChoX datasets. Molrep<sup>162</sup> as well as Phaser<sup>168-170</sup> were used to perform the molecular replacement trials. For ChoX in

complex with acetylcholine at first Phaser<sup>168-170</sup> and Molrep<sup>162</sup> were utilized for molecular replacement. After rigid body refinement in Refmac5<sup>161</sup> the Phaser output showed the better R-factors as well as a higher FOM (35/35.4 % for R/R<sub>free</sub> and a FOM of 62.2 % for Phaser output versus 37/38.3 % R/R<sub>free</sub> and a FOM of 55.3 % for the Molrep output). Thus, the refinement was carried out with the Phaser output. The Phaser Z-scores of the two top solutions obtained in the translation function and the Likelihood-gain of the final two monomers is shown in

Table 3.9.

Phaser for the two top solutions	Acetylcholine
Z-score	43.78 / 44.69
likelihood-gain	2517.92 / 2280.77

**Table 3.9** Molecular replacement statistics for the two top solutions obtained with Phaser<sup>168-170</sup> on the dataset of ChoX in complex with acetylcholine.

Molecular replacement with the unliganded ChoX was solved with Molrep<sup>162</sup>, which resulted in a clear cut solution containing two monomers (see Rf/sigma from second to third peak) for the resulting model of the asymmetric unit. The final R-factor and the correlation coefficient were 41.8% and 39.2%, respectively. The statistics are summarized in

Table 3.10.

Molrep for the two top solutions	unliganded ChoX
R-factor	41.8%
Correlation coefficient	39.2%
Rf/sigma 2 <sup>nd</sup> to 3 <sup>rd</sup> peak	8.09 → 3.73

**Table 3.10** Molecular replacement statistics for the two top solutions obtained with Molrep<sup>162</sup> on the dataset of unliganded ChoX.

However, obtaining an acceptable molecular replacement solution for the semi-closed unliganded ChoX was less straightforward. Using the previously generated input model on the dataset processed in P2<sub>1</sub> did not yield a clear cut result (Rf/sigma for 2<sup>nd</sup> to 3<sup>rd</sup> peak 4.46 → 4.10). Therefore, domains I and II were used separately as input model. But again no clear solution could be obtained (Rf/sigma for 2<sup>nd</sup> to 3<sup>rd</sup> peak 4.23 → 3.92 and 3.76 → 3.74 for domain I and II, respectively). As a next step, molecular replacement with the data processed in P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, the apparent, twinned space group, was carried out. The drop of Rf/sigma between

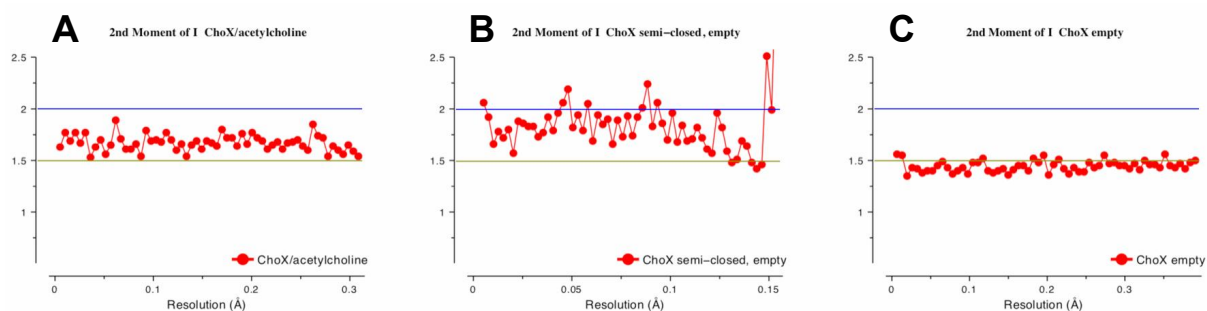
the first and second solution was more pronounced ( $R_f/\sigma$  4.8  $\rightarrow$  4.1). The output model was further refined in Refmac5<sup>161</sup> and then used as a new input on the dataset processed in P2<sub>1</sub>.

### *Detecting twinned data*

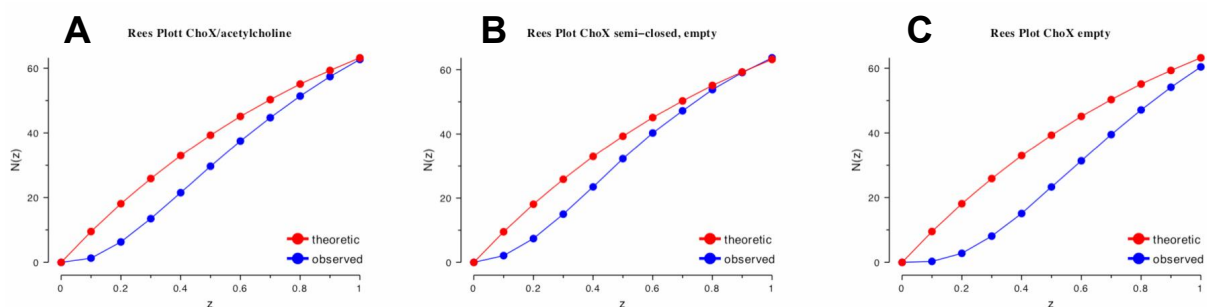
The refinement progress of ChoX/acetylcholine showed slow improvement and R-factors were stuck at 24.5/29% for  $R/R_{\text{free}}$  with a FOM of 79.3%. As the unit cell parameters showed a value of roughly 90° for the  $\beta$  angle, it was reassured that the determination of the space group P2<sub>1</sub> was correct. Therefore, twinning analyses were performed to detect possible pseudo-merohedral twinning. Statistical analysis showed high pseudo-merohedral twinning for the ChoX/acetylcholine dataset. As this problem was detected with ChoX/acetylcholine crystals, which grew from microseeding experiments, subsequently all data obtained from crystals grown in a similar setup were analyzed for twinning.

Statistical analysis for ChoX/acetylcholine (**A**), unliganded, semi-closed ChoX (**B**) and unliganded ChoX (**C**) are presented in figures 3.11-3.15. The second moments of the intensities (Figure 3.13) already give a first indication that the data is derived of a twinned crystal<sup>140</sup>. The value for untwinned data should be 2.0, whereas a perfect twinned crystal shows values around 1.5. Figure 3.13 thus indicates a perfect and nearly perfect twin for the unliganded ChoX and ChoX/acetylcholine. It is rather hard to estimate the twinning fraction for unliganded, semi-closed ChoX, however, twinning is definitely encountered in this case as well. The Rees plots<sup>141</sup> (Figure 3.14) of the three datasets corroborate these findings. According to the different percentage of weak intensities, unliganded ChoX shows the highest twinning fraction followed by ChoX in complex with acetylcholine. Furthermore, the Britton plots<sup>143</sup> (Figure 3.15) indicate twinning as well, however, when  $\alpha$  is evaluated it shows a higher twinning fraction for ChoX/acetylcholine than for unliganded ChoX, with approximate values of 0.39 and 0.35, respectively. The dataset of semi-closed ChoX shows a twinning fraction of only 28%. As the Britton plot<sup>143</sup> is not completely reliable concerning the estimation for  $\alpha$ , the data was further analyzed using the Yeates plot<sup>145</sup> (Figure 3.16). Analysis according to Yeates provides estimates for  $\alpha$  of  $\sim 0.41$ , 0.30 and 0.40 for ChoX/acetylcholine, semi-closed ChoX and unliganded ChoX, respectively. However, to avoid possible errors in the estimated twinning fraction, which can occur if a crystal shows anisotropic diffraction, analysis using the L-function according to Padilla and Yeates<sup>147</sup> was also performed (Figure 3.1). The highest twinning fraction is exhibited by the unliganded ChoX, estimated to be a

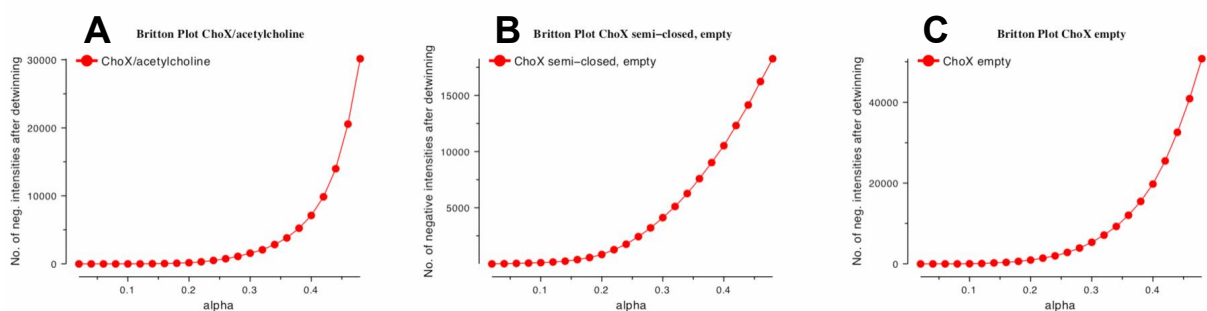
perfect twin. ChoX in complex with acetylcholine shows an estimated  $\alpha$  of  $\sim 0.48$ . And the semi-closed ChoX displays the lowest twinning fraction with an estimated  $\alpha$  of 0.45%.



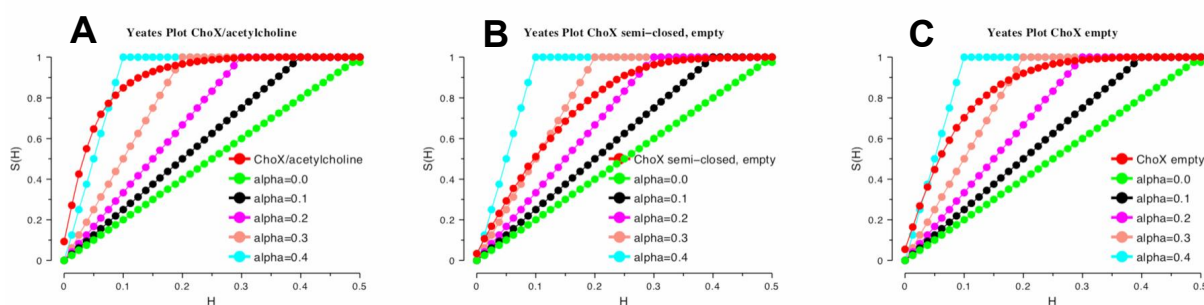
**Figure 3.13** Second moments of intensities. **A** Data of ChoX acetylcholine. **B** Data of unliganded semi-closed ChoX. **C** Data of unliganded ChoX.



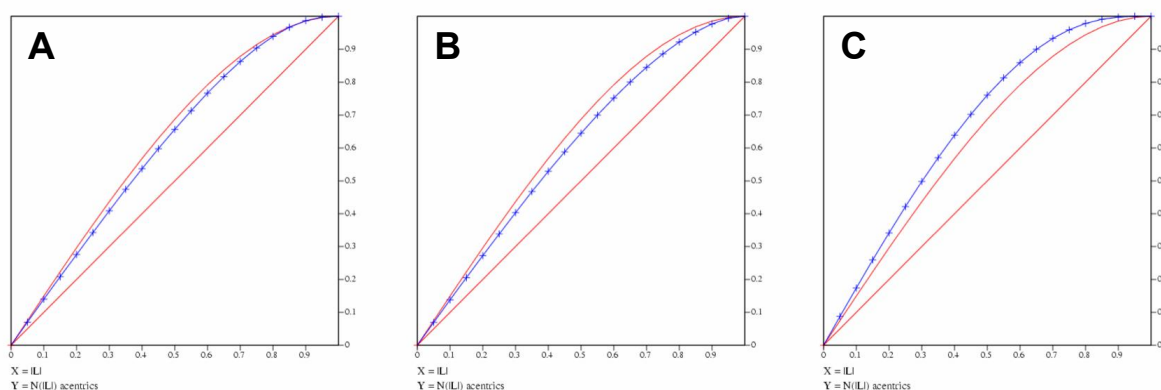
**Figure 3.14** Rees plot of intensities. **A** Data of ChoX acetylcholine. **B** Data of unliganded semi-closed ChoX. **C** Data of unliganded ChoX.



**Figure 3.15** Britton plot of intensities. **A** Data of ChoX acetylcholine. **B** Data of unliganded semi-closed ChoX. **C** Data of unliganded ChoX.



**Figure 3.16** Yeates plot of intensities. **A** Data of ChoX acetylcholine. **B** Data of unliganded semi-closed ChoX. **C** Data of unliganded ChoX.



**Figure 3.17** The L-function of intensities. **A** Data of ChoX acetylcholine. **B** Data of unliganded semi-closed ChoX. **C** Data of unliganded ChoX.

### *Refinement of twinned datasets*

In cases where the twinning factor  $\alpha$  is low, the twinning usually is neglected as it has only minor effect on the final model. If the twinning factor gets higher it has to be accounted for the twinned data. Data can be deconvoluted if the twinning fraction does not approach 50%. All three datasets showed twinning fractions of at least 40%, a value too high for deconvolution of the data and proceeding with usual refinement methods. Datasets of ChoX/acetylcholine and ChoX in its unliganded state were of high resolution (1.8 and 1.6 Å, respectively). Thus, refinement was pursued using SHELXL<sup>149</sup>, which allows for individual refinement of  $\alpha$ . Here, manual rebuilding in COOT<sup>160</sup> was alternated with least-squares refinement cycles. As for choline, the coordinate file and geometric parameters for acetylcholine were generated with PRODRG<sup>166</sup>.

However, the dataset of semi-closed ChoX was too low in resolution (2.6 Å) for successful refinement with SHELXL<sup>149</sup>. Therefore, iterative cycles of maximum likelihood restrained

refinement using isotropic B-factors in Refmac5<sup>161</sup> and manual building in COOT<sup>160</sup> was conducted. As a final step the data was detwinned in CNS<sup>150</sup> with an assumed  $\alpha$  of 0.5. Thus, it was possible to obtain detwinned R-factors. The final refinement statistics and validation of the calculated models are given in

Table 3.11.

	Acetylcholine	Unliganded closed	Unliganded semi-closed
<b>Resolution (Å)</b>	20-1.8	20-1.6	20-2.6
<b><math>R_{\text{work}} / R_{\text{free}}</math></b>	15.1/21.0	18.3/23.6	26.1/27.0
<b>No. atoms</b>	4616	4553	4352
Protein	4360	4384	4352
Ligand/ion	20	-	-
Water	236	169	0
<b>B-factors</b>			
Protein	8.2/8.3	14.8/15.1	30.6/32.2
Ligand/ion	8.5/17.4	-	-
Water	4.6	12.3	-
<b>R.m.s. / <math>\sigma</math> e.s. deviations</b>			
Bond lengths (Å)	$\sigma$ 0.007	$\sigma$ 0.014	0.009
Bond angles ( $^\circ/\sigma$ Å)	$\sigma$ 0.027	$\sigma$ 0.012	1.301
<b>Ramachandran Plot (%)</b>			
most favored	86.5 (424)	80.0 (399)	83.3 (408)
additionally allowed	12.9 (63)	19.6 (88)	13.9 (68)
generously allowed	0.6 (3)	0.4 (5)	1.6 (8)
disallowed regions	0.0 (0)	0.0 (0)	1.2 (6)
<b>Twinning fraction</b>	0.490	0.5025	0.500

**Table 3.11** Refinement statistics obtained for the models of ChoX/acetylcholine ChoX in unliganded state and ChoX in unliganded, semi-closed state at 1.8 Å, 1.6 Å and 2.6 Å, respectively. For the Ramachandran plot the values in brackets give the number of amino acids found in the respective region.

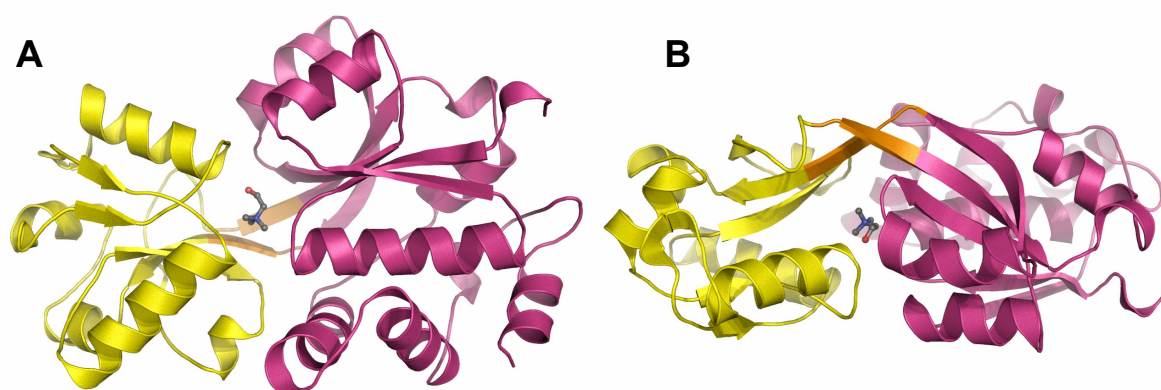
### 3.2.7 Structures of ChoX complexed with ligands

#### *Choline*

After secretion into the periplasm the secretion sequence, which is placed at the N-terminal and targets the substrate binding protein to the periplasmic space, is cleaved off. Thus, ChoX in crystallization setups only comprises the residues 28-325, which includes a C-terminal hexahistidine tag. In all structures two monomers were present in the ASU. Comparing the root mean square deviations (rmsds) for the C $\alpha$  atoms of the two molecules in the ASU it is revealed that, except for ChoX in its empty closed state, the rmsds lie below 0.5 Å. ChoX in its empty closed state exposes higher rmsds for the two molecules in the ASU.

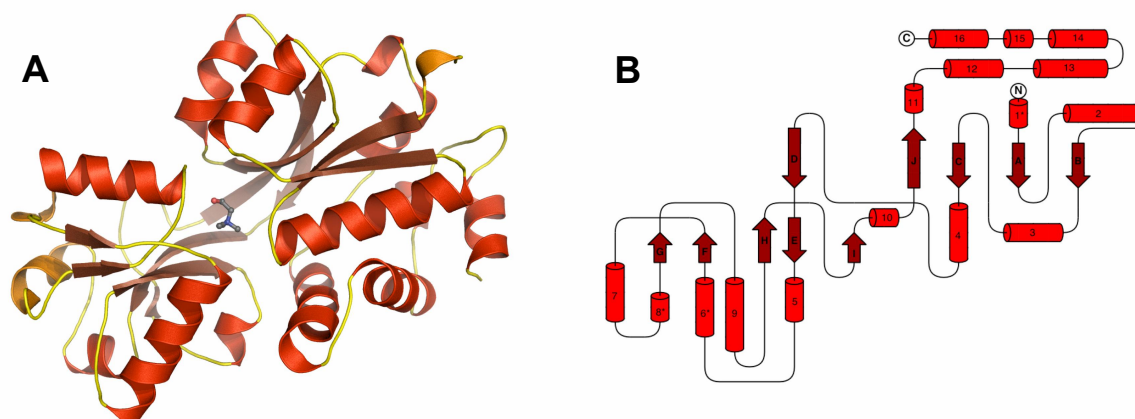
The structure of ChoX in complex with choline, consisting of residues 29-318 A/29-317 B, exhibits an elliptical shape with dimensions of 60x40x25 Å. Figure 3.18 shows ChoX in

complex with choline (colored grey). The protein can be divided into two globular domains which are highlighted in yellow and pink in Figure 3.18. Domain I, is comprised of residues 29-114 and 235-318 (colored pink in Figure 3.18) with a disulfide bridge between Cys33 and Cys247 connecting the N-terminal and the C-terminal parts of this domain. Residues 120-230 build up domain II which is colored yellow in Figure 3.18. Two strands (residues 115-119 and 231-234, orange in figure 3.16) are forming a linker connecting the two globular domains.



**Figure 3.18** **A** Overall structure of ChoX in complex with choline (grey). Domain I is colored pink and domain II yellow. The linker connecting the domains is depicted in orange. **B** Rotation of 90° towards the front of the page.

Regarding the secondary structure of ChoX both domains show common features (see Figure 3.19). A central anti-parallel 5-stranded  $\beta$ -sheet (brown) forms the core of each domain. As commonly found in binding proteins, helices (red) are flanking those  $\beta$ -sheets<sup>184</sup>. However, not all of these helices are  $\alpha$ -helices but some exhibit a  $3_{10}$ -helical fold. To distinguish them, these are colored in orange in Figure 3.19. A topological presentation of the overall fold of ChoX, marking the  $3_{10}$ -helices with an asterisk, is depicted in Figure 3.19b.

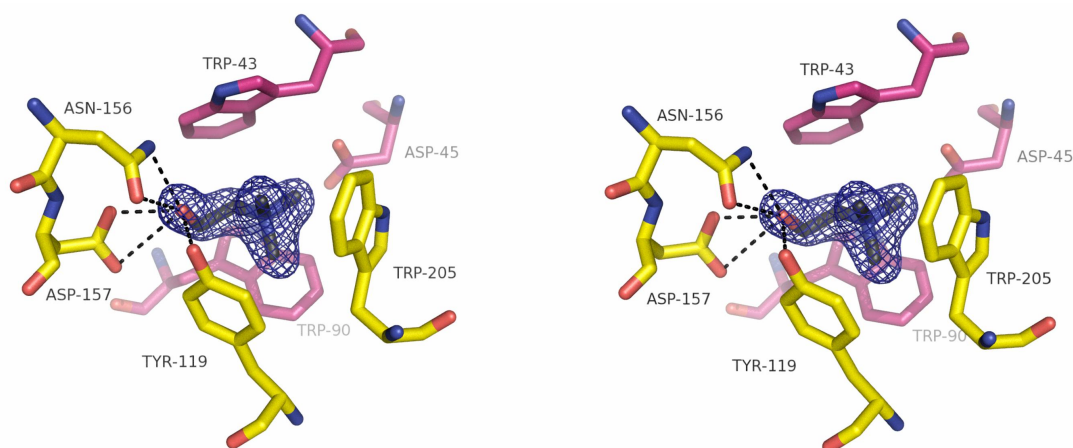


**Figure 3.19** **A** Overall structure of ChoX in complex with choline (grey). Secondary structure elements are highlighted.  $\beta$ -sheets are shown in brown,  $\alpha$ -helices are depicted red and  $3_{10}$ -helices are colored orange. **B** A topological presentation of the secondary structure elements.  $\beta$ -sheets are presented as brown arrows and  $\alpha$ -helices as red cylinders. The asterisks mark  $3_{10}$ -helices.

One remarkable feature of ChoX is the difference in sizes of its two domains. Domain II consisting of ~110 residues is much smaller in size than domain I, which comprises ~170 residues. Comparing the secondary structure of both domains reveals, that in domain II,  $\beta$ -strands F, G and I are less pronounced than the corresponding  $\beta$ -strands in domain I.

Closer attention should be drawn on  $\beta$ -strands D (residues 108-115) and E (residues 117-123). The C-terminal part of  $\beta$ -strand D and the N-terminal part of  $\beta$ -strand E both contribute to the linker that connects the two globular domains. Basically, those two  $\beta$ -strands can be regarded as one strand, which is solely disrupted by a single residue, Gly116. Glycines, however, are known to contribute a high flexibility to the backbone. Furthermore, the residues flanking Gly116 are alanines, a residue which holds a small and non-bulky side chain. The glycine residue combined with flanking alanines generating the possibility of a flexible region. The rotational freedom of the backbone in the linker region might be of major importance for a movement leading to the “Venus-fly trap mechanism”<sup>67,185</sup>.

The structure of ChoX complexed with choline reveals the substrate being buried in a gorge between the two globular domains of the protein (see Figure 3.18). A presentation of the substrate binding site is shown in Figure 3.20. The binding of choline is mediated by residues of domain I, depicted in pink, as well as by residues of domain II, colored yellow. The trimethylammonium cation moiety of the choline molecule is harbored in an aromatic box<sup>186</sup> that consists of three tryptophans, Trp43 (N-terminally of helix 2, domain I), Trp90 ( $\beta$ -sheet C, domain I) and Trp205 ( $\beta$ -sheet H, domain II) as well as Tyr119 ( $\beta$ -sheet E, at the N-terminus of domain II). Besides, Asp45 makes additional interactions with this cation moiety (distances between 3.3 and 3.8 Å respectively). The “tail” of the choline molecule, the hydroxyl group, establishes strong interactions with Asp157 and Asn156 (distance of 2.7 Å and 3.0 Å, respectively). In addition to these strong interactions, weak hydrogen bonds (distance ~ 3.4 Å) are found to Asp157 and Asn156 as well as Tyr119. Regarding the trimethylammonium cation group as the “headgroup” of the choline molecule, the interactions of the substrate can be summarized as follows. The “headgroup” interacts with residues of both, domain I (pink) and II (yellow), whereas the “tail” is binding exclusively to residues of domain II (yellow) (see Figure 3.20). To show the electron density of the substrate, a  $1F_oF_c$  map, calculated after choline has been omitted from the refinement, is contoured at 3 sigma and presented in Figure 3.20.



**Figure 3.20** Stereorepresentation of the choline binding site. A 1FoFc-omit map contoured at 3 sigma is shown around the choline molecule. The trimethylammonium group is bound by the aromatic box, consisting of residues of domain I (pink) and residues of domain II (yellow). The hydroxyl group of choline is solely bound by residues of domain II.

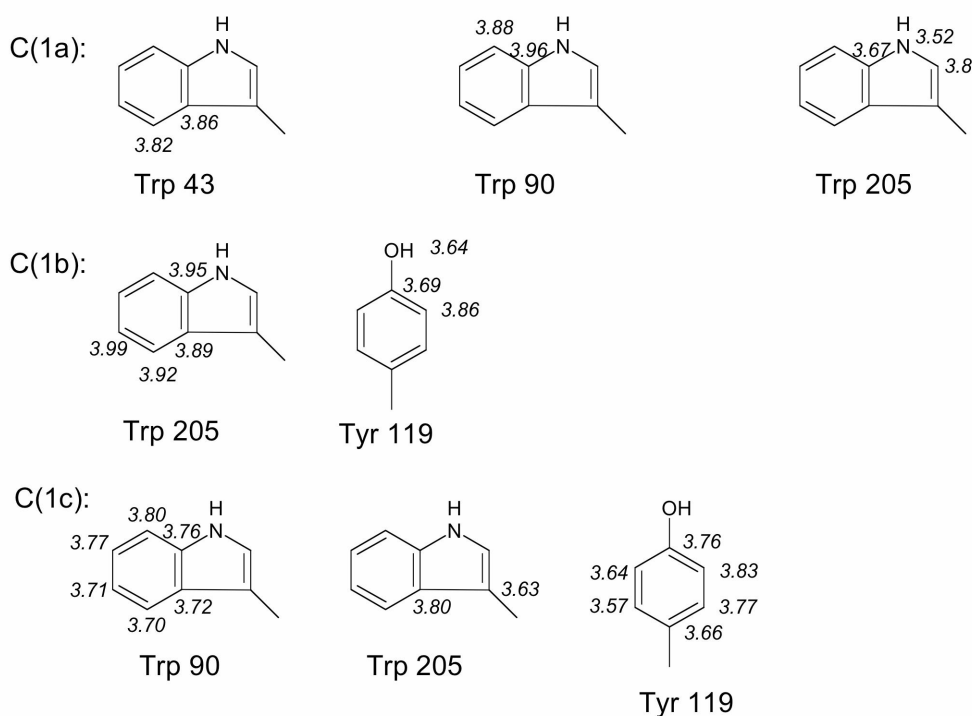
Due to the delocalization of the positive charge onto the three methyl groups, the trimethylammonium group can be regarded as a bulky cation. Interactions between this group and the surrounding aromatic residues have been described before for other binding proteins<sup>186-188</sup> and involve cation- $\pi$ , van der Waals, cation dipole interactions as well as non-classical hydrogen bonds<sup>186</sup>. To consider these interactions more thoroughly, van der Waals radii of the possibly interacting atoms, of methyl carbons for the methyl groups in the trimethylammonium moiety and of all Trp and Tyr side chain atoms forming the aromatic box, are added. Assuming a standard deviation of  $\sim 0.5$  Å, a distance interaction range can be calculated. Within this range the atoms contact each other and establish attractive forces. Distances, smaller than this calculated range, possibly establish repulsive forces. According to Li and Nussinov<sup>189</sup> van der Waals radii of 1.92 Å for the methyl carbon atom in choline can be assumed. Atoms of the aromatic side chains would exhibit van der Waals radii of 1.74 Å for the  $C_\gamma$  atoms, 1.82 Å for all other aromatic carbon atoms in a tyrosine and 1.54 Å for the hydroxyl oxygen atom in a tyrosine. The aromatic atoms in tryptophan all show van der Waals radii of 1.8 Å with the exception for the N atom (1.66 Å) and the  $sp^1$ -hybridized  $C_\gamma$ ,  $C_{\delta 2}$  and  $C_{\epsilon 2}$  (1.74 Å).

Table 3.12 shows the range of distances within which an attractive interaction is possible.

	Tyrosine			Tryptophan		
	$C_\gamma$	$C_{\delta x}, C_{\epsilon x}, C_\zeta$	$O_\eta$	$C_\gamma, C_{\delta 2}, C_{\epsilon 2}$	$C_{\delta 1}, C_{\epsilon 3}, C_{\zeta x}, C_{\eta 2}$	$N_{\epsilon 1}$
$C_{\text{methyl}}$	3.41-3.91 Å	3.49-3.99 Å	3.21-3.71 Å	3.41-3.91 Å	3.49-3.99 Å	3.33-3.83 Å

**Table 3.12** Distance ranges for possible attractive interactions for the trimethylammonium group in the aromatic box.

A schematic representation of the side chains of the aromatic residues taking part in those interactions is depicted in Figure 3.21. Here, distances (in Å) from the methyl carbon atom (1a, 1b and 1c) to the residues of the aromatic box are noted next to the specific aromatic atom. Neglecting minor deviations from the calculated distances, all together 28 interactions can be found. Interestingly, most interactions are established from the C(1c) atom.



**Figure 3.21** Schematic representation of interactions mediated by the side chains of the aromatic box and the three methyl groups of the trimethylammonium group in choline. Distances from the methyl carbon atoms (1a, 1b, 1c) to the atoms in the aromatic groups are noted next to the respective aromatic atom.

### Acetylcholine

The overall fold of ChoX in complex with choline and acetylcholine is essentially the same exhibiting root mean square deviations of 0.47 Å and 0.41 Å for molecules A and B, respectively (281 Cα, residues 35-315). Minor differences are caused by crystal contacts. Being buried in a gorge between the two domains of ChoX, acetylcholine is bound at the same position as choline. Comparing the two monomers in the asymmetric unit with respect to substrate binding, differences can be detected. To highlight these differences Figure 3.22 shows a superimposition of the substrate binding site of both monomers (monomer B in lighter colors) with residues of domain I and II colored in turquoise and blue, respectively.



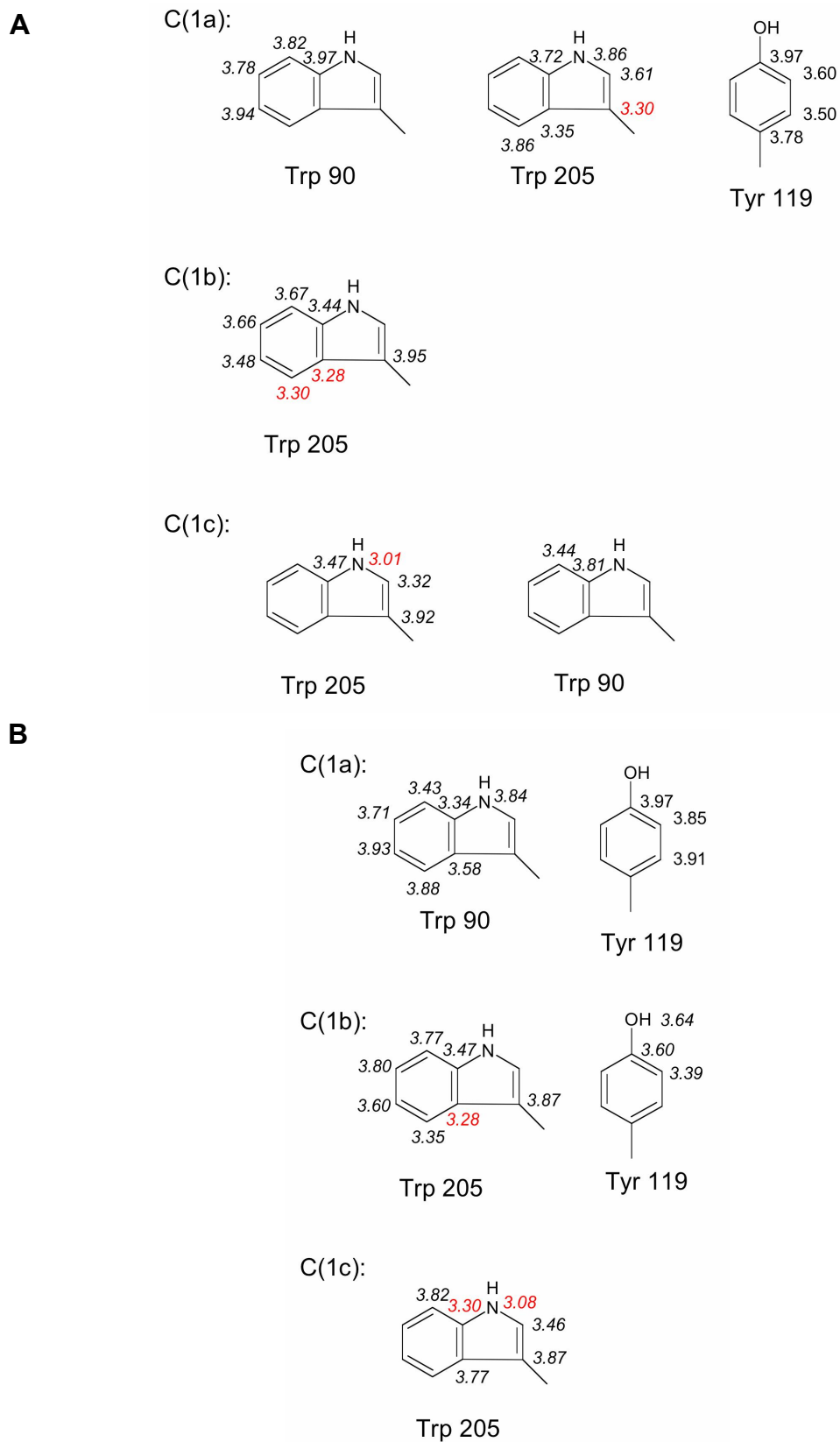
**Figure 3.22** Stereorepresentation of the acetylcholine binding site. Monomer A (darker colors) and B (lighter colors) of the asymmetric unit are superimposed to highlight the differences in the acetylcholine positions. The trimethylammonium group is bound by the aromatic box, consisting of residues of domain I (turquoise) and residues of domain II (blue). The acetyl group of the substrate is bound solely by residues of domain II.

The first variation lies within the binding site, which reveals marginal deviations of the side chain positions. Furthermore, the substrate geometry differs in the two monomers. While the torsion angle of the bond from  $_3\text{C}-_2\text{C}$  is  $134^\circ$  in both monomers, the torsion angles from  $_5\text{C}$  to  $_2\text{C}$  are  $92^\circ$  and  $55^\circ$  for the acetylcholine in monomer A and B, respectively. Therefore, acetylcholine in monomer B exhibits a more tense conformation. Regarding the position of both acetylcholine molecules it becomes evident that the place of the trimethylammonium group is rather similar but the “tails” lie in different orientations. Interestingly, while the methyl carbon in the acetyl group of monomer A shows towards Trp43 the same atom in monomer B points more towards Try119. Therefore, the H-bonds, established between the carbonyl oxygen of acetylcholine and the protein, vary slightly from monomer A to monomer B. In both cases an interaction to the hydroxyl group of Tyr119 is built up (3.1 and 3.2 Å for monomer A and B, respectively). Besides, interactions to the amide group of Asn156 can be

established, which are stronger in monomer B. If Asp157 exposes a protonated state, the acetyl groups of both acetylcholine molecules are positioned such that interactions with this amino acid are possible. However, if the amino acid is deprotonated, additional repulsive interactions are established. Furthermore, in monomer A the carbonyl oxygen lies within H-bonding distance (2.9 Å) to the backbone amid group of Gly158.

The difference in binding of both acetylcholine molecules is further manifested in the mean B-factors. In monomer B, acetylcholine exhibits an elevated value of 17.4 Å<sup>2</sup>, whereas in monomer A, this value is reduced to 8.5 Å<sup>2</sup>, thereby reflecting the mean B-factors of the two ChoX monomers (8.2 Å<sup>2</sup> and 8.3 Å<sup>2</sup> for monomer A and B, respectively). The lower B-factor of acetylcholine in monomer A also becomes evident in the better quality of the electron density map around this molecule.

Taken together, the binding of the acetylcholine molecules is mediated by similar interactions as those found for the choline binding. Again, the aromatic box establishes the framework for trimethylammonium binding and Asp45 further contacts the bulky cation (2.8 Å and 3.2/3.55 Å for monomer A and B, respectively). Slight variations in monomer A and B for the positions of the trimethylammonium groups as well as the carboxylate group of the Asp45 account for the differences in the binding lengths. For both acetylcholine molecules, 23 interactions between the trimethylammonium group and the aromatic box are discovered, which are marked in detail in Figure 3.23. Assuming that short distances (3.0-3.3 Å) may possibly impose repulsive interactions (colored red in Figure 3.23), the acetylcholine molecule in monomer A should be less tightly bound than in monomer B (4 versus 3 repulsive interactions). With respect to the aromatic box, monomer B establishes slightly stronger interactions with the substrate than monomer A. In contrast to the binding of choline to the aromatic box, no interactions with Trp43 are established neither by the acetylcholine in monomer A nor in monomer B.



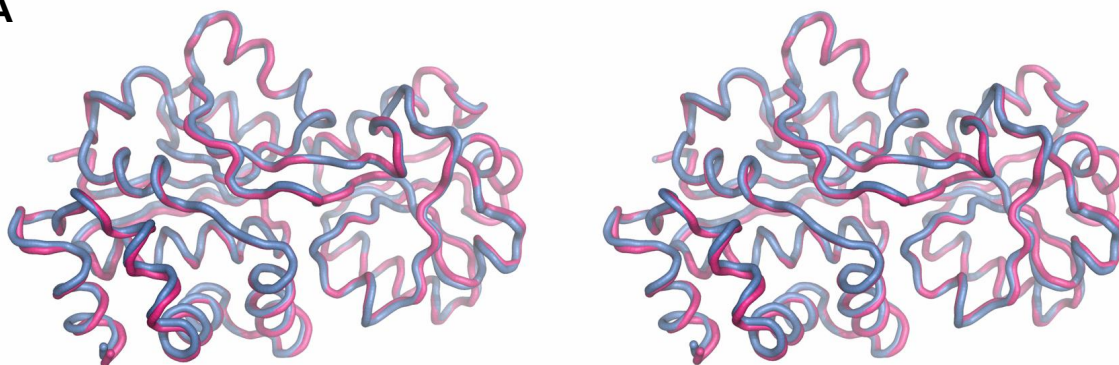
**Figure 3.23** Schematic representation of interactions mediated by the side chains of the aromatic box and the three methyl groups of the trimethylammonium group in acetylcholine. Distances from the methyl carbon atoms (1a, 1b, 1c) to the atoms in the aromatic groups are noted next to the respective aromatic atom. **A** Acetylcholine in monomer A. **B** Acetylcholine in monomer B.

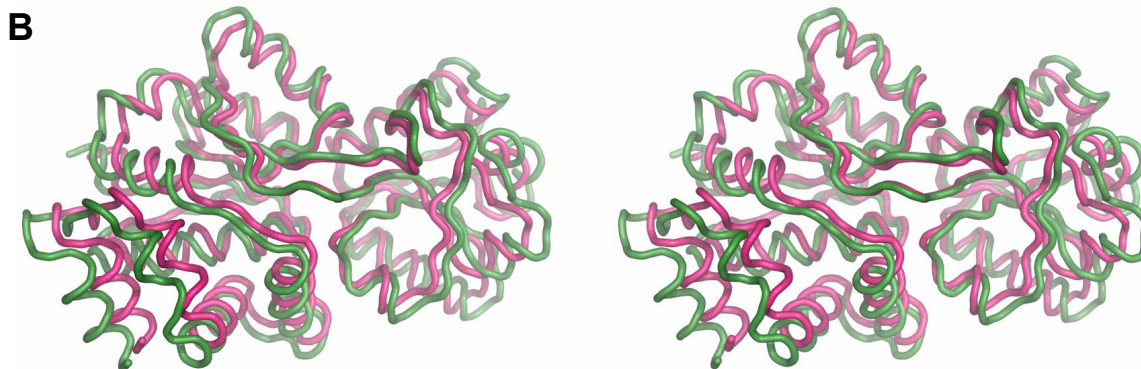
### 3.2.8 Structures of ChoX in various unliganded states

The microseeding method allowed for crystallization of ChoX in different substrate free forms, in an unliganded closed as well as an unliganded semi-closed conformation. The two monomers in the ASU of the unliganded closed structure reveal a deformation of the N-terminal  $3_{10}$ -helix in one monomer. Also the secondary structure of helix 10 and 16 in monomer A and helix 15 in monomer B is partly lost. Here, crystal contacts account for these deformations. This is reflected in the rather high rmsd of the two monomers 0.88 Å for C $\alpha$  31-318. In comparison, the rmsd of the two monomers in the semi-closed structure show better agreement (0.47 Å for same residue range).

To show the differences between substrate bound ChoX and unliganded ChoX a stereorepresentation of the superimposition of both structures (ChoX/choline in pink and ChoX in unliganded closed conformation in blue) is presented in Figure 3.24a. Thorough investigation reveals only small differences between the two structures. For example, in monomer A of the closed unliganded structure  $\beta$ -strands D and E, which in the ligand bound structure are disrupted by a single glycine residue (Gly116), show a larger gap (residues 115-119 in monomer B). With respect to the binding site also minor changes can be detected between the choline complexed structure and the unliganded closed structure (Figure 3.25). The residues that are taking part in substrate binding show slight displacements of their C $\alpha$  atoms (rmsd for binding site residues 0.19 Å). Also, in the orientations of the side chains only minor changes are detectable. Here, Trp43 and Trp90 show marginal displacements of their indole moieties. Thus, the general framework for the substrate binding site is fully accomplished in the unliganded closed species.

**A**

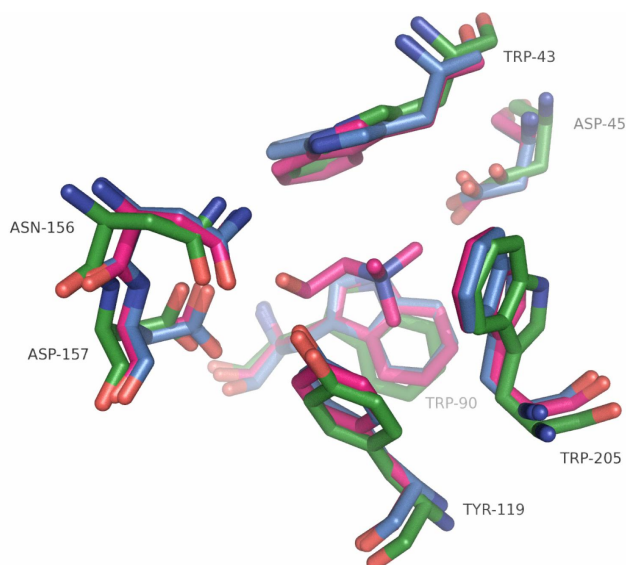




**Figure 3.24** Stereorepresentation of the superimposition of choline bound ChoX (pink) with different unliganded structures. **A** Superimposition with unliganded closed structure (blue). **B** Superimposition with unliganded semi-closed structure (green).

The superimposition of the structure of ChoX/choline (pink) and the unliganded semi-closed form (green) of ChoX is presented in Figure 3.24b. In contrast to the above mentioned comparison with the unliganded closed state, major differences are revealed. The position of the individual domains is altered, which is probably due to a gain in flexibility in the linker region, as the secondary structure elements are lost in this part. In the semi-closed structure the protein exhibits a more open conformation.

Whereas residues contributing to the binding site overlay well in the ChoX/choline and the closed structure (rmsd 0.19 Å), the alignment using the same residues with ChoX/choline and the unliganded semi-closed conformation yield a fit with rmsd of 0.77 Å. A superimposition (Figure 3.25) showing the binding site of all three structures clearly indicates that in the semi-closed conformation (green) the binding residues are all positioned away from the putative substrate position. Thus, the binding site of the semi-closed structure appears wider than the binding sites of the closed and choline bound form. Interestingly, the Asp45, which mediates the interaction to the choline headgroup in choline as well as acetylcholine complexed conformations, is placed ~1 Å further away from the putative substrate position. This finding indicates an important role for Asp45 which will be further discussed later.



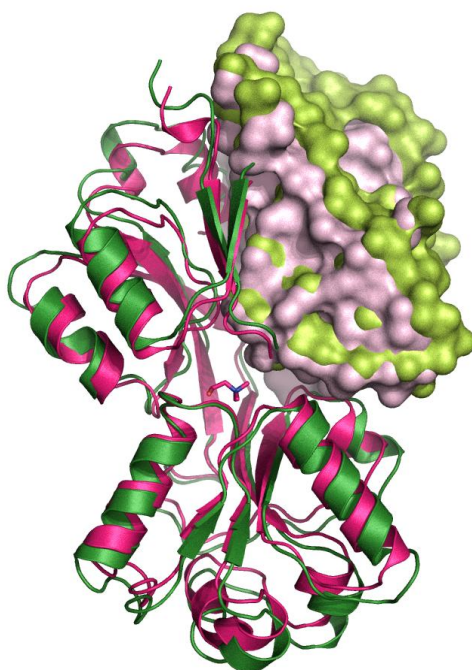
**Figure 3.25** Superimposition of the binding site residues of ChoX/choline (pink), unliganded ChoX (blue) and unliganded semi-closed ChoX (green). The choline molecule is shown in pink.

The superimposition of the choline complexed and the semi-closed structure reveals that the two domains are more apart in the latter structure (see Figure 3.24b). This is also manifested in the binding site, where the semi-closed conformation exhibits a wider binding pocket. Further investigations on domain I/domain II interactions show that the distance between the carboxylate group of Glu183 and the hydroxyl group of Thr44 is increasing from  $\sim 2.7$  Å to at least 3.4 Å. Thus, one strong interaction, holding domain I and II in close proximity, is greatly reduced. A weak hydrogen bond ( $\sim 3.3$  Å) between the amide group of Gln184 and the backbone carbonyl of Gly42 is completely lost in the semi-closed conformation. This results in an almost open pathway to the choline binding site.

Interactions between Glu206 and Asn265 also disappear in the semi-closed conformation. Besides Asn213 and Asn273 form a strong interaction in the choline complexed structure (distance  $\sim 2.9$  Å), which is missing in the semi-closed structure (distance at least 5.8 Å). By weakening the interaction between domain I and II the individual domains gain more flexibility. Further displacement of the two domains relative to each other appears possible leading to a more opened structure. Thus, the here presented semi-closed form might represent an intermediate state between a fully-opened and fully-closed conformation.

Analyzing hinges and rigid bodies of a protein could reveal interesting insight into possible motions. To conduct such an investigation on the semi-closed as well as the choline bound structure appeared rather interesting as it was considered possible to obtain information about movements of the protein upon ligand binding. Utilizing the HingeProt-Server<sup>173</sup> enabled the

analysis of putative hinges and rigid bodies of the semi-closed as well as the choline complexed structure. In accordance to the “Venus-Fly Trap mechanism” domains I and II are detected as rigid groups. Additionally, a subdomain within domain I is revealed as a rigid body. This domain, which further is called subdomain Ia, comprises residues 44-63 (~helix 2) and residues 256-318. In the choline bound structure this subdomain Ia is extended by residues 249-255. To highlight the subdomain Ia in the protein Figure 3.26 represents a superimposition of ChoX/choline (pink) and the semi-closed conformation (green), where the subdomain Ia (residues 44-63 and 256-318) is shown as surface presentation in lighter colors.



**Figure 3.26** Superimposition of the ChoX/choline (pink) and the unliganded semi-closed ChoX (green) structures. Subdomain Ia is depicted as surface representation in lighter colors.

Having detected rigid bodies within the protein it further appeared interesting to analyze how these rigid bodies are oriented in the different structures (choline bound and semi-closed conformation). Thus, a rotational movement was investigated using the DynDom Server<sup>174</sup>. A rotational outward movement of  $\sim 2^\circ$  of the subdomain Ia relative to the remaining domains (residues 70-115 and 120-231) was detected.

Such a rotational movement of subdomain Ia, which forms part of a tight network within domain I, must lead to the loss of interactions. The change in the H-bonding network at the interface between subdomain Ia and the remaining protein was analyzed. Table 3.13 provides distances of the interactions present in the choline complexed and the semi-closed conformation. Interactions between the subdomain Ia and residues of domain II are highlighted in blue.

from residue (subdomain Ia)	atom	to residue	atom	Distance (Å) choline complexed (monomer A and B)	Distance (Å) in semi- closed (monomer A and B)
44	O <sub>γ1</sub>	183	O <sub>ε2</sub>	2.68/2.65	3.40/3.55
45	O <sub>δx</sub>	choline	CAA	3.31/3.28	-
45	O <sub>δ2</sub>	90	N <sub>ε1</sub>	3.38/3.40	5.10/4.61
45	N	43	O	3.25/3.32	3.43/3.46
46	N	43	O	3.37/3.44	3.80/3.73
47	O <sub>γ1</sub>	43	O	2.96/2.93	2.98/2.57
47	O <sub>γ1</sub>	40	O <sub>δ2</sub>	2.95/3.08	3.32/3.95
47	O <sub>γ1</sub>	40	O <sub>δ1</sub>	3.07/3.11	3.46/4.13
47	N	43	O	3.02/3.03	3.61/3.41
58	O <sub>εx</sub>	64	O <sub>γ1</sub>	2.76/2.64	3.02/5.03, 6.80
256	N <sub>δ2</sub>	252	O	3.26/3.41	3.74/4.19
257	N	253	O	2.82/2.85	3.67/3.76
257	N	113	O	2.67/2.78	3.50/3.47
259	O	111	N <sub>η1</sub>	3.71, 5.28/4.85, 5.34	3.75, 4.92/2.55
265	N <sub>δ2</sub>	206	O <sub>ε1</sub>	3.19/7.19, 8.03	8.85, 7.48/5.76, 7.44
265	O <sub>δ1</sub>	208	N	3.01/3.25	4.12/4.89
273	O <sub>δ1</sub>	213	N <sub>δ2</sub>	2.89/2.90	5.86/6.50
298	O	252	N <sub>ζ</sub>	3.12/7.73	3.99/3.0
301	N	249	O <sub>δ1</sub>	2.73/2.82	3.76/3.93
301	O	249	N <sub>δ2</sub>	2.70/2.68	3.63/3.56

**Table 3.13** List of interactions between subdomain Ia and the remaining protein. Distances for these interactions are given for the choline bound and the unliganded semi-closed conformation. Interactions to domain II are highlighted in blue.

Interestingly, Asp45 and Trp90, which both take part in substrate binding via the bulky cation, lose the interaction between them in the unliganded semi-closed conformation. Also the interaction between Thr44 and Glu183, strongly pronounced in the choline bound form, is weakened in the semi-closed form. This appears especially interesting considering that Thr44 is the first residue of the subdomain Ia. Furthermore, interactions between domain II (colored blue in Table 3.13) and subdomain Ia are well pronounced in the choline complexed structure but decrease substantially in the semi-closed form.

## 4 Discussion

### 4.1 The ABC transporter HlyB from *Escherichia coli*

The 1023 amino acid comprising toxin HlyA is secreted via the inner and the outer membrane of *E. coli* in a single step<sup>85</sup>. This process is fueled by the ABC transporter HlyB, which resides in the inner membrane. Recently, structural information on fully assembled ABC transporters has been greatly increased<sup>10,30,54,57</sup>. Structures of importers as well as one exporter have been solved which undoubtedly help to understand some of the biochemical data gained on ABC transporters. The substrates of these ABC transporters are rather small in comparison to HlyA. Although the transport mechanism of ABC transporters can be investigated in more detail with the available structures and biochemical experiments can be designed on the basis of these structures, it still remains a mystery how such a large protein as HlyA is transported with the help of an ABC transporter. It is very likely that the arrangement of the transmembrane region of HlyB shows a novel composition to account for the translocation of such a large protein. The aim of this thesis was to obtain structural information on HlyB. This would provide great insight on how ABC transporters master such a complex task of fueling the transport of a 1023 amino acid toxin.

#### 4.1.1 The solubilization and purification of HlyB

Structural determination of membrane proteins still represents a complicated task. To date no general method has been developed that facilitates crystallization, and especially the crystallization of membrane proteins shows only little success. The first crystal structures of membrane proteins were determined using proteins that have been isolated from their natural sources<sup>5,190-192</sup>. The establishment of recombinant protein expression techniques has had great impact on the success of membrane protein structure determination<sup>193</sup>. The employment of affinity tags<sup>194</sup> often allows a quick purification protocol, thereby reducing the risk of denaturing the protein during the procedure. However, the solubilization, a prerequisite for purification, still represents an essential and often crucial step in membrane protein purification as certain detergents may destroy the integrity of the protein or are not at all able to extract the protein from its natural environment. Therefore, it is very useful to find several different detergents successful in solubilizing the protein of interest.

In an initial attempt to characterize the solubilization behavior of HlyB, 32 different detergents were assayed in a solubilization screen. The tested detergents are representatives of three different classes; these are non-ionic, zwitterionic and ionic detergents. Nine non-ionic detergents were chosen from the polyoxyethylene group. Furthermore, nine sugar based non-ionic detergents were tested with maltose as well as sucrose constituting the sugar moiety. As tail these detergents either had an alkyl chain or contained a cyclohexyl group. Seven representatives were chosen from the zwitterionic class. These either exhibited alkyl chains or cyclohexyl groups. Most of the chosen zwitterionic detergents are lipid like and contain a phosphatidylcholine head group. Besides, three partially zwitterionic detergents were tested. These contained head groups exposing a partial charge on the oxide moiety. The solubilization screen further comprised four ionic detergents. These were mainly alkyltrimethylammonium cations.

The detergents were analyzed for their ability to extract HlyB from the *L. lactis* membrane. The results clearly revealed that non-ionic detergents, sugar based as well as the polyoxyethylene detergents, are unsuccessful in solubilizing HlyB. The ionic detergent tetradecyltrimethylammoniumchloride showed the best results at solubilizing HlyB. However, it has been discovered before (Nils Hanekop, personal communication) that this detergent is very unspecific and solubilizes a high amount of various membrane proteins from *L. lactis* membranes. This finding and the fact that ionic detergents often denature membrane proteins and their complexes<sup>178</sup> further suggested to take a zwitterionic detergent for solubilization and purification instead of an ionic detergent. In the class of zwitterionic detergents, best results were obtained with detergents comprising a phosphatidylcholine headgroup. Longer alkyl chains revealed better solubilization, which might be due to the fact that these chain lengths most effectively shielded the hydrophobic part of HlyB. This, in turn, influences the possibility of forming crystal contacts and also allows for a higher water concentration within the crystal, thereby reducing crystal quality<sup>178</sup>. Therefore, a purification which aims for crystallization and structure determination is best performed with detergents exhibiting shorter alkyl chains. Also phosphatidylcholine detergents containing a cyclohexyl tail (e.g. CYCLOFOS7) exposed good solubilization abilities. Due to the cyclohexyl group in the tail these detergents comprise more hydrophobicity in a shorter chain length. Furthermore, they expose higher CMC values, a useful characteristic when detergent exchange via dialysis is anticipated.

HlyB was purified by affinity purification followed by a dialysis step. Therefore, a detergent

should be utilized which shows not only good solubilization abilities but also exhibits not too low CMC values. FOS-CHOLINE14 shows these characteristics. The wt HlyB purification in FOS-CHOLINE14 yielded a 95% pure sample, which showed a HlyA inducible ATPase activity (Stefan Jenewein, unpublished data). Thus the protein keeps its integrity during the purification process. However, HlyA inducible ATPase activity varied from sample to sample.

Using high amounts of detergent in the purification step can result in disruption of protein-protein interactions. ABC transporters expose two TMDs and two NBDs and therefore are especially susceptible to this process. In this case, using an ATPase deficient mutant and supplementing all buffers used in purification with ATP may help to preserve the “dimeric” intact species. This effect may also be achieved by utilizing non-hydrolyzable ATP analogs with wt HlyB. To guarantee the purification of dimeric HlyB, it might also be useful to express two HlyB species exposing two different tags. In a two step affinity purification only dimeric species are isolated. One drawback of this method is the low amount of dimeric species probably obtained as not all formed dimers expose both tags.

The variations in protein yield are possibly due to a variation in expression level, which can be explained as follows. The expression level is influenced by the amount of nisin A used for induction of the expression. However, Nisin A was expressed by *L. lactis* and secreted into the medium. And this medium was used for induction of HlyB expression. Therefore, it is likely that the supernatant used for induction shows varying amounts of nisin A and thereby influences the expression level. Expression experiments with purified nisin A showed a much higher robustness. Therefore, to achieve a reproducible protein preparation, it is useful to utilize purified nisin A.

Possible variations in the expression level may further influence the solubilization behavior of HlyB. It may be possible that the amount and composition of the copurified, bound lipids is varying. This may influence ATPase activity as well as the reproducibility of crystallization results<sup>195</sup>.

In a detergent exchange screen, FOS-CHOLINE14 micelles were to be substituted by different detergent micelles. Here, the protein was immobilized on an IMAC column and subsequently washed with buffer containing the new detergent. The detergent exchange experiments, however, did not yield successful results. Mostly, the protein was aggregated. Therefore, it can be summarized that a detergent exchange in an IMAC step may be too harsh and the protein is denatured and aggregated during this process. Another possibility is to exchange the detergent on a gel filtration chromatography step<sup>179</sup>. Mostly, this procedure does

not allow for a complete detergent exchange. Nevertheless, mixed detergents may influence the copurified, bound lipids<sup>179</sup> and thereby may help to obtain conditions a suitable sample not only for protein crystallization but also for ATPase activity measurements.

#### **4.1.2 Crystallization of HlyB**

Crystallization of HlyB was performed with wt HlyB and a mutant, HlyB H662A. The H662A mutation resulted in an ATPase deficiency in the isolated NBD<sup>44</sup>. Therefore, it was anticipated that this mutation in the full length transporter may as well result in an ATPase deficient species. Thus, a dimerization of the NBD may be induced after adding ATP to purified H662A HlyB. Due to lack of hydrolysis this conformation may be more stable and a decay of the NBD dimer is reduced.

Commercially available screens were utilized for the experiments with both species. However, only little differences were revealed and none of the species resulted in an increased amount of crystal formation. Thus, the flexibility of the NBD may not be as crucial for crystallization as expected. Indeed, when comparing the structures of AMP-PNP bound (Sav1866<sup>30</sup>) and unliganded (ModABC<sup>60</sup>) ABC transporters it is discovered that the NBDs make extensive contacts with their respective TMDs, thereby possibly being restricted in flexibility. However, this is not necessarily true for the degenerated C39 domain, which may expose high flexibility.

The influence of benzamidine, an additive which had been successfully used in membrane protein crystallization experiments with fumarate reductase<sup>196</sup>, was analyzed in crystallization setups with HlyB. It was shown that setups containing this additive exposed less strong precipitation. In contrast, 2-methylpentandiol, results in brown precipitate. This additive, as well as 1,2,3-heptantriol, is supposed to reduce the size of the micelle<sup>197</sup>. The micelle could become too small to accommodate HlyB and that may possibly result in denaturation of HlyB. Although 2-methylpentandiol influences the crystallization process greatly, 1,2,3-heptantriol shows hardly any influence on the setup. Thus, to orchestrate micelle size it may be useful to explore the milder additives like bezamidine and 1,2,3-heptantriol.

Salt as precipitant always resulted in no or very little precipitant. On the one hand this may point to an inverse phase behavior, which could be assessed by dialyses of the protein solution against water. The other explanation may be that the salt is increasing the micelle size, thereby preventing crystal formation<sup>198</sup>. Most useful for precipitation are PEGs which increase the CMC values but leave the volume of the micelles unaffected<sup>198</sup>. Increased phase

separation was exhibited in setups with low pH values. An acidic environment influences the micelle size, especially if zwitterionic detergents are used.

Crystals of HlyB were obtained in setups using cacodylate buffer at a moderate pH value. However, reproducibility was very low. This may be due to the fact that they are grown in a batch setup. As explained in 2.4.1, unlike a vapor diffusion setup the batch setup only explores a single point in the phase diagram of the protein solution. Thus, reproducibility is very low. In comparison to the purification of soluble proteins, membrane protein purification inherently results in bigger variations in detergent and lipid concentration, thereby further decreasing the reproducibility.

Nevertheless, in addition to testing different crystallization methods it is worthwhile optimizing the obtained conditions.

#### **4.1.3 Outlook**

Membrane protein crystallization follows no rules and no blueprint can be developed which leads to successful crystallization. In unsuccessful cases it is often difficult to analyze the reason for the negative outcome. Even though crystals of HlyB could be generated the problems in reproducibility and their low quality necessitate further exploration of the crystallization space.

Although the solubilization behavior of HlyB has been characterized in this work, other conditions might be revealed which lead to extraction of HlyB from membranes. Special care should be taken to utilize mild, non-ionic detergents, as these usually keep the integrity of the protein<sup>178</sup>. In this respect it might be useful to analyze different salt/pH concentrations as well as detergents mixtures of sugar based detergents. Detergent mixtures may also show an impact on crystal formation as shown for the crystallization of GlpT<sup>179,199</sup>.

HlyB exposes two different hydrophilic domains, the NBD and a degenerated C39 domain. The NBDs can be stabilized by formation of a dimeric state. When ATP is added to the hydrolysis deficient mutant H662A a dimeric state of the NBD is locked. However, the N-terminal degenerated C39 domain may expose a high flexibility and may hinder protein crystal formation. Therefore, it might be valuable to analyze the crystallization behavior of a C39 deletion mutant.

Another approach should be the evaluation of endogenous lipids present after purification<sup>179,200</sup>. Gradually, the notion that lipids are only forming transient interactions with membrane proteins is changed and the importance of lipids for membrane protein integrity and activity is accepted<sup>201</sup>. Some membrane protein structures expose lipids which are either

endogenous or added during protein preparation<sup>195,202</sup>. Therefore, it is reasonable to evaluate the binding of lipids to HlyB by thin layer chromatography<sup>200</sup> and possibly substitute lipids, preferably from *E. coli*, which were lost during purification.

A different approach may be to crystallize HlyB not in detergent micelles but in a different environment. Lipidic cubic phases<sup>124</sup> may form an alternative to the conventional micelle based crystallization. Until recently, this method exposed some drawbacks; it utilized a high amount of protein and was very tedious to be set up, as the lipidic cubic phase exposes a rather high viscosity. However, development of robotics especially designed for this method, have made it possible not only to perform setups quickly but also to allow for the utilization of low amounts of protein<sup>203</sup>. A modified version of the lipidic cubic phase is the so-called lipidic sponge phase. The sponge phase exposes a less curved lipid bilayer and larger aqueous pores which allow the incorporation of membrane proteins with larger hydrophilic domains<sup>204</sup>. Most interesting, the sponge phase is liquid and can be used in conventional vapor diffusion setups.

Recently, it has been shown that membrane proteins may also be crystallized in bicelles. These bicelles are formed by a mixture of lipids and detergents which can be handled at low temperatures while maintaining a bilayered structure<sup>205,206</sup>. Furthermore, different amphiphiles (peptidetergents<sup>207</sup>, peptergents<sup>208</sup>, lipopeptide detergents<sup>209</sup> and tripod amphiphiles<sup>210</sup>) have been developed which expose a self assembled micelle formation. They are able to gently extract membrane proteins from their natural environment and may be useful in crystallization processes.

A different approach uses antibody fragments which help to enlarge the hydrophilic part of the membrane protein and thus enable stable crystal contacts<sup>211</sup>. Impressively, due to the employment of antibody fragments it was possible to crystallize the  $\beta$ -adrenergic receptor<sup>212,213</sup>, a member of the G-protein coupled receptor family known to be notoriously flexible and extremely hard to crystallize.

## **4.2 The substrate binding protein ChoX from *Sinorhizobium meliloti***

### **4.2.1 Binding affinities for choline, acetylcholine and glycine-betaine**

The affinities of choline, acetylcholine and glycine-betaine to ChoX were analyzed by determining the  $K_d$  values by a filter assay using radioactively labeled substrates. Choline

exposed the highest affinity to ChoX ( $K_d = 0.3 \pm 0.1 \mu\text{M}$ ) whereas glycine-betaine ( $K_d = 3.8 \pm 0.6 \mu\text{M}$ ) showed a ten times lower affinity and acetylcholine showed a hundred times lower binding affinity than choline ( $K_d = 22.8 \pm 6.9 \mu\text{M}$ ). These results, however, stand in opposition to the findings of Dupont and coworkers<sup>113</sup>, which present affinities of  $2.7 \mu\text{M}$  and  $145 \mu\text{M}$  for choline and acetylcholine. Furthermore, in their study they were not able to reveal glycine-betaine binding. However, to assess glycine-betaine binding a single point determination of an inhibition assay was utilized. Due to the inherent problems of this assay<sup>214</sup> a wrong picture of glycine-betaine ChoX interaction may have been generated. In contrast to the published results from Dupont and coworkers<sup>113</sup>, it could be shown that ChoX indeed binds glycine-betaine. The discrepancy in the determined binding affinities may have different causes. Dupont and coworkers used a protein concentration of  $5 \mu\text{M}$  in the filter assay with the purified ChoX. At this concentration a receptor correction, as described in 3.2.2, may need to be applied to correct for the high protein concentration. Besides, to allow for successful fitting, the binding curve should reach a maximum value. However, following this rule a determination of a  $K_d$  value for acetylcholine of  $145 \mu\text{M}$  is not appropriate when  $120 \mu\text{M}$  acetylcholine is utilized as the maximum substrate concentration. Finally, Dupont and coworkers conducted all binding experiments at  $10 \text{ mM}$  Tris/HCl pH 7.4. However, in our hands, ChoX formed higher oligomers (analyzed by gel filtration chromatography) when a buffer at this ionic strength was used. Higher salt concentrations (e.g.  $200 \text{ mM}$ ) exposed a monomeric peak in the gel filtration run (data not shown). The buffer utilized by Dupont for the binding assays may induce oligomer formation, which may hinder the correct determination of the  $K_d$  values.

#### 4.2.2 Seeding and twinning – is there a correlation?

Acetylcholine presents an important neurotransmitter both in the central nervous system as well as the peripheral nervous system. Already in 1936 the Nobel Prize was awarded for studies on acetylcholine as agent in the chemical transmission of nerve impulses. Despite the importance of this substance, to date there are hardly any protein-acetylcholine complex structures present in the database. As it was shown that acetylcholine can bind to ChoX<sup>113</sup>, it seemed especially interesting to characterize the binding interaction. However, the crystallization experiments with acetylcholine in ChoX only showed electron density for choline in the binding site. Therefore it was anticipated that acetylcholine was subjected to hydrolysis during crystal growth. The quest for a method to obtain crystals within a few days led to the establishment of microseeding experiments. Although it was possible to acquire

crystals within a few hours, they all showed a high twinning fraction unlike the crystals grown under conventional conditions thus suggesting a correlation of twinning and seeding.

The formation of a surface in the nucleation process is guided by the nucleation barrier. However, in a seeding experiment this barrier is of minor importance as a certain amount of nuclei are already available. Therefore, usually, seeding experiments are conducted under less saturated conditions as the original condition of which the seeds were received<sup>127,130</sup>. The crystallization of ChoX in complex with acetylcholine required extremely quick crystal growth to prevent hydrolysis of the substrate. Thus, in order to accelerate the growth process conditions were chosen which would guarantee a high saturation. A quick crystal formation implies that the equilibrium between the segregation of a solid phase (crystal) and the dissolving of the newly formed phase is shifted. This, in turn, is significant for the occurrence of crystal growth disorders, such as twinning, as possible defects in crystal development are less likely to be dissolved and the possibility of a defect crystal formation is enhanced.

Vice versa, these ideas suggest that crystals which required more time for their growth would possibly exhibit less crystal defects. This hypothesis was tested with a ChoX mutant (D157L). Utilizing microseeding, crystals were grown in complex with choline under less saturated conditions. The crystals appeared after three days and indeed showed a lower twinning fraction ( $\alpha \sim 0.35$ ). The advantage of a lower twinning fraction is the possibility to deconvolute the reflections and use detwinned reflections for structure determination. With higher twinning fractions deconvolution of the reflections cannot be performed.

In conclusion it could be shown that seeding can help to obtain crystals complexed with intact substrates that are normally easily subjected to hydrolysis. By carefully choosing the degree of saturation, i.e. varying precipitant or protein concentration, it also appears to be possible to balance the minimization of twinning and substrate decay, in this case hydrolysis.

#### **4.2.3 Structural aspects of choline and acetylcholine binding**

The structure of ChoX in complex with choline as well as with acetylcholine reveals the binding of the substrate in a cleft between domains I and II. The substrates are bound in a very similar fashion. An aromatic box, composed of residues of domain I and domain II, generates the framework for binding of the trimethylammonium group, a bulky cation. The binding of the tail of the substrate is accomplished solely by residues of domain II.

Comparing the interactions that facilitate the binding of the bulky cation reveals a lower amount of interactions in case of acetylcholine binding (23 and 28 for acetylcholine and choline, respectively). Due to the slightly bigger size of acetylcholine the trimethylammonium

group is positioned more towards the aromatic box. Hence, it is possible that repulsive interactions are established, thus, lowering the affinity for acetylcholine to ChoX. Interestingly, as acetylcholine is placed further into the box, interactions between the trimethylammonium group and Trp43 are no longer possible. In contrast, choline fits better into the aromatic box and is positioned such that more attractive interactions can be established and repulsive interactions are avoided. However, as acetylcholine is placed more at the back of the box, Asp45 is able to make tighter interactions with the trimethylammonium cation (2.8/3.2 Å and 3.3/3.3 Å for acetylcholine and choline, respectively).

The binding of the “tail” of the substrate is better established with choline than with acetylcholine. Choline exposes a hydroxyl group which allows function as H-bond donor as well as acceptor. In contrast, the acetyl group can only function as H-bond acceptor. Thus, acetylcholine is able to interact with Asp157 only if this residue is protonated. Furthermore, the hydroxyl group of choline can establish interactions with the amide group and the carbonyl group of Asn156, whereas the carbonyl group of acetylcholine can solely interact with the amide group of this amino acid. Thus, not only the “headgroup” of acetylcholine opposes repulsive interactions also the “tail” possibly establishes repulsive interactions, if Asp157 is protonated or if Asn156 is positioned with its carbonyl group towards the carbonyl of the acetylcholine. Therefore, from a structural point of view, due to more attractive and less repulsive interactions, the binding of choline should be tighter than the binding of acetylcholine to ChoX.

Although crystals of ChoX with glycine-betaine are available, the quality of these is not sufficient to calculate a reliable model which could give detailed insight into substrate binding. Thus, up to now, a structure of ChoX with glycine-betaine is unavailable. Glycine-betaine, possessing a carboxylate group instead of a hydroxyl group, in principal is as big as choline and therefore should fit as well in the aromatic box. However, it can be speculated that the “tail” of glycine-betaine possibly also opposes repulsive interactions like the ones described for acetylcholine. Taking the available structures into account, it can be concluded, that choline possibly binds strongest to ChoX followed by glycine-betaine. Of the three possible substrates acetylcholine should expose the lowest affinity. This finding corroborates the experimental data obtained by measurements with radioactively labels substrates.

#### **4.2.4 Binding mode of acetylcholine**

As mentioned before acetylcholine plays an important role in neurotransmission. However, only few structures are available allowing for an analysis of protein-ligand interaction with

this important molecule. To contribute to the understanding of acetylcholine binding on a molecular level ChoX was crystallized with acetylcholine and structural analysis was performed.

The two monomers of the asymmetric unit of the ChoX/acetylcholine structure show the same binding site for acetylcholine, however, slightly different positions are revealed for the ligand in the two monomers. In both cases the acetylcholine head groups are placed in the aromatic box, composed of one tyrosine and three tryptophans. Interestingly, although Trp43 is an essential part of this box, no interactions from the trimethylammonium group to this residue are established. Furthermore, Asp45 contacts the bulky cation. The acetyl group interacts with Asn156 and Tyr119 and possibly with Asp157, if this residue is protonated.

Comparing the uncovered interactions with the binding mode of acetylcholine in the acetylcholine esterase structure<sup>215</sup> the following differences can be found: the ligand in the acetylcholine esterase structure is bound at two different places. One acetylcholine molecule lies near the surface with the acetyl group pointing inside a cavity of the protein. Here, the trimethylammonium group interacts with a tyrosine and a tryptophan, whereas the “tail” forms an interaction to the backbone of the protein via a water molecule. Furthermore an H-bond is established between the ester bonded oxygen and the hydroxyl group of the tyrosine. The second molecule lies within the active site and performs interactions with a tryptophan as well as a tyrosine. Even though many aromatic residues are present in the active site an aromatic box like the one found in the ChoX structure is not revealed. Therefore, a different binding mode for acetylcholine is established with ChoX.

The structure of L-ficolin was solved in complex with acetylcholine<sup>216</sup>. L-ficolin is a trimeric protein that exposes a lectin-like activity and binds solely to the acetyl group of acetylcholine. Within this structure the torsion angles of acetylcholine are not restricted due to the lack of binding of the trimethylammonium group. Therefore, the acetylcholine molecules are solely restricted by sterical hindrance. To reveal if the acetylcholine molecules bound in ChoX exhibit a sterically unfavorable conformation the torsion angles of the molecules are compared. The L-ficolin structure shows torsion angles between 129-162° and 82-154° for the <sub>1</sub>N-<sub>4</sub>O and the <sub>2</sub>C-<sub>3</sub>C torsions, respectively. In the ChoX/acetylcholine structure these torsion angles expose values of 134°/134° and 92°/55° respectively. Monomer A, exposing the better defined acetylcholine electron density, shows torsion values for acetylcholine that lie perfectly within this limit. Thus, it is possible for acetylcholine to bind into the binding site of ChoX without the need to create additional sterical strain.

#### 4.2.5 Binding of acetylcholine in ChoX compared to carbamylcholine binding in the AChBP of the nicotinic acetylcholine receptor of *Lymnaea stagnalis*

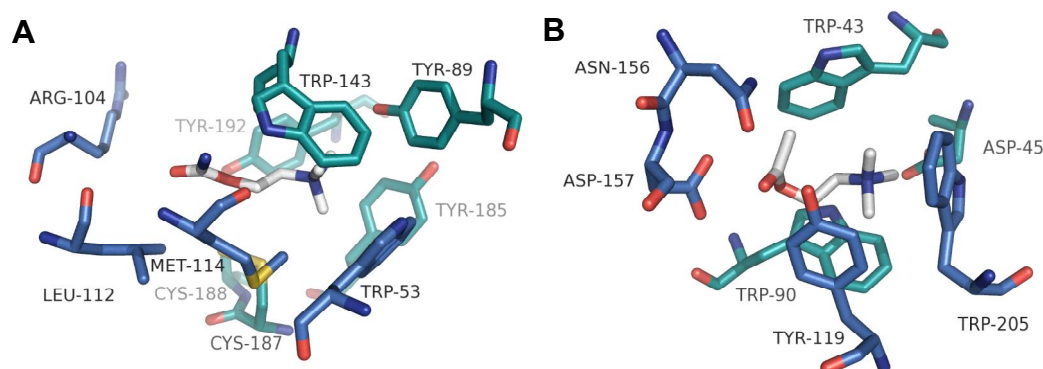
It is well known that substrate binding proteins are homologous to membrane receptors that contain extracellular binding domains<sup>217</sup>. This has been shown for metabotropic glutamate receptors<sup>218</sup> and NMDA receptors<sup>219</sup>. The homology to ionotropic glutamate receptors (iGluR) is also reflected in the high structural homology (DALI-server: Z-score 12.9) of ChoX to the ligand binding core of the glutamate receptor 2 (PDB accession code: 1MY0)<sup>220</sup>. Furthermore, homologs of acetylcholine receptor type ligand gated ion channels have been identified in prokaryotes<sup>221</sup> thus making it especially interesting to compare binding of acetylcholine in ChoX with the binding in acetylcholine receptors.

Nicotinic acetylcholine receptors are prominent members of the ligand gated ion channel superfamily. These transmembrane proteins exhibit a homo- or heteropentameric arrangement with each monomer being composed of a ligand binding domain, a transmembrane domain and an intracellular domain. The extracellular ligand binding domains control the opening and closing of the ion channel. Hereby the ligand is bound at the interface between two subunits, the principal and the complementary side. Nicotinic acetylcholine receptors are associated with various diseases the most prominent being Alzheimer's disease<sup>222</sup> or Parkinson's disease<sup>223</sup>.

The acetylcholine binding protein (AChBP) of the mollusk *Lymnaea stagnalis* (*L. stagnalis*) shows 24% sequence identity to the extracellular domain of human nicotinic acetylcholine receptors. As the structure of the AChBP from *L. stagnalis* in complex with carbamylcholine has been solved to 2.5 Å<sup>224</sup> (PDB accession code: 1UV6), this structure can be used for comparison of the binding modes of acetylcholine to nAChRs and to ChoX.

Figure 4.1 pictures the binding site of carbamylcholine (A) in the AChBP and of acetylcholine to ChoX (B). In both cases the bulky trimethylammonium cation is bound to an aromatic box. For binding of acetylcholine to ChoX this box consists of three tryptophan residues and one tyrosine (Trp43, Trp90, Trp205 and Tyr119). The aromatic box in AChBP is built up by three tyrosine residues and two tryptophans (Trp53', Trp143 and Tyr89, Tyr185, Tyr192). Here Trp53' belongs to the complementary side whereas the other aromatic residues are part of the principal side. The aromatic box is composed of both sides, which reflects the situation in ChoX, where the aromatic box is build up of residues of domain I (Trp90 and Trp43) as well as of residues of domain II (Tyr119 and Trp205). The carbamoyl group, however, is bound by residues of the complementary side. In the binding site of ChoX, the acetylcholine tail is also bound by residues of domain II (Asp157, Asn156 and Tyr119). Hence, it can be summarized

that domain I and domain II in ChoX reflect the principal and the complementary side of AChBP, respectively. The fact that the binding sites of the AChBP and ChoX show similarities is possibly due to a common ancestor.



**Figure 4.1** Comparison of the acetylcholine binding pockets. **A** The binding of carbamylcholine to *L. stagnalis* AChBP (PDB accession code: 1UV6). The principal side is depicted in turquoise and the complementary side in blue. **B** The binding of acetylcholine to ChoX. Domain I is colored in turquoise and domain II in blue.

#### 4.2.6 The rotational subdomain Ia

##### *Identification of a “transport triggering mechanism”*

ABC transporter present a huge family of transporters associated with export as well as import functions. In comparison to exporters, importers exhibit an additional subunit, the substrate binding protein. In import systems, the substrate specificity is ascribed solely to this subunit. Thus, it is not astonishing, that many structural investigations had been carried out focusing on the molecular determinants of substrate recognition. These investigations present a solid base for protein engineering and development of biosensors<sup>225</sup>. However, only little is known about the interactions mediated by the substrate binding protein with its cognate transporter, possibly due to the shortage of structures presenting a complex of both. Even less is known about how these interactions, brought about on the outer side of the cellular membrane, trigger ATP binding and hydrolysis on the cytoplasmic side. Although it is very challenging to address these questions purely with the help of structural information on the substrate binding protein, the possibility of analyzing substrate bound as well as different unliganded conformations of ChoX provides a helpful starting point.

Analysis of possible hinges and rigid bodies within the semi-closed unliganded conformation of ChoX indicates the existence of a subdomain Ia within domain I. Comparing the position of subdomain Ia in the semi-closed unliganded conformation with the choline bound

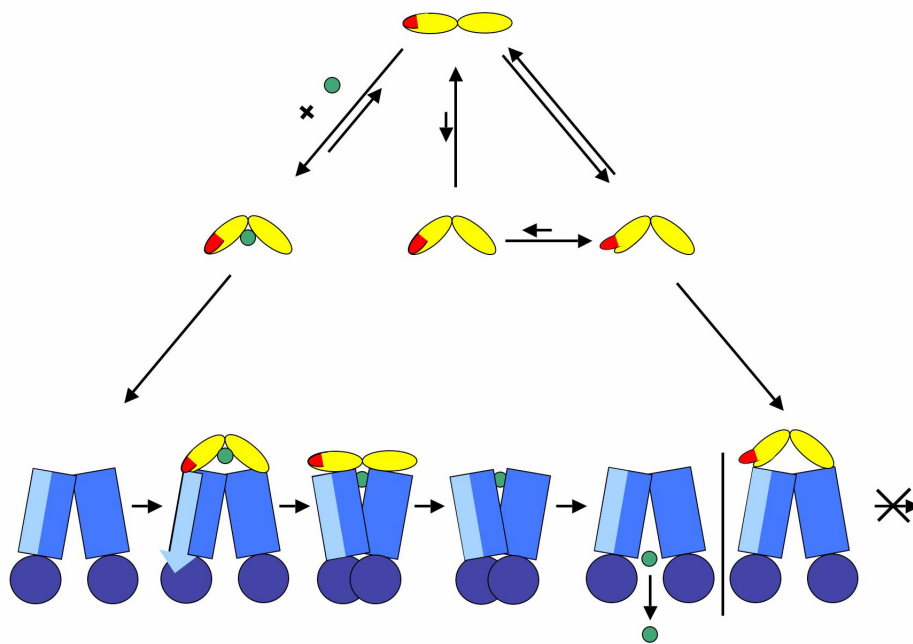
conformation of ChoX reveals a rotational outward movement of subdomain Ia. Vice versa, it can be concluded that upon ligand binding this subdomain Ia rotates inward. Upon the inward rotation an extensive H-bonding network is established between subdomain Ia and the rest of the protein. The idea of a rotational movement brought about by ligand binding is further corroborated by the fact that in the ChoX/choline structure Asp45, integral part of the rotational subdomain Ia, is making interactions with the bulky cation moiety (distance  $\sim 3.3$  Å) of the substrate. The superimposition of the binding site residues (figure 3.23) indicates the displacement of Asp45 in the semi-closed unliganded conformation (green) when compared to the choline bound (pink) conformation, positioning Asp45 in a place, where interactions with the substrate are impossible.

Having identified a subdomain which possibly rotates inwards upon substrate binding automatically generates the question for its implication. Up to date, substrate binding proteins are thought of as molecules exposing two globular domains which are connected via a flexible linker. In absence of substrate these two domains expose an equilibrium between an open and a closed conformation<sup>64,71</sup>. The substrate is bound in a cleft between the two globular domains. Upon substrate binding the equilibrium between open and closed conformation is shifted towards the closed conformation<sup>64</sup> according to a “Venus-Fly Trap mechanism”<sup>67,185</sup>.

In 1989, Ames and coworkers first detected induction of ATPase activity due to the presence of substrate binding protein<sup>75,226</sup>. In 1992 it was shown by Davidson and coworkers<sup>74</sup> that the addition of maltose binding protein to the maltose permease induces ATP hydrolysis. But only if substrate is bound to the SBP, ATPase activity is fully developed. This finding is further corroborated by investigations on the histidine transporter<sup>76,77</sup> from *Salmonella typhimurium* (*S. typhimurium*) and the glycine/proline-betaine transporter from *Bacillus subtilis*<sup>73</sup>. Of further interest is the identification of substrate binding protein independent mutants of the histidine<sup>227</sup> and the maltose<sup>228</sup> transport systems. In case of the maltose importer these mutations lie within the TMD, whereas it is the NBD of the histidine importer which holds the respective mutations.

However, these findings lead to other questions. How is it possible that the binding of liganded and unliganded substrate binding protein to the transporter induces different levels of ATPase hydrolysis? If the binding of SBP leads to an induction of ATPase activity how can the cell prevent unfruitful hydrolysis cycles which would occur when unliganded closed substrate binding protein interacts with the transporter? And finally, how is the presence of substrate binding protein possibly signaled through the transmembrane domains to the nucleotide binding domains?

The identification of the rotational subdomain Ia may help answering these questions. In essence the rotational subdomain presents a sensor for the substrate and is rotated inwards if ligand is bound. In an unliganded state the two domains of the binding protein undergo an opening and closing motion. Hereby, the closed state can adopt a semi-closed conformation with subdomain Ia rotated outwards or a closed conformation with the subdomain Ia rotated inwards. However, as the rotation of subdomain Ia is induced by substrate binding the unliganded closed state shows little population. This is schematically illustrated in Figure 4.2 where the rotational subdomain is indicated by a red half-ellipse.



**Figure 4.2** Schematic representation of the “transport triggering mechanism”. The substrate binding protein is presented as two yellow ellipses. In an unliganded state the substrate binding protein exhibits a equilibrium between an open (top), a closed and a semi closed state (middle and right in the middle panel). These states are populated to different extends. In the semi closed state subdomain Ia, indicated as red half-ellipse, is rotated outwards. The equilibrium is shifted towards the closed state (shown on the left) upon substrate (green circle) binding. Binding of the liganded substrate binding protein to the transporter triggers a signal (light blue arrow) in a part of the transmembrane domain (left, bottom panel). The signal is transferred to the nucleotide binding protein (dark blue circles), where subsequently ATP is hydrolyzed and transport can occur. In the unliganded form the subdomain Ia exposes a semi-closed state and cannot make the interactions necessary to induce the signaling (right, bottom panel).

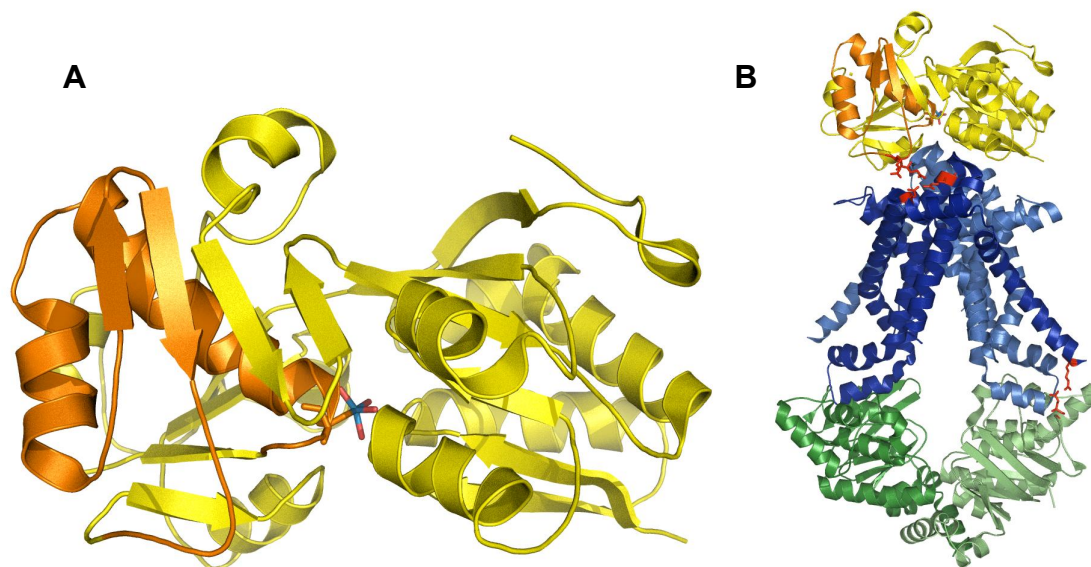
The protein can exhibit three different closed states (middle panel), the liganded conformation shown to the left in Figure 4.2, the unliganded semi-closed conformation shown to the right in Figure 4.2 and the unliganded closed conformation (in the middle). The fact that ATPase activity is induced to higher levels when liganded binding protein is bound to the transporter can now be explained as follows. As the liganded state exhibits the rotational subdomain Ia in an inward facing conformation it can establish interactions with the transporter (left side

lower panel) which are impossible to be established if the subdomain is rotated outwards (right side lower panel). In a positive interaction of the subdomain with the transporter, as anticipated in the liganded state, the signal is transmitted (shown as a light blue arrow) through the TMD (blue) and passed on to the NBD (shown in dark blue). This further induces ATP binding and/or hydrolysis, thus, leading to the import of the substrate (green circle). If no substrate is present, the subdomain remains rotated outward and the interactions with the TMD leading to a signaling are circumvented, and no ATP binding and/or hydrolysis can occur. In cases where the substrate binding protein exhibits a closed unliganded conformation, a positive interaction with the TMD, as expected in a liganded state, is possible. This explains why little ATPase activity is detected when unliganded binding protein is added to the transporter. However, the low population of this state prevents that high amounts of ATP are hydrolyzed. Thus, the rotational subdomain presents a sensor-transmitter switch, which helps the transporter to differentiate between liganded and unliganded substrate binding protein. With the aid of the identified “transport triggering mechanism” the hydrolysis of ATP in an unfruitful manner, i.e. when no substrate is present in the binding protein, can be effectively reduced in the cell.

#### *Structural evidence supporting a “transport triggering mechanism”*

The above presented mechanism sheds more light on the two questions of how different levels of substrate binding protein induced ATPase activity can be achieved and how the cell can greatly circumvent unfruitful ATP hydrolysis. However, the question of signal transmission can only be answered in an unsatisfying way. Here, a structure of ChoX in complex with its ABC transporter ChoVW would make an immense contribution. Unfortunately, this structural information is not available. Structural information of ABC transporters in complex with their binding proteins is accessible for the BtuCDF<sup>54</sup>, MalEFGK<sup>57</sup> and ModABC<sup>10</sup>. However, only the ModABC structure exhibits a closed, liganded conformation of the binding protein in complex with its ABC transporter. Therefore, the investigation of the structure of ModABC helps to manifest the concept of a “transport triggering mechanism”. The structure of the liganded molybdate binding protein ModA in complex with its ABC transporter ModBC from *Archaeoglobus fulgidus* was solved recently (PDB accession code: 2ONK)<sup>10</sup>. Here, ModB annotates the TMD and ModC the NBD, which in this structure exhibits an ATP free state (see Figure 4.3). To verify the idea of a “transport triggering mechanism” hinges and rigid bodies of ModA were analyzed like previously described for ChoX. Indeed, ModA shows a

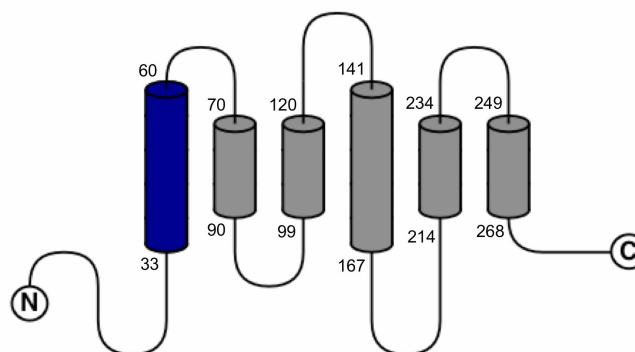
possible rotational subdomain consisting of residues 149-207. Like in ChoX this subdomain comprises a binding site residue, Asp153.



**Figure 4.3** Cartoon representation of the ModABC structure (PDB accession code: 2ONK). **A** The substrate binding protein ModA colored in yellow with the possible rotational subdomain depicted in orange. **B** Structure of the full-length transporter in complex with the binding protein. The transmembrane domains are shown in blue and the nucleotide binding domains are colored green. Interactions possibly involved in signaling according to the “transport triggering mechanism” are highlighted in red.

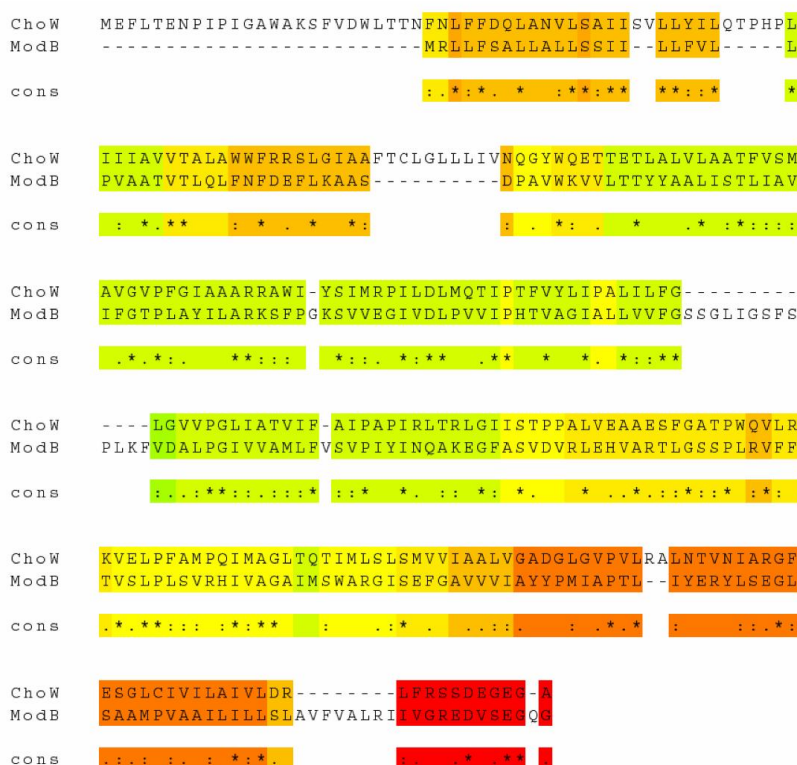
To address the question of signal transmission the complex of ModABC was investigated. The structure reveals interactions between the identified subdomain (149-207, shown in orange in Figure 4.3) and the TMD. These interactions are established by ModA Glu203 and ModB Arg220 (2.5/2.7 Å) as well as ModA Arg204 and ModB Asp44 (3.3 Å) (colored red in Figure 4.3). Especially interesting is the interaction with Asp44, positioned in helix 1b which lies in the periplasm and forms the connection between the transmembrane helices 1 and 2. Connecting two transmembrane helices via a periplasmic helix opposed high rigidity within this arrangement. Thus, it can be anticipated that an interaction from the substrate binding protein with helix 1b of the TMD is influencing the position of helix 1 or helix 2, as helix 1b forms a part within this rigid arrangement (helix1-1b-2). Furthermore, helix 1 stands in contact with the NBD (colored green in Figure 4.3) at the cytoplasmic side of the transmembrane part. A strong salt bridge (colored red in Figure 4.3) is created between Glu2 (ModB) and Arg68 (ModC) (3.2 Å). Interestingly, Arg68 is positioned nine residues N-terminally of the Q-loop glutamine, which is part of the NBD-TMD transmission interface. Transmembrane helix 1 of the molybdate importer presents a rather long helix, composed of 30 amino acids. Hydrophobicity analysis of the transmembrane part of the choline importer

reveals a long, N-terminal helix as well (see blue cylinder in Figure 4.4). The predicted helix 1 of ChoW is composed of a similar amount of residues, 28 amino acids.



**Figure 4.4** Schematic presentation of the predicted transmembrane helices of ChoW (TMpred<sup>229</sup>). A long N-terminal helix is depicted in blue.

This fact in conjunction with the finding, that the N-terminal part as well as the C-terminal part of ChoW and ModB show similarities (see dark orange and red underlined part of Figure 4.5) indicates the possibility of similar signaling pathways for the choline importer and the molybdate importer.



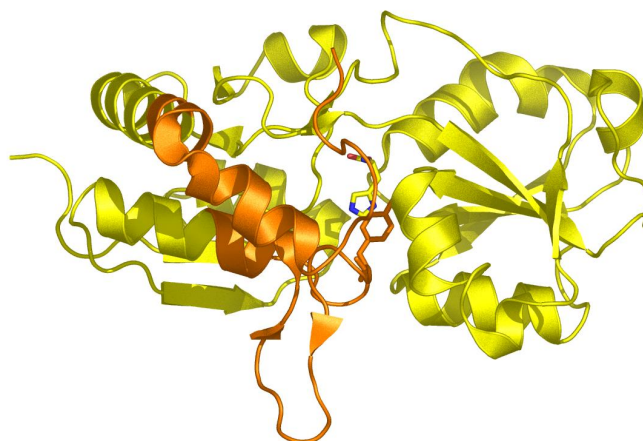
**Figure 4.5** Sequence alignment of the transmembrane domains ChoW and ModB (T-COFFEE<sup>230</sup>). Regions of high similarities are colored orange and red.

The interactions between the putative rotational subdomain of the substrate binding protein and helix 1b of the TMD, and further the interaction between transmembrane helix 1 and the

NBD present a possible signaling pathway from the periplasmic side to the cytoplasm, thus, supporting the idea of a “transport triggering mechanism”.

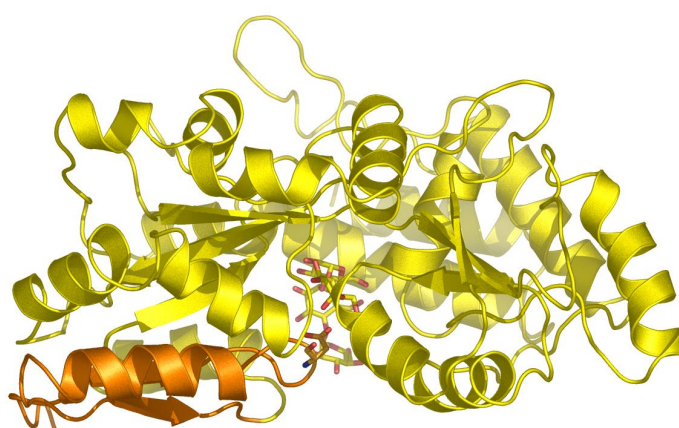
#### *Biochemical evidence for the “transport triggering mechanism”*

Undoubtedly, the contribution of structural information is of enormous advantage when revealing mechanistic processes. Biochemical experiments, allowing for the verification of ideas based on a structural knowledge, are most beneficial to complement the structural data. Within this thesis it was impossible to investigate the proposed mechanism by biochemical means. Therefore, biochemical data on other binding protein dependent systems was analyzed to corroborate the “transport triggering mechanism”. Biochemical data on mutant histidine binding protein from *Salmonella typhimurium* revealed a defect in histidine transport ( $K_m$  values 3-150 times worse than for the wild type protein)<sup>231</sup>. The mutant Y14H shows changes in the characteristics of histidine transport. As tyrosine 14 is taking part in substrate binding it is not astonishing that the  $K_d$  value is changed, from 0.03  $\mu\text{M}$  in the wild type to 2  $\mu\text{M}$  in Y14H mutant. Analyzing the conformation of the substrate binding protein with the help of a monoclonal antibody, which specifically binds to the closed form, indicates that the protein exhibits an abnormal closed conformation. In this study it is anticipated that the Y14H mutant can exhibit a closed conformation, which, however, is different from the normally closed conformation. This abnormal closed conformation could reflect a semi-closed conformation as found for ChoX. In this case the Y14 should be part of a rotational subdomain. Possible rigid bodies in the histidine binding protein, which may undergo a rotational movement, are analyzed and, indeed, one subdomain is found comprising residues 10-33 and 215-238. The identified subdomain and the substrate binding tyrosine 14 are presented in Figure 4.6, where the *E. coli* structure (PDB accession code: 1HSL) is shown which is 98% identical to the *S. typhimurium* protein.



**Figure 4.6** Cartoon presentation of the *E. coli* histidine binding protein (PDB accession code: 1HSL). The possible rotational subdomain is depicted in orange. Tyrosine 14, which interacts with the substrate, is presented in sticks.

Another interesting finding corroborating the importance of a rotational subdomain was made on the well studied maltose binding protein MalE from *E. coli*. In this protein Asp14, takes part in substrate binding by hydrogen bonding to the sugar moiety. The mutation of this residue to Tyr does not affect the affinity for maltose (9.4  $\mu\text{M}$  and 8.5  $\mu\text{M}$  for wild type and D14Y MalE, respectively)<sup>232</sup>. Nevertheless, substrate transport is reduced by ~80 % utilizing the mutant MalE binding protein<sup>233</sup>. Analysis of the possible hinges and rigid domains in MalE indicates a subdomain within MalE comprising residues 1-40 (see Figure 4.7). The identified mutation D14Y lies within this subdomain. This residue may function like the Asp45 in ChoX; substrate binding to MalE may induce a rotation of the subdomain in MalE mediated by binding of Asp14 to the substrate and an ATPase activity can be triggered.



**Figure 4.7** Cartoon presentation of the *E. coli* maltose binding protein (PDB accession code: 4MBP). The possible rotational subdomain is depicted in orange. Aspartate 14, which interacts with the substrate, is presented in sticks.

The facts that the two mutations, identified in HisJ (Y14H) and MalE (D14Y), both show reduced transport activities and that the mutated amino acid taking part in substrate binding

are situated in a possible rotational subdomain add to the idea that a “transport triggering mechanism”, induced by the rotational subdomain, is possible not only for the choline importer but also for other ABC import systems. These results suggest an inability of the mutant to trigger ATPase activity and corroborate the “transport triggering mechanism”.

### *Evolutionary aspects*

The identified rotational subdomain, which influences the triggering of ATP binding and/or hydrolysis, establishes a powerful tool to minimized wasteful ATP usage. However, the question arises if all substrate binding proteins exhibit such a subdomain. In general all substrate binding proteins can be divided into two globular lobes, both exposing a central  $\beta$ -sheet surrounded by  $\alpha$ -helices. The substrate is bound in the cleft between those lobes. In 1999 Fukami-Kobayashi and coworkers investigated the evolutionary history of type I and II substrate binding proteins<sup>184</sup>. SBP that belong to the third class showing a single helical linker segment like BtuF<sup>65</sup> and were not included in this study. For the analyzed SBP the authors present a sequence based tree that, in conjunction with the structural analysis of 15 SBPs, leads to the proposal of a genealogical chart. This chart assumes a common ancestor for the development for type I as well as type II binding proteins. The proteins along the evolutionary path show changes in their secondary elements and their structural arrangement. Interestingly, along their evolutionary path the binding proteins loose some of their secondary elements but also gain some extra elements. For example, HisJ, the histidine binding protein, gains two N-terminal  $\beta$ -strands positioned after the first  $\beta$ -strand of the central  $\beta$ -sheet. Furthermore, an additional helix is revealed at the C-terminus. Strikingly, the additional structural elements in HisJ (residues ~11-28 and 221-238) represent the biggest part of the before described rotational subdomain (residues 10-33 and 215-238). The additionally gained structural elements in binding proteins might take the role of a rotational subdomain, functioning as a sensor for the substrate and transmitting this information according to the “transport triggering mechanism”. It might be possible that in evolution the substrate binding proteins may have developed to minimize wasteful ATP hydrolysis. SBP lacking the additional structural elements might trigger transport via different interactions. This could be possible by a periplasmic loop, like the periplasmic loop between TM 5 and 5a of the BtuCDF complex, which in the complex structure is placed in the substrate binding site of BtuF. However, in cases where the substrate is highly abundant a “transport triggering mechanism” might not be needed as the liganded binding protein is the prevailing species.

#### 4.2.7 Outlook

The structural characterization of the choline binding protein ChoX presents a first step for understanding the binding mode of different ligands to this protein. However, to further pinpoint the importance of individual amino acids and their contribution to the affinity for different ligands mutants have to be designed and their binding properties have to be characterized biochemically. Definitely, with the help of the presented structures the choice of mutation can be rationalized more easily.

Furthermore, it is extremely interesting to evaluate the “transport triggering mechanism”. To gain further insight to this newly proposed mechanism, not only the substrate binding protein ChoX but also the transmembrane component ChoW and the NBD ChoV and the interplay of these parts have to be analyzed. *In vitro* experiments with a reconstituted ChoWVX system or *semi-in vivo* systems with isolated, intact membranes and purified ChoX could form the basis to uncover the effects of different mutations on ATP hydrolysis and substrate transport.

Undoubtedly, the structural characterization of the choline importer in complex with its binding protein, as shown for ModABC<sup>10</sup>, BtuCDF<sup>54</sup> and MalEFGK<sup>57</sup>, and in various catalytic states would greatly enhance our understanding for the mechanism underlying substrate translocation in ABC importers.

## References

1. Mellman, I. & Warren, G. The road taken: Past and future foundations of membrane traffic. *Cell* **100**, 99-112 (2000).
2. Singer, S.J. & Nicolson, G.L. The fluid mosaic model of the structure of cell membranes. *Science* **175**, 720-731 (1972).
3. Saier, M.H., Jr. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* **64**, 354-411 (2000).
4. Kaback, H.R., Sahin-Toth, M. & Weinglass, A.B. The kamikaze approach to membrane transport. *Nat Rev Mol Cell Biol* **2**, 610-620 (2001).
5. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of the protein subunits in the photosynthetic reaction center of *Rhodospseudomonas viridis* at 3 Å resolution. *Nature* **318**, 618-624 (1985).
6. Linton, K.J. & Higgins, C.F. The Escherichia coli ATP-binding cassette (ABC) proteins. *Mol Microbiol* **28**, 5-13 (1998).
7. Saier, M.H., Hvorup, R.N. & Barabote, R.D. Evolution of the bacterial phosphotransferase system: from carriers and enzymes to group translocators. *Biochem Soc Trans* **33**, 220-224 (2005).
8. Higgins, C.F. ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* **8**, 67-113 (1992).
9. Blight, M.A. & Holland, I.B. Structure and function of haemolysin B, P-glycoprotein and other members of a novel family of membrane translocators. *Mol Microbiol* **4**, 873-880 (1990).
10. Hollenstein, K., Frei, D.C. & Locher, K.P. Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213-216 (2007).
11. Riordan, J.R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066-1073 (1989).
12. Riordan, J.R. Assembly of functional CFTR chloride channels. *Annu Rev Physiol* **67**, 701-718 (2005).
13. Bavoil, P., Hofnung, M. & Nikaido, H. Identification of a Cytoplasmic Membrane-Associated Component of the Maltose Transport-System of Escherichia-Coli. *J Biol Chem* **255**, 8366-8369 (1980).
14. Albers, S.V. *et al.* Glucose transport in the extremely thermoacidophilic *Sulfolobus solfataricus* involves a high-affinity membrane-integrated binding protein. *J Bacteriol* **181**, 4285-4291 (1999).
15. Locher, K.P., Lee, A.T. & Rees, D.C. The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**, 1091-1098 (2002).
16. Gottesman, M.M. & Pastan, I. Biochemistry of multidrug resistance mediated by the multidrug transporter. *Annu Rev Biochem* **62**, 385-427 (1993).
17. Schmitt, L. & Tampe, R. Affinity, specificity, diversity: a challenge for the ABC transporter TAP in cellular immunity. *Chembiochem* **1**, 16-35 (2000).
18. Hinsä, S.M., Espinosa-Urgel, M., Ramos, J.L. & O'Toole, G.A. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. *Mol Microbiol* **49**, 905-918 (2003).
19. Wada, M. *et al.* Mutations in the canalicular multispecific organic anion transporter (cMOAT) gene, a novel ABC transporter, in patients with hyperbilirubinemia II Dubin-Johnson syndrome. *Hum Mol Genet* **7**, 203-207 (1998).
20. Allikmets, R. Simple and complex ABCR: Genetic predisposition to retinal disease. *Am J Hum Genet* **67**, 793-799 (2000).

21. Khwaja, M., Ma, Q.H. & Saier, M.H. Topological analysis of integral membrane constituents of prokaryotic ABC efflux systems. *Res Microbiol* **156**, 270-277 (2005).
22. Biemans-Oldehinkel, E., Doeven, M.K. & Poolman, B. ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett* **580**, 1023-1035 (2006).
23. van der Heide, T. & Poolman, B. ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep* **3**, 938-943 (2002).
24. Higgins, C.F. *et al.* A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature* **323**, 448-450 (1986).
25. Oswald, C., Holland, I.B. & Schmitt, L. The motor domains of ABC-transporters. What can structures tell us? *Naunyn Schmiedebergs Arch Pharmacol* **372**, 385-399 (2006).
26. Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J* **1**, 945-951 (1982).
27. Vetter, I.R. & Wittinghofer, A. Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. *Q Rev Biophys* **32**, 1-56 (1999).
28. Bianchet, M.A., Ko, Y.H., Amzel, L.M. & Pedersen, P.L. Modeling of nucleotide binding domains of ABC transporter proteins based on a F1-ATPase/recA topology: structural model of the nucleotide binding domains of the cystic fibrosis transmembrane conductance regulator (CFTR). *J Bioenerg Biomembr* **29**, 503-524 (1997).
29. Schmitt, L., Benabdelhak, H., Blight, M.A., Holland, I.B. & Stubbs, M.T. Crystal structure of the nucleotide binding domain of the ABC-transporter haemolysin B: Identification of a variable region within ABC helical domains. *J Mol Biol* **330**, 333-342 (2003).
30. Dawson, R.J.P. & Locher, K.P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180-185 (2006).
31. Mourez, M., Hofnung, M. & Dassa, E. Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *Embo J* **16**, 3066-3077 (1997).
32. Dassa, E. & Hofnung, M. Sequence of gene malG in E. coli K12: homologies between integral membrane components from binding protein-dependent transport systems. *Embo J* **4**, 2287-2293 (1985).
33. Hung, L.-W. *et al.* Crystal structure of the ATP-binding domain of an ABC transporter. *Nature* **396**, 703-707 (1998).
34. Diederichs, K. *et al.* Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon *Thermococcus litoralis*. *Embo J* **19**, 5951-5961 (2000).
35. Chen, J., Lu, G., Lin, J., Davidson, A.L. & Quirocho, F.A. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol Cell* **12**, 651-661 (2003).
36. Lu, G., Westbrook, J.M., Davidson, A.L. & Chen, J. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc Nat Acad Sci U S A* **102**, 17969-17974 (2005).
37. Ose, T., Fujie, T., Yao, M., Watanabe, N. & Tanaka, I. Crystal structure of the ATP-binding cassette of multisugar transporter from *Pyrococcus horikoshii* OT3. *Proteins* **57**, 635-638 (2004).
38. Yuan, Y.R. *et al.* The crystal structure of the MJ0796 ATP-binding cassette. Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter. *J Biol Chem* **276**, 32313-32321 (2001).

39. Smith, P.C. *et al.* ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell* **10**, 139-149 (2002).
40. Karpowich, N. *et al.* Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. *Structure* **9**, 571-586 (2001).
41. Gaudet, R. & Wiley, D.C. Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. *EMBO J.* **20**, 4964-4972 (2001).
42. Verdon, G., Albers, S.V., Dijkstra, B.W., Driessen, A.J. & Thunnissen, A.M. Crystal structures of the ATPase subunit of the glucose ABC transporter from *Sulfolobus solfataricus*: nucleotide-free and nucleotide-bound conformations. *J Mol Biol* **330**, 343-358 (2003).
43. Verdon, G. *et al.* Formation of the productive ATP-Mg<sup>2+</sup>-bound dimer of GlcV, an ABC-ATPase from *Sulfolobus solfataricus*. *J Mol Biol* **334**, 255-267 (2003).
44. Zaitseva, J., Jenewein, S., Jumpertz, T., Holland, I.B. & Schmitt, L. H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *Embo J* (2005).
45. Zaitseva, J. *et al.* A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer. *Embo J* **25**, 3432-3443 (2006).
46. Lewis, H.A. *et al.* Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. *Embo J* **23**, 282-293 (2004).
47. Lewis, H.A. *et al.* Impact of the Delta F508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. *J Biol Chem* **280**, 1346-1353 (2005).
48. Scheffel, F. *et al.* Structure of the ATPase subunit CysA of the putative sulfate ATP-binding cassette (ABC) transporter from *Alicyclobacillus acidocaldarius*. *FEBS Lett* **579**, 2953-2958 (2005).
49. Story, R.M. & Steitz, T.A. Structure of the recA protein-ADP complex. *Nature* **355**, 374-376 (1992).
50. Abrahams, J.P., Leslie, A.G., Lutter, R. & Walker, J.E. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621-628 (1994).
51. Ambudkar, S.V., Kim, I.W., Xia, D. & Sauna, Z.E. The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding. *FEBS Lett* **580**, 1049-1055 (2006).
52. Kim, I.W. *et al.* The conserved tyrosine residues 401 and 1044 in ATP sites of human P-glycoprotein are critical for ATP binding and hydrolysis: evidence for a conserved subdomain, the A-loop in the ATP-binding cassette. *Biochemistry* **45**, 7605-7616 (2006).
53. Ernst, R., Koch, J., Horn, C., Tampe, R. & Schmitt, L. Engineering ATPase activity in the isolated ABC cassette of human TAP1. *J Biol Chem* **281**, 27471-27480 (2006).
54. Hvorup, R.N. *et al.* Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. *Science* **317**, 1387-1390 (2007).
55. Dawson, R.J.P. & Locher, K.P. Structure of the multidrug ABC transporter Sav1866 from *Staphylococcus aureus* in complex with AMP-PNP. *FEBS Lett* **581**, 935-938 (2007).
56. Pinkett, H.W., Lee, A.T., Lum, P., Locher, K.P. & Rees, D.C. An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science* **315**, 373-377 (2007).
57. Oldham, M.L., Khare, D., Quirocho, F.A., Davidson, A.L. & Chen, J. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* **450**, 515-521 (2007).

58. Chang, G. *et al.* Structure of MsbA from E. coli: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters (Retraction of vol 293, pg 1793, 2001). *Science* **314**, 1875-1875 (2006).
59. Ward, A., Reyes, C.L., Yu, J., Roth, C.B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A* **104**, 19005-19010 (2007).
60. Hollenstein, K., Dawson, R.J. & Locher, K.P. Structure and mechanism of ABC transporter proteins. *Curr Opin Struct Biol* **17**, 412-418 (2007).
61. Quijcho, F.A., Gilliland, G.L. & Phillips, G.N., Jr. The 2.8-A resolution structure of the L-arabinose-binding protein from Escherichia coli. Polypeptide chain folding, domain similarity, and probable location of sugar-binding site. *J Biol Chem* **252**, 5142-5149 (1977).
62. Yao, N., Trakhanov, S. & Quijcho, F.A. Refined 1.89-A structure of the histidine-binding protein complexed with histidine and its relationship with many other active transport/chemosensory proteins. *Biochemistry* **33**, 4769-4779 (1994).
63. Oh, B.H. *et al.* The bacterial periplasmic histidine-binding protein. structure/function analysis of the ligand-binding site and comparison with related proteins. *J Biol Chem* **269**, 4135-4143 (1994).
64. Oh, B.H. *et al.* Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J Biol Chem* **268**, 11348-11355 (1993).
65. Borths, E.L., Locher, K.P., Lee, A.T. & Rees, D.C. The structure of Escherichia coli BtuF and binding to its cognate ATP binding cassette transporter. *Proc Natl Acad Sci U S A* **99**, 16642-16647 (2002).
66. Karpowich, N.K., Huang, H.H., Smith, P.C. & Hunt, J.F. Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein mobility upon ligand binding. *J Biol Chem* **278**, 8429-8434 (2003).
67. Mao, B., Pear, M.R., McCammon, J.A. & Quijcho, F.A. Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J Biol Chem* **257**, 1131-1133 (1982).
68. Newcomer, M.E., Lewis, B.A. & Quijcho, F.A. The radius of gyration of L-arabinose-binding protein decreases upon binding of ligand. *J Biol Chem* **256**, 13218-13222 (1981).
69. Flocco, M.M. & Mowbray, S.L. The 1.9 Angstrom X-Ray Structure of a Closed Unliganded Form of the Periplasmic Glucose/Galactose Receptor from Salmonella-Typhimurium. *J Biol Chem* **269**, 8931-8936 (1994).
70. Duan, X.Q. & Quijcho, F.A. Structural evidence for a dominant role of nonpolar interactions in the binding of a transport/chemosensory receptor to its highly polar ligands. *Biochemistry* **41**, 706-712 (2002).
71. Sack, J.S., Saper, M.A. & Quijcho, F.A. Periplasmic binding protein structure and function. *J Mol Biol* **206**, 171-191 (1989).
72. Ravindranathan, K.P., Gallicchio, E. & Levy, R.M. Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *J Mol Biol* **353**, 196-210 (2005).
73. Horn, C., Bremer, E. & Schmitt, L. Functional overexpression and in vitro re-association of OpuA, an osmotically regulated ABC-transport complex from Bacillus subtilis. *FEBS Lett* **579**, 5765-5768 (2005).
74. Davidson, A.L., Shuman, H.A. & Nikaido, H. Mechanism of maltose transport in Escherichia coli: transmembrane signaling by periplasmic binding proteins. *Proc Natl Acad Sci U S A* **89**, 2360-2364 (1992).

75. Prossnitz, E., Gee, A. & Ames, G.F. Reconstitution of the Histidine Periplasmic Transport-System in Membrane-Vesicles - Energy Coupling and Interaction between the Binding-Protein and the Membrane Complex. *J Biol Chem* **264**, 5006-5014 (1989).
76. Liu, C.E., Liu, P.Q. & Ames, G.F. Characterization of the adenosine triphosphatase activity of the periplasmic histidine permease, a traffic ATPase (ABC transporter). *J Biol Chem* **272**, 21883-21891. (1997).
77. Petronilli, V. & Ames, G.F. Binding protein-independent histidine permease mutants. Uncoupling of ATP hydrolysis from transmembrane signaling. *J Biol Chem* **266**, 16293-16296 (1991).
78. Chen, J., Sharma, S., Quioco, F.A. & Davidson, A.L. Trapping the transition state of an ATP-binding cassette transporter: evidence for a concerted mechanism of maltose transport. *Proc Natl Acad Sci U S A* **98**, 1525-1530 (2001).
79. Dawson, R.J., Hollenstein, K. & Locher, K.P. Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism. *Mol Microbiol* **65**, 250-257 (2007).
80. Holland, I.B. Translocation of bacterial proteins--an overview. *Biochim Biophys Acta* **1694**, 5-16 (2004).
81. Yen, M.R., Tseng, Y.H., Nguyen, E.H., Wu, L.F. & Saier, M.H., Jr. Sequence and phylogenetic analyses of the twin-arginine targeting (TAT) protein export system. *Arch Microbiol* **177**, 441-450 (2002).
82. Stephenson, K. Sec-dependent protein translocation across biological membranes: evolutionary conservation of an essential protein transport pathway. *Mol Membr Biol* **22**, 17-28 (2005).
83. Holland, I.B., Schmitt, L. & Young, J. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway. *Mol Membr Biol* **22**, 29-39 (2005).
84. Desvaux, M., Hebraud, M., Henderson, I.R. & Pallen, M.J. Type III secretion: what's in a name? *Trends Microbiol* **14**, 157-160 (2006).
85. Gray, L., Mackman, N., Nicaud, J.M. & Holland, I.B. The carboxy-terminal region of haemolysin 2001 is required for secretion of the toxin from Escherichia coli. *Mol Gen Genet* **205**, 127-133 (1986).
86. Delepelaire, P. Type I secretion in gram-negative bacteria. *Biochim Biophys Acta* **1694**, 149-161 (2004).
87. Baumann, U., Wu, S., Flaherty, K.M. & McKay, D.B. Three-dimensional structure of the alkaline protease of Pseudomonas aeruginosa: a two-domain protein with a calcium binding parallel beta roll motif. *Embo J* **12**, 3357-3364 (1993).
88. Goebel, W. & Hedgpeth, J. Cloning and functional characterization of the plasmid-encoded hemolysin determinant of Escherichia coli. *J Bacteriol* **151**, 1290-1298 (1982).
89. Ghigo, J.M. & Wandersman, C. A carboxyl-terminal four-amino acid motif is required for secretion of the metalloprotease PrtG through the Erwinia chrysanthemi protease secretion pathway. *J Biol Chem* **269**, 8979-8985 (1994).
90. Thanabalu, T., Koronakis, E., Hughes, C. & Koronakis, V. Substrate-induced assembly of a contiguous channel for protein export from E.coli: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *Embo J* **17**, 6487-6496 (1998).
91. Letoffe, S., Delepelaire, P. & Wandersman, C. Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein-mediated exporters is ordered and promoted by substrate binding. *Embo J* **15**, 5804-5811 (1996).
92. Balakrishnan, L., Hughes, C. & Koronakis, V. Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by Escherichia coli. *J Mol Biol* **313**, 501-510 (2001).

93. Benabdelhak, H. *et al.* A specific interaction between the NBD of the ABC-transporter HlyB and a C-terminal fragment of its transport substrate haemolysin A. *J Mol Biol* **327**, 1169-1179 (2003).
94. Vakharia, H., German, G.J. & Misra, R. Isolation and characterization of *Escherichia coli* tolC mutants defective in secreting enzymatically active alpha-hemolysin. *J Bacteriol* **183**, 6908-6916 (2001).
95. Pimenta, A.L., Racher, K., Jamieson, L., Blight, M.A. & Holland, I.B. Mutations in HlyD, part of the type 1 translocator for hemolysin secretion, affect the folding of the secreted toxin. *J Bacteriol* **187**, 7471-7480 (2005).
96. Welch, R.A., Dellinger, E.P., Minshew, B. & Falkow, S. Haemolysin contributes to virulence of extra-intestinal *E. coli* infections. *Nature* **294**, 665-667 (1981).
97. Mackman, N. & Holland, I.B. Secretion of a 107 K dalton polypeptide into the medium from a haemolytic *E. coli* K12 strain. *Mol Gen Genet* **193**, 312-315 (1984).
98. Mackman, N. & Holland, I.B. Functional characterization of a cloned haemolysin determinant from *E. coli* of human origin, encoding information for the secretion of a 107K polypeptide. *Mol Gen Genet* **196**, 129-134 (1984).
99. Mackman, N., Nicaud, J.M., Gray, L. & Holland, I.B. Genetical and functional organisation of the *Escherichia coli* haemolysin determinant 2001. *Mol Gen Genet* **201**, 282-288 (1985).
100. Nicaud, J.M., Mackman, N., Gray, L. & Holland, I.B. Characterization of HlyC and Mechanism of Activation and Secretion of Hemolysin from *Escherichia-Coli*-2001. *FEBS Lett* **187**, 339-344 (1985).
101. Zaitseva, J. *et al.* Functional characterization and ATP-induced dimerization of the isolated ABC-domain of the haemolysin B transporter. *Biochemistry* **44**, 9680-9690 (2005).
102. Dinh, T., Paulsen, I.T. & Saier, M.H., Jr. A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of gram-negative bacteria. *J Bacteriol* **176**, 3825-3831 (1994).
103. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914-919 (2000).
104. Sohlenkamp, C., Lopez-Lara, I.M. & Geiger, O. Biosynthesis of phosphatidylcholine in bacteria. *Prog Lipid Res* **42**, 115-162 (2003).
105. Lopez-Lara, I.M., Sohlenkamp, C. & Geiger, O. Membrane lipids in plant-associated bacteria: Their biosyntheses and possible functions. *Mol Plant Microbe Interact* **16**, 567-579 (2003).
106. Kent, C. Eukaryotic phospholipid biosynthesis. *Annu Rev Biochem* **64**, 315-343 (1995).
107. Kaneshiro, T. & Law, J.H. Phosphatidylcholine Synthesis in *Agrobacterium Tumefaciens*. I. Purification and Properties of a Phosphatidylethanolamine N-Methyltransferase. *J Biol Chem* **239**, 1705-1713 (1964).
108. de Rudder, K.E.E., Sohlenkamp, C. & Geiger, O. Plant-exuded choline is used for rhizobial membrane lipid biosynthesis by phosphatidylcholine synthase. *J Biol Chem* **274**, 20011-20016 (1999).
109. Arakawa, T. & Timasheff, S.N. The stabilization of proteins by osmolytes. *Biophys J* **47**, 411-414 (1985).
110. Kortstee, G.J. Aerobic Decomposition of Choline by Microorganisms .1. Ability of Aerobic Organisms, Particularly Coryneform Bacteria, to Utilize Choline as Sole Carbon and Nitrogen Source. *Arch Microbiol* **71**, 235-& (1970).
111. Bernard, T., Pocard, J.A., Perroud, B. & Lerudulier, D. Variations in the Response of Salt-Stressed *Rhizobium* Strains to Betaines. *Arch Microbiol* **143**, 359-364 (1986).

112. Pocard, J.A., Bernard, T., Smith, L.T. & Lerudulier, D. Characterization of 3 Choline Transport Activities in Rhizobium-Meliloti - Modulation by Choline and Osmotic-Stress. *J Bacteriol* **171**, 531-537 (1989).
113. Dupont, L. *et al.* The Sinorhizobium meliloti ABC transporter Cho is highly specific for choline and expressed in bacteroids from Medicago sativa nodules. *J Bacteriol* **186**, 5988-5996 (2004).
114. Kuipers, O.P., de Ruyter, P., Kleerebezem, M. & de Vos, W.M. Quorum sensing-controlled gene expression in lactic acid bacteria. *J Biotechnol* **64**, 15-21 (1998).
115. de Ruyter, P.G., Kuipers, O.P. & de Vos, W.M. Controlled gene expression systems for Lactococcus lactis with the food- grade inducer nisin. *Appl Environ Microbiol* **62**, 3662-3667 (1996).
116. Kunji, E.R., Slotboom, D.J. & Poolman, B. Lactococcus lactis as host for overproduction of functional membrane proteins. *Biochim Biophys Acta* **1610**, 97-108 (2003).
117. Chayen, N.E. Methods for separating nucleation and growth in protein crystallisation. *Prog Biophys Mol Biol* **88**, 329-337 (2005).
118. Chayen, N.E., Stewart, P.D.S., Maeder, D.L. & Blow, D.M. An Automated-System for Microbatch Protein Crystallization and Screening. *J Appl Crystallogr* **23**, 297-302 (1990).
119. Chayen, N.E., Stewart, P.D.S. & Blow, D.M. Microbatch Crystallization under Oil - a New Technique Allowing Many Small-Volume Crystallization Trials. *J Cryst Growth* **122**, 176-180 (1992).
120. D'Arcy, A., Elmore, C., Stihle, M. & Johnston, J.E. A novel approach to crystallising proteins under oil. *J Cryst Growth* **168**, 175-180 (1996).
121. D'Arcy, A., MacSweeney, A. & Haber, A. Practical aspects of using the microbatch method in screening conditions for protein crystallization. *Methods* **34**, 323-328 (2004).
122. Barends, T.R.M. & Dijkstra, B.W. Oils used in microbatch crystallization do not remove a detergent from the drops they cover. *Acta Crystallogr, Sect D: Biol Crystallogr* **59**, 2345-2347 (2003).
123. Jancarik, J. & Kim, S.H. Sparse-Matrix Sampling - a Screening Method for Crystallization of Proteins. *J Appl Crystallogr* **24**, 409-411 (1991).
124. Landau, E.M. & Rosenbusch, J.P. Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci U S A* **93**, 14532-14535 (1996).
125. Bergfors, T.M. *Protein Crystallization - Techniques, Strategies, and Tips*, (International University Line, La Jolla, 1999).
126. Birtley, J.R. & Curry, S. Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy. *Acta Crystallogr, Sect D: Biol Crystallogr* **61**, 646-650 (2005).
127. Bergfors, T. Seeds to crystals. *J Struct Biol* **142**, 66-76 (2003).
128. van Aalten, D.M.F. *et al.* Binding site differences revealed by crystal structures of Plasmodium falciparum and bovine acyl-CoA binding protein. *J Mol Biol* **309**, 181-192 (2001).
129. Luft, J.R. & DeTitta, G.T. A method to produce microseed stock for use in the crystallization of biological macromolecules. *Acta Crystallogr Sect D: Biol Crystallogr* **55**, 988-993 (1999).
130. Stura, E.A. & Wilson, I.A. Applications of the Streak Seeding Technique in Protein Crystallization. *J Cryst Growth* **110**, 270-282 (1991).
131. Bergfors, T. <http://xray.bmc.uu.se/~terese/crystallization/pics/streakseed.jpeg>.

132. Chayen, N.E., Saridakis, E. & Sear, R.P. Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium. *Proc Natl Acad Sci U S A* **103**, 597-601 (2006).
133. Drenth, J. *Principles of Protein X-ray Crystallography*, (Springer Verlag, New-York, 1999).
134. Massa, W. *Kristallstrukturbestimmung*, (Teubner, Stuttgart, 1996).
135. Blow, D. *Outline of Crystallography for Biologists*, (Oxford University Press, Oxford, 2004).
136. Hendrickson, W.A. Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science* **254**, 51-58 (1991).
137. Hendrickson, W.A. Maturation of MAD phasing for the determination of macromolecular structures. *J Synchrotron Radiat* **6**, 845-851 (1999).
138. Hendrickson, W.A., Horton, J.R. & Lemaster, D.M. Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (Mad) - a Vehicle for Direct Determination of 3-Dimensional Structure. *Embo J* **9**, 1665-1672 (1990).
139. Yeates, T.O. & Fam, B.C. Protein crystals and their evil twins. *Structure* **7**, R25-29 (1999).
140. Stanley, E. The Identification of Twins from Intensity Statistics. *J Appl Crystallogr* **5**, 191-194 (1972).
141. Rees, D.C. The Influence of Twinning by Merohedry on Intensity Statistics. *Acta Crystallogr, Sect A* **36**, 578-581 (1980).
142. Bailey, S. The CCP4 Suite - Programs for Protein Crystallography. *Acta Crystallogr, Sect D: Biol Crystallogr* **50**, 760-763 (1994).
143. Britton, D. Estimation of Twinning Parameter for Twins with Exactly Superimposed Reciprocal Lattices. *Acta Crystallogr, Sect A* **28**, 296-& (1972).
144. Fisher, R.G. & Sweet, R.M. Treatment of Diffraction Data from Protein Crystals Twinned by Merohedry. *Acta Crystallogr, Sect A* **36**, 755-760 (1980).
145. Yeates, T.O. Simple Statistics for Intensity Data from Twinned Specimens. *Acta Crystallogr, Sect A* **44**, 142-144 (1988).
146. Yeates, T.O. Detecting and overcoming crystal twinning. in *Macromol Crystallogr, Pt A*, Vol. 276 344-358 (1997).
147. Padilla, J.E. & Yeates, T.O. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr, Sect D: Biol Crystallogr* **59**, 1124-1130 (2003).
148. Sultana, A. *et al.* Structure determination by multiwavelength anomalous diffraction of aclacinomycin oxidoreductase: indications of multidomain pseudomerohedral twinning. *Acta Crystallogr, Sect D: Biol Crystallogr* **63**, 149-159 (2007).
149. Sheldrick, G.M. & Schneider, T.R. SHELXL: High-resolution refinement. in *Macromol Crystallogr, Pt B*, Vol. 277 319-343 (1997).
150. Brünger, A.T. & Warren, G.L. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Cryst* **D54**, 905-921 (1998).
151. Redinbo, M.R. & Yeates, T.O. Structure Determination of Plastocyanin from a Specimen with a Hemihedral Twinning Fraction of One-Half. *Acta Crystallogr, Sect D: Biol Crystallogr* **49**, 375-380 (1993).
152. Hope, H. Cryocrystallography of Biological Macromolecules - a Generally Applicable Method. *Acta Crystallogr, Sect B: Struct Sci* **44**, 22-26 (1988).
153. McFerrin, M.B. & Snell, E.H. The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *J Appl Crystallogr* **35**, 538-545 (2002).
154. Bourenkov, G.P. & Popov, A.N. A quantitative approach to data-collection strategies. *Acta Crystallogr, Sect D: Biol Crystallogr* **62**, 58-64 (2006).

155. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. in *Methods Enzymol*, Vol. 276 (eds. Carter, C.W. & Sweet, R.M.) (Academic Press, London, 1997).
156. Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J Appl Cryst* **26**, 795-800 (1993).
157. Terwilliger, T.C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr D* **55**, 849-861 (1999).
158. Terwilliger, T.C. Maximum-likelihood density modification. *Acta Crystallogr, D: Biol Crystallogr* **56**, 965-972 (2000).
159. Terwilliger, T.C. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr, Sect D: Biol Crystallogr* **59**, 38-44 (2003).
160. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr, Sect D: Biol Crystallogr* **60**, 2126-2132 (2004).
161. Murshudov, G., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr, Sect D: Biol Crystallogr* **53**, 240-255 (1997).
162. Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56 Pt 12**, 1622-1624 (2000).
163. Schomaker, V. & Trueblood, K. On Rigid-Body Motion of Molecules in Crystals. *Acta Crystallogr, Sect B: Struct Crystallogr Cryst Chem B* **24**, 63-& (1968).
164. Winn, M.D., Isupov, M.N. & Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr, Sect D: Biol Crystallogr* **57**, 122-133 (2001).
165. Brunger, A.T. Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* **355**, 472-475 (1992).
166. Schuttelkopf, A.W. & van Aalten, D.M.F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr, Sect D: Biol Crystallogr* **60**, 1355-1363 (2004).
167. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein models. *Acta Crystallogr, Sect D: Biol Crystallogr* **49**, 129-147 (1993).
168. Read, R.J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr, Sect D: Biol Crystallogr* **57**, 1373-1382 (2001).
169. Storoni, L.C., McCoy, A.J. & Read, R.J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr, Sect D: Biol Crystallogr* **60**, 432-438 (2004).
170. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallogr, Sect D: Biol Crystallogr* **61**, 458-464 (2005).
171. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283-291 (1993).
172. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566-579 (1995).
173. Emekli, U., Schneidman-Duhovny, D., Wolfson, H.J., Nussinov, R. & Haliloglu, T. HingeProt: Automated prediction of hinges in protein structures. *Proteins* (2007).
174. Lee, R.A., Razaz, M. & Hayward, S. The DynDom database of protein domain motions. *Bioinformatics* **19**, 1290-1291 (2003).
175. Kleywegt, G.J. Experimental assessment of differences between related protein crystal structures. *Acta Crystallogr, Sect D: Biol Crystallogr* **55**, 1878-1884 (1999).
176. DeLano, W.L. The PyMol Molecular Graphics System. <http://www.pymol.org> (2002).

177. Hanekop, N. Strukturbiologische Charakterisierung des ABC-Transporters LmrA aus *L. lactis* und des Substratbindepoteins EhuB aus *S. meliloti*. *Dissertation* (2006).
178. Iwata, S. *Methods and Results in Crystallization of Membrane Proteins*, (International University Line, La Jolla, 2003).
179. Lemieux, M.J. *et al.* Three-dimensional crystallization of the Escherichia coli glycerol-3-phosphate transporter: A member of the major facilitator superfamily. *Protein Sci* **12**, 2748-2756 (2003).
180. Abele, R., Keinänen, K. & Madden, D.R. Agonist-induced isomerization in a glutamate receptor ligand-binding domain - A kinetic and mutagenetic analysis. *J Biol Chem* **275**, 21355-21363 (2000).
181. Phillips, J.C. & Hodgson, K.O. The Use of Anomalous Scattering Effects to Phase Diffraction Patterns from Macromolecules. *Acta Crystallogr, Sect A* **36**, 856-864 (1980).
182. Matthews, B.W. Solvent content of protein crystals. *J Mol Biol* **33**, 491-497 (1968).
183. Brünger, A.T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475 (1992).
184. Fukami-Kobayashi, K., Tateno, Y. & Nishikawa, K. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol* **286**, 279-290 (1999).
185. Quijoch, F.A. & Ledvina, P.S. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol Microbiol* **20**, 17-25 (1996).
186. Schiefner, A. *et al.* Cation-pi interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from Escherichia coli. *J Biol Chem* **279**, 5588-5596 (2004).
187. Schiefner, A., Holtmann, G., Diederichs, K., Welte, W. & Bremer, E. Structural basis for the binding of compatible solutes by ProX from the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *J Biol Chem* **279**, 48270-48281 (2004).
188. Horn, C. *et al.* Molecular determinants for substrate specificity of the ligand-binding protein OpuAC from *Bacillus subtilis* for the compatible solutes glycine betaine and proline betaine. *J Mol Biol* **357**, 592-606 (2006).
189. Li, A.J. & Nussinov, R. A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins* **32**, 111-127 (1998).
190. Weiss, M. *et al.* Molecular architecture and electrostatic properties of a bacterial porin. *Science* **254**, 1627-1630 (1991).
191. Allen, P.M. *et al.* Enhanced immunogenicity of a T cell immunogenic peptide by modifications of its N and C termini. *Int. Immunol.* **1**, 141-150 (1989).
192. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H. & Rees, D.C. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: the cofactors. *Proc Natl Acad Sci U S A* **84**, 5730-5734 (1987).
193. Grishammer, R. Understanding recombinant expression of membrane proteins. *Curr Opin Biotechnol* **17**, 337-340 (2006).
194. Mohanty, A.K. & Wiener, M.C. Membrane protein expression and production: effects of polyhistidine tag length and position. *Protein Expression and Purif* **33**, 311-325 (2004).
195. Guan, L., Smirnova, I.N., Verner, G., Nagamori, S. & Kaback, H.R. Manipulating phospholipids for crystallization of a membrane transport protein. *Proc Natl Acad Sci U S A* **103**, 1723-1726 (2006).
196. Lancaster, C.R., Kroger, A., Auer, M. & Michel, H. Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. *Nature* **402**, 377-385 (1999).

197. Thiagarajan, P. & Tiede, D.M. Detergent Micelle Structure and Micelle-Micelle Interactions Determined by Small-Angle Neutron-Scattering under Solution Conditions Used for Membrane-Protein Crystallization. *J Phys Chem* **98**, 10343-10351 (1994).
198. Littrell, K., Urban, V., Tiede, D. & Thiagarajan, P. Solution structure of detergent micelles at conditions relevant to membrane protein crystallization. *J Appl Crystallogr* **33**, 577-581 (2000).
199. Huang, Y.F., Lemieux, M.J., Song, J.M., Auer, M. & Wang, D.N. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301**, 616-620 (2003).
200. Lemieux, M.J., Reithmeier, R.A. & Wang, D.N. Importance of detergent and phospholipid in the crystallization of the human erythrocyte anion-exchanger membrane domain. *J Struct Biol* **137**, 322-332 (2002).
201. Qin, L., Sharpe, M.A., Garavito, R.M. & Ferguson-Miller, S. Conserved lipid-binding sites in membrane proteins: a focus on cytochrome c oxidase. *Curr Opin Struct Biol* **17**, 444-450 (2007).
202. Zhang, H., Kurisu, G., Smith, J.L. & Cramer, W.A. A defined protein-detergent-lipid complex for crystallization of integral membrane proteins: The cytochrome b6f complex of oxygenic photosynthesis. *Proc Natl Acad Sci U S A* **100**, 5160-5163 (2003).
203. Cherezov, V. & Caffrey, M. Picolitre-scale crystallization of membrane proteins. *J Appl Crystallogr* **39**, 604-606 (2006).
204. Wadsten, P. *et al.* Lipidic sponge phase crystallization of membrane proteins. *J Mol Biol* **364**, 44-53 (2006).
205. Faham, S. & Bowie, J.U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J Mol Biol* **316**, 1-6 (2002).
206. Faham, S. *et al.* Crystallization of bacteriorhodopsin from bicelle formulations at room temperature. *Protein Sci* **14**, 836-840 (2005).
207. Schafmeister, C.E., Miercke, L.J. & Stroud, R.M. Structure at 2.5 Å of a designed peptide that maintains solubility of membrane proteins. *Science* **262**, 734-738 (1993).
208. Yeh, J.I., Du, S., Tortajada, A., Paulo, J. & Zhang, S. Peptergents: peptide detergents that improve stability and functionality of a membrane protein, glycerol-3-phosphate dehydrogenase. *Biochemistry* **44**, 16912-16919 (2005).
209. McGregor, C.L. *et al.* Lipopeptide detergents designed for the structural study of membrane proteins. *Nat Biotechnol* **21**, 171-176 (2003).
210. Yu, S.M. *et al.* An improved tripod amphiphile for membrane protein solubilization. *Protein Sci* **9**, 2518-2527 (2000).
211. Hunte, C. & Michel, H. Crystallisation of membrane proteins mediated by antibody fragments. *Curr Opin Struct Biol* **12**, 503-508 (2002).
212. Cherezov, V. *et al.* High-Resolution Crystal Structure of an Engineered Human {beta}2-Adrenergic G Protein Coupled Receptor. *Science* (2007).
213. Rosenbaum, D.M. *et al.* GPCR Engineering Yields High-Resolution Structural Insights into {beta}2 Adrenergic Receptor Function. *Science* (2007).
214. Segel, I.H. *Enzyme Kinetics*, (Wiley-Interscience, New York, 1993).
215. Bourne, Y. *et al.* Substrate and product trafficking through the active center gorge of acetylcholinesterase analyzed by crystallography and equilibrium binding. *J Biol Chem* **281**, 29256-29267 (2006).
216. Garlatti, V. *et al.* Structural insights into the innate immune recognition specificities of L- and H-ficolins. *Embo J* **26**, 623-633 (2007).

217. Felder, C.B., Graul, R.C., Lee, A.Y., Merkle, H.P. & Sadee, W. The venus flytrap of periplasmic binding proteins: An ancient protein module present in multiple drug receptors. *Aaps Pharmsci* **1**, art. no.-2 (1999).
218. O'Hara, P.J. *et al.* The Ligand-Binding Domain in Metabotropic Glutamate Receptors Is Related to Bacterial Periplasmic Binding-Proteins. *Neuron* **11**, 41-52 (1993).
219. Kuryatov, A., Laube, B., Betz, H. & Kuhse, J. Mutational Analysis of the Glycine-Binding Site of the Nmda Receptor - Structural Similarity with Bacterial Amino Acid-Binding Proteins. *Neuron* **12**, 1291-1300 (1994).
220. Jin, R.S. & Gouaux, E. Probing the function, conformational plasticity, and dimer-dimer contacts of the GluR2 ligand-binding core: Studies of 5-substituted willardiines and GluR2 S1S2 in the crystal. *Biochemistry* **42**, 5201-5213 (2003).
221. Tasneem, A., Iyer, L.M., Jakobsson, E. & Aravind, L. Identification of the prokaryotic ligand-gated ion channels and their implications for the mechanisms and origins of animal Cys-loop ion channels. *Genome Biol* **6**, R4 (2005).
222. Terry, A.V. & Buccafusco, J.J. The cholinergic hypothesis of age and Alzheimer's disease-related cognitive deficits: Recent challenges and their implications for novel drug development. *J Pharmacol Exp Ther* **306**, 821-827 (2003).
223. Picciotto, M.R. & Zoli, M. Nicotinic receptors in aging and dementia. *J Neurobiol* **53**, 641-655 (2002).
224. Celie, P.H.N. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907-914 (2004).
225. Dwyer, M.A. & Hellinga, H.W. Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr Opin Struct Biol* **14**, 495-504 (2004).
226. Ames, G.F.L. & Joshi, A.K. Energy Coupling in Bacterial Periplasmic Permeases. *J Bacteriol* **172**, 4133-4137 (1990).
227. Speiser, D.M. & Ames, G.F. Salmonella-Typhimurium Histidine Periplasmic Permease Mutations That Allow Transport in the Absence of Histidine-Binding Proteins. *J Bacteriol* **173**, 1444-1451 (1991).
228. Treptow, N.A. & Shuman, H.A. Genetic-Evidence for Substrate and Periplasmic-Binding-Protein Recognition by the Malf and Malg Proteins, Cytoplasmic Membrane-Components of the Escherichia-Coli Maltose Transport-System. *J Bacteriol* **163**, 654-660 (1985).
229. Hoffmann, K. [http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html).
230. Notredame, C., Higgins, D.G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-217 (2000).
231. Wolf, A. *et al.* Structure-Function Analysis of the Periplasmic Histidine-Binding Protein - Mutations Decreasing Ligand Binding Alter the Properties of the Conformational Change and of the Closed Form. *J Biol Chem* **270**, 16097-16106 (1995).
232. Treptow, N.A. & Shuman, H.A. Allele-specific malE mutations that restore interactions between maltose-binding protein and the inner-membrane components of the maltose transport system. *J Mol Biol* **202**, 809-822 (1988).
233. Hor, L.I. & Shuman, H.A. Genetic-Analysis of Periplasmic Binding-Protein Dependent Transport in Escherichia-Coli - Each Lobe of Maltose-Binding Protein Interacts with a Different Subunit of the Malfgk(2) Membrane-Transport Complex. *J Mol Biol* **233**, 659-670 (1993).

## Abbreviations

ABC	ATP binding cassette
AchBP	Acetylcholine binding protein
ADP	Adenosine diphosphate
APS	Ammonium persulfate
ASU	Asymmetric unit
ATP	Adenosine triphosphate
Bicine	N,N-Bis(2-hydroxyethyl) glycine
BSA	Bovine serum albumine
CDP	Cytidine 5'-diphosphate
CFTR	Cystic fibrosis transmembrane conductance regulator
CMC	Critical micelle concentration
CPS	Counts per second
CV	Column volume
DDM	Dodecyl- $\beta$ -D-maltopyranoside
DM	Decyl- $\beta$ -D-maltopyranoside
DNA	Desoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylene diamine tetracetic acid
EG	Ethylene glycol
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmatic reticulum
FOM	Figure of merit
HEPES	N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid
HRP	Horseradish peroxidase
IDA	Iminodiacetic acid
IMAC	Immobilized metal ion affinity chromatography
<i>L. lactis</i>	<i>Lactococcus lactis</i>
MAD	Multiple wavelength anomalous diffraction
MDR	Multidrug resistance
MES	2-(N-morpholino)ethansulfonic acid
MFP	Membrane fusion protein
Min	Minutes
MIR	Multiple isomorphous replacement
MPD	Methyl pentandiol
MRP	Multidrug resistance related protein
MWCO	Molecular weight cutoff
NBD	Nucleotide binding domain
NMDA	N-methyl-D-aspartic acid
OD	Optical density
OMP	Outer membrane protein
PAGE	Polyacrylamide gel electrophoresis
PC	Posphatidylcholine
PDB	Protein data base
PEG	Polyethylen glycol
PMSF	Phenylmethylsulfonyl fluoride
PVDF	Polyvinylidene difluoride
RMSD	Root mean square deviation
RTX	Repeats in toxin
<i>S. meliloti</i>	<i>Sinorhizobium meliloti</i>

SBP	Substrate binding protein
SDS	Sodium dodecylsulfate
TAP	Transporter associated with antigen processing
TDAO	Tetradecyl-N,N-dimethylamine-N-oxide
TEMED	N,N,N',N'-Tetramethylethylenediamine
TM	Transmembrane
TMD	Transmembrane domain
Tris	Trishydroxymethylaminomethane
WT	Wild type

## **Acknowledgement**

The last years have certainly been an interesting period filled with lots of good feelings and learning experiences as well as feelings of frustration. I would like to thank all people I have met along this way, in Frankfurt as well as in Düsseldorf, who have made this time more joyful. More specifically I would like to thank....

My supervisor Prof. Lutz Schmitt, for the interesting projects, his radiant enthusiasm and support.

Prof. Hanns Weiss, Dr. Ulrich Schulte and their crew, for supporting the “newcomers” in all ways.

Prof. Erhard Bremer and Marina Höing, for the fruitful cooperation on the ChoX project.

My present and past colleagues for enjoying coffee-breaks and beer-hours together. The coffee and beer fans are: Robert, Sander, Stefan, Silke, Yasmin, Ruta, Annett, Katrin and Kim.

Sander, Stefan, Robert, Nils, Justin and Silke, for many discussions and their generous help with projects.

Stefan, for waking my interest in climbing. It had been a bit sleepy the last years but it is brought back to life now. Thanks to Mensch and all my climbing partners.

Robert, for being him.

My deep gratitude goes to Sander, for taking care of absolutely everything, teaching me crystallography, generating a cheerful environment and happily joining the synchrotron trips. It will be boring without you. I would also like to thank Nicki and the boys for their sacrifices.

Thanks to the new and old friends for putting up with me: Annett, Carola, Cissi, Emese, Robert, Ruta and Simone.

My family, for their love and support of what I do.

# CURRICULUM VITAE

---

## CHRISTINE OSWALD

### PERSONAL DETAILS

---

Date of birth: 14.03.1978  
Place of birth: Memmingen, Germany  
Nationality: German

### EDUCATION

---

- 03/2004-      PhD project at Johann Wolfgang Goethe University, Frankfurt and  
                 Heinrich Heine University, Düsseldorf, Germany (since Aug. 2005)  
                 Supervisor: Prof. Lutz Schmitt  
                 Title: "Structure-function studies of ABC transporters  
                 HlyB from *Escherichia coli* and ChoX from *Sinorhizobium meliloti*"
- 04/2003-      Project work at Chalmers Technical University Göteborg, Sweden  
03/2004      Supervisors: Prof. Bengt Göran Karlsson and Prof. Ute Krengel
- 05-11/2002    Diploma thesis in Chemistry  
                 Chalmers Technical University Göteborg, Sweden and Technical  
                 University Munich, Germany  
                 Title: "Interaction studies of the hydrophilic domains of *Escherichia coli*  
                 transhydrogenase by NMR"  
                 Supervisors: Prof. Bengt Göran Karlsson and Prof. Ute Krengel  
                 (Göteborg); Prof. Horst Kessler (Munich)
- 1997 - 2002    Studies at Technical University Munich, Germany  
                 Subject: Chemistry
- 1997            High school diploma (Abitur) at the Vöhl Gymnasium  
                 Memmingen, Germany

## PROFESSIONAL EXPERIENCE

---

11-12/2001, 08-10/1999, 08-09/1998

GPC Biotech, Martinsried, Germany

PCR, cloning and control of semi-automated system for large-scale microbiological analysis of complex cDNA libraries

## WORKING FIELD AND EXPERIENCE

---

Biochemistry: membrane protein expression, purification and functional analysis

Crystallography: crystallization, data collection (mainly synchrotron radiation sources) and processing, structure determination using both MAD data and molecular replacement programs, structure refinement and validation

## CONFERENCES AND WORKSHOPS

---

2007 11th SBNet Meeting, Tällberg, Sweden

2006 8th International School on the Crystallography of Biological Macromolecules, Como, Italy  
EMBO workshop, PEPC5, Hamburg, Germany (poster award)

2005 Course in Practical Protein Crystallization, Uppsala, Sweden  
5<sup>th</sup> Transport Colloquium Rauischholzhausen, Germany  
4<sup>th</sup> International Summerschool on Membrane Biochemistry, Hirschegg, Austria

2004 8th SBNet Meeting, Tällberg, Sweden  
3<sup>th</sup> International Summerschool on Membrane Biochemistry, Hirschegg, Austria

2003 7th SBNet Meeting, Tällberg, Sweden,  
Workshop on weak protein interactions; Transient Network, Sevilla, Spain

2002 6th SBNet Meeting, Tällberg, Sweden

## PUBLICATIONS

---

- Oswald C., Johansson T., Tornroth S., Okvist M., Krengel U.; Crystallization and preliminary crystallographic analysis of NAD(H)-binding domain of *Escherichia coli* transhydrogenase; Acta Cryst. D, 2004, 743-745.
- Johansson T., Oswald C., Pedersen A., Tornroth S., Okvist M., Karlsson B.G., Rydstrom J., Krengel U.; X-ray structure of domain I of the proton pumping membrane protein transhydrogenase from *Escherichia coli*; JMB, 2005, 299-312.
- Zaitseva J., Jenewein S., Oswald C., Jumpertz T., Holland I.B., Schmitt L.; A molecular understanding of the catalytic cycle of the nucleotide-binding domain of the ABC transporter HlyB; Biochem. Soc. Trans., 2005, 990-995.
- Oswald C., Holland I.B., Schmitt L.; The motor domain of ABC-transporters. What can structures tell us?; Naunyn Schmiedebergs Arch. Pharmacol.; 2006, 385-399.
- Zaitseva J., Oswald C., Jumpertz T., Jenewein S., Wiedenmann A., Holland I.B., Schmitt L.; A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer; EMBO J., 2006, 3432-3443.
- Oswald C., Jenewein S., Holland I.B., Schmitt L.; Water-mediated protein-fluorophore interactions modulate the affinity of an ABC-ATPase/TNP-ADP complex; J. Struct. Biol., in press.
- Oswald C., Smits S.H.J., Hönig M., Bremer E., Schmitt L.; A “transport triggering” mechanism in the choline binding protein ChoX from *Sinorhizobium meliloti*; Structure, under revision.

Hiermit versichere ich, dass ich die vorgelegte Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.  
Zitate wurden kenntlich gemacht.

Christine Oswald

Düsseldorf, den 18.12.2007