X-ray Structure of the Na^+-coupled Glycine-Betaine Symporter BetP from *Corynebacterium glutamicum*

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

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

von
Susanne Ressl
aus Vímperk
Tschechische Republik

Frankfurt am Main (2009)

(D30)

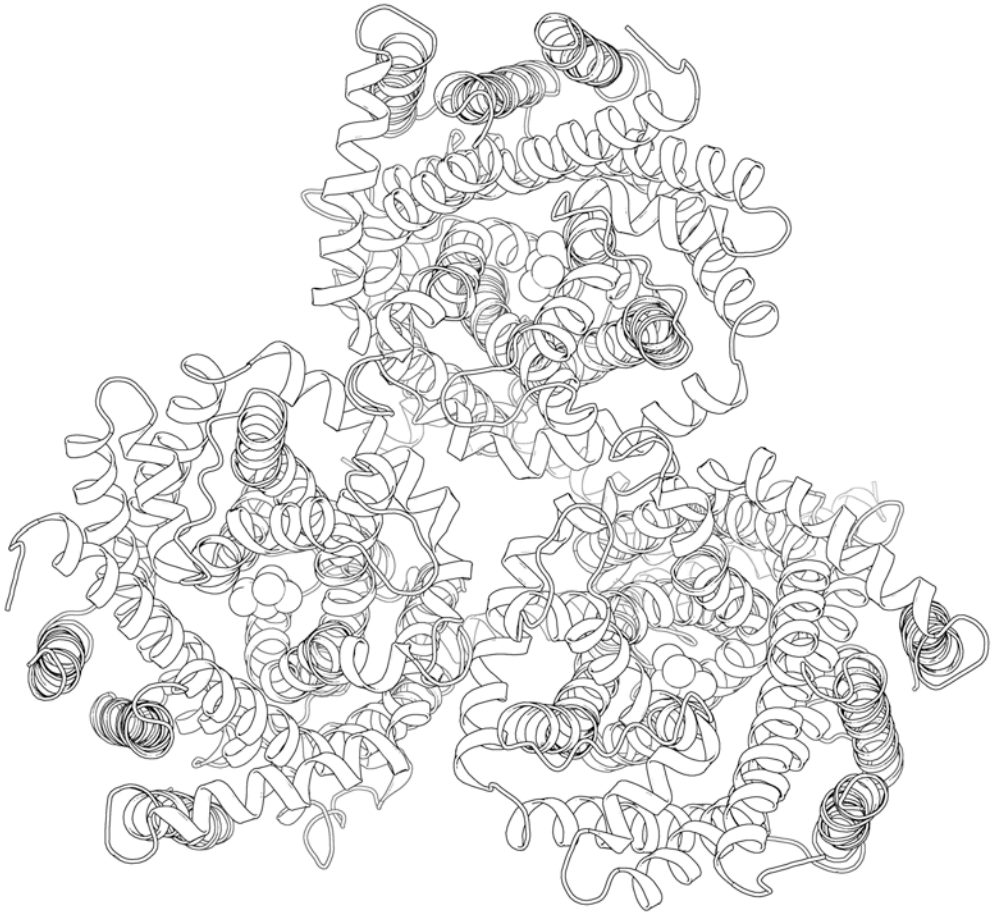
Diese Arbeit wurde in der Abteilung Strukturbiologie des Max-Planck-Instituts für Biophysik in Frankfurt am Main durchgeführt und vom Fachbereich 14 Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe Universität Frankfurt als Dissertation angenommen.

Diese Arbeit wurde von Dr. Christine Ziegler und Dr. Anke Terwisscha van Scheltinga betreut.

Dekan: Prof. Dr. Dieter Steinhilber
1. Gutachter: Prof. Clemens Glaubitz
2. Gutachter: Prof. Werner Kühlbrandt

Datum der Disputation:

Ein Gelehrter in seinem Laboratorium ist nicht nur ein Techniker;
er steht auch vor den Naturgesetzen
wie ein Kind vor der Märchenwelt.
Marie Curie, *07.11.1867 - †04.07.1934



** Pro moje milí rodiče **

Parts of this thesis have been published in the following articles:

Ressl, S., Terwisscha van Scheltinga, A.C., Vonrhein, C., Ott, V. & Ziegler, C.: Molecular basis of transport and regulation in the Na+/betaine symporter BetP. (2009) Nature, 458, 47-52

All relevant coordinates and structure factors have been deposited in the RCSB Protein Data Bank under the access code:

2WIT Crystal structure of the sodium-coupled glycine-betaine symporter BetP from *Corynebacterium glutamicum* with bound substrate

TABLE OF CONTENTS

ABSTRACT

Cellular membranes are important sites of interaction between cells and their environment. Among the multitude of macromolecular complexes embedded in these membranes, transporters play a particularly important role. These integral membrane proteins perform a number of vital functions that enable cell adaptation to changing environmental conditions.

Osmotic stress is a major external stimulus for cells. Bacteria are frequently exposed to either hyperosmotic or hypoosmotic stress. Typical conditions for soil bacteria, such as *Corynebacterium glutamicum*, vary between dryness and sudden rainfall. Physical stimuli caused by osmotic stress have to be sensed and used to activate appropriate response mechanisms.

Hypoosmotic stress causes immediate and uncontrolled influx of water. Cells counteract by instantly opening mechanosensitive channels, which act as emergency valves leading to fast efflux of small solutes out of the cell, therebydiminishing the osmotic gradient across the cell membrane.

Hyperosmotic stress, on the other hand, results in water efflux. This is counterbalanced by an accumulation of small, osmotically active solutes in the cytoplasm, the so-called compatible solutes. They comprise a large variety of substances, including amino acids (proline), amino acid derivatives (betaine, ectoine), oligosaccharides (trehalose), and heterosides (glucosylglycerol). Osmoregulated transporters sense intracellular osmotic pressure and respond to hyperosmotic stress by facilitating the inward translocation of compatible solutes across the cell membrane, to restore normal hydration levels.

This work presents the first X-ray structure of a member of the Betaine-Choline-Carnitine-Transporter (BCCT) family, BetP. This Na^+-coupled symporter from *Corynebacterium glutamicum* is a highly effective osmoregulated and specific uptake system for glycine-betaine.

X-ray structure determination was achieved using single wavelength anomalous dispersion (SAD) of selenium atoms. Selenium was incorporated into the protein during its expression in methione auxotrophic *E. coli* cells, grown in media supplemented with selenomethionine. SAD data with anomalous signal up to 5 Å led to the detection of 39 selenium sites, which were used to calculate the initial electron

density map of the protein. Medium resolution and high data anisotropy made the structure determination of BetP a challenging task. A specific strategy for data anisotropy correction and a combination of various crystallographic programs were necessary to obtain an interpretable electron density map suitable for model building.

The crystal structure of BetP shows a trimer with glycine-betaine bound in a three-fold cation-π interaction built by conserved tryptophan residues. The bound substrate is occluded from both sides of the membrane and aromatic side chains line its transport pathway. Very interestingly, the structure reveals that the α-helical C-terminal domain, for which a chemo- and osmosensory function was elucidated by biochemical methods, interacts with cytoplasmic loops of an adjacent monomer. These unexpected monomer-monomer interactions are thought to be crucial for the activation mechanism of BetP, and a new atomic model combing biochemical results with the crystal structure is proposed.

BetP is shown to have the same overall fold as three unrelated Na^+-coupled symporters. While these were crystallised in either the outward- or inward-facing conformation, BetP reveals a unique intermediate state, opening new perspectives on the alternating access mechanism of transport.

ZUSAMMENFASSUNG

Transportsysteme der Biomembran

Der Glycinbetain Transporter BetP, der Gegenstand dieser Arbeit ist, gehört zur Gruppe der membranständigen Transporterproteine. Membrantransporter werden nach der Eigenschaft ihrer Energiekopplungsmechanismen als Primärtransporter und Sekundärtransporter klassifiziert. Der Hauptanteil der Primärtransporter gehört zur ATP-Binde-Cassette Superfamilie (van der Does and Tampe, 2004) und der P-Typ ATPasen (Apell, 2004), welche beide die ATP Hydrolyse als treibende Kraft nutzen.

Sekundärtransporter hingegen nutzen entweder Ionen- oder Solutegradienten als treibende Kraft für ihren Translokationsmechanismus über die Membran hinweg. Sie stellen eine große und vielseitige Gruppe von Membranproteinen dar, deren Substratdiversität sich in über 100 Familien widerspiegelt. Sekundärtransporter sind sowohl bei Pro- als auch bei Eukaryoten in viele physiologische Prozesse involviert. In Prokaryoten spielen sie bei der Aufnahme von osmoprotektiven Substanzen (kompatible Solute) aus der Umgebung nach osmotischem Stress eine bedeutende Rolle (Poolman et al., 2002). Die meisten osmoreaktiven Sekundärtransporter gehören der *major facilitator superfamlily* (MFS) an, wie beispielsweise ProP aus *E. coli* (Culham et al., 2003; Racher et al., 2001). Eine weitere Familie stellt die *sodium-solute symporter family* (SSF) mit beispielsweise OpuE aus *B. subtilis* dar (von Blohn et al., 1997). Das am besten untersuchte Mitglied der Familie der *betaine-carnitine-choline transporter* (BCCT) ist der osmoregulierte Glycinbetain Symporter BetP aus *Corynebacterium glutamicum* (Farwick et al., 1995). *C. glutamicum* ist ein nicht-sporulierendes, Gram-positives Bodenbakterium. Innerhalb der Gram-positiven Bakterien zählt es in der großen Untergruppe der GC-reichen Bakterien zu den Actinomyceten (Abe et al., 1967).

Reaktionen der Zelle auf osmotischen Stress

C. glutamicum ist durch wechselnde Umweltbedingungen ständig einer Vielzahl verschiedenartiger Stressfaktoren ausgesetzt. Eine der am häufigsten auftretenden Stresssituationen sind Schwankungen in der Osmolarität des umgebenden Mediums. So sind die Zellen ständig der Gefahr ausgesetzt zu platzen oder auszutrocknen. Um

die Integrität der Zelle zu bewahren, haben prokaryotische Zellen wirksame Strategien entwickelt osmotischem Stress rasch entgegenzuwirken.

Bei einem hyperosmotischen Schock führt die hohe Osmolarität des externen Mediums zum Ausstrom von intrazellulärem Wasser und die Zelle plasmolysiert. Die Zelle reagiert mit Erhöhung ihrer zytoplasmatischen Osmolarität durch Akkumulation kompatibler Solute. Kompatible Solute sind als niedermolekulare Substanzen definiert, die in sehr hohen Konzentrationen im Zytoplasma akkumuliert werden können, ohne dabei die Physiologie der Zelle zu stören. Kompatible Solute sind hochlöslich, polar und tragen bei physiologischem pH-Wert bis auf wenige Ausnahmen keine Nettoladung. Im Gegensatz zu *chaotropen* Molekülen (z. B. Harnstoff) sind die *kosmotropen* kompatiblen Solute in der Lage, die native Proteinstruktur zu stabilisieren, indem sie von der Hydrathülle der Polypeptide ausgeschlossen werden und so die Wechselwirkung zwischen den Wassermolekülen und der Proteinoberfläche weiter verstärken (Arakawa and Timasheff, 1985; Wood, 1999).

Die Stressantwort teilt sich in mehrere Phasen auf. Nach dem Efflux von Wasser ist die erste schnelle Reaktion die Aufnahme von Kalium-Ionen, die sowohl in Gram-positiven als auch in Gram-negativen Bakterien beobachtet wird. Da die Kaliumaufnahme zu einer stark erhöhten Ionenkonzentration im Cytoplasma führt, welche unvorteilhaft für die Aufrechterhaltung nativer Proteinkonformationen ist, werden die K^+-Ionen im Folgenden durch neutrale kompatible Solute ersetzt. Dies kann durch *de novo* Synthese geschehen, oder - falls verfügbar - durch Aufnahme aus dem umgebenden Medium. *C. glutamicum* ist in der Lage Prolin, Glutamat, Glutamin und Trehalose zu synthetisieren (Ronsch et al., 2003; Wolf et al., 2003). Schneller und energetisch günstiger als die *de novo* Synthese ist die Aufnahme von kompatiblen Soluten aus dem Medium. Hierzu sind Bakterien mit verschiedenen osmoregulierten Aufnahmesystemen mit unterschiedlichen Substratspektren ausgestattet.

Das Aufnahmesystem für Glycinbetain:

Die Betain Permease, BetP aus *Corynebacterium glutamicum*

Der Glycinbetain Transporter BetP (Betain Permease) aus *C. glutamicum* besteht aus 595 Aminosäuren und besitzt ein Molekulargewicht von 64,2 kDa (Peter et al., 1996). Computergestützte Sekundärstrukturanalysen sagen für den Transporter

zwölf Transmembranhelizes sowie ~60 Aminosäuren lange zytoplasmatische Extensionen am N- und C-Terminus voraus (Nicklisch, 2008).

Als Mitglied der BCCT-Familie besitzt BetP ein hoch konserviertes Motiv in der achten Transmembranhelix, für das eine Beteiligung an der Substratbindung angenommen wird (Kappes et al., 1996; Peter et al., 1996; Saier, 2000). BetP transportiert ausschließlich Glycinbetain (Betain) und zeigt eine sehr hohe Substrataffinität (Betain K_m: 8,6 µM), sowie einen hohen V_{max} Wert von bis zu 110 nmol/min*mg TG. Zudem ermöglicht BetP die Akkumulation von Betain bis zu einem extrem hohen Gradienten von $4x\ 10^6$ (intern/extern). Der Transport von Betain ist an das elektrochemische Na^+-Potential gekoppelt und wird durch den Symport von 2 Na^+-Ionen begleitet (Farwick et al., 1995).

Nach einem hyperosmotischen Schock geht BetP in weniger als einer Sekunde in den aktivierten Zustand über. Der Schwellenwert zur BetP-Aktivierung liegt hierbei in *C. glutamicum* bei 300-400 mosmol/kg. Unterhalb dieses Wertes ist BetP inaktiv. Seine maximale Aktivität erreicht BetP bei 1200-1300 mosmol/kg (Peter et al., 1996). Wenn der hyperosmotische Stress durch die akkumulierten kompatiblen Solute kompensiert ist, wird die BetP-Aktivität reduziert, es kommt zur Aktivitätsanpassung, dieser im Detail noch wenig erforschte Vorgang wird als "activity adaptation" bezeichnet.

Mit Hilfe des artifiziellen Proteoliposomen Messsystems konnte gezeigt werden, dass die erhöhte Konzentration an K^+-Ionen der aktivierende Stimulus in diesem System für BetP ist (Rubenhagen et al., 2001). Demnach wurde dem Transporter die Funktionen eines Osmosensors und eines Chemosensors zugeschrieben (Rubenhagen et al., 2001; Schiller et al., 2004a). Mögliche Sensorbereiche liegen voraussichtlich in zytoplasmatischen Bereichen des Transporters. Die Deletion von 25 bzw. 45 Aminosäuren der zytoplasmatischen C-terminalen Domäne führte zu einer konstitutiven BetP Aktivität (Peter et al., 1996). Dies belegt eine direkte oder indirekte Beteiligung der C-terminalen Domäne an die Stimulusdetektion. Ohne hyperosmotischen Stimulus scheint die Domäne dafür verantwortlich zu sein, BetP in seiner inaktiven Konformation zu halten. Neuste Ergebnisse durch ortsgerichtete Mutagenese innerhalb der putativen, C-terminalen α-helikalen Sensordomäne haben gezeigt, dass nicht einzelne Aminosäurereste, sondern vielmehr die korrekte Konformation oder dessen relative räumliche Orientierung für

die Regulation von BetP entscheidend ist (Ott et al., 2008). Mittels Oberflächenplasmonresonanz-Spektroskopie wurde die Interaktion verschiedener C-terminaler BetP-Varianten in verschiedenen Membranumgebungen untersucht und führte zum Ergebnis, dass die Konformation der α-helikalen C-Domäne über Interaktionen mit negativ geladenen Membranlipiden stabilisiert wird. Ergänzende biochemische Methoden führten zur Identifizierung weiterer zytoplasmatischer Proteinbereiche, welche mit der C-terminalen Domäne interagieren können (Ott et al., 2008). Folgende Schlussfolgerungen konnten gezogen werde: a) die C-terminale Domäne interagiert mit der Lipidoberfläche; b) unterschiedliche Membranumgebungen nehmen Einfluss auf mögliche strukturstabilisierende Effekte; c) die Konformation oder relative Orientierung der C-terminalen Domäne sind wichtig für den Aktivierungsprozess; d) biochemische Untersuchungen haben proteininterne Interaktionen belegt.

Wird BetP heterolog in *E. coli* exprimiert, so zeigt sich zwar ein grundsätzlich ähnliches Aktivierungsprofil wie in *C. glutamicum*, das Aktivitätsoptimum tritt jedoch schon bei 600-800 mosmolol/kg auf (Peter et al., 1996). Durch Messungen im Proteoliposomensystem konnte belegt werden, dass die beobachtete Optimum-Verschiebung in *E. coli* auf die unterschiedlichen Membranzusammensetzungen von *E. coli* und *C. glutamicum* zurückzuführen ist. Während die *E. coli* Membran zu 80% aus dem ungeladenen Phosphatidylethanolamin besteht, hat die *C. glutamicum* Membran einen Anteil von 87% an negativ geladenem Phosphatidylglycerin (Ozcan et al., 2007). So wurde bei einem höheren Anteil an negativ geladenen Lipiden auch eine höhere K^+-Konzentration zur BetP Aktivierung beobachtet.

Röntgenkristallographische Strukturanalyse von BetP

Diese Arbeit präsentiert die erste atomare Struktur eines Transporters der BCCT Familie. Anormale Röntgenstreuung (single wavelength anomalous dispersion, SAD) von Selen Atomen wurde zur Phasenbestimmung verwendet. Mittels Zugabe von Selenomethionin in das Medium wurde Selen während der Proteinbiosynthese in BetP eingebaut. Der SAD Datensatz zeigte anormales Signal bis 5Å und führte zur Ermittlung der Selen Koordinaten, welche die anschließende Berechnung der Proteinelektronendichte ermöglichte. Der Auflösungsbereich von 3.35Å und die hohe Anisotropie des Datensatzes erschwerte die Strukturaufklärung in beachtlichem

Maße. Deshalb war die in dieser Arbeit angewendete Strategie und Kombination verschiedener Programme nötig, um am Ende zu einer interpretierbaren Elektronendichte zu gelangen, in welche die BetP Struktur modelliert werden konnte.

Die Kristallstruktur von BetP zeigt ein Trimer mit jeweils einem gebundenen Betain pro Monomer. Dabei ist das Substrat von keiner Seite der Membran zugänglich, und folglich liegt der Transporter in einer geschlossenen Konformation vor. Des Weiteren ist Betain über Kationen-π Interaktionen zwischen konservierten Tryptophanen gebunden. In der Kristallstruktur ist die C-terminale Domäne, welcher eine zentrale regulatorische Funktion zugeschrieben wird, fast vollständig aufgelöst. Die Kristallstruktur zeigt demnach zum ersten Mal, auf welche Weise die C-terminale Domäne mit zytoplasmatischen Regionen des Transporters interagiert. Dies ermöglichte eine atomare Beschreibung des Aktivierungsprozesses von BetP. Das atomare Aktivierungsmodell kombiniert biochemische Ergebnisse mit der trimeren Architektur des Tranporters und weist auf äußerst wichtige Interaktionsregionen hin. Die Kernaussage des Aktivierungsmodells basiert auf der Detektion des osmotischen Signals über die C-terminale Domäne und die Übermittlung dieses Signals zur Substrattransportregion des jeweils benachbarten BetP Monomers.

BetP weist eine ähnliche Faltung wie drei nicht miteinander verwandte Na^+-gekoppelte Symporter, $LeuT_{Aa}$, Mhp1 und vSGLT, auf. Während diese in der jeweilig nach außen offenen (outward-facing) oder innen offenen (inward-facing) Transporterkonformation kristallisierten, weist BetP eine intermediäre Konformation auf. Innerhalb der Strukturvergleiche dieser vier Transporter, welche die Gemeinsamkeit eines invertierten fünf-plus-fünf Transmembranhelixmotifs tragen, steuerte die Strukturaufklärung von BetP zum weiteren Verständnis des "alternating access" Transportmechanismus bei. Zusätzlich zu dem bereits beschriebenen wechselseitigen Kippmechanismus des fünf-plus-fünf Transmembranhelixmotifs, konnte mittels der BetP Struktur eine zusätzliche Rotationsbewegung beschrieben werden. Die Rotationsbewegung kann als eine Irisbewegung beschrieben werden, die zum Schließen oder Öffnen der periplasmatischen oder zytoplasmatischen Seite des Transportes führt.

Untersuchungen mittels analytischer Ultrazentrifugation, Gefrierbruch (*Freeze-fracture*) und Elektronenmikroskopie von 2D-Kristallen und der daraus resultierenden 7.5 Å Projektionsmappe haben gezeigt, dass BetP als Trimer vorliegt (Ziegler et al., 2004). Unterschiede in der Elektronendichteverteilung zwischen den einzelnen Monomeren des BetP Trimers weisen auf ein asymmetrisches Trimer hin. Eine darauf resultierende 3D-Elektronendichtekarte welche mittels kryo-elektronenmikroskopischen Aufnahmen von 2D-Kristallen unter verschiedenen Kippwinkeln erstellt wurde, zeigte eine noch deutlichere Asymmetrie des BetP Trimers (Tsai, 2008). Mit Hilfe der Kristallstruktur von BetP konnte die 3D Elektronendichte der kryoelektromikroskopischen (EM) Daten interpretiert werden. Die Interpretation der asymmetrischen Anordung des Trimers in der 3D EM-Mappe deutet auf einen kooperativen Transportmechanismus innerhalb des Trimers hin. Anhand der Kristallstruktur können spezifische Bereiche wie die amphiphatische und horizontal liegende Helix 7 als mögliche kooperative Kontaktpunkte angesprochen werden. Demnach weist die trimere Architektur des BetP Transporters möglicherweise auf funktionelle Kopplungsmechanismen zwischen den Monomeren innerhalb des Aktivierungs- und Transportprozesses hin.

Mit Hilfer der vorliegenden Kristallstrukur von BetP sind weitere Erkenntnisse auf dem Gebiet der Substrat- und Ionenbindung der Kopplungsmechanismen sowie des Transportmechanismus möglich. Die Kristallstruktur ermöglicht nun gezieltere ortsspezifische Mutagenesexperimente, welche nicht nur zum weiteren Verständnisses über BetP selbst, sondern auch über weitere Mitglieder der BCC-Transporter Familie führen könnten.

ABBREVIATIONS

2D	two-dimensional
3D	three-dimensional
AHT	anhydrotetracyclin
BCCT	betaine/carnitine/choline transporter
C-terminal	carboxyl-terminus of an amino acid sequence
CL	cardiolipin
CMC	critical micelle concentration
CV	column volume
Cy-5	cymal-5
DDM	β-dodecyl maltoside
dw	dry weight
EM	electron microscope
FFT	fast Fourier transformation
HABA	4-hydroxy-azobenzen-2-carboxylic acid
kDa	kilodalton
K_m	Michaelis-Menten-constant
Kd	dissociation constant
LSSR	Local Structure Similarity Restraints
MAD	multiwavelength anomalous diffraction
MR	molecular replacement
MWCO	molecular weight cutoff
NCS	non-crystallographic symmetry
N-terminal	amino-terminus of an amino acid sequence
OD600	optical density at 600 nm
PAGE	polyacrylamide-gel-electrophoresis
PE	phosphatidyl ethanolamine
PEG-X	polyethylene glycol with a mean molecular mass of X g/mol
PG	phosphatidyl glycerol
R_{cryst}	crystallographic *R* factor
R_{free}	„free“ crystallographic *R* factor
R_{meas}	redundancy-weighted *R* factor for symmetry-related intensities

r.m.s.d.	root-mean square deviation
RT	room temperature
SAD	single wavelength anomalous dispersion
SDS	sodium dodecylsulfate
SeMet	*L*-selenomethionine
SIRAS	single isomorphous replacement with anomalous scattering
TLC	thin-layer chromatography
TM	transmembrane
TRIS	2-amino-hydroxymethylpropane-1,3-diol
v/v	volume per volume
V_{max}	maximum velocity
w/v	weight per volume
WT	wild type

1. Introduction

Membranes are the site of cellular entry and exit of solutes and proteins, and a site of signal transduction. Thus, they represent a prime target for the regulation of cell physiology. Rather than being merely a lipid matrix in which the proteins responsible for these activities float randomly and independently, complex processes involving several components seem to be efficiently organized at and in cell membranes. In addition to roles in transport, signal and energy transduction, integral membrane proteins are important for retaining membrane architecture, and oligomerisation of membrane proteins provides a possible mechanism for the specific regulation and stability of membrane structure (Strauss et al., 2008; Van Dort et al., 2001). Presumably for space-economy reasons, the majority of membrane proteins could form oligomers. Protein co-localisation and defined orientation in the two-dimensional space of the membrane enhance the likelihood of self-association (Grasberger et al., 1986).

In terms of energy-coupling mechanisms, two classes of membrane transporters, comprising the primary and secondary transport systems, are abundant in all known species of eukarya, bacteria and archaea. Primary transport systems convert light or chemical energy into electrochemical energy, such as solute concentration gradients across membranes. The majority of primary solute-transport systems belong to the ATP-binding cassette (ABC) superfamily. Secondary transport systems use the free energy difference that is stored in the electrochemical gradients of protons, Na^+ ions or other solutes across the membrane to drive translocation reactions. Depending on the direction of transport and whether a coupling ion is used, the systems are termed symporters, antiporters and uniporters. Symporters transport a substrate into the cell (or organelle) against its concentration gradient by co-transport with another solute, most often an H^+ or Na^+ ion, whereas antiporters use the coupling ion (solute) in a counter transport mechanism. Uniporters can be either channels or carrier proteins that facilitate free flow of specific molecules with the solute gradient by response to a stimulus.

1.1. Active transport of solutes across a membrane

Active transport enables a cell to accumulate solutes against electrical and/or concentration gradients by making use of energy sources to promote their thermodynamically "uphill" movement. Transport of molecules across the lipid bilayer is obviously an important function of membrane proteins and uses a variety of mechanisms.

Transport proteins (transporters) are required for all cells to take up or dispose of nutrition or waste compounds, respectively. In eukaryotic cells, transporters mediate the flux of metabolites between intracellular compartments. In the cell envelope of Gram-negative bacteria, active transport across the outer membrane is facilitated by porins that contain water-filled channels with varying specificity for solutes. Transporters of the inner membrane tend to be highly specific for their substrate or a certain class of substrates.

Transmembrane sodium-ion gradients provide energy that is used by secondary transporters to drive the translocation of solute molecules into the cell. Sodium-coupled transporters are involved in many physiological processes and several crystal structures of sodium-coupled transporters have been reported: the Na^+/alanine transporter $LeuT_{Aa}$ from *Aquifex aeolicus* of the neurotransmitter sodium symporter (NSS) family (Yamashita et al., 2005), the Na^+/galactose transporter vSGLT from *Vibrio parahaemolyticus* of the sodium solute symporter (SSS) family (Faham et al., 2008) and the benzyl-hydantoin symporter Mhp1 from *Microbacterium liquefaciens* of the nucleobase cation symport 1 (NCS-1) family (Weyand et al., 2008).

1.2. Fluctuations in the environment – osmotic stress

Fluctuations of external osmolarity are one of the most basic types of environmental stress factors affecting prokaryotic and eukaryotic cells. Bacterial cells have developed a number of strategies to counteract osmotic stress to overcome the damaging effects of hypo- and hyperosmotic conditions in their environment. Hypoosmotic stress leads to an immediate and uncontrolled influx of water, whereas hyperosmotic stress results in water efflux from the cell. For both of these stress situations cells must rapidly sense and react to re-establish their osmotic equilibrium. During hypoosmotic stress an opening of mechanosensitive channels act as

emergency valves and facilitate rapid efflux of small solutes out of the cell thereby relieving the physical stress situation.

As a response to hyperosmotic stress, compatible solutes are transported and accumulated in the cytoplasm. Compatible solutes are of polar, highly soluble and often zwitterionic chemical nature, such as betaine (glycine-betaine), amino acids (proline), amino acid derivatives (ectoine) and polyols (glycerol, trehalose). They can all be accumulated to very high concentrations in the cytoplasm without interfering with normal physiological functions of the cell. The accumulation is achieved either by biosynthesis or by uptake. The latter process is faster and more favourable for the cell, in terms of energy and carbon cost. These processes have to be effectively regulated in their response to osmotic conditions to guarantee an optimally adapted response to varying hyperosmotic stress; this process is called osmoregulation.

Both, primary and secondary mechanisms for osmoregulated uptake of compatible solutes have been studied in detail in various Gram-negative and Gram-positive bacteria (Morbach and Kramer, 2002; Morbach and Kramer, 2004; Poolman et al., 2002; Wood, 1999). The main purpose of these systems is to respond to hyperosmotic stress by regulation on the level of both gene expression and protein activity. In *Corynebacterium glutamicum*, a gram-positive soil bacterium, many uptake carriers for compatible solutes have been identified and characterised. The most important and thus far best-studied system in *C. glutamicum* is the glycine-betaine carrier BetP.

How a transporter senses osmotic stress and transduces the signal into an activity change represents one of the major challenges in the field of cellular osmoregulation.

1.3. The Na^+-coupled glycine-betaine BetP from *Corynebacterium glutamicum*

BetP is a member of the BCCT family of secondary transporters. It consists of 595 amino acids and is predicted to span the membrane twelve times. Furthermore it contains two hydrophilic domains of ~60 amino acids at the N- and the C-terminal ends, both facing the cytoplasm (Rubenhagen et al., 2001). Information about the oligomeric properties of BetP comes from by analytical ultracentrifugation studies and a 7.5Å projection map of BetP, which was obtained by electron cryo-microscopy

of 2D crystals. Both indicate BetP is a trimer in its functional and native state (Ziegler et al., 2004).

The functional properties of BetP can be summerised as follows:
(i) The catalytic function is betaine transport driven by the electrochemical Na^+ potential; (ii) the regulatory function refers to its ability to efficiently adapt its catalytic activity during hyperosmotic stress; (iii) the sensory function is its ability to directly sense the extent of osmotic stress without the help of additional proteins or cofactors.

1.3.1. Electrochemical properties of BetP

The uptake of glycine-betaine by BetP is coupled to the electrochemical Na^+ potential, which drives the influx of betaine by the co-transport of two Na^+ ions (Farwick et al., 1995). Extremely high steady-state accumulation ratios of up to 4 x 10^6 (internal/external concentration) have been determined, and the transporter is highly specific for glycine-betaine. The affinity of BetP for glycine-betaine in *C. glutamicum* is high (K_m of 8.6 μM), whereas the Na^+ affinity is low (K_m of 4 mM). BetP is, together with acetate and glucose uptake, one of the fastest transporters in *C. glutamicum*, characterised by V_{max} values up to 110 nmol/min*mg(dw) (Farwick et al., 1995). This high activity of BetP makes, together with the high abundance of glycine-betaine in soil, glycine-betaine uptake dominant in comparison to the other osmoregulated carriers in *C. glutamicum*.

1.3.2. Osmoregulatory properties of BetP

The accumulation of glycine-betaine in the cell must be adapted to the cells environmental circumstances. Glycine-betaine accumulation depends thus on the internal and external osmolality and consequently on the turgor pressure of the cell. BetP has been measured to respond instantly to osmotic stress and its activity adapts to the actual hyperosmotic stress level. BetP is practically inactive in the absence of hyperosmotic stress or presence of hypoosmotic stress. However, upon an osmotic upshift, BetP becomes activated in less than 1 second (Figure 1). If the hyperosmotic stress is compensated by accumulation of glycine-betaine, the activity is reduced and

therefore adapted to prevent an excess of internal solute accumulation (Kramer and Morbach, 2004).

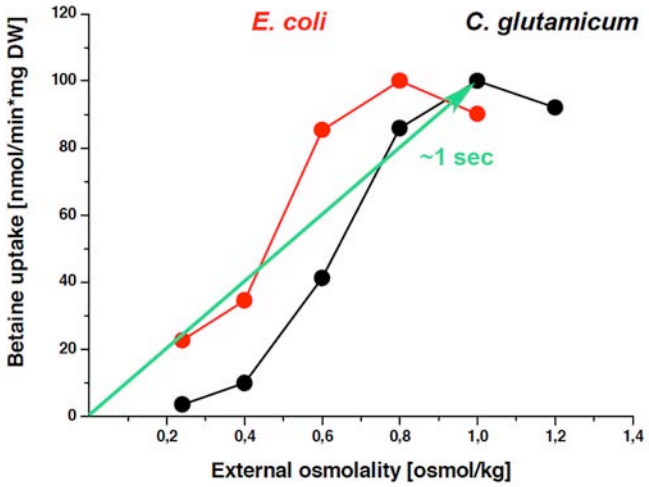


Figure 1|Activation profile of BetP by hyperosmotic stress. BetP was expressed and betaine uptake measured in *C. glutamicum* cells (black curve) and in *E. coli* cells (red curve) upon hyperosmotic stress simulation adjusted with NaCl in the external buffer (Ott, 2008). The green arrow indicates the fast activation time of ~1 second from about no transport to the maximum transport rate in *C. glutamicum* cells.

The threshold of activation is observed at about 300 – 400 mosmol/kg, and BetP reaches an optimum activity around 1300 mosmol/kg (Figure 1). The membrane cell wall of *E. coli* is composed to 80% of neutral charged lipids (Peter et al., 1996) and of *C. glutamicum* to 87% of negative charged phosphatidylglycerol (Ozcan et al., 2007). The activation profile of BetP is very similar when measured in intact cells of *C. glutamicum* (Figure 1) but with a shift towards lower osmolalities in *E. coli* cells (Figure 1). A similar activation profile was measured in the proteoliposome system (Rubenhagen et al., 2000). Based on experiments in proteoliposomes this shift in activation level towards lower osmolalities was found to depend on the different phospholipid composition of the membrane (Rubenhagen et al., 2000).

Up to now properties of BetP have been shown to be an autonomous sensory and regulatory system working independently of additional effectors. This is supported by its full functional competence observed both in *E. coli* and in proteoliposomes and all three properties, transport, osmoregulation and osmosensing can be addressed to characteristics of BetP. Furthermore, BetP senses the stimulus

responsible for osmoregulation in all systems, *C. glutamicum*, *E. coli* cells and proteoliposomes. In addition, differences in membrane composition resulted in a modulation of the BetP activation profile.

1.3.3. Osmosensing properties of BetP

During a hyperosmotic stress response, an osmosensor element transduces a signal to the transport protein, which results in an osmotic activation or activity adaptation of the transporter. BetP possesses all functions for signal transduction. Therefore, it should harbour at least one sensor domain, which reliably transduces an appropriate signal to the catalytic transport domain, thus regulating its activity. Possible stimuli sensed by bacterial cells are divided into four different categories (Wood, 1999). Those are stimuli directly from the external osmolality, ionic strength, or concentration of particular solutes. Similar parameters are proposed to be relevant at the cytoplasmic site too, because a change in internal water activity is the consequence of a change in external osmolality. This leads to molecular crowding of cytoplasmic macromolecules and might be one relevant signal. Furthermore, all membrane related parameters such as cell turgor and membrane strain may well be important for osmoregulated transporters. Finally, changes in the external osmolality might also directly influence soluble and membrane embedded proteins by changing their surface hydration and thus might result in conformational changes.

When reconstituted into proteoliposomes, BetP showed to be more effectively activated by an increase of the luminal monovalent cations K^+, Rb^+ and Cs^+. Several possible triggers were excluded from the experimental set up, such as changes of external solutes, internal solutes (except K^+, choline$^+$, $NH4^+$, Na^+) and membrane strain and cell turgor, because proteoliposomes are lacking turgor pressure. Na^+, NH_4^+ and choline$^+$ were shown to be less efficient in BetP activation and BetP was therefore concluded to be a K^+-specific chemosensor (Rubenhagen et al., 2001).

The N- and the C-terminal domains of BetP have been shown to influence activation strongly in intact *C. glutamicum* cells (Peter et al., 1998) (Figure 2). Truncation of the N-terminal domain did not significantly change the catalytic activity of BetP, however, the activation profile was shifted to higher osmolalities.

Truncations of the C-terminal domain have a more drastic effect. Truncation of 25 to 45 amino acids from the C-terminus result in a deregulation of the transporter. Truncation of just 12 C-terminal amino acids led only to a partial deregulation. These mutant forms were found to be constitutively active in intact *C. glutamicum* cells even in the absence of osmotic stress, although at one quarter reduced V_{max} (Kramer and Morbach, 2004).

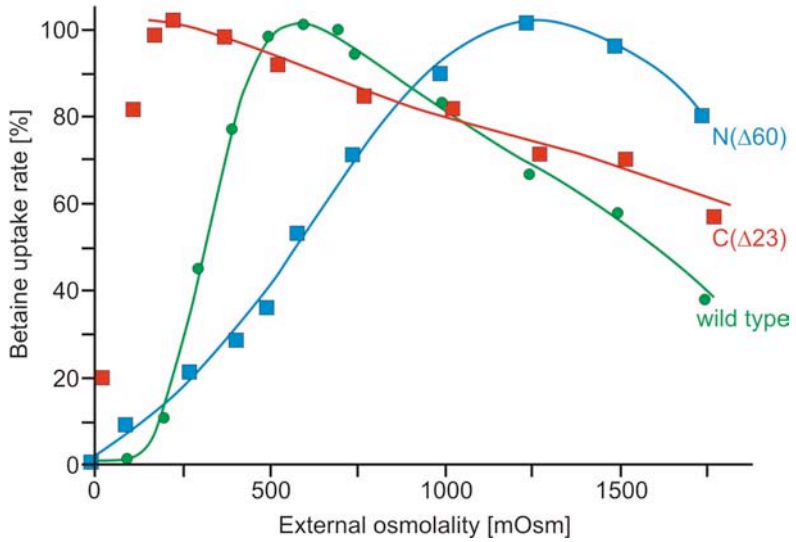


Figure 2| Consequence of N- and C-terminal truncations of BetP on transport activity. Glycine-betaine (betaine) uptake activity in wild type BetP *C. glutamicum* cells (yellow curve) increases in response to increasing hyperosmotic stress. Truncation of 60 amino acids at the N-terminal domain of BetP resulted in an activation profile that is shifted to higher osmolalities (blue curve). Truncation of 25 amino acids at the C-terminal domain leads to deregulation of BetP (red curve). Its increase at low Na^+ concentration is caused by the dependence on Na^+ as a co-substrate. (Adapted from http://www.kraemerlab.uni-koeln.de/osmosensing.php).

From these results it was concluded that the C-terminal domain is involved in osmosensing. Interestingly, they also indicate that the C-terminal domain contains an inhibitory element as well. Hence, this domain seems to be required to keep BetP in an inactive state in the absence of osmotic stress. Furthermore, this property was observed as being influenced by the identity of the lipids in which BetP was embedded (Schiller et al., 2006). By surface plasmon resonance spectroscopy (Ott et al., 2008) it was determined that the C-terminal domain interacts with the membrane surface. Lipid effects on osmodependent regulation have been observed with other

transporters as well, for example ProP of *E. coli* and OpuA of *L. lactis*. They respond with substrate transport at higher levels of hyperosmotic stress when the fraction of negatively charged phospholipids is increased in proteoliposomes or cells (Tsatskis et al., 2005; van der Heide et al., 2001).

Recent studies (Ott et al., 2008) have given new insights on the N-terminal domain of BetP, which in previous findings was already deemed to have a contribution to the regulation of BetP (Peter et al., 1998). Truncations of the N-terminal domain not only lead to a shift in sensitivity towards osmotic stress, but the surrounding lipids also affect the impact of N-terminal truncations on the regulatory properties of BetP (Ott et al., 2008). Since this was also observed in mutants that lacked the N-terminal domain completely, this was thought to have an indirect effect on BetP regulation. The regulatory influence of the N-terminal domain is therefore not based on direct peptide-membrane interaction but rather on protein-protein interaction with the opposing C-terminal domain, which in turn interacts directly with the membrane.

Based on these observations, a model for a molecular mechanism of the activation process of BetP was formulated (Figure 3) (Ott et al., 2008):

The membrane bound state of the C-terminal domain can be regarded as the inactive state, since an increased binding of the C-terminal domain to the membrane surface in the presence of negatively charged lipids made BetP activation increasingly difficult. This protein-lipid interaction of the C-terminal domain is an important feature of the inactivation of BetP and critically depends on the phospholipid charge composition of the membrane. The change from the inactive to the active state of the transporter is modulated by interaction changes of the C-terminal domain with part of loops at the cytoplasmic side (protein-protein interaction) as well as with the membrane (protein-lipid interaction). This is described as a functional switch model of BetP between the inactive and the active state depending on the location of the C-terminal domain.

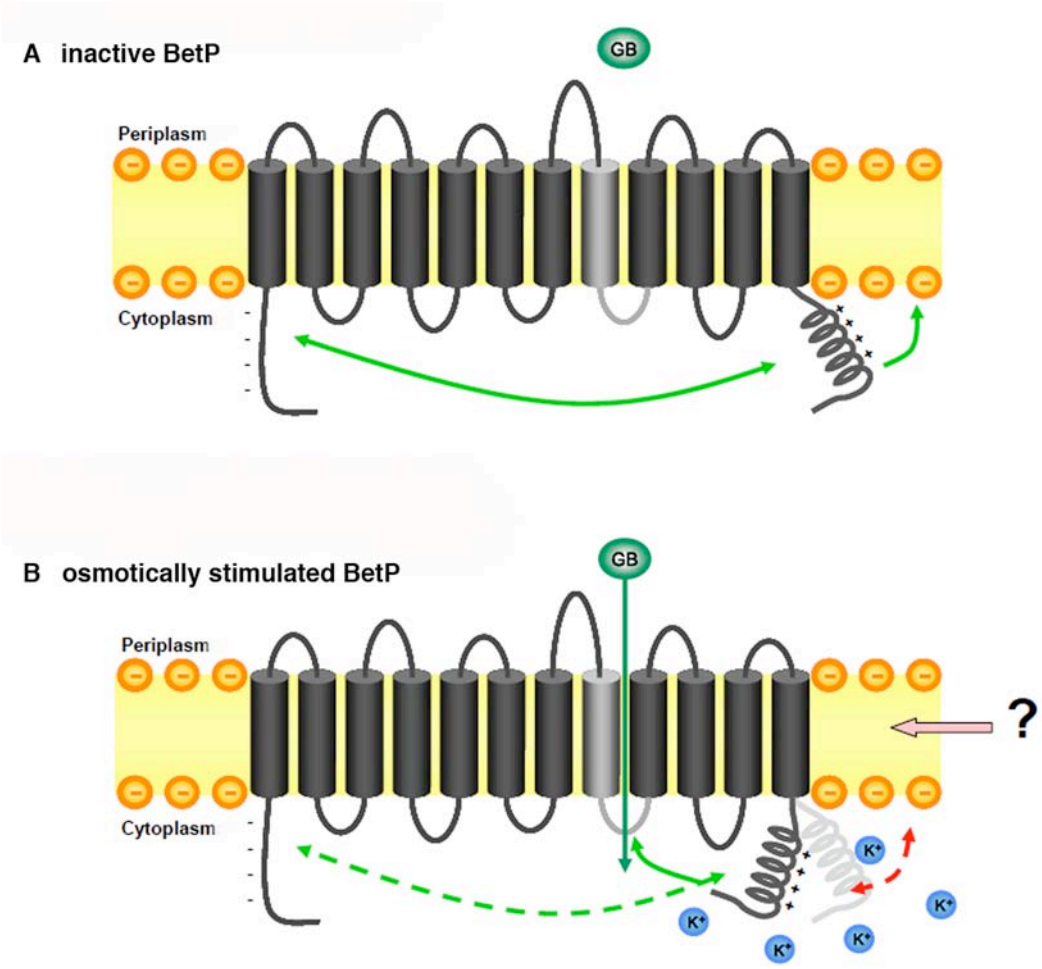


Figure 3| Model of the regulation mechanism of BetP. A, inactive BetP**.** The green arrows indicate the interaction between the N-terminal and C-terminal domain (protein-protein interaction) and the interaction of the C-terminal domain with the lipids of the membrane (protein-lipid interaction). **B,** osmotically stimulated BetP. An unknown stimulus, indicated with the pink arrow in the plane of the membrane, together with the increase in K^+ concentration in the cytoplasm lead to the activation of BetP. The direct or indirect sense of this K^+ stimulus by the C-terminal domain induces a conformational change that leads to other protein-protein interactions between the C-terminal domain and cytoplasmic loops, in particular loop 8 (light grey). The modulated interactions are indicated with dashed lines. (Adapted after (Ott, 2008)

Obviously this model cannot describe all aspects of BetP activation, since it has been shown that also the composition of the hydrophobic part of the surrounding membrane and/or its physical state seems to influence the activity state of BetP (Ozcan et al., 2007; Ozcan et al., 2005).

In contrast to the BetP activation mechanism after an osmotic upshift, the downregulation after the point of osmotic compensation is unclear, concerning the

stimuli and the responsible signal transduction. As the adapted state is not equivalent to the inactive/resting state before osmotic activation (Figure 1), stimuli such as K^+ and a change in external parameters were excluded. The regulatory mechanism might therefore be related to stimuli pertaining to the membrane properties, turgor or membrane strain.

1.3.4. Structural studies on BetP

That BetP is a trimer in its native state was ascertained by analytical ultracentrifugation. The first structural information for BetP was derived from a 7.5Å projection map, which was calculated from 2D crystals that were analysed by electron cryo-microscopy (Ziegler et al., 2004). BetP monomers were observed to associate as crystallographic dimers of trimers. In each monomer 10–12 transmembrane α-helices were distinguishable as well as pore-like features suggesting potential transport pathways. The projection map of wildtype BetP trimer appeared asymmetric. Although showing differences between the monomers, the projection map has been averaged (Figure 74a). Furthermore, a projection map of CaiT, a L-carnitine/γ-butyrobetaine antiporter of the BCCT family and an *E. coli* homologue to BetP, showed CaiT to be a trimer in 2D crystals, as well as by single particle microscopy (Vinothkumar et al., 2006).

A 3D electron density map was generated from a 2D crystal tilt series collected by cryo-electron microscopy of a BetP mutant lacking 45 C-terminal amino acids (ΔC45). The 3D electron density map is of 8Å resolution in the plane of the membrane and 16Å perpendicular to the membrane. This density allowed the assignment of helices but not of any sequence assignment, because loop regions were not resolved; and therefore the assigned 12 helices were not connectable to each other. The trimer in the 3D electron density map for the ΔC45 mutant appeared asymmetric as well; yet averaging of the monomers was not possible this time (Figure 74c).

Both sets of structural information show BetP as an asymmetric homotrimer proposing each of the three monomers representing a different state of the transporter.

1.4. Aim of this work

The aim of this thesis was to determine a high-resolution structure of the secondary transporter BetP from *Corynebacterium glutamicum* by X-ray crystallography. The only structural information available at the beginning of this work had been obtained by cryo-electron microscopy of two-dimensional crystals (Tsai, 2008; Ziegler et al., 2004). Both studies found BetP to be an asymmetric homotrimer in the membrane.

An N-terminally truncated and residue engineered mutant of BetP has been chosen for 3D crystallisation and structure determination. This mutant, ΔN29BetPE44/45/46A, is active at a higher osmolarity compared to wildtype BetP but is otherwise fully regulated and active in both, cells and proteoliposomes (Ott et al., 2008).

It is expected that an atomic model of BetP will help to answer many open questions about the function of the transporter, opened by the extensive amount of biochemical data available at present. Of particular interest is the architecture and function of the C-terminal domain: What does it look like? What interactions is it involved in? Can the recent molecular switch model of activation be confirmed or even described in more detail? Having a structure in hand will also greatly assist rationalising the trimeric architecture of BetP. If BetP can be crystallised with bound substrate, an atomic model may show where and how it is bound. Finally, an atomic model of BetP may give further insights into the global topic of sodium-coupled transport by contributing to the understanding of the alternating access mechanism of transport.

2. Materials

2.1. Detergents and phospholipids

Detergents (Anatrace, US-Maumee; Glycon, D-Luckenwalde), and phospholipids (Avanti, USA, Alabaster) used are summarised in the tables below.

Table 1| Detergents used in this work.

Detergent	Chemical name	critical micelle concentration (cmc) [%] in water
n-Dodecyl-β-D-Maltoside (DDM)	n-Dodecyl-β-D-Maltopyranoside	0.15
Cymal 5	5-Cyclohexyl-1-Pentyl-β-D-Maltoside	0.12
Cymal 4	4-Cyclohexyl-1-Pentyl-β-D-Maltoside	0.37
Cymal 3	3-Cyclohexyl-1-Pentyl-β-D-Maltoside	1.6
Cymal 2	2-Cyclohexyl-1-Pentyl-β-D-Maltoside	5.4

Table 2| Lipids used in this work.

Lipid	Chemical name	MW	FA chains	Chemical formula
POPG	1-Palmitoyl-2-Oleoyl-sn-Glycero-3-[Phospho-rac-(1-Glycerol)]	766.04	16:0	$C_{40}H_{80}NO_{10}P$

2.2. Bacteria and plasmids

The bacteria strains used in this work are listed in Table 3, and plasmids used are listed in Table 4.

Table 3| *Escherichia coli* strains used in this work.

E. coli strain	Genotypes	Reference
DH5α™-T1	F- Δ 80lacZ Δ M15 Δ (lacZYA-argF)U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44thi-1 gyrA96 relA1 tonA (confers resistance to phage T1)	(Killmann et al., 1996)
JM109	F`traD36 $lacI^q$ Δ(lacZ)M15 proA B / e14- (McrA-) Δ(lac-proAB) thi-1 gyrA96 (Nx^R) endA1 hsdR17 ($r^-_k m^-_k$)relA1 supE44 recA1	(Yanisch-Perron et al., 1985)
B834 (DE3)	F- ompT hsdS(r_B- m_B-) gal dcm met- (DE3)	(Wood, 1966)
BL21 RIL-X	F- ompT hsdS(r_B- m_B-) dcm- Tet' gal λ(DE3) endA metA::Tn5(kan') Hte [argU ileY leuW Cam']	Stratagene
BL21 RP-X	F- ompT hsdS(r_B- m_B-) dcm- Tet' gal λ(DE3) endA metA::Tn5(kan') Hte [argU proL Cam']	Stratagene
MKH13	araD39 (argF-lac) U169 relA51 rps150 flbB5301 deoC ptsF25 Δ(putPA)101 Δ(proP)2 Δ(proU)	(Haardt et al., 1995)

Table 4| Plasmids used in this work.

Plasmid	Resistance	Properties	Reference
pASK-IBA5 betP	Amp^R	pASK-IBA5 with betP cloned over the BsaI/HindIII restriction site	(Rubenhagen et al., 2000)
pAcl1 ΔN29	Amp^R	pAcl1 with betP cloned between the BsaI and HindIII restriction sites; 29 N-teminal amino acids deleted	(Ziegler, unpublished)
pASK-IBA7 ΔN29EEE44/45/46AAA	Amp^R	pASK-IBA7 with betP cloned between the BsaI and HindIII restriction sites; 29 N-teminal amino acids deleted and codon exchange at E44A, E45A and E46A,	(Ressl et al., 2009)

3. Methods

Standard methods used in this work will be described briefly, whereas methods of X-ray crystallography will be introduced and described in more detail.

3.1. Production, purification and characterisation of BetPΔN29EEE44/45/46AAA$_{StrepII}$

3.1.1. Competent cells and plasmid DNA transformation

The plasmid *pASK-IBA7-ΔN29EEE44/45/46AAA* was transformed into the competent *E. coli*-One Shot®Invitrogen DH5α™-T1 strain using the following protocol. 100µl of competent DH5α™-T1 (Invitrogen) cells was thawed on ice and 1.5µl of plasmid DNA was mixed and incubated for 30 minutes on ice. Cells were then heat shocked at 42°C for 60 seconds followed by incubation on ice for 10 minutes. Agar plates with 50 µg/ml carbenicillin were inoculated with 20 µl to 50 µl of transformed cells and were left overnight at 37°C in the incubator.

3.1.2. Cell growth and *betP* gene expression

For expression of the BetP mutant BetPΔN29EEE44/45/46AAA$_{StrepII}$ (BetA) in DH5α™-T1 cells, 50 ml of Luria Bertani (LB)-medium (10 g/l bacto trypton, 5 g/l bacto yeast extract, 10 g/l NaCl) (Sambrook et al., 1989) with 50 µg/ml carbenicillin were inoculated with several colonies of cells from one agar plate harbouring the plasmid *pASK-IBA7-ΔN29EEE44/45/46AAA*. The 50 ml preculture was grown for about five hours at 37°C with 200 rpm in a shaker. The cells were then sub cultured in 2 l of LB-medium 50 µg/ml with carbenicillin and grown at 37°C at 115 rpm to an OD_{600} of about 2. 50 ml of this culture was used to inoculate 2 l of 2YTx2S medium (16g/l bacto trypton, 10 g/l bacto yeast extract, 20 g/l NaCl). BetP synthesis is under the control of the tetracycline promoter and was induced with 200 µg/l anhydrotetracyclin (AHT) at the OD_{600} of about 1. After induction cells were grown either at 37°C for 3 hours or overnight at 25°C at 115 rpm.

3.1.3. Production of selenomethionine BetA

Selenomethionine-labelled BetA (SeMet-BetA) was produced by the pathway inhibition method (Doublie, 1997). In this method, methionine biosynthesis in the cell is halted by the addition of certain amino acids to the growth medium. Instead of methionine, L-selenomethionine is present in the medium, which is then taken up by the cell and incorporated into newly synthesized protein. Incorporation rates of up to 95% selenomethionine can be achieved using this method (Doublie, 1997).

The plasmid *pASK-IBA7-ΔN29EEE44/45/46AAA* was transformed into *E. coli* BL21 RIL-X cells (Stratagene) as described in Chapter 3.1.1. 5 ml LB medium with 50 μg/ml carbenicillin were inoculated with one colony from the agar plate and cells grew for seven hours at 37°C at 200 rpm. 100 ml SeMet MD medium with Nutrient Mix (both Molecular Dimensions) were inoculated with this 5 ml of preculture and 40 mg/l Methionine (Molecular Dimensions) and grown overnight at 30°C.

The cells were further cultured in 2 l *SelenoMet* media (Molecular Dimensions) containing 40 mg/l L-selenomethionine and protein was produced after induction with 200 μg/l anhydrotetracyclin (AHT) at an OD_{600} of about 0.7 for 6 h at 37°C.

3.1.4. Membrane preparation and solubilisation

Cells were harvested by centrifugation at 4200 rpm at 4°C for 15 min in a low-speed Beckman J6-MI (Avanti J20 XPI) centrifuge. The cell pellet was resuspended and homogenised in cold 100 mM Tris-Cl (pH 7.5) buffer and broken with a cell disrupter (Constant Systems) at 1.8-1.9 kbar. Broken cells were centrifuged at 12,500 rpm for 30 min at 4°C in a Beckman Avanti J-30 I centrifuge. The supernatant was further centrifuged at 45,000 rpm for one hour at 4°C using a Beckman Optima L-70K ultracentrifuge. The pellet containing the membrane was resuspended and homogenised using a potter in solubilisation buffer (50 mM Tris-Cl pH 7.5; 17.4 % glycerol) and adjusted to about 10 mg/ml of total protein content.

Membranes were solubilised in 2-3% β-D-dodecyl maltoside (DDM) for 40-50 minutes on ice. Selenomethionine samples were solubilised under a permanent nitrogen stream with 5 mM DTT and 0.5 mM EDTA added to the solubilisation buffer to reduce oxidation processes. After solubilisation an ultracentrifugation step of 50 minutes at 45,000 rpm at 4°C was used to separate the soluble fraction from the

insoluble parts of the membrane sample. The supernatant was diluted 1:4 with dilution buffer (50 mM Tris-Cl pH 7.5; 4.35 % glycerol; 266 mM NaCl).

3.1.5. Isolation of BetA with StrepTactin®-affinity chromatography

BetA was isolated by its N-terminal *Strep*-tag®II-peptide (NH2-WSHPQFEK-COOH; IBA, Göttingen, (Schmidt et al., 1996). The *Strep*-tag®II binds with high affinity to StrepTactin® (Voss and Skerra, 1997). This bond can be counteract with desthiobiotin which has a higher affinity to the StrepTactin® resin than the *Strep*-tag®II tag. The following steps were performed at 4°C. For SeMet-BetA purification 5 mM DTT and 5 mM EDTA were added to the subsequent steps to minimise oxidation.
The StrepTactin® resin was equilibrated with five column volumes (CV) of buffer A (50 mM Tris-Cl with pH 7.5, 8.7% glycerol, 200 mM NaCl) with 0.1% β-DDM. The usual StrepTactin® CV was between 1.5 and 3 ml. The diluted soluble fraction after the last ultracentrifugation step was loaded on the affinity column overnight with a flow rate of about 1 ml/min.

The following different washing steps, each representing a separate purification strategy, were applied to the sample after it was loaded on the StrepTactin® affinity column:

a) 1. Wash buffer A with 0.01% β-DDM and 500mM NaCl, 6 to 8 CV
   2. Wash buffer A with 0.6% Cymal-5, 6 to 8 CV

b) 1. Wash buffer A with 0.01% β-DDM and 500mM NaCl, 6 to 8 CV
   2. Wash with POPG 0.25 mg/ml in 0,05 % β-DDM, 6 to 8 CV
   3. Wash buffer A with 0.6% Cymal-5, 4 CV

c) 1. Wash buffer A with 0.01% β-DDM and 500mM NaCl, 6 to 8 CV
   2. Wash buffer A with 0.6% Cymal-5 overnight with very slow flow, 6 to 8 CV

d) 1. Wash buffer A with 0.1% β-DDM and 500mM NaCl, 6 to 8 CV
   2. Wash buffer A with 0.6% Cymal-5, 200 mM NaCl, 6 to 8 CV

BetA was eluted adding 5mM of desthiobiotin to buffer A with 0.6% Cymal-5. Fractions of 200 to 500 µl were collected, and the protein content of each elution was checked with the Bradford assay (Bradford, 1976).

3.1.6. Size exclusion chromatography

For a pure and monodisperse sample for 3D crystallisation, size exclusion chromatography (SEC) was introduced as a further purification step.
A Superose 6 10/300 column (Amersham-Pharmacia) was connected to a ÄKTA system (GE Healthcare) and equilibrated with gel filtration buffer (20-25 mM Tris-Cl pH 7.5, 200 mM NaCl; 0.6% Cymal-5). The flow rate was usually set to 0.3 ml/min depending on the column pressure. The absorption at 280nm was recorded and the size of the elution fractions was set to 0.25 ml.

3.1.7. Concentrating protein samples

BetA solutions were concentrated in Vivaspin centrifugal concentration devices (Millipore) with a molecular weight cutoff of 100 kDa. The sample was spun at 3000g at 4°C to reach a final concentration of ~10 mg/ml.

3.1.8. Protein concentration estimation

A standard absorbance curve calibrated with bovine serum albumin (BSA) by Coomassie dye binding (Bradford, 1976) was carried out for each new batch of Bradford reagent. The gradient m and intercept value b were determined from the standard curve and used in Equation 1 to calculate the protein concentration x (mg/ml). One or two µl of protein solution was pipetted directly to 1 ml of Bradford reagent into a cuvette, mixed and incubated for 3 min. The absorption at 595 nm was read and the protein concentration was calculated by

Equation 1

$$x = \frac{y - b}{m \cdot B_v \cdot s_v}$$

with
- x protein concentration mg/ml
- y absorption at 595 nm
- b intercept value for Bradford batch
- m gradient value for Bradford batch
- B_v volume of used Bradford reagent in ration 1=1ml/1ml or 2=1ml/0.5ml
- s_v sample volume in µl

3.1.9. SDS-polyacrylamide gel electrophoresis

For SDS-polyacrylamide gel electrophoresis (SDS-PAGE) 12% polyacrylamide separation gels and 5% stacking polyacrylamide gels were cast according to (Roche, 2007). The stacking gel was modified in terms of its blue colouring with 2ml H_2O and 700µl of Coomassie brilliant blue G250 solution. The sample pockets were loaded with 4-30 µg of protein in 10-30 µl solution containing a minimum of one third of sample buffer (125 mM Tris-HCl, pH 6.8, 50% (v/v) glycerol, 10 mM β-mercaptoethanol and 0,01% bromphenolblue). The electrophoresis was carried out with the Mini Gel System from Biorad, first at 90 V until the protein left the stacking gel and then at 140 V for the separation gel.

3.1.10. Blue and clear native gel electrophoresis

Blue and clear native gel electrophoresis (BNE and CNE) was performed according to (Wittig et al., 2007). For BNE and CNE, Novex® 4-12% Tris-glycine gradient gels from Invitrogen were used. The Tris-glycine gels are based on the Laemmli System (Laemmli, 1970) with minor modifications. The separating and stacking gels of Novex® 4-12% Tris-glycine gels have a pH of 8.65 unlike traditional Laemmli gels that have a stacking gel pH of 6.8 and separating gel pH of 8.8.

The protein sample (3-4 µg) was mixed with either sample buffer containing 5% Coomassie brilliant blue G250, 100 mM Bis-Tris pH 7.5 and 0.5 M 6-aminocaprocacid, or 5% Glycerol and 0.01% Ponceau S. The electrophoresis Xcell SureLock (Invitrogen) chamber was put on ice and the power supply was set initially to 100 V until the sample entered the gel. The voltage was then increased to 300 V.

The same anode buffer (25 mM imidazole/HCl, pH 7.0) was used for both BNE and CNE. For BNE the cathode buffer contained 50 mM tricine, 7.5 mM imidazole, pH 7.0, 0.02% Coomassie brilliant blue G250 in the first half of the run and was then changed to the cathode buffer containing just 0.002% Coomassie brilliant blue G250. The cathode buffer for the CNE contained 50 mM tricine and 7.5 mM imidazole pH 7.0. For successful BNE or CNE, the NaCl concentration in the sample buffer should not be higher than 50 mM. After BNE the gel was destained in a solution containing 10% acetic acid and 10% ethanol.

3.1.11. Staining of polyacrylamide gels

Gels were stained by the protocol from (Studier, 2005). The gel was suspended in a covered box with ~50 ml solution 1 (50% (v/v) ethanol, 10% (v/v) acetic acid) and heated up to almost boiling point in the microwave. After rocking for at least 5 min rocking at room temperature solution 1 was discarded and ~50 ml solution 2 (5% (v/v) ethanol, 7.5% (v/v) acetic acid) and 200 µl Coomassie brilliant blue solution (0.005g brilliant blue R 250 in 2 ml of 95% ethanol) was added. The gel in solution 2 was heated again to almost boiling. The bands appeared within the next minutes and final results were obtained after about 3-4 hours or overnight.

3.1.12. Western blotting and immuno-detection

A sandwich of four layers of blot paper (Whatman) soaked in transfer buffer (100 mM Tris-ac pH 8.3; 0.2% SDS; 10% methanol) was put on the anode of the blot apparatus (Trans-Blot SD, Biorad München). The polyvinylidene difluoride (PVDF) membrane was shortly activated in methanol and the SDS gel was applied on top, followed by a second layer of blot paper soaked in transfer buffer. The voltage was adjusted to 15 V for 35 minutes.

For immuno detection of the target protein on the PVDF membrane, unspecific binding was blocked with 3 % BSA in TBS buffer (200 mM Tris-Cl pH 7.5; 1.5 M NaCl) for two to three hours rocking at 4°C or room temperature (RT). The PVDF membrane was then incubated overnight at 4°C with the primary anti-StrepTagII IgG1 mouse antibody (Ab) (IBA), 1:4000 diluted with TBS buffer. A wash step with gently added TBS buffer containing the secondary anti-mouse IgG rabbit Ab with a conjugated alkaline phosphatase (Sigma) 1:4000 diluted with TBS buffer was applied to the PVDF membrane and left rocking overnight at 4°C or for two to three hours at RT. Specific Ab binding was detected by the 5-bromo-4-chloro-3-indolyl phosphate/nitroblue tetrazolium (BCIP/NBT, SIGMA *FAST*™ Tablets) detection system. Alkaline phosphatase dephosphorylates the BCIP, which then reduces the NBT resulting in a purple colour.

3.1.13. Two-dimensional thin-layer-chromatography

Two-dimensional thin-layer-chromatography (2D TLC) was performed using Silica Gel 60 pre-coated glass-backed plates (10 cm x 10 cm) (Merck). In the first dimension, the running buffer contained 13 ml of chloroform, 5 ml of methanol and 0.8 ml of water. In the second dimension, the running buffer contained 13 ml of chloroform, 7 ml of methanol and 1 ml of 25% H_2O-ammonia. Both TLC runs were performed in a sealed glass chamber. To detect the lipid and detergent content in the sample, two different colour detection methods were carried out. For unspecific detection of detergent and lipid, the iodine vapour method (Gasser et al., 1977), and for detection of phospholipids, the molybdenum blue method (Muthing and Radloff, 1998) was used. 2D TLC was performed by Ulrike Geldmacher (MPIBP Frankfurt).

3.2. Crystallography

The following chapters give some background on X-ray crystallography methods and introduces into strategies and programs used for structure determination. If not referenced otherwise, background and equations are based on the text books of (Blundell and Johnson, 1976; Drenth, 1994; Messerschmidt, 2007).

3.2.1. Crystal lattices and symmetry

A crystal is a periodic arrangement of identical repeating units, called unit cells. A unit cell in a crystal is a box with six faces, eight edges and three angles, that can be related to a neighbouring unit cell by a pure translation. The unit cell is described by the lengths of the three cell axes a, b and c together with the corresponding angles α, β and γ. These are the unit cell parameters; alternatively referred to as vectors a, b and c. The smallest part of the unit cell that cannot be further reduced through symmetry operations is called the asymmetric unit. The whole crystal lattice can be described by the symmetry elements occurring within the crystal together with lattice translations. 32 ways to combine symmetry operations are possible in crystals, resulting in 230 distinct space groups (International Tables for Crystallography, 1996). Every space group is defined by a specific set of symmetry operations. These symmetry operations generate the complete unit cell when applied to the asymmetric unit. Therefore, during structure analysis, it is only necessary to consider the asymmetric unit because applying all symmetry operators to it, generate the full unit cell. Taking the 32 possible point groups into account, seven crystal systems exist: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic. The conditions imposed on the unit cell geometry increase from triclinic to the cubic system. Apart from the primitive cell, which has $1/8^{th}$ of a lattice point in each corner, other crystal systems have face and body centered lattices, leading to 14 possible Bravais lattices. The number of point groups allowed for chiral biological macromolecules is reduced to 11 and the number of space groups to 65, which reduces symmetry operations to rotation and screw axes.

3.2.2. Macromolecular crystallography

X-ray structure analysis depends on single protein crystals of sufficient size and quality. Crystallisation methods for biological macromolecules rely on the principal of slowly raising the concentration of a precipitant that binds solvent molecules, thereby removing them from the hydration sphere of the protein molecules in solution. This leads to the association of protein molecules with one another and, ideally, to crystal formation. A phase diagram can describe the crystallisation process; as the precipitant concentration rises, the protein solution becomes supersaturated. A metastable state is reached in which existing crystal nuclei grow but new ones do not appear.

It is therefore desirable to reach the metastable state early in the process, so that just a few crystal nuclei form in order to grow a few single crystals in one drop (McPherson, 1990).

Crystallisation conditions for a given protein are rationally hard to predict. In practice one searches for suitable conditions using screens. The most important screens are bifactorial grid screens and multifactorial sparse matrix screens (Carter and Carter, 1979). Typically, three to four factors such as the buffer ion, pH value, precipitant type and concentration, and an additional salt, are varied (Jancarik et al., 1991). For membrane proteins, the same principals apply, but the presence of detergent and/or lipids further complicate the phase diagram. For most detergents, the cmc depends on ionic strength, temperature and pH of the solution, among other factors. The addition of different detergent types to the crystallisation drop referred to as the mixed micelle approach, can also help membrane protein crystal growth.

In this work vapour diffusion was the method used for 3D crystallisation, in both hanging and sitting drops. Both entail a droplet containing purified protein, buffer, and precipitant being allowed to equilibrate with a larger reservoir volume containing similar buffers and precipitants in higher concentrations. The hanging drop differs from the sitting drop method in the vertical orientation of the protein solution drop within the system.

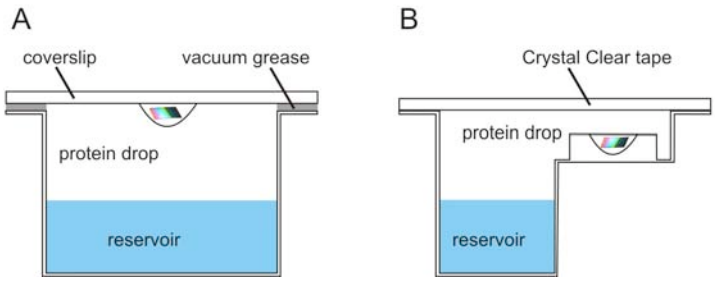


Figure 4| Schematic representation of the hanging (A) and the sitting drop (B) protein crystallisation.

To find the right crystallisation buffer condition, the 3D crystallisation started by screening with sparse matrix screens (Carter and Carter, 1979; Jancarik et al., 1991) using either the Mosquito robot (TTP LabTech) or the Cartesian 3D pipeting-robot (Cartesian technologies). Those robots allow very quick and precise set up of screens with 96 different conditions on one crystallisation plate (Greiner, Axygen) with a drop size of 40 nl to 1 µl which additionally minimising evaporation. Special initial screens for membrane protein crystallisation from Hampton Research, Jena Bioscience and Qiagen were used. After finding positive hits, a screen with 24 conditions around this condition was created and set up in pre-greased VDXm™-24 well plates (Hampton Research, Jena Bioscience). The usual drop size for hanging drops was 2 – 4 µl, and for sitting drops 5 – 20 µl. Each screen was tested at different temperatures: 4°C, 18°C, 21°C and 26°C with a variation in protein concentration from 5 to 15 mg/ml. To further optimise crystallisation conditions, various additive screens (Additive screen and detergent screens 1 to 3, both from Hampton Research) were tested as well.

Seeding methods

Often it is not trivial to obtain large, single, diffraction-quality crystals for X-ray analysis. If only small crystals are obtained during crystallisation it might be necessary to uncouple nucleation and crystal growth in order to grow large single crystals. Macroseeding and microseeding (McPherson, 1990) are the methods of choice for this purpose. In macroseeding, large crystals are washed and then placed in a freshly set up drop where the old crystal will continue growing. In microseeding, an existing crystal is finely crushed and the resulting seed stock used in dilution to inoculate a fresh crystallisation drop, where the microscopic nuclei will then grow into large crystals.

Heavy atom derivatisation

Generally macromolecular crystals contain a large amount of solvent forming a discrete network of solvent channels. Molecules or ions smaller than the diameter of these channels can penetrate the crystal and reach specific protein binding sites by diffusion. This is used to prepare heavy atom derivative and protein-ligand complexes by soaking in appropriate solutions of the compound (Petsko, 1985). Binding of the compound can induce conformational changes in the protein or alter crystal packing. This often leads to disturbance of the crystal lattice and loss of diffraction, but can also improve diffraction if binding of the compound leads to stronger lattice contacts. In cases where ligand-binding sites are obstructed through crystal packing, or binding results in conformational changes, co-crystallisation is often the only viable way to obtain heavy metal derivatives.

Most of the heavy atom derivatives in Table 5 were either prepared by soaking crystals in a solution of 1-300 mM heavy atom compound in crystallisation buffer with slightly higher PEG400 concentration than the original mother liquor. Soak times varied from some seconds to 1 or 2 days. Classical back soaking, the short transfer of derivative crystals back to a heavy atom free buffer in order to remove excess unbound heavy atom compound from the crystal, was performed to minimize radiation sensitivity of derivative crystals.

Table 5| Heavy atom compounds used in this thesis.

Name	Formula	MW (g/mol)
Potassium tetrachloroplatinate	K_2PtCl_4	415.11
Potassium hexachloroplatinate	K_2PtCl_6	486.01
Potassium tetranitroplatinate	$K_2Pt(NO_2)_4$	457.32
Potassium tetracyanoplatinate	$K_2Pt(CN)_4$	377.36
Dichloro(ethylenediamine)platinum	$PtCl_2(H_2NCH_2CH_2CH_2NH_2)$	326.10
Potassium tetrabromoplatinate	K_2PtBr_4	592.93
Platinum potassium iodide	K_2PtI_6	1035
Di-μ-iodobis(ethylenediamine)diplatinum	$[Pt_2I_2(H_2NCH_2CH_2NH_2)_2](NO_3)_2$	888.2
Mercury (II) chloride	$HgCl_2$	271.5
Mercury (II) acetate	$Hg(OOCCH_3)_2$	318.7
Mercury (II) potassium iodide	$K_2(HgI_4)$	786.48
Ethylmercury chloride	C_2H_5HgCl	265.13
Mercury (II) iodide	HgI_2	454.45
Hexatantalum tetradecabromide	$[Ta_6Br_{12}]_2^+$ x 2 Br^-	2204.3
Selenomethionine	$C_5H_{11}NO_2Se$	196.106
Xenon	Xe	131.293

For Xenon (Xe) derivatives of BetP, native crystals were mounted in a loop, and then transferred to the Xe pressure chamber (Xcell, Oxford Cryosystems). A Xe pressure up to 10 bar was applied for about 10 to 20 minutes. After reducing the pressure carefully in 10-20 sec, crystals were directly plunged into liquid nitrogen.

3.2.3. Crystal mounting and cryo protection

Cryo-data collection strongly reduces the effects of secondary radiation damage, which is time- and temperature-dependent. Primary radiation damage, on the other hand is solely dose-dependent (Garman and Owen, 2006). Therefore, measurement at low temperatures generally leads to better diffraction and data quality (Garman, 1999). But the additional handling steps during cryoprotection may damage the crystal, resulting in high mosaicity, crystal splitting, or even loss of diffraction. An other factor is formation of ice crystals upon freezing, which can impose mechanical stress on the protein crystal and give rise to strong diffraction when exposed to X-rays. This obscures protein diffraction and may cause problems in data evaluation.

Usually cryoprotection can be achieved through the addition of cryo-additives e.g. ethylene glycol, glycerol or small polyethyleneglycol (PEG), which cause the buffer to vitrify upon freezing. The alternative method of immersion of the crystal in oils, such as paraffin or a mix of paraffin with Paratone-N, was used in this work. Single crystals were picked from the crystallisation drop with a nylon loop or preferably a LithoLoop™ (Molecular Dimensions) of appropriate size. Afterwards the loop was tipped carefully with the side on the cover slide to reduce the oil volume around the crystal. This procedure was repeated until excess mother liquor was removed and no phase separation was observed. Finally the crystal was quickly plunged into liquid nitrogen.

Another mild method of cryoprotection was to apply 1 ml of 50% PEG 400 solution to the reservoir two to three days before crystal freezing. As a consequence, the amount of water in the hanging crystallisation drop is further reduced, hence to a cryogenic concentration of PEG400 in the drop.

3.3. X-ray structure analysis

3.3.1. Theory of X-ray diffraction

X-rays are electromagnetic radiation with wavelengths between 10^{-7} - 10^{-11} m (1-124 keV). Convential X-ray generators produce X-rays when a beam of high-energy electrons, accelerated through a voltage in a vacuum, hit a metal (Cu) target; usually a rotating copper anode. Radiation of higher intensity is produced in a synchrotron, where the oscillation of high-energy electrons in a magnetic field gives rise to intense radiation over a broad frequency range (Rosenbaum et al., 1971). The desired wavelength is selected using a double diamond crystal monochromator (Girard et al., 2006).

Upon interaction of X-rays with matter, electrons oscillate in the rapidly changing electromagnetic field. As oscillating dipoles, they emit coherent radiation with the same wavelength as the incident radiation, a type of scattering called elastic or Thomson scattering, used in X-ray structure analysis. Effects due to inelastic Compton scattering are not taken into account in X-ray structure analysis. Compton scattering mainly contributes to the background and contributes to radiation damage. The intensity distribution is determined by the arrangement of the scatterers within the repeating units of the crystal, in other words the electron density distribution. In certain directions, the scattered waves interfere constructively; giving rise to detectable reflections, in all other directions they interfere destructively and cancel each other out. The symmetry of the resulting reflection pattern directly reflects the underlying symmetry of the crystal lattice.

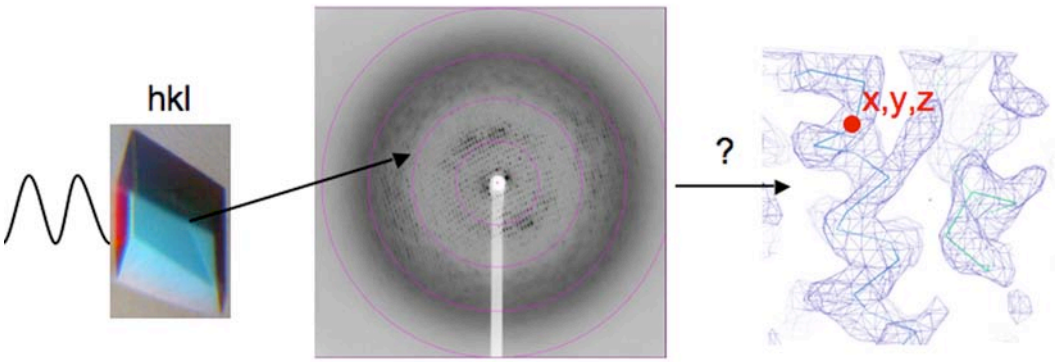


Figure 5| The diffraction experiment. In the diffraction experiment, the intensities of the incident waves are scattered from the planes (*hkl*) in the crystal. The amplitude of the wave *F(hkl)* is proportional to the square root of the spot intensity measured on the detector. Calculation for a position *(xyz)* in the electron density map requires the phase $\alpha(hkl)$ information, which is associated with the structure factor amplitude *F(hkl)*. The experimental set up allows the recording of amplitude component of the scattered wave but not its phase component; this is the phase problem. The question mark indicates the unknown phase information, which is needed to generate the electron density map.

The difference vector between the incident and diffracted waves is the diffraction vector $\vec{s}$. The amplitude of the diffracted wave E at the position R is obtained by integrating over all diffracting volume elements V with associated electron density $\rho_{el}(\vec{r})$ (Equation 2). The terms in front of the integral describe the physical properties of the diffracted wave; the integral itself contains information on the electron density distribution in the diffracting volume element and is called the structure factor **F**. In mathematical terms, the amplitude and phases of the diffracting waves represent the Fourier transform of the scattering density (Fourier analysis, Equation 2).

Equation 2

$$E = const.\exp\left[i\left(\omega t - \vec{k}\vec{R}\right)\right]\int_V \rho_{el}(r)\exp(2\pi irs)dr$$

with $\omega=2\pi c/\lambda$ angular frequency ω, speed of light c and wavelength λ

t time

$\vec{s}$ diffraction vector $2\pi s = k - k_0$

$\vec{k}$ wave vector of the diffraction beam

$\rho_{el}(\vec{r})$ electron density at location r

Constructive interference between two diffracting rays only occurs if the phase difference between the incident and diffracted waves ($2\pi rs$) is a multiple of 2π. This gives rise to the Laue conditions listed in (Equation 3).

Equation 3

$$\mathbf{F}(\bar{s}) = \int_V \rho_{el}(r)\exp(2\pi i\bar{r}\bar{s})d\bar{r}$$

Equation 4

$$\begin{aligned} &\vec{a}\cdot\bar{s} = h \\ &\vec{b}\cdot\bar{s} = k \qquad \text{when } h, k, l \text{ are integers} \\ &\vec{c}\cdot\bar{s} = l \end{aligned}$$

The integers h, k and l are called Miller indices and describe a set of lattice planes. Each set of lattice planes h, k, l corresponds to a specific order of diffraction and pass through a set of lattice points. If the positional vector $\vec{r}$ is expressed in unit cell fractional coordinates x, y and z, the structure factor equation over the volume of the unit cell V_{UC} takes the form:

Equation 5

$$\mathbf{F}(hkl) = V_{UC}\int_{x=0}^{1}\int_{y=0}^{1}\int_{z=0}^{1}\rho_{el}(x,y,z)\exp\left[2\pi i\left(hx+ky+lz\right)\right]dxdydz$$

Just as the structure factor of a specific diffraction order can be calculated by Fourier analysis of the electron density $\rho_{el}(x,y,z)$, the electron density can in turn be calculated by an inverse Fourier transformation (Fourier synthesis) of the structure factors $\mathbf{F}(hkl)$. Because diffraction only occurs along discrete directions, the integration can be replaced by a summation over all hkl (Equation 6).

Equation 6

$$\rho_{el}(x,y,z) = \frac{1}{V_{EZ}}\sum_h\sum_k\sum_l F(hkl)\exp\left[-2\pi i(hx+ky+lz)\right]$$

A structure factor $\mathbf{F}(hkl)$ can be described by a complex number with magnitude $F(hkl)$ and the phase $\alpha(hkl)$ (Equation 7).

Equation 7

$$\boldsymbol{F}(hkl) = F(hkl)\exp\left[i\alpha(hkl)\right]$$

Structure factor amplitudes are derived from observed diffraction intensities (Equation 8), but the corresponding phase information is lost in the diffraction experiment, as no phase-sensitive X-ray detectors exist. Therefore, indirect methods for phase determination were developed to solve the phase problem (3.3.8).

Equation 8

$$I_{obs}(hkl) \propto F^2(hkl) = \mathbf{F}(hkl) \cdot \mathbf{F}^*(\mathrm{hkl})$$

A simplified description of the diffraction process by crystal lattices is given by Bragg's law (Equation 9). Here, the diffraction process is described as reflection of incident rays at a set of parallel lattice planes *hkl* that only occurs at specific glancing angles θ_{hkl}.

Equation 9

$$n\lambda = 2d_{hkl}\sin\theta_{hkl}$$

with n 1, 2, 3, …..
λ wavelength (Å)
d_{hkl} spacing of the lattice planes *hkl* (Å)
θ_{hkl} glancing angle of the reflection *hkl* (°)

3.3.2. Reciprocal space and the Ewald construction

Diffraction by a set of lattice planes *hkl* gives rise to a corresponding reflection hkl. If their respective diffraction vectors $\vec{s}_{hkl}$ are placed on a common origin, the set of their end points P(hkl) defines the so-called reciprocal lattice with the basis vectors $\vec{a}^* = \vec{s}_{100}$, $\vec{b}^* = \vec{s}_{010}$ and $\vec{c}^* = \vec{s}_{001}$. Using the Laue conditions, an expression relating real and reciprocal space is obtained (Equation 10).

Equation 10

$$\vec{s} = h\frac{\vec{b}\times\vec{c}}{\vec{a}(\vec{b}\times\vec{c})} + k\frac{\vec{c}\times\vec{a}}{\vec{a}(\vec{b}\times\vec{c})} + l\frac{\vec{a}\times\vec{b}}{\vec{a}(\vec{b}\times\vec{c})} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$$

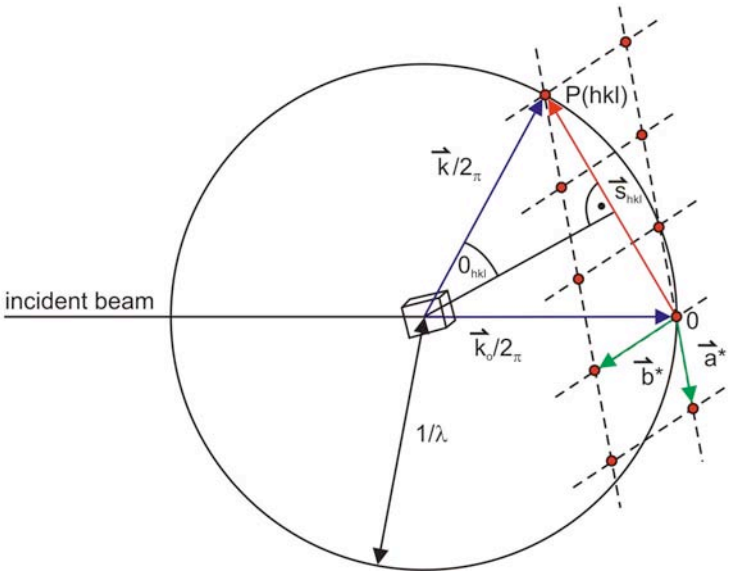


Figure 6| The Ewald construction showing the relationship between real and reciprocal space in a diffraction experiment using monochromatic radiation. The crystal lies in the middle of the Ewald sphere with radius 1/λ. The origin O of reciprocal space is defined as the interception of the incident (primary) beam with the Ewald sphere. A reflection is recorded if the end point of a diffraction vector $\vec{s}_{hkl}$ (reciprocal lattice point, red) coincides with the surface of the Ewald sphere.

The Ewald construction (Figure 6) is a graphical representation of Bragg's law (Ewald, 1921). The crystal is located in the middle of the Ewald sphere with radius $1/\lambda=|\vec{k}|/2\pi$, and the origin of reciprocal space defined at the intersection of the incident beam with the Ewald sphere. A reflection is recorded if the end point of a diffraction vector $\vec{s}_{hkl}$ (reciprocal lattice point) coincides with the surface of the Ewald sphere. Under these conditions, the diffraction vector is perpendicular to the reflecting lattice plane, its length being $1/d_{hkl}=2\sin\theta_{hkl}/\lambda$. This relationship exactly corresponds to the Bragg condition (Equation 9). In the rotation method of data collection, the crystal, and with it the reciprocal lattice, is rotated around an axis, bringing different lattice points into reflecting position at different rotation angles.

Under normal conditions, the intensities of the reflections *hkl* and $\overline{hkl}$ are identical. This gives rise to an inversion center at the origin of the reciprocal lattice not

present in real space. The diffraction by a crystal is identical whether it occurs from top or bottom side of a lattice plane. This relationship is described by Friedel's law (Equation 11). Accordingly, the two reflections hkl and $\overline{hkl}$ are called Friedel mates. In case of significant anomalous scattering, Friedel's law no longer applies and the intensities of the two Friedel mates differ. These anomalous differences can be used to solve the phase problem (3.3.8).

Equation 11

$|\mathbf{F}(hkl)|^2 = |\mathbf{F}^*(\overline{hkl})|^2$, from which follows I(hkl)=I($\overline{hkl}$)

3.3.3. Temperature factors

The structure factor $\boldsymbol{F}$(hkl) can be written as the sum of scattering from all atoms in the unit cell (Equation 12). Both the actual thermal motion of atoms in the cell as well as static disorder give rise to a fall-off in the intensity of diffracted X-rays towards higher glancing angles (Debeye, 1914). This is taken into account by the Debye-Waller factor correction term $\exp(B_i/4d_{hkl}^2)$, a Gaussian function of $(1/\mathrm{d}hkl)$. The quantity B is generally called the temperature factor, even though in real crystals the effect of disorder, and not thermal vibration, account for most of the intensity falloff with resolution. B is proportional to the mean square displacement $\langle u^2 \rangle$ of an atom from its resting position (Equation 13). The atomic form factor f_i also decreases with resolution, further reducing the intensity of the reflection hkl.

Equation 12

$$\boldsymbol{F}(hkl) = \sum_i^n f_i^0 \exp\left(-B_i \frac{1}{4d_{hkl}^2}\right) \exp(2\pi i \bar{r}_i \bar{s})$$

with f_i^0 atomic scattering factor of atom i

B_i temperature factor (B factor) of atom i

Equation 13

$$B = 8\pi^2 \langle u^2 \rangle$$

Determination of scale and temperature factors

The scale factor *C*, which places structure factor amplitudes on an absolute scale, and the mean temperature factor *B* can be determined using a Wilson plot (Wilson, 1949b). First, the data are divided into resolution bins. A graph is drawn of the natural logarithm of the quotient of the mean measured intensity in the bin $\langle I(s) \rangle$ and the absolute intensity calculated from the sum of all scattering factors at rest $\langle I_{abs}(s) \rangle = \sum f_i^0$ against $(\sin\theta/\lambda)^2$. The scale factor *C* and the temperature factor *B* can then be determined by linear regression analysis (Equation 14) provided the approximate contents of the unit cell are known.

Equation 14

$$\ln \frac{\langle I(\bar{s}) \rangle}{\sum_i \left(f_i^0 \right)^2} = \ln C - 2B \frac{\sin^2 \theta}{\lambda^2}$$

3.3.4. The Patterson function

The Patterson function (Equation 15) is a Fourier transformation of the measured reflection intensities. It can be directly calculated using measured data without knowledge of any phase angles. Because it represents a set of vectors, the Patterson function is written as a function of the components *u,v,w* belonging to the vector $\bar{u}$, instead of the real space coordinates *x,y,z*. The Patterson function represents a convolution of electron density with its inverse (Equation 16). In this form, the Patterson synthesis can be interpreted as a summation of all vectors between all atoms in the unit cell, since the product of the electron densities at the locations $\bar{r}_1$ and $r_1 + \bar{u}$ is only non-zero if *u* corresponds exactly to the vector between two given atoms. A Patterson map of a structure with *N* atoms per unit cell has *N(N*-1) non-origin peaks. Because there are always two opposed vectors between two atoms, the Patterson function is centrosymmetric, and screw axis in the crystal are converted into rotation axis in the Patterson map. The peak heights depend on the product of the number of electrons of both atoms.

Equation 15

$$P(u,v,w) = \frac{1}{V_{EZ}} \sum_{h} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)]$$

Equation 16

$$P(\vec{u}) = \int_{\vec{r}_1} \rho(\vec{r}_1)\rho(-[\vec{u} - \vec{r}_1])d\vec{r}_1$$

In small molecule crystallography, it is often possible to de-convolute the Patterson map and solve the structure directly by reconstruction of the original atom positions from all sets of difference vectors. For proteins, yet, both the large number of atoms and the typically low data resolution usually render the direct interpretation of a Patterson map impossible. However, the Patterson function can be used to determine heavy atom positions using anomalous or isomorphous differences, as only atoms that give rise to such differences will be visible in the resulting map. Furthermore, the Patterson function can be used to determine non-crystallographic symmetry (NCS) (3.3.9).

3.3.5. Data collection

In data collection it is necessary to collect a complete data set containing all Bragg reflections up to a given resolution. Errors or inaccuracies during data collection lead to difficulties in space group determination, phasing, and downstream calculations, and may ultimately prevent structure determination. Therefore, the positions and intensities of all reflections must be measured as accurately as possible, making data collection a crucial point in X-ray structure analysis (Dauter, 1999; Dauter, 2005).
In this work crystal diffraction data was collected at beamlines at Synchrotron facilities listed in Table 6. Data collection was performed under a permanent cryo-stream and intensities were recorded using the rotation method (Arndt et al., 1973); where the crystal is rotated around a fixed axis in small intervals. During this rotation, a position-sensitive detector accumulates the diffracted X-rays, resulting in an image (*frame*) containing positional coordinates with their intensities and information regarding the geometry of the detector. In order to achieve the most complete data sets from a

minimum rotation, the strategy program STRATEGY (Leslie, 1992) was used prior to data collection.

Table 6| Synchrotron facilities. The following Synchrotron facilities with their beamlines and corresponding detectors were used for crystal screening and data collection. The specific detector type and beam size for each beamline is listed.

Synchrotron facility	Beamline	Characteristics	Detector & beam size (horizontal x vertical)
European Synchrotron Radiation Facility (ESRF) France 2nd generation	ID 14-1	λ=0,934 Å	ADSC Q210 CCD, 100 x 100 μm
	ID 14-2	λ=0,933 Å	ADSC Q4 CCD, 100 x 100 μm
	ID 14-4	λ=0,85 – 1.3 Å	ADSC Q315r CCD, 100 x 100 μm
	ID 29	λ=0.62 – 2.4 Å	ADSC Q315r CCD, 90 x 60 μm
	ID23-1	λ=0.62 – 2.48 Å	ADSC Q315r CCD, 50 x 50 μm
Swiss Light Source (SLS) Switzerland 3rd generation	X10SA (PXII)	λ=0.6199 – 1.9075 Å	mar225 CCD, 50 x 10 μm

3.3.6. Indexing, scaling and data reduction

All diffraction data were processed with the XDS program package (Kabsch, 1988; Kabsch, 1993). The first step in data processing is the determination of the dimensions and orientation of a crystal lattice relative to the laboratory coordinate system. Using the strongest reflections, the reciprocal lattice is indexed and the possible Laue symmetries are determined. Frequently reflections are only partially recorded on a single frame, therefore three-dimensional spot profiles are determined from consecutive frames. This process is repeated for several regions of the detector and batches of data. In the integration step, all reflections are integrated by fitting them to the best available profile (profile fitting). After integration, geometric parameters such as cell parameters, detector distance and orientation of the crystal are refined using reflections that lie close enough to their predicted positions (post-refinement). Integration and post-refinement are usually repeated in an iterative fashion until convergence (Leslie, 2006). The obtained raw reflection intensities and their standard deviations are then corrected for absorption and Lorentz polarisation effects. At this

stage of data processing, it is possible to determine the correct space group by comparing the intensities of symmetry-equivalent reflections in a given space group and analyzing systematic absences caused by lattice centering or the presence of screw axis. Experimental conditions can change during data collection, e.g. variation of the X-ray intensity, or weakening diffraction due to radiation damage or anisotropic diffraction. Intensities are scaled within the dataset to reduce the mean variation in intensity between multiple measurements of equivalent reflections (Evans, 2006). The scale factors vary with resolution and rotation angle. During scaling, it is also possible to at least partially compensate for the effects of radiation damage (Diederichs, 2006). After scaling, equivalent reflections are merged and reduced to the asymmetric unit of the appropriate space group. This was done with the programs XSCALE of the XDS package.

Alternatively to data scaling with XSCALE the resulting XDS output file was converted with POINTLESS (Evans, 2006) to a MTZ file for further scaling with SCALA and amplitudes were calculated from intensities using TRUNCATE using the CCP4 interface 6.0.1 (CCP4, 1994).

Data quality

Several statistical indicators assure data quality. They are calculated in resolution bins since the characteristics of intensity measurements are strongly dependent on resolution. The R (reliability) factor measures accuracy as the ratio between the mean difference of a value that should be the same and the mean magnitude of the measured value. The most commonly used R factor is R_{sym}, which describes the variations in the intensity of symmetry-related reflections that should be identical according to crystal symmetry (Equation 17).

Equation 17

$$R_{sym} = \frac{\sum_{hkl}\sum_{i=1}^{n_{hkl}}\left|I(hkl)_i - \langle I(hkl)\rangle\right|}{\sum_{hkl}\sum_{i=1}^{n_{hkl}}\left|I(hkl)_i\right|}$$

with $I(hkl)_i$ intensity of reflection i(hkl)

$\langle I(hkl)\rangle$ mean intensity of all symmetry-related reflections (hkl) in the group.

A similar R factor, R_{merge}, describes the agreement between multiple intensity measurements of the same reflection from the same or different crystals, but is rarely distinguished from R_{sym}. All R factors describe the average fractional error with respect to intensities. The values of both R_{sym} and R_{merge} depend on the multiplicity of observations (also called redundancy), and become larger if reflections are measured several times. In effect, collecting highly redundant data, which actually improves data quality, will lead to high R_{sym} values. Therefore, a redundancy-independent R factor R_{meas} (also called $R_{r.i.m.}$) was introduced (Equation 18) which allows a more realistic assessment of data quality (Diederichs and Karplus, 1997).

Equation 18

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n_{hkl}}{n_{hkl}-1}} \sum_{i=1}^{n_{hkl}} \left| I(hkl)_i - \langle I(hkl) \rangle \right|}{\sum_{hkl} \sum_{i=1}^{n_{hkl}} \left| I(hkl)_i \right|}$$

with n_{hkl} number of independent measurements of the reflection hkl

In addition to the R factors, the ratio of the mean intensity and the mean standard deviation of intensities $\langle I \rangle / \langle \sigma(I) \rangle$ (usually abbreviated to $I/\sigma(I)$), the fractional data completeness and the multiplicity of observations are given as indicators of data quality. In this work $I/\sigma(I)$ was not used as a quality criterium because all reflection were included in data processing. At medium resolution with high data anisotropy, it is necessary to include all, even the weak reflections as long as they are accurately measured (3.3.7).

3.3.7. Correction for diffraction data anisotropy

Diffraction anisotropy is directional dependence of diffraction quality of a crystal, and its characteristic shape of diffraction pattern is ellipsoidal rather than spherical (Figure 7). This results in incomplete data in the higher resolution shells. Anisotropy derives from crystal packing interactions being stronger in one direction than another. Frequently, the directional dependence coincides with reciprocal cell

directions *a**, *b** and *c**. Anisotropic data can affect various steps in structure determination.

The directional dependence of the intensity falloff with resolution is measured first and then the directionality is parameterised by three principal components, one for each direction of the crystal. When the falloff is nearly the same in all three directions (Figure 7B), the three principal components will be approximately equal and close to zero. If the falloff has a strong directional dependence (Figure 7A), then the component describing the weakest diffracting direction will be large and positive and the component describing the strongest diffracting direction will be large and negative. Regardless of the degree of anisotropy, the sum of the three components is constrained to add up to zero. The anisotropic ΔB is defined as the difference between the two components with the most extreme values and as such indicated by the anisotropic ΔB statistic. An anisotropic ΔB of 10 Å^2 indicates mild anisotropy; a ΔB over 25 Å^2 strong and a ΔB over 50 Å^2 indicate severe anisotropy.

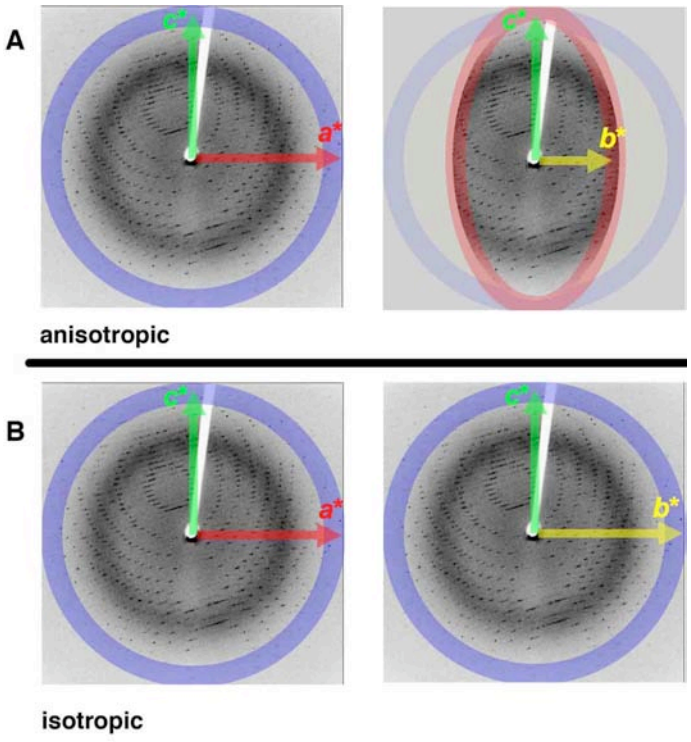


Figure 7| Anisotropic diffraction and isotropic diffraction. The directional dependence at diffraction anisotropy is shown in **A**; the crystal diffracts to high resolution in the vertical direction along *c** and the horizontal direction along *a** but diffracts to lower resolution in the horizontal direction along *b**. **B**; Isotropic diffraction shows the same resolution in all three directions along *a**, *b** and *c**.

Anisotropic data can be corrected fast by using the server http://www.doe-mbi.ucla.edu/~sawaya/anisoscale/ (Strong et al., 2006). After data analysis on merged data, the server program performs first an ellipsoidal truncation, where reflections falling outside the ellipsoid defined by the three specified resolution limits will be discarded from the data set. This truncation algorithm is rather basic and one might want to keep an eye on how much data is actually discarded. After ellipsoidal truncation, the anisotropic scale factors are calculated and applied. After these two steps an additional negative isotropic *B*-factor is applied to the corrected anisotropic data, to increase the magnitude of the high-resolution reflections to their original values.

Anisotropy makes data quality difficult to judge, especially for the higher resolution shells, where anisotropy has its biggest impact (Figure 7). In this work, correction for data anisotropy using the server program was not sufficient. Therefore, the data was analysed for its internal variation in resolution using the CORRECT.LP output file from XDS after running CORRECT. Subsequently, data were defined into different batches using $I/\sigma(I)$ criterion for data batch composition. These different data batches were then merged and scaled together. Data were further converted with POINTLESS (CCP4, 1994) from the XDS format to the MTZ format. Unique batch assignment using SORTMTZ (CCP4, 1994) allowed continuous numbering for later sorting of the batches of images. Data were scaled and merged with SCALA (CCP4, 1994), further analysis for data anisotropy was done in PHASER (McCoy, 2007), using its PHASER_MR mode (script see 10.1), or with SFCHECK (CCP4, 1994) (script see 10.2). These two programs gave the values for the anisotropic tensor. The values $b11$, $b22$ and $b33$ are the three principal components of the anisotropic tensor and act as *B*-factors along a^*, b^*, c^*, respectively. These specific values were applied for anisotropy correction of the data using the program CAD (CCP4, 1994) (script see 10.3). A negative isotropic *B*-factor was also applied to the data in CAD (script see 10.4). Further corrections of data anisotropy used in this work are continued in the phasing chapter 3.3.8.

3.3.8. Phase determination by anomalous scattering

As the phase information is lost at the point of diffraction intensity recording, the determination of the phase is the basic problem in any structure determination. Solving the phase problem by the anomalous scattering approach is, due to the great developments at synchrotron facilities, nowadays the method of choice in most structure determinations.

If the X-ray energy used in a diffraction experiment is in a similar range as one of the electron-binding energies in the scatterer, the assumption of elastic scattering (3.3.1) no longer holds true. Instead of simply exciting the electrons in the object to vibrate with the same frequency as the incident radiation, electronic transitions can be induced. The coherently scattered radiation is thereby emitted with a shifted phase compared to the incident wave, and also has diminished energy, equal to the exact difference between the electronic states. This diffraction of X-rays with altered phase and intensity is called anomalous diffraction, and occurs close to all absorption edges. In practice, anomalous diffraction effects at the K- and L- edges of the heavier elements are most often used (Table 11 in section 6).

Anomalous diffraction effects are accounted for by using additional wavelength-dependent terms to modify the atomic scattering factor f (Equation 19). Friedel's law states that the two reflections in a Friedel pair have the same amplitude, but phases of opposite sign (Equation 11). If a protein crystal contains anomalous scatterers A, their contribution is described by a wavelength-dependent structure factor $\boldsymbol{F}_A(+)$ or $\boldsymbol{F}_A(-)$, which includes the wavelength-independent structure factor $\boldsymbol{F}_A$. The wavelength-dependent part of the structure factor of the anomalous scatterers is itself composed of two components, a real component $\boldsymbol{F'}_A$ and an imaginary component $\boldsymbol{F''}_A$ whose phase is shifted by 90° relative to the real component. The difference in the anomalous component $\boldsymbol{F''}_A$ between the two Friedel mates gives rise to a difference in the magnitudes of $\boldsymbol{F}(+)$ and $\boldsymbol{F}(-)$, thus breaking Friedel's law (Equation 20 and Figure 8).

Equation 19

$$f = f_0 + f'(\lambda) + if''(\lambda)$$

with f_0 wavelength-independent atomic scattering factor
$f'(\lambda)$ real part of the wavelength-dependent atomic scattering factor
$f''(\lambda)$ imaginary part of the wavelength-dependent atomic scattering factor

Equation 20

$$\boldsymbol{F}(+) = \boldsymbol{F}_P + \boldsymbol{F}_A(+) = \boldsymbol{F}_P + \boldsymbol{F}_A + \boldsymbol{F'}_A + \boldsymbol{F''}_A$$

$$\boldsymbol{F}(-) = \boldsymbol{F}_P + \boldsymbol{F}_A(-) = \boldsymbol{F}_P + \boldsymbol{F}_A + \boldsymbol{F'}_A - \boldsymbol{F''}_A$$

with

$\boldsymbol{F}_P$	structure factor of normally scattering atoms
$\boldsymbol{F}_A$	wavelength-independent part of the structure factor of the anomalous scatterers
$\boldsymbol{F'}_A=(f'(\lambda)/f_0)\,\boldsymbol{F}_A$	normal (real) component of the wavelength-dependent part of the structure factor of the anomalous scatterers
$\boldsymbol{F''}_A=(f''(\lambda)/f_0)\,\boldsymbol{F}_A$	anomalous (imaginary) component of the wavelength-dependent part of the structure factor of the anomalous scatterers

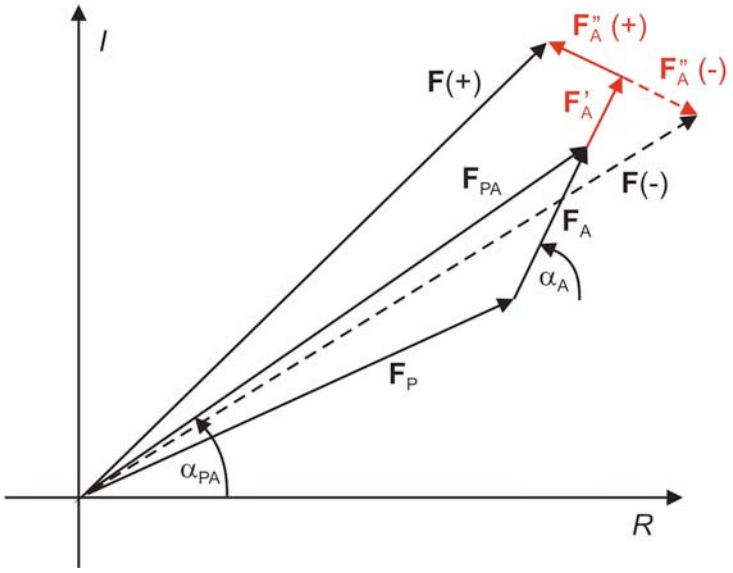


Figure 8| Violation of Friedel's law in the case of anomalous diffraction. The structure factor belonging to the (-) Friedel mate has been mirrored with respect to the real axis for illustrative purposes. The symbols are explained in Equation 20. While $\boldsymbol{F}_P$, $\boldsymbol{F}_A$ and $\boldsymbol{F'}_A$ are identical in both Friedel mates, the sign difference in $\boldsymbol{F''}_A$ leads to different phases and amplitudes of $\boldsymbol{F}(+)$ and $\boldsymbol{F}(-)$ and to a violation of Friedel's law. In the diagram, the magnitude of the anomalous contribution is strongly exaggerated.

In order to use anomalous differences for phasing, it is necessary to choose an X-ray wavelength close to an absorption edge. The K edges of most of the elements occurring in biology (C, N, O, P) are not accessible. From the fourth period of the periodic table onwards, the transition energies in the K and L absorption edges are within the energy range of synchrotron sources. Even when using strong anomalous scatterers like lanthanides and actinides, the anomalous contribution to the measurable

diffraction intensities rarely exceeds 3-5%. Therefore, highly accurate and redundant intensity measurements are essential, and radiation damage can become especially troublesome since the heavy atoms giving rise to anomalous differences are often also the first ones to suffer from radiation damage because they absorb more energy (3.3.5).

An elegant way of introducing anomalous scatterers into protein crystals is the direct incorporation of selenium in the form of selenomethionine during protein synthesis (Hendrickson et al., 1990). The SeMet variant can often be purified and crystallised (3.1.3, 3.2.2) under the same conditions as the native protein, but care must be taken to avoid oxidation of the selenium atoms, which may alter the energy of the K absorption edge.

Anomalous scattering at multiple wavelengths

In a MAD experiment, a crystal containing anomalous scatterers is used to collect several datasets at wavelengths close to a suitable absorption edge. At the edge, the values of the wavelength-dependent atomic scattering factors f' and f'' change in a characteristic fashion (Figure 9). The values of f'' are determined experimentally by performing an X-ray fluorescence scan and plotting the absorption μ against the X-ray energy (Equation 21).

Equation 21

$$f''(E) = \frac{m_e c}{2he^2} E \cdot \mu(E)$$

with m_e, e electron mass and charge
c speed of light
h Planck constant
μ absorption

The values of the real part of anomalous scattering f' are obtained from the f'' curve. The curves for the f' and f'' possess characteristic features close to an absorption edge that are useful for MAD phasing (Figure 9). At the edge, the value of f'' rises abruptly. The value of f' on the other hand drops and approaches a minimum at the inflection point of the f'' curve before rising again.

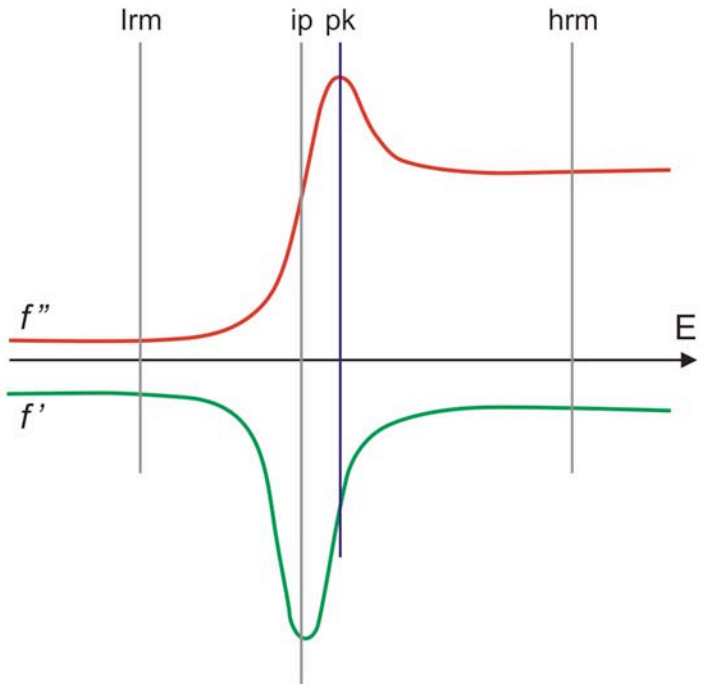


Figure 9| The scattering factors *f'* and *f"* as a function of the X-ray energy E. lrm, low energy remote; ip, inflection point; pk, peak; hrm, high energy remote.

For a MAD experiment, a sensible combination of wavelengths must be chosen for phase determination, and ideally all datasets must be collected from the same crystal to minimise changes in intensity. One data set is always collected at a wavelength corresponding to the peak *f"* because the anomalous signal is strongest there. An advantage of the MAD method is that differences in the normal scattering contribution *f'* between two wavelengths give rise to an isomorphous signal. In practice, a second dataset is therefore collected at a wavelength corresponding to the minimum in *f'*, the inflection point. This ensures that the differences in *f"* between the peak and other datasets are maximal. Data collection at wavelengths far away from the absorption edge (high or low energy remote) should be considered, because *f"* is always higher on the high energy side of an edge and a significant anomalous signal is still attainable, which can be used for phasing. In a MAD experiment, the choice of wavelengths depends on the desired anomalous or dispersive contributions and the diffracting power of the crystal over the course of the measurement.

Radiation damage is a major limiting factor, especially since the collected data must be highly redundant, leading to long exposure times. Due to radiation sensitivity,

it was preferred to collect several peak *f"* data sets and to maximise the redundancy of the anomalous data.

Determination of heavy atom positions by anomalous scattering

For every reflection *hkl*, the anomalous difference can be determined from the difference between the measured amplitudes of the (+) and (-) Friedel mates (Equation 22)

Equation 22

$$\Delta F_{ano} = {}^{\lambda}F(+) - {}^{\lambda}F(-)_i$$

with ${}^{\lambda}F(+)$ and ${}^{\lambda}F(-)$ structure factor amplitudes of the Friedel mates at wavelength λ

If structure factor amplitudes are measured at multiple wavelengths, the difference between the mean amplitude of the Friedel pair at two different wavelengths is called the dispersive difference (Equation 23). Dispersive differences arise from the change in the real part of the wavelength-dependent atomic scattering factor of the anomalous substructure between two wavelengths. Scattering factors of all normally scattering atoms in the protein do not change.

Equation 23

$$\Delta F_{disp} = \left| {}^{\lambda 1}\overline{F} - {}^{\lambda 2}\overline{F} \right|$$

with ${}^{\lambda 1}\overline{F}$ mean value of ${}^{\lambda}F(+)$ and ${}^{\lambda}F(-)$ at λ_1

${}^{\lambda 2}\overline{F}$ mean value of ${}^{\lambda}F(+)$ and ${}^{\lambda}F(-)$ at λ_2

In this work, scaling of the SAD data using a local scaling procedure (Matthews and Czerwinski, 1975) and data analysis was performed with XPREP (Bruker Nonius) and XSCALE (Kabsch, 1988; Kabsch, 1993). For the assessment of data quality, the anomalous signal-to-noise ratio (cutoff criterion > 1.3 σ) and correlation coefficient of anomalous differences between datasets (cutoff criterion > 0.3) were determined. XPREP was also used to calculate ΔF and F_A values for subsequent substructure determination using SHELXD (Schneider and Sheldrick, 2002). SHELXD differs in particular in the extensive use it makes of the Patterson function involving $\Delta \boldsymbol{F}$ data. An

advantage of the Patterson function is that it provides a good noise filter for the $\Delta \boldsymbol{F}$ data; meaning that negative regions of the Patterson function can simply be ignored. The SHELXD algorithm works best when the number of sites is known, therefore it is very applicable for selenium-methionine (SeMet) data, as used for structure determination in this work. SHELXD was used through the graphical interface HKL2MAP (Pape and Schneider, 2004) using the SHELXD file generated in XSCALE.

Anomalous substructure refinement and phase determination

After determination of the anomalous substructure its structure factors $\boldsymbol{F}_A(+)$ and $\boldsymbol{F}_A(-)$ can be calculated using the atomic positions and appropriate scattering factors for the given wavelength (Equation 12). The structure factors of the protein $\boldsymbol{F}_{PA}$ can be determined using the Harker construction allowing the determination of the native protein phase α_P. Because Friedel's law is valid for the normally scattering part, the phase angle of F_{PA} is given by the intersection of the two circles. Working with SAD data results in two intersections and is called the twofold phase ambiguity (Figure 10).

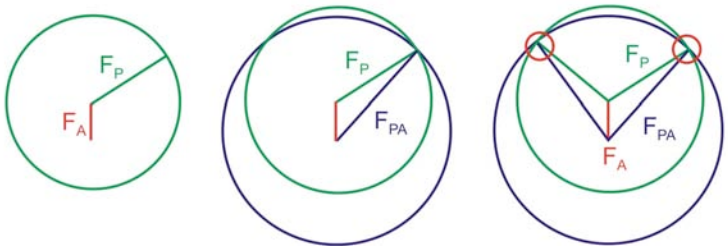


Figure 10| Harker construction for the SAD case and the twofold phase ambiguity. Because of the unknown phase angle, structure factors vectors are rotated around 360° and are therefore depicted as circles in he Harker construction. From left to right: a circle is drawn with the radius of the structure factor of the protein F_P (green) and the anomalous contribution - F_A (red) is drawn from the center; the second circle is drawn to the resulting structure factor F_{PA} (blue) at the end of - F_A. The red circles at the two intersections of the green and blue circle represent the two equally probable phase angles in the SAD case.

In this work, SHARP/autoSHARP (de La Fortelle and Bricogne, 1997) was used for heavy atom parameter refinement and phasing. SHARP/autoSHARP uses the maximum likelihood-based target functions for the refinement of the coordinates,

occupancies and *B* factors of the heavy atoms as well as the relative overall scale and *B* factors of each derivative. Decisions regarding the usefulness of a given derivative for the phasing process and the modification of its heavy atom substructure was made based on the values of quality indicators and the inspection of the resulting electron density for the native protein. Several quality indicators are commonly used in heavy atom phasing. The figure of merit (FOM) is obtained from the phase probability distribution for a given reflection and is used for weighting during for the calculation of initial electron density maps. It can be shown (Drenth, 1994) that the value of FOM represents the mean cosine of the phase error of a given reflection *hkl*.

The resulting phasing statistics from SHARP/autoSHARP are given for both hands, the right and left hand. The determination of anomalous scatterer does not provide information about the correct hand, meaning that one has to analyse these statistics for both hands, which usually are better for the correct hand. It is best to judge the correct hand from the resulting electron density map, especially by looking at helices, which should clearly follow a right-handed course. After determination of the correct hand, the phase was recalculated subsequently in SHARP including NCS averaging and solvent flattening using SOLOMON. This can lead to higher values for phasing power and FOM and lower values for R_{Cullis}, meaning the difference between the observed F_{PA} and the calculated F_{PA}, the "lack of closure" (R_{Cullis}=(phase-integrated lack of closure)/($|F_{PA}-F_P|$)).

The Cullis *R* factor R_{Cullis} (Cullis et al., 1962) relates the lack of closure to the difference between the measured derivative and native structure factor amplitudes (Equation 24). Values below 0.6 suggest a very good derivative, with values up to 0.8 still being useful for phasing. The phasing power (*PP*) is defined as the ratio of the calculated heavy atom structure factor amplitude and the lack of closure (Equation 25). Values above unity indicate an isomorphous signal larger than the measurement error, and thus a useful derivative for phasing.

Equation 24

$$R_{Cullis} = \frac{\sum_{hkl} \left| |F_{PA}| - |F_P + F_A| \right|}{\sum_{hkl} |F_{PA} - F_P|}$$

Equation 25

$$PP = \frac{\sum_{hkl} F_A}{\sum_{hkl} \left| F_{PA} - \left| F_P + F_A \right| \right|}$$

However, if the measured isomorphous difference $|F_{PA}-F_P|$ is small due to weak substitution, and correspondingly a substructure with low occupancies is modelled ($F_A \cong 0$), both R_{cullis} and PP assume very good values even though the resulting native protein phase angles may be completely wrong. Therefore it is always necessary to inspect the resulting protein electron density maps calculated with experimental phases and observed native structure factor amplitudes as coefficients. Adding further correct heavy atom sites to a derivative or including another good derivative in the phasing procedure should always result in better defined protein density and stronger contrast between protein and solvent because of decreasing phase errors.

3.3.9. Non-crystallographic symmetry

The asymmetric unit of a crystal can contain more than one copy of a molecule. In this case, the symmetry relation between molecules in the same asymmetric unit is called non-crystallographic symmetry (NCS). Symmetry relations in the asymmetric unit cannot be always applied due to differences between the molecules of the asymmetric unit. In protein crystals possible NCS operations involve rotation and translation. Rotational NCS can either be of the closed type (proper NCS), with a rotation angle corresponding to a rotation of $(360/n)°$, or open (improper) in all other cases. The former case is frequently found in crystals of oligomeric proteins, which form assemblies with n-fold rotational symmetry in solution (Cn point groups). In addition to the rotation component, a translation operation is needed to describe the position of the rotated copy relative to the original one. NCS operators relating two copies of a protein are usually given as a 3x3 matrix describing the rotation and three parameters describing the translation in orthogonal coordinates. The rotation component of the NCS operator can be determined from the peaks in the so-called self rotation function, a superposition of the native Patterson function with a rotated copy of itself (Rossmann and Blow, 1962).

In contrast to the rotational component, the translational component of NCS cannot be determined without additional phase information. If an initial electron density map and the orientation of a proper *n*-fold rotation axis are available, the position of the axis in the unit cell can be determined by translating it throughout the cell and calculating the correlation of electron density values in *n* equivalent masked regions around the axis. If the masks are chosen correctly, the correlation coefficient will be high at the actual position of the rotation axis (Vonrhein and Schulz, 1999). Nevertheless, this procedure only works for proper rotational symmetry. Both the rotational and translational components of the NCS operator can be determined from heavy atom sites, provided the number and positions of the heavy atoms remains roughly the same in each protomer.

3.3.10. Density modification

After phase determination, the initial electron density map is often still noisy and difficult to interpret due to phase errors. Only in rare cases the initial phases are of such quality that the resulting map is interpretable, allowing a complete model to be built. Therefore, methods have been developed to improve the quality of the electron density by imposing restraints based on common features of electron density maps. One example of such a feature is non-negativity, which demands that no regions with negative density may exist in electron density maps. Another common feature of high-quality electron density maps is that both the mean value and the variance of the electron density in regions belonging to protein are high, while both values should be low in the solvent region as this is disordered. Imposing such restraints on the electron density in real space will lead to improved phases in reciprocal space upon back-transformation of the map (Wang, 1985). This procedure can be repeated iteratively and also used to extend phases to the full resolution of the dataset.

Two principally different approaches to density modification presently exist. In the conventional density modification protocol, the electron density is modified in real space as described above, and new phases are obtained from back-transformation. These phases are combined with the original experimental phases, resulting in an improved set of phase estimates that can in turn be used to calculate a map. Yet, this method has two major problems. When combining the two phase sets, they must be

weighted, but the choice of an appropriate weighting scheme is difficult. The second problem is the lack of a robust criterion for ending the cyclic density modification procedure, as conventional phase quality indicators such as the FOM will continue rising (and be overestimated) if enough density modification cycles are calculated. A more recent approach is the use of statistical density modification procedures that treat density modification as an optimisation procedure in reciprocal space involving all available sources of phase information to obtain new structure factors (Terwilliger, 1999).

In this work, the following density modification methods were used as implemented in SHARP/autoSHARP and CCP4:

Solvent flattening and solvent flipping

Protein crystals contain a large amount of loosely bound or disordered solvent (bulk solvent). Both the mean value and the variance of the electron density should be low in solvent regions, while the opposite is true for regions belonging to protein. Imposing these general features on maps will lead to phase improvement in reciprocal space. The process consists of two steps. The first step is the determination of an appropriate mask to separate protein from solvent regions. This can be done automatically in real or reciprocal space (Leslie, 1987; Terwilliger and Berendzen, 1999; Wang, 1985). After defining a mask, the electron density in unmasked regions is either set to a constant value (solvent flattening) or inverted (solvent flipping). In statistical density modification, the variance of electron density around a given point is also taken into consideration. Solvent flattening procedures generally become more powerful when the solvent content of the crystal is high.

NCS averaging

If several copies of a molecule are present in the asymmetric unit, the NCS relationships between them can be used to improve the electron density map. Improved phases from both, solvent flattening and NCS averaging are combined with the original amplitudes, which result in an improved electron density map. As in solvent flattening, a mask must first be defined for each copy. Then, the NCS operators are used to map equivalent regions of electron density in the other copies onto the first one

and average them. The averaged electron density is then mapped back onto the original positions, which now all contain the same averaged density at equivalent positions. These equivalencies in real space impose restraints on the range of possible phase values after back-transformation of the map. Especially in the case of high-order NCS, averaging can lead to substantial phase improvement. NCS averaging only works well as long as the regions inside the NCS masks are very similar. If the individual copies show significant changes in structure and therefore electron density features, averaging over them will lead to phase deterioration. Therefore, it is important to monitor the correlation coefficients between equivalent densities inside the NCS masks during averaging. If necessary, weights can be applied during averaging that depend on the similarity among the NCS copies (Abrahams and Leslie, 1996). In statistical density modification, deviations from the average density are accounted for by analysing the values of the electron density probability distribution for each grid point (Terwilliger, 2003).

B-factor sharpening

B-factor sharpening is a useful tool for the enhancement of low-resolution electron density maps (Bass et al., 2002; Brunger et al., 2009a; DeLaBarre and Brunger, 2003; DeLaBarre and Brunger, 2006). It entails the use of a negative B_{sharp} value in a resolution-dependent weighting scheme applied to a particular electron density map (Equation 26).

Equation 26

$$F_{sharpened\_map} = \exp(-B_{sharp} \sin^2 \theta / \lambda^2) \times F_{map}$$

with

F_{map}	structure factor of the particular electron density map
$F_{sharpened\_map}$	structure factor of the sharpened map
θ	reflecting angle
λ	wavelength of the X-ray radiation

A good choice for B_{sharp} is the negative Wilson B value of the diffraction data. Still, one should always check the resulting $F_{sharpened\_map}$ electron density map. Applying a negative B_{sharp} value effectively up-weights higher resolution terms. This weighting scheme results in increased detail for higher resolution features such as side-chain conformations. The cost of the increased detail is increased noise throughout the

electron density map. Thus, negative B_{sharp} sharpening is a density modification technique that is only as good as the diffraction data and the phases available. In this work, negative B_{sharp} values between -90Å^2 to -130Å^2 were applied to the data using the program CAD (CCP4, 1994) (script see 10.4). It should be noted that the up-scaled noise in the phases, resulting in a noisy electron density map, can be reduced if the negative B_{sharp} is applied to the data prior to density modification.

3.3.11. Model building and electron density maps

After solving the crystallographic phase problem, the resulting electron density map must be interpreted by building a model of the molecule. If resolution and quality of the experimental density are high enough, this task can at least partially be performed in an automated fashion.

In this work, automatic model building was performed with BUCCANEER version *0.9.9* (Cowtan, 2006). Manual building was done using COOT (Emsley and Cowtan, 2004b) and *O* (Jones et al., 1990). Planar peptide restraints were applied in COOT throughout model building. During structure regularisation of newly built parts, the weighting value for stereochemistry was set to a strict value of 20 to support ideal stereochemistry of the model; compared to the default value of 60 where the regularisation is more dependent on the electron density. When building secondary structure elements a specific restraints mode for either α-helix or ß-sheet was used. *O* was used specifically to improve loop regions with the lego_cα and lego_side_chain commands. These allow rebuilding of segments in the structure using stereochemistry information from a high-resolution structure data bank.

The following types of electron density map were used:

($\boldsymbol{F}_{obs}$ * *FOM*)

Maps calculated with ($\boldsymbol{F}_{obs}$ * *FOM*)*exp*($i\alpha_{obs}$) as Fourier coefficients are used directly after experimental phasing for initial model building. The measured structure factor amplitudes $\boldsymbol{F}_{obs}$ are weighted by the *FOM* from phasing.

$(\boldsymbol{F}_{obs}- \boldsymbol{F}_{calc})$

Difference Fourier maps with the coefficients $(\boldsymbol{F}_{obs}- \boldsymbol{F}_{calc})exp(i\alpha_{calc})$ show differences between the model and the true structure. Missing parts of the model appear as positive and falsely positioned parts as negative difference density. $(\boldsymbol{F}_{obs}- \boldsymbol{F}_{calc})$ maps are used for the placement of not yet built portions of the protein, water molecules, or small-molecule ligands, and is usually viewed at a contour level of 3 σ.

$(2\boldsymbol{F}_{obs}- \boldsymbol{F}_{calc})$

A map calculated with $(2\boldsymbol{F}_{obs}- \boldsymbol{F}_{calc})exp(i\alpha_{calc})$ as Fourier coefficients can be viewed as a superposition of an $(\boldsymbol{F}_{obs})exp(i\alpha_{calc})$ and an $(\boldsymbol{F}_{obs}- \boldsymbol{F}_{calc})exp(i\alpha_{calc})$ map. It is commonly viewed at a contour level of 1.0-1.5 σ and shows the features of the modelled protein. Because of the strong influence of model bias, weighted $(2\boldsymbol{F}_{obs}- \boldsymbol{F}_{calc})$ maps are used instead.

σ_A-*weighted* $(m\boldsymbol{F}_{obs}- D\boldsymbol{F}_{calc})$ $(2m\boldsymbol{F}_{obs}- D\boldsymbol{F}_{calc})$

In cases where parts of the model have not yet been built or wrongly placed, there will be errors in both the phase and amplitude of the calculated structure factor. Therefore, the structure factor amplitudes used to calculate a map must be weighted accordingly, reflecting the fact that both the observed and calculated structure factor amplitudes have errors. In $\sigma_A$weighting (Read, 1986; Read, 1990), corrections are introduced that account for errors in the model, incompleteness and disorder, and errors in scaling. σA-weighted Fourier coefficients were calculated with BUSTER-TNT (Blanc et al., 2004; Roversi et al., 2000; Tronrud et al., 1987), and the resulting weighted electron density maps used throughout refinement.

3.3.12. Macromolecular refinement

The aim of every crystal structure determination is to build a chemically and biologically sensible model of a macromolecule that is in good agreement with experimental diffraction data. Using an inverse Fourier transformation (Equation 12), it is possible to calculate structure factors for the model, given atomic positions, occupancies and *B* factors. The calculated structure factor amplitudes F_{calc} can then be

compared to the observed values F_{obs}. A statistical measure for the fit of the model to the observed data is the crystallographic *R* factor R_{cryst} (Equation 27). Well-refined protein structures will generate *R* values of 10-25% depending on resolution. *R* values of around 50% are common for initial or severely incomplete models. Models with 59% are obtained for randomly placed atoms in the unit cell (Wilson, 1949a).

Equation 27

$$R_{cryst} = \frac{\sum_{hkl} |F_{obs} - kF_{calc}|}{\sum_{hkl} |F_{obs}|}$$

with k resolution-dependent scale factor

The aim of macromolecular refinement is to minimise a residual function of the observed structure factor amplitudes by changing the parameters of the model. The residual used in standard least squares refinement is given in Equation 28**.**

Equation 28

$$f(P) = \sum_{hkl} \frac{\left[F_{obs}(hkl) - F_{calc}(hkl,P)\right]^2}{\sigma_{obs}(hkl)^2}$$

with P set of model parameters
$F_{calc}(hkl,P)$ calculated structure factor amplitude for the reflection *hkl* using a set of parameters *P*

The more observables (independent reflections) are available, the easier it is to determine the optimal set of parameters *P* that results in the minimum value of the residual function. In theory it should be possible to find the minimum using an equal number of observables and parameters. In practice, however, the ratio of observables to parameters should be much higher than unity because the various minimisation programs use approximations to the fullmatrix method (Murshudov et al., 1997; Tronrud, 2004). It is good refinement practice to begin the refinement process with a high "observable-to-parameter" ratio, which can be lowered as refinement progresses and errors in the model become smaller (Kleywegt and Jones, 1995).

The number of observables mainly depends on data resolution and solvent content, with resolution being the more important factor, and is usually constant in a given refinement. However, the observable-to-parameter ratio can be widely varied by using different parameterisations of the model (see below). Depending on the type of parameterisation, this has different effects.

When a constraint is used, a condition that defines a rigid relationship between two parameters so that shifts in one parameter value directly lead to shifts in another, parameters are effectively removed from the refinement process and their number is in effect reduced. In the case of restraints, parameters are allowed to deviate from a target value, but a penalty is imposed if they do so. This allows further observational equations to be written that relate parameters to one another, and effectively increases the number of observations. Nonetheless, the effect of introducing restraints is not as easy to calibrate as the effect of constraints because the new observations are not necessarily independent of one another.

During refinement, care must be taken to avoid over-fitting, *i.e.* the introduction of too many parameters into the model that are not supported by experimental data and in effect may just be modelling noise. In order to prevent overfitting, a small subset of the experimental data called the test set (usually 5-10% of all data, but at least 500 reflections) are set aside and not used for crystallographic refinement or map calculation. Using just the test set of reflections, a "free" R factor (Brunger, 1993) can be calculated in the same fashion as R_{cryst} (Equation 29). The value of R_{free} correlates with the mean error of the model phases and is used as an indicator of refinement progress (Kleywegt and Brunger, 1996). Changes in model parameters that lead to an increase in R_{free} are probably not improving the model but rather overfitting the data.

Equation 29

$$R_{free} = \frac{\sum_{hkl \in T} |F_{obs}(hkl) - kF_{calc}(hkl)|}{\sum_{hkl \in T} |F_{obs}(hkl)|}$$

Model parameterisation: Coordinates

In refinement, a protein molecule is defined as a collection of atoms, each defined by positional coordinates x, y, and z, the B factor, and an occupancy value.

Since the occupancy is usually assumed to be unity (except in cases where double conformations are modelled by two conformers, each with half occupancy), four parameters are needed per atom (nine in the case of anisotropic *B* factors, see below) to parameterise the model. Several other possibilities exist to use a different set of parameters to describe the model.

Rigid body refinement can be used in cases where it is assumed that the position and orientation of a protein has changed, but not its internal structure. Therefore three rotational and three translational parameters are sufficient to describe the model, as the internal structure of the rigid group is not refined. Because only six parameters must be refined per rigid body, this type of refinement can be performed at very low resolution, and has a high radius of convergence.

At intermediate resolution, as it is the case in this work, there is much more data than needed for rigid body refinement, but not enough for individual positional refinement. Instead, a torsion-angle description can be used to describe the polypeptide, as the complete protein fold can be described by two main chain torsion angles per residue plus a variable number of side-chain torsion angles. The bond angles and distances are taken from standard libraries (Engh and Huber, 1991) and constrained. Because the average residue has about eight atoms and five torsion angles, only five parameters need to be refined in torsion-angle refinement for an average residue, as compared to 24 in conventional positional refinement, resulting in a five-fold decrease in the number of parameters. Torsion-angle refinement is therefore used at medium resolution, as implemented in *phenix.refine* (Afonine et al., 2005), CNS (Brunger et al., 1998; Rice and Brunger, 1994) and BUSTER-TNT (Blanc et al., 2004; Roversi et al., 2000; Tronrud et al., 1987).

If NCS is present in the crystal, refining just one protomer and using the NCS relationships to generate the other copies can reduce the number of parameters. That way, only one set of atomic coordinates has to be refined, and the number of positional parameters drops to $1/n$ if n-fold NCS is used. Yet, this is only true for NCS constraints, which do not allow any variation in structure between NCS copies. If NCS restraints are used instead, variations between the individual protomers are allowed, and all atomic coordinates in the respective NCS copies must again be refined separately. Therefore NCS restraints do not reduce the number of parameters but can be treated as additional observables, depending on their strength.

In this work, NCS constraints were used throughout the refinement.

Model parameterisation: B factors

The vibration of individual atoms around their central position is described by the *B* factor (3.3.3). Different parameterisations of the *B* factor are used depending on the number of observables available for refinement. The simplest temperature factor model is to refine a single *B* factor for the whole model. This is commonly done in rigid body refinement at low resolution, with one *B* factor per rigid group. A more complicated model is the use of group *B* factors with one *B* factor per residue, or separate *B* factors for the main chain and side chain atoms in each residue. Atoms inside the groups are constrained to have the same value of *B*, reducing the number or parameters. At high resolution (around 2.0 Å), individual *B* factors can be refined, with one isotropic *B* factor per atom. At very high to medium resolution, individual anisotropic *B* factors can be refined. Because six parameters are now used to describe the displacement instead of one, this leads to a considerable increase in parameters.

In refinement, it is somewhat of a paradox that isotropic *B* factors must be used at medium to low resolution, even though it is very probable that strongly anisotropic motion of atoms within the crystal lattice is the reason for the weak diffracting power of the crystal. Therefore, a different parameterisation that allows modelling of anisotropic motion without the need for many new parameters was introduced with the *TLS* (translation, liberation, screw) model (Schomaker and Trueblood, 1968; Winn et al., 2001). In *TLS* refinement, the model is divided into rigid groups called *TLS* groups. The rigid-body displacement of atoms in the group is described by the three 3x3 tensors *T, L* and *S*. The *T* and *L* tensors are symmetric and describe the translation (in units of $Å^2$) and the rotational component (in units of rad^2) of the displacement. The third tensor *S* is not symmetric and describes the correlation between the translation and rotation movement. Thus, 20 new parameters are needed to describe a *TLS* group. This is considerably less than the number of new parameters introduced by anisotropic *B* factor refinement. Thus, *TLS* refinement can be used at quite low resolution and still offer good results (Painter and Merritt, 2006; Winn et al., 2004). In this work, refinement using *TLS* parameterisation was done with *phenix.refine*. In this procedure, the residual isotropic *B* factors after *TLS* refinement are passed on to *phenix.refine* and

conventionally refined. The related domains were grouped to distinct *TLS* groups by analysing the fold of the protein.

Maximum-likelihood refinement

In contrast to the methodology used by early refinement programs such as PROLSQ (Hendrickson and Konnert, 1980), current refinement programs minimise a maximum likelihood residual (Equation 30) instead of a least squares residual (Equation 28).

Equation 30

$$f(P) = \sum_{hkl} \frac{\left[F_{obs}(hkl) - \left\langle F_{calc}(hkl,P)\right\rangle\right]^2}{\sigma_{obs}(hkl)^2 + \sigma_{calc}(hkl,P)^2}$$

with P set of model parameters
$\left\langle F_{calc}(hkl,P)\right\rangle$ expectation value of F calculated from all plausible models similar to P
$\sigma_{calc}(hkl,P)$ width of the distribution of values for $F_{calc}(hkl,\mathrm{P})$
$\sigma_{obs}(hkl)$ experimental standard deviation of $F(hkl)$

The main difference between the least-squares and maximum likelihood residuals is that errors in the model are explicitly included and added to the uncertainties in the experimental data in the weighting term. Stereochemical restraints are included in the likelihood distribution for the observed data, which synchronises the treatment of restraints as additional observables in the least-squares methodology (see above).

The programs used for structure refinement in this work, CNS (Brunger, 2007; Brunger et al., 1998; Rice and Brunger, 1994), *phenix.refine* (Afonine et al., 2005), REFMAC-5 (Murshudov et al., 1997) and BUSTER-TNT (Blanc et al., 2004; Roversi et al., 2000; Tronrud et al., 1987), all use maximum-likelihood targets. At the end of the refinement, mainly *phenix.refine* and BUSTER-TNT were used subsequently. The BUSTER-TNT version, a developer version and not released, was used in collaboration with Dr. Clemens Vonrhein (Global Phasing Ltd.).

Bulk solvent correction

Protein crystals contain large amounts of disordered solvent (bulk solvent) that does not contribute much to diffraction at resolutions higher than 6-9 Å. However, the bulk solvent contribution at low resolution is appreciable and must be accounted for in order to calculate accurate structure factors from the model.

In the bulk solvent model, this introduces two additional parameters: k_{sol}, the ratio of the mean electron densities of solvent and protein, and B_{sol}, a measure for the fall-off of the solvent scattering with resolution and its diffuseness (Glykos and Kokkinidis, 2000). The use of bulk solvent scaling enables all reflections to be used in refinement, which improves refinement and results in electron density maps with higher contrast. Bulk scaling is implemented in all refinement programs used in this work. In BUSTER-TNT, a more refined scaling model makes an additional correction for missing parts of the model (Blanc et al., 2004).

Simulated annealing

In the simulated annealing (SA) method, each point s of the search space is analogous to a state of some physical system, and the function $E(s)$ to be minimised is analogous to the internal energy of the system in that state. The goal is to bring the system, from an arbitrary initial state, to a state with the minimum possible energy. This method is also used in molecular dynamics studies and refinement (Brunger and Adams, 2002; Brunger and Rice, 1997). SA was done with *phenix.refine* at a point of refinement where the R factors did not drop further, which could be a sign for the model being trapped in local energy minimum. The SA algorithm in *phenix.refine* works as temperature-controlled molecular dynamics. The applied temperature to the system is not a physical measure in SA refinement. The temperature is a control parameter, which determines how readily the system can escape local minima.
In order to improve the stereo-chemical properties of the model the phenix.geometry-minimization option in *phenix.pdbtools* was used.

3.3.13. Structure validation

The final model obtained after model building and refinement will inevitably contain errors. These can be global, like errors in the symmetry operators or overall

scaling, or local errors due to misplaced residues, wrong sequence assignment or false stereochemistry, to name a few. A variety of statistical indicators of model quality exist to track down errors, both local and global.

Reliability factors R_{cryst} and R_{free}

The crystallographic R factor R_{cryst} describes the agreement of observed structure factor amplitudes and those calculated from the whole model (Equation 27). As such, it can only be a global indicator of the ability of the modelled distribution of atoms to explain the measured X-ray data and does not say anything about model geometry or chemical plausibility. As the value of R_{cryst} depends on the observable to parameter ratio, it is of little use in the independent validation of a protein structure (Kleywegt and Jones, 1995). In contrast, the free R factor (Equation 29) is sensitive to overfitting and offers some measure of model plausibility (Brunger, 1993). In cases where the statistical basis of the free R definition is compromised, such as the presence of high-order NCS (Fabiola et al., 2006), the value of R_{free} may be artificially low. Nevertheless, R_{free} remains a useful tool to monitor refinement progress.

Stereochemistry

During refinement, various stereochemical parameters such as bond distances, bond angles, planar groups and improper torsion angles are restrained to a set of ideal values obtained from the analysis of small-molecule crystal structures (Engh and Huber, 1991). The RMS deviation from ideality is commonly reported as a measure for model quality. Because the ideal values for the employed stereochemical restraints are included in the target function during refinement, adherence of the model to the ideal values is not an indicator of model quality, but rather an indictor of how strictly the restraints were enforced during refinement. In this work, deviations from standard geometry were calculated with PROCHECK (Laskowski et al., 1993) and MOLPROBITY (Lovell et al., 2003).

Main chain and side chain torsion angles

Because of the rigidity of the peptide bond, only two rotational degrees of freedom per residue exist in the main chain. The two main-chain angles φ and ψ are restricted to certain values by the geometry of the peptide group. For all residues except

for glycine, the presence of a bulky side chain at Cα would otherwise lead to clashes with main chain amide hydrogen or carbonyl oxygen atoms. In a Ramachandran diagram, φ is plotted against ψ for all residues in the model. The residues should come to lie in allowed regions originally defined by modelling (Ramachandran and Sasisekharan, 1968) and since updated using libraries of observed main chain torsion angles in high resolution structures (Lovell et al., 2003). Special allowed regions are defined for glycine residues. Because main chain torsion angles are not restrained during refinement, they are an excellent source of information for structure validation. For structures refined at resolutions around 2.0 Å, at least 98% of all Ramachandran angles are expected in the allowed region. The lower the resolution the less easy it is to achieve these statistics. A similar situation exists for side chain torsion angles. Because of steric hindrance, only certain combinations of side chain torsion angles are found. Therefore, the side chain angles χ1, χ2, χ3 etc. can be compared to rotamer libraries compiled by analyzing high-resolution structures (Lovell et al., 2000). Only a limited number of side chain rotamers occurs in proteins, so if a set of side chain torsion angles corresponds to a frequently occurring rotamer, it is likely to be correct. During structure refinement, torsion angles cannot be restrained to a single ideal value because usually several equivalent ideal values exist for a given torsion angle, usually separated by 120 °. This causes minimization problems as several minima exist, and side chain torsion angles are either weakly restrained during refinement or not refined at all. Therefore, they are very good independent quality indicators in structure validation. In this work, main chain torsion angle analysis (Ramachandran analysis) was done with PROCHECK (Laskowski et al., 1993), side chain torsion angle (rotamer) analysis with COOT (Emsley and Cowtan, 2004a) and MOLPROBITY (Lovell et al., 2003).

All-atom contact analysis

Residues in the protein interior are in a tightly packed environment and must form specific interactions with other residues in order for the protein to be stable. Hydrophobic residues must associate with one another, charged residues must have a suitable polar partner nearby and hydrogen bond donors or acceptors will be integrated in hydrogen-bonding networks. In addition, interatomic distances may not be less than the sum of the individual van der Waals radii in the absence of specific bonding. All-atom contact analysis takes all these conditions into account by generating a protein

model with "riding" hydrogens and then analyzing all interatomic distances. This allows the identification of problematic residue environments, such as clashes (close contacts), buried hydrogen bond donors that do not participate in a hydrogen bond network, or His, Asn or Gln residues that are flipped by 180 ° around χ2 (Lovell et al., 1999; Word et al., 1999b). In this work, all-atom contact analysis was performed using PROBE (Word et al., 1999a) as implemented in MOLPROBITY (Lovell et al., 2003).

3.3.14. Graphical presentation of protein structures

A graphical representation of a protein showing all atoms in the structure is only useful if just a small part of the structure is shown. When presenting whole protein models, the overall fold can be shown in simplified form by plotting just the Cα positions (Cα trace) or using an abstract representation of the fold (ribbon representation) that may include secondary structure elements (cartoon representation) (Richardson, 1985a, 1985b). Protein structure representations were created with PyMOL (http://PyMOL.sourceforge.org) and MOLSCRIPT (Kraulis, 1991). Contacts between protein molecules or between ligand and protein were analysed with CONTACT (CCP4, 1994) with 3.4 Å and 5.5 Å cut-off distances for polar and hydrophobic interactions, respectively. Contact surface areas were calculated with PISA (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) (Krissinel and Henrick, 2007). Funnels and cavities were calculated with Caver (http://loschmidt.chemi.muni.cz/caver/index.php).

3.3.15. Structure comparison

Most proteins do not have a unique fold, but rather belong to a family of proteins with related structure. Structure superposition is used to compare different structures in order to understand relationships between family members. When comparing different structures and substructures, the number of structurally equivalent residues, a distance cutoff in Å (if used) and the RMS deviation of structurally equivalent atoms (usually just Cα atoms) in Å are typically reported.

The programs used in this work use different methods for generating structure superposition: alignment of distance matrices in DaliLite

(http://www.ebi.ac.uk/Tools/dalilite/) (Holm and Park, 2000) matching secondary structure elements in SSM (Krissinel and Henrick, 2007) and brute-force methods in LSQMAN (Kleywegt, 1996). All programs compute Z scores indicating the statistical significance of the match.

4. Expression, purification and crystallisation

At the onset of this work, Dr. Christine Ziegler provided the plasmid *pASK-IBA7-ΔN29EEE44/45/46AAA*. The protocols for BetP over-expression and purification had been established in the lab of Prof. Reinhard Krämer (Institute of Biochemistry, University Cologne) and in the lab of Dr. Christine Ziegler (Department of Structural Biology, MPI of Biophysics, Frankfurt). In the Ziegler lab, the over-expression had been optimised such that the yield was high enough for structural studies of BetP. Nevertheless, in this work several modifications were necessary in order to optimise protein expression and purification for X-ray crystallography. These modifications included expression time and growth temperature, specific detergent exchange on the affinity column and purification *via* size exclusion chromatography.

4.1. Expression and purification of BetA

Expression of the gene for the mutant BetA was carried out as described in 3.1.1 and 3.1.2. For BetA production, DH5α™-T1 cells were cultivated on a 12 to 24 l scale at 37°C for ~6 h or at 25°C overnight. A typical growth curve is shown in Figure 11.

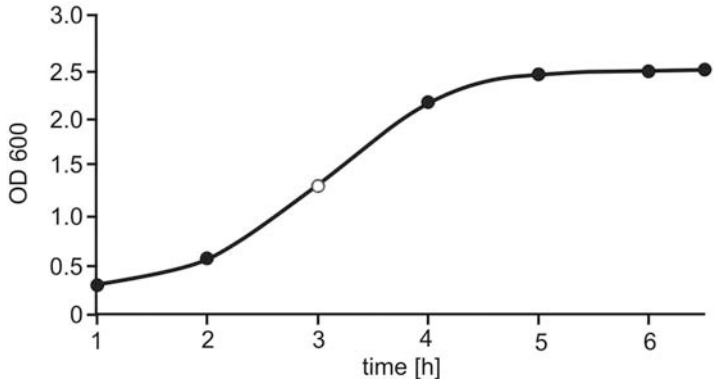


Figure 11| Growth curve for DH5α™-T1 cells expressing the gene for BetA (27-11-2007) after induction (open circle) with 200µg/L AHT. The optical density (OD) is plotted against time (h).

After harvest, cell breakage, sequential ultra-centrifugation, membrane solubilisation (3.1.2, 3.1.4) and a final ultra-centrifugation step to separate the insoluble fraction, the solubilisate was loaded onto a StrepTactin®-affinity chromatography column, with a resin volume of 2-3 ml (3.1.5). After washing steps (a, b, c and/or d, see 3.1.5), which always included a detergent exchange from 0.01% β-DDM to 0.6% Cymal-5 (3.1.5, Table 1), the sample was eluted from the affinity column (3.1.5). The eluted protein was practically free of impurities, as judged by Coomassie and silver stained SDS-PAGE gels (lanes E in Figure 12). The usual amount of protein loaded on the SDS-PAGE varied between 5µg and 10µg. The preparations differed in the duration of detergent exchange and salt concentration. A slow flow (~0.5ml/min), equal to longer duration of the detergent exchange, had a positive influence on the homogeneity of the sample. High salt concentration had a positive effect on sample purity.

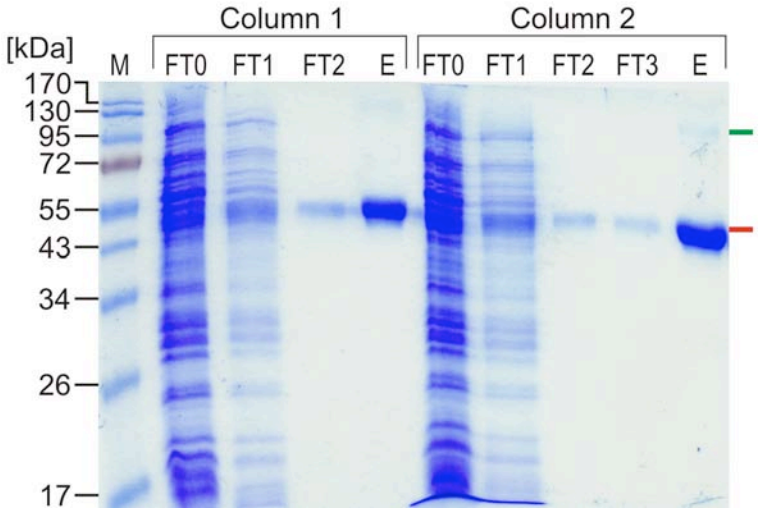


Figure 12| SDS-PAGE of a typical purification with StrepTactin®-affinity chromatography. The BetA solubilisate (03-11-2006) was loaded on two columns. Column 1 was treated with washing steps a) (3.1.5) and column 2 was treated according to washing steps b) (3.1.5), which included an incubation step with POPG lipid. **M**, marker lane (Fermentas – pre-stained PAGE ruler); Column 1: **FT0**, flow-through fraction of the loading step; **FT1**, flow-through of the high salt wash with 500mM NaCl; **FT2**, flow-through of the detergent exchange step from 0.01% β-DDM to 0.6% Cymal-5; **E**, of concentrated (~5 mg/ml) eluate of the BetA sample with 5mM desthiobiotin. Column 2: **FT0** and **FT1** are as described for column 1; **FT2**, flow-through of the lipid incubation step with POPG (0.25mg/ml in 0.05% β-DDM); **FT3**, flow-through after detergent exchange; **E**, concentrated (~5mg/ml) eluate of BetA. Red bar, BetA monomer; green bar, BetA trimer.

Most of the impurities were in the flow-through (FT) fractions, and a small loss (approx. 5% by Western blot), of BetA occurs during the washing steps (Figure 12, FT1, FT2, FT3). The elution (E) fraction shows a single band at 55kDa (red bar) with an additional band at ~120kDa in the position of a probable BetA trimer (green bar). Both bands were confirmed by Western blot analysis to be BetA (Figure 16). As membrane proteins are frequently observed to run at slightly lower apparent molecular mass in the SDS-PAGE gel than predicted, the bands in the gel most likely reflect the monomer at 55kDa and the trimer at ~100kDa. The elution sample from the affinity column was concentrated (3.1.7) to ~5 mg/ml (3.1.8) and loaded onto a size exclusion chromatography (SEC) Superose 6 10/300 column, which separated aggregates, nonamer, hexamer, trimer and rarely monomers of BetA (Figure 13, Figure 14). Additional SEC purification was necessary to obtain good quality 3D crystals.

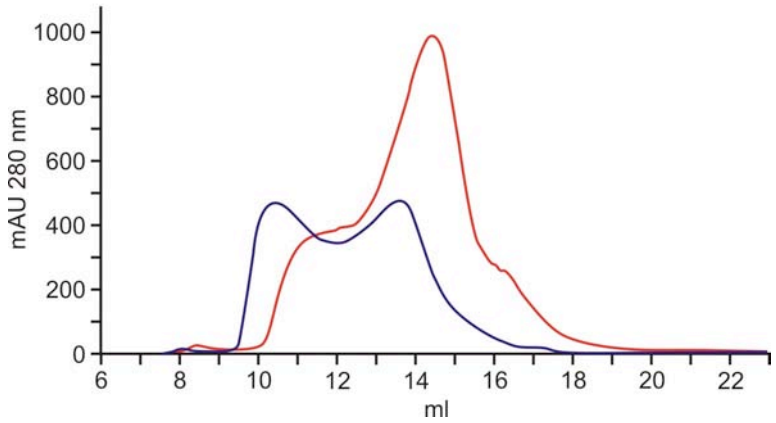


Figure 13| SEC chromatogram of BetA (03-11-2006). The blue line shows the SEC chromatogram of the eluate of column 1 (C1) (Figure 12), the red line shows the SEC chromatogram of the eluate from column 2 (C2) (Figure 12).

The specific samples are named according to their date of expression and SEC fraction, which is either from the first peak (p1) between 9.5 and 12.5 ml or the second peak (p2) between 12.5 and 16 ml (Figure 13). The two BetA 03-11-2006 samples were each treated differently. The solubilised fraction loaded on column 1 (C1) was treated with a) (3.1.5) and the one loaded on column 2 (C2) was treated with b) (3.1.5). The C2 sample was treated with POPG and appeared to be more

monodisperse than the sample from C1 without the POPG treatment. The washing step with POPG lipid (Table 2) incubation was omitted in preparations as it was found that it did not improve the 3D crystals (Ressl, 2006).

Whereas no differences were observed on the SDS-PAGE (Figure 12), the SEC profiles (Figure 13) show differences in the two samples C1 and C2. While sample C1 separates almost into two distinct peaks, one at 10.5 ml and the other at 13.5 ml, the sample from C2 shows one major peak at 14.5 ml with shoulders at 11.5 ml and 16.5 ml. Comparison with a BetP homologue, the trimeric carnitine transporter CaiT from *E. coli* (Vinothkumar et al., 2006), shows that the peak around 12-14ml contains the BetA trimer and fractions eluted earlier contain higher oligomers accordingly.

BetP has been determined to be a trimer in its native state by various techniques (Ressl, 2006; Tsai, 2008; Ziegler et al., 2004), and trimer fractions were observed in the SEC chromatograms (Figure 13, Figure 14) and by CN-/BN-PAGE (Figure 20, Figure 19). Low amounts of BetP monomers were increased at higher temperatures or when washing the protein on the affinity column without detergent (Figure 20).

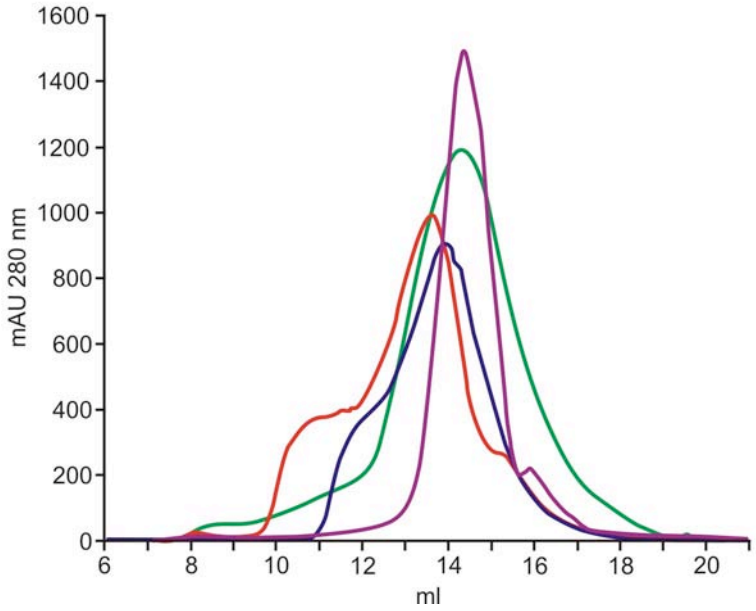


Figure 14| SEC chromatograms of four different BetA purifications. Red line, C2-03-11-2006; blue line, 23-11-2006; green line, 07-08-2008; violet line, 10-11-2008

Four different SEC profiles of BetA purification, each from a separate over-expression, are shown in Figure 14. The profiles 23-11-2006 (blue) and C2-03-11-2006 (red) featured a main peak at ~13ml with shoulders at higher and lower volumes. The two other purifications 07-08-2008 (green) and 10-11-2008 (violet) show one major single peak in the range of 14 – 15 ml. The latter two purifications are derived from overnight expressions at 24°C. The effect of better monodispersity of protein expressed overnight at lower temperatures was reproducible. The typical final yield for BetA was ~1.2 mg per liter of medium.

It turned out that the monodisperse protein batches crystallised better (see Figure 24). The different types or amount of detergent and lipids present may explain the variability of the SEC profiles, reflecting a range of detergent-protein and detergent-lipid-protein-complexes. The separation of different BetA oligomers was only achieved in an approximately quantitative manner. Another reason for the obtained higher monodisperse protein after slower expression at 24°C may result in more proper folded BetP and therefore in more intact and stable trimers.

A longer SEC column than a Superose 6 30/100 SEC and/or using different column materials could presumably separate a mixture of BetP oligomers.

4.2. Expression, purification and crystallisation of SeMet-BetA

The growth curve for BL21 RIL-X cells producing SeMet-BetA is shown in Figure 15.

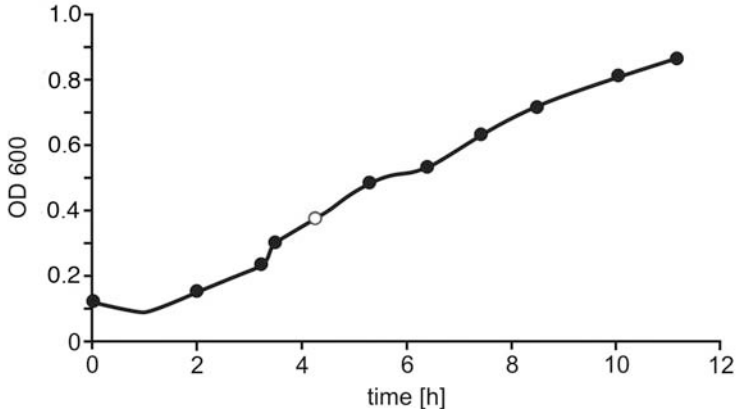


Figure 15| Growth curve for BL21 RIL-X cells expressing the gene for SeMet-BetA after induction (open circle) with 200µg/L AHT.

Apart from maintaining strictly reducing conditions, the overall purification procedure for SeMet-BetA was similar to that of native BetA (4.1). SeMet-BetA was purified by StrepTactin®-affinity chromatography and identified by immunoblotting (Western blotting) after SDS-PAGE (Figure 16) (3.1.12). The lanes for the elution fraction E in Figure 16 and Figure 17 show bands on a gel at the level of the monomer as well as the trimer and revealed no visible impurities.

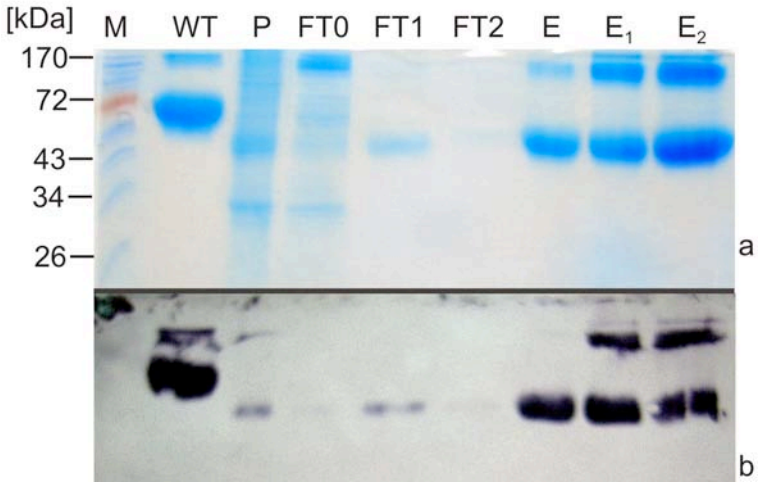


Figure 16| SDS PAGE (a) and Western blot (b) for SeMet-BetA purification from 08-08-2006. M, marker lane (Fermentas – pre-stained PAGE ruler); **WT**, BetP wildtype as a control; **P**, pellet after solubilisation; **FT0**, flow-through fraction of the loading step; **FT1**, flow-through of the high salt wash (500mM NaCl); **FT2**, flow-through of the detergent exchange step (0.01% β-DDM to 0.6% Cymal-5); **E**, eluate of SeMet-BetA; **E_1**, SeMet-BetA after desalting on a PD10 column; **E_2**, SeMet-BetA after desalting on a PD10 column.

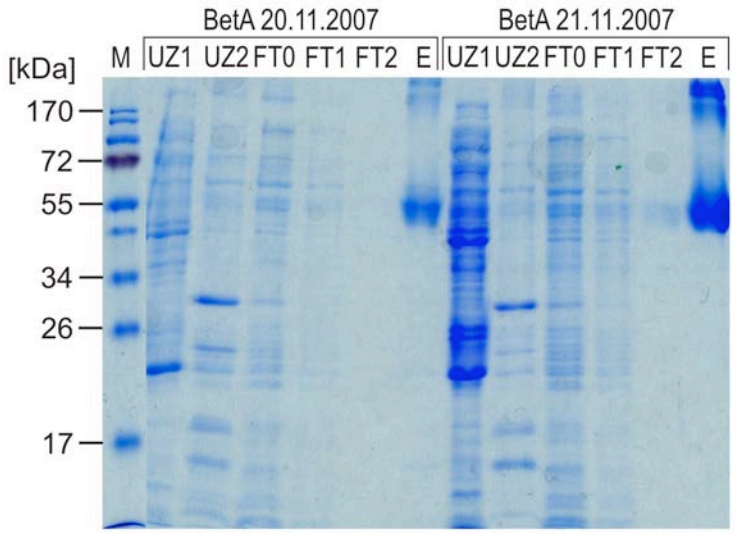


Figure 17| SDS PAGE for SeMet-BetA purification from 20-11-2007 and 21-11-2007. M, marker lane (Fermentas – pre-stained PAGE ruler); **UZ1**, supernatant after the 1st ultra-centrifugation step; **UZ2**, pellet after solubilisation and the 2nd ultra-centrifugation; **FT0**, flow-through fraction of the loading step; **FT1**, flow-through of the high salt wash (500mM NaCl); **FT2**, flow-through of the detergent exchange step (0.01% β-DDM to 0.6% Cymal-5); **E**, eluate of SeMet-BetA.

SEC was also necessary to achieve successful 3D crystallisation of SeMet-BetA. SEC profiles for four SeMet-BetA purifications are shown in Figure 18. The SEC profiles for SeMet-BetA showed wider elution peaks (Figure 18) but no multiple peaks were observed, similarly to native BetA (Figure 13). However, the yield of SeMet-BetA, on average ~0.2 mg per liter of SeMet-media, was lower compared to the yield for native BetA.

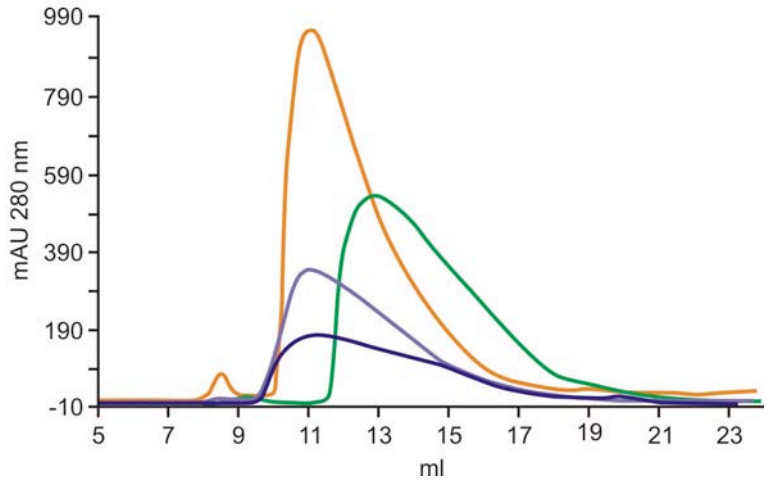


Figure 18| SEC for SeMet-BetA samples. Green SeMet-BetA 09-04-2007, orange SeMet-BetA 30-05-2007, light blue SeMet-BetA 20-11-2007, dark blue SeMet-BetA 18-12-2007.

4.3. Clear-native- and blue-native-PAGE of BetA and SeMet-BetA

4-12% clear-native(CN)-/blue-native(BN)-PAGE was used to check the monodispersity of the pooled SEC fractions. These 4-12% CN-/BN-gels always revealed a combination of higher oligomers of BetA or SeMet-BetA, which suggested that an exact separation of oligomers is not possible on the Superose 6 10/300 column. The amount of protein loaded onto CN-/BN-PAGE gels varied between 3µg and 6µg. Fraction pooled between 10 and 12 ml are indicated as peak1 (P1) and fractions pooled between 12 and 15 ml indicated as peak2 (P2) were loaded on the 4-12% CN-/BN-PAGE gels.

The 4-12% CN-/BN-gels (Figure 19, Figure 20) showed a repetitive pattern of BetA oligomers. For almost all samples, regardless of which pool of SEC peak or shoulder fractions they derived from, distinct bands for nonameric (N) at ~9x65 kDa, hexameric (H) at ~6x65 kDa, trimeric ~3x65 kDa and monomeric (M) BetP at ~65 kDa were detected on the gel. The only difference was the relative intensity of the bands. Lane d in Figure 20 shows a high content of BetA monomer. This sample was washed on the affinity column without detergent in the wash buffer, which could have resulted in a loss of stabilising lipid and/or detergent bound to the BetA trimer, potentially resulting in its dissociation. Lane e in Figure 20 show a SeMet-BetA

sample, which ran slightly higher than BetA, possibly due to incorporation of selenium in this sample.

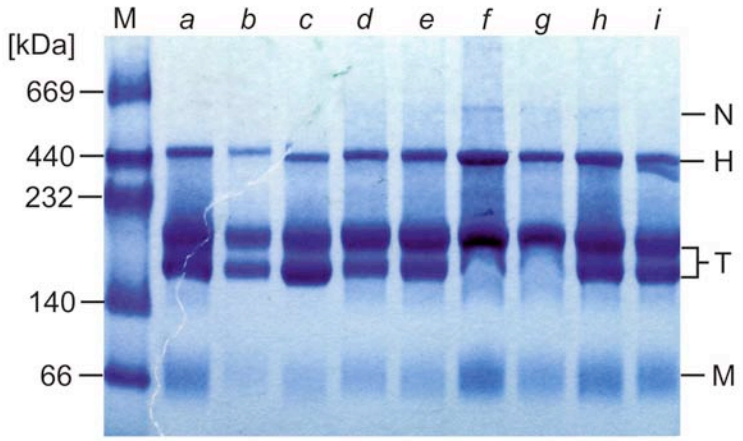


Figure 19| BN-PAGE 4-12% of BetA samples after SEC. M, native marker (Amersham Biosciences); **a**, BetA-C1P1-23-11-2006; **b**, BetA-C2P1-03-11-2006; **c**, BetA-C2P12-03-11-2006; **d**, BetA-C2P1-07-11-2006; **e**, BetA-C2P2-07-11-2006; **f**, BetA-C1P1-13-11-2006; **g**, BetA-C2P1-13-11-2006; **h**, BetA-C1P2-20-11-2006; **i**, BetA-C2P1-20-11-2006. N=nonamer ~585kDa, H=hexamer ~390kDa, T=trimer ~195kDa, M=monomer ~65kDa.

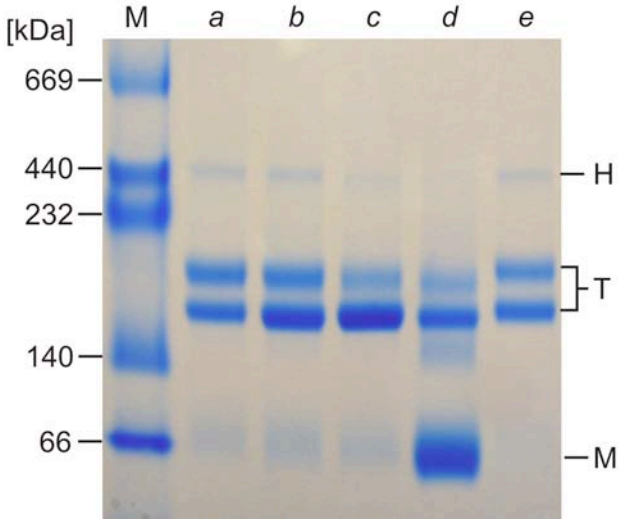


Figure 20| CN-PAGE 4-12% of BetA samples after SEC. M, native marker (Amersham Biosciences); **a**, BetA-C1P1-27-11-2006; **b**, BetA-C1P2-27-11-2006; **c**, BetA-C1P1-11-12-2006; **d**, BetA-C1P2-11-12-2006; **e**, SeMet-BetA-C1P1-18-12-2006. H=hexamer ~390kDa, T=trimer ~195kDa, M=monomer ~65kDa.

CN-/BN-PAGE of native BetA (Figure 19, Figure 20), or of SeMet-BetA showed similar pattern of oligomer distribution (Figure 20, Figure 21). A direct comparison of SeMet-BetA and native BetA on a BN-PAGE is shown in Figure 21. The most significant difference is the appearance of two trimer bands. This might be due to a different amount of detergent and/or lipid bound to the respective trimer population leading to this shift in molecular weight.

The SeMet-BetA samples from 03-07-2006 were purified with different DTT and EDTA concentrations on the StrepTactin® column and were further checked on the BN-PAGE in order to see any influence of DTT or EDTA on the samples. The sample from column 1 (C1) was washed with 15 mM DTT and 5 mM EDTA, C2 with 5 mM DTT and 2 mM EDTA, C3 with 10 mM DTT and 5 mM EDTA and C4 with 20 mM DTT and 10 mM EDTA. As seen in Figure 21, the difference in concentration had no detectable influence on the diversity of the samples. However, storage of a sample at room temperature resulted in degradation or dissociation (Figure 21, lane d).

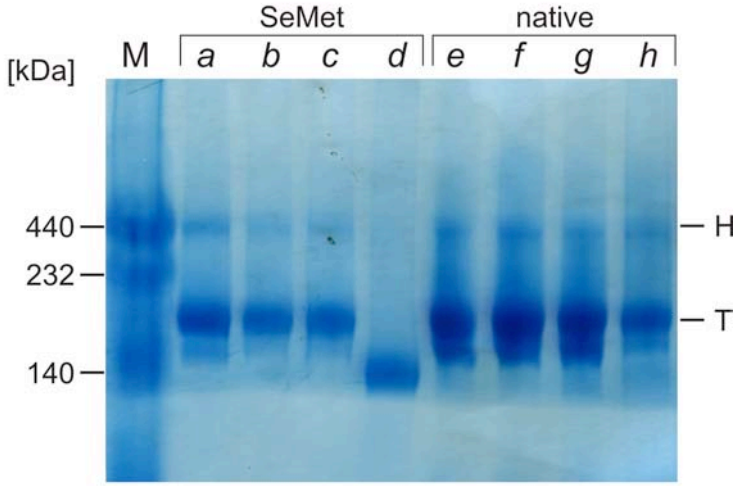


Figure 21| BN-PAGE 4-12% of SeMet-BetA and BetA treated in different ways. M, native marker (Amersham Bioscience); **a**, SeMet-BetA C1-03-07-2006 15 mM DTT and 5 mM EDTA stored at -20°C; **b**, SeMet-BetA C2-03-07-2006 5 mM DTT and 2 mM EDTA stored at -20°C; **c**, SeMet-BetA C4-03-07-2006 20 mM DTT and 10 mM EDTA stored at -20°C; **d,** SeMet-BetA C1-03-07-2006 15 mM DTT and 5 mM EDTA stored at room temperature (RT); **e**, BetA C1-17-07-2006 desalted on PD10 column; **f**, BetA C1-17-07-2006 resalted with 200mM NaCl; **g**, BetA C2-17-07-2006 desalted on PD10 column; **h**, BetA C2-17-07-2006 resalted with 200mM NaCl. H=hexamer ~390 kDa, T=trimer ~195 kDa.

4.4. Lipid and detergent content by thin-layer chromatography

Detergent and/or lipid in BetA samples were analysed by 2-dimensional (2D) thin-layer chromatography (2D TLC). The 2D TLC plates in Figure 22 show distinct detergent and lipid content and reflect the variation between the different purifications as observed by SEC and CN-/BN-PAGE (Figure 19, Figure 20, Figure 24). The detergents Cymal-5 or β-DDM, indicated by red stars in Figure 22, cannot be distinguished by TLC. Phospholipids, detected by molybdenum blue were present in samples e and f, as indicated by the blue star in Figure 22. Sample e is from the SeMet-BetA preparation that yielded the crystals from which the structure was determined.

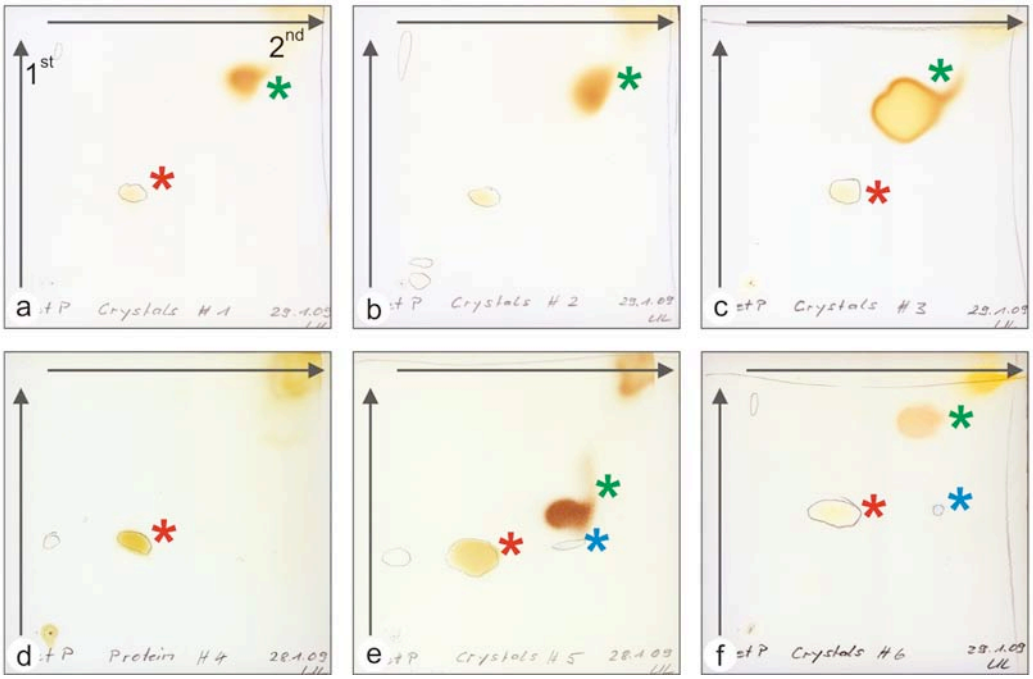


Figure 22| Two-dimensional thin-layer-chromatography for BetA and SeMet-BetA protein and crystal samples. a crystals from the sample BetA 07-09-2007, **b** crystals for SeMet-BetA 07-09-2007, **c** crystals from BetA 07-08-2008, **d** protein sample from BetA 06-11-2008, **e** crystals from SeMet-BetA 18-12-2006, **f** crystals from BetA 2005 (FS1 plate). The red stars, detergent, either Cymal-5 or β-DDM; the green stars, PEG400; the blue stars, phospholipids. The arrows indicate the first and second dimension.

4.5. Crystallisation of BetA and SeMet-BetA

The crystallisation conditions (Ressl, 2006) with PEG 400, NaCl and Na_3-citrate buffer in a pH range between 5.3 and 5.6 at 18°C reproducibly gave crystals, which diffracted to 2.6 Å in the best direction. These crystals were usually of rhomboid shape. Those diffracting best derived from a clear drop as a single crystal with a size of about 400 µm across (a-1, a-2, b-1, b-2 in Figure 23). Crystallisation drops were usually set up with freshly purified protein concentrated after pooling SEC fractions between 12 and 15 ml. The first crystals usually were observed after three days. Drops that remained clear for more than one week were either macro-seeded, or 300-500µl of 50% PEG400 or 5M NaCl were added to the reservoir solution. Crystals usually grew over a time period of up to three months. BetA crystals that were up to 4-5 months old were robust when handled with crystallisation tools. Older crystals had a jelly-like consistency and did not diffract. Reproducible crystal growth was essential for generating heavy atom derivatives of the native BetA crystals (6.1, 6.2). The most important factor for good crystal growth was the decision about which SEC fractions were pooled. As outlined above, SEC revealed different BetA oligomers. Fractions containing higher oligomers, derived from a volume around 10 to 12ml, often resulted in poor crystal growth (Figure 24). On the other hand, pooled trimer fractions from around 13 to 15ml, usually resulted in good crystal growth. Monodispersity of the protein is thus an important consideration for the crystallisation of BetP.

Table 7| Crystallisation conditions for BetA and SeMet-BetA

Protein concentration[a]	5 to 15 mg/ml
Reservoir solution[b]	100 mM Na_3-citrate buffer pH 5.35 to 5.6, 10 to 27% PEG400, 50 to 200 mM NaCl 50 to 150 mM betaine (optional)
Drop volume[b]	400 nl to 10 µl[c]
Temperature	18°C

[a] prior to setting up the drop
[b] equal volumes of reservoir buffer and protein solution were mixed
[c] volumes larger than 5 µl were set up as sitting drops

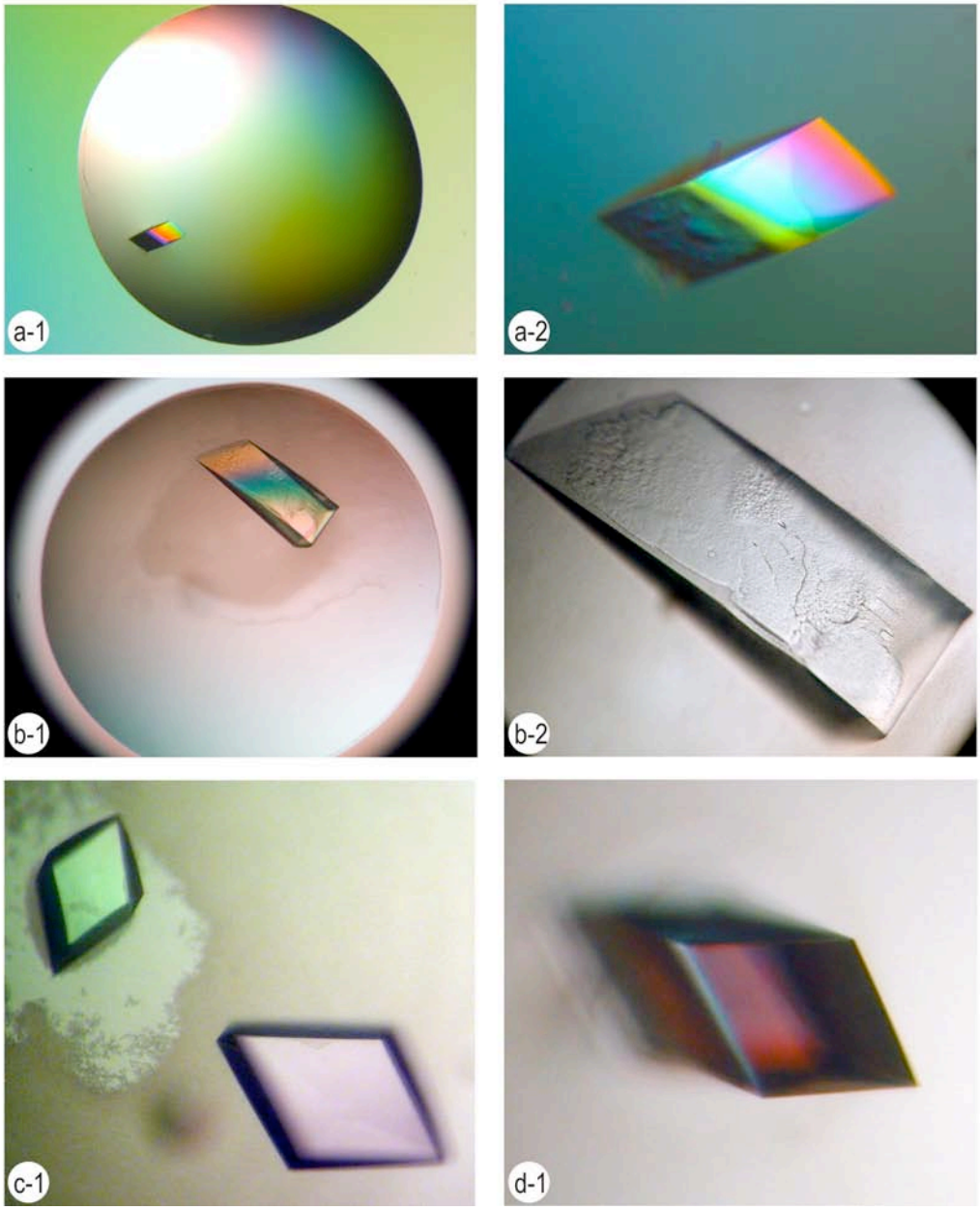


Figure 23| Gallery of BetA crystals from different purifications. As a guide, the drop size of 2-3µl is about 2 mm in diameter. **a-1** and **a-2** show the crystal Native2 in the crystallisation condition: 21% PEG400, 100 mM Na_3-citrate buffer pH 5.55, 150 mM NaCl, **b-1** and **b-2** show the crystal from purification BetA-23-11-2006 (~12 mg/ml) grown in 19% PEG400, 100 mM Na_3-citrate buffer pH 5.5, 100 mM NaCl, **c-1** from BetA 23-11-2006 (~12 mg/ml) appeared in 20% PEG400, 150 mM Na_3-citrate buffer pH 5.45, 100 mM NaCl and **d-1** from BetA 23-11-2006 (~12 mg/ml) grown in 22% PEG400, 100 mM Na_3-citrate buffer pH 5.4, 100 mM NaCl. All crystals grew at 18°C.

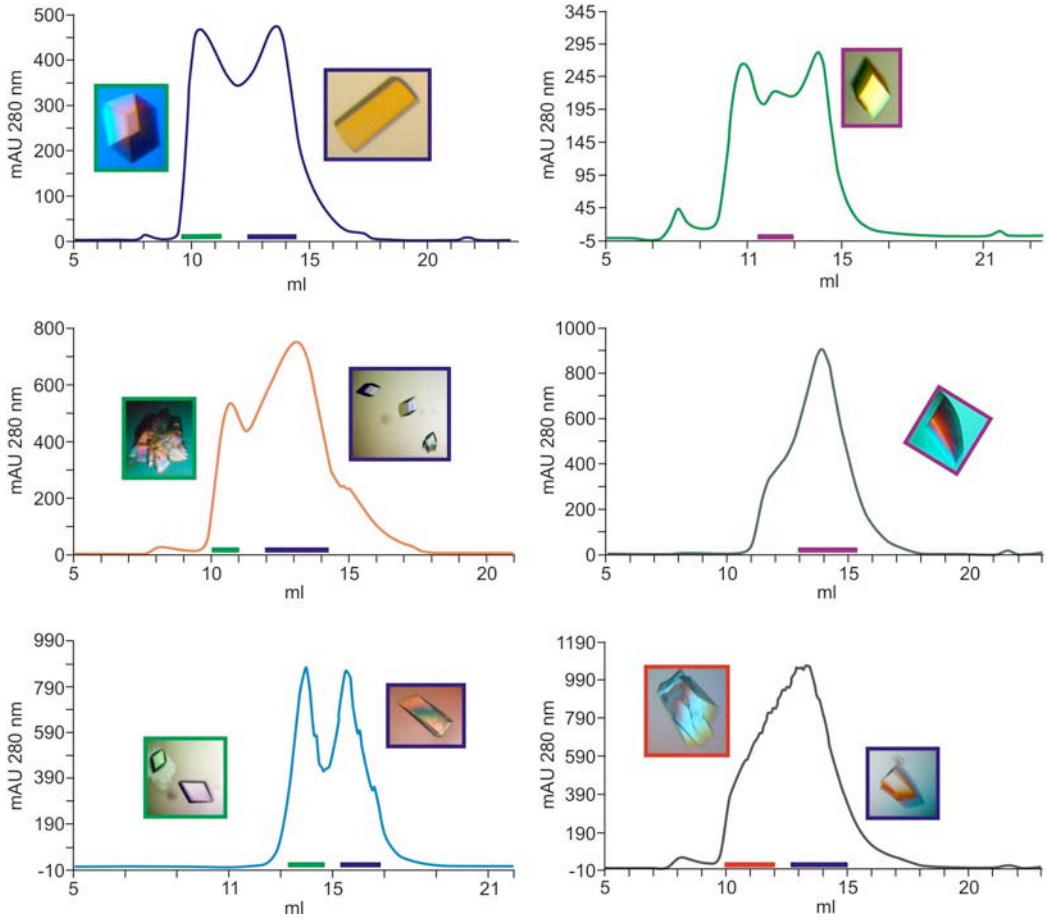


Figure 24| SEC chromatograms and crystals of BetA grown from pooled SEC fractions. Coloured bars in the peak volume indicate the fraction used for the crystal next to the peak, framed with the same line colour.

As for many SeMet-derivative proteins, SeMet-BetA crystallised under similar conditions as the native protein but required a slightly higher percentage of PEG400. The crystal shape and size was similar to those of native BetA (Figure 23). Optimal SeMet-BetA crystals grew in a clear drop. Crystal growth of SeMet-BetA followed the same timeline as for native BetA (4.2). Several times, in conditions with low PEG400 concentration between 10-16% no growth of crystal nuclei was observed for more than one week. The addition of 500µl 5M NaCl to the reservoir solution induced crystal growth and resulted in either one or more single crystals in a clear drop (Figure 25 d-1 to e-2). Crystals growing in a clear drop proved to be the best diffracting ones. Further optimisation for SeMet-BetA crystallisation led to the conditions shown in Table 7.

To test for possible different and better crystallisation conditions, SeMet-BetA samples at concentrations from 5 to 10 mg/ml were screened with several initial sparse matrix screens (3.2.2) by setting 400 nl sitting drops, with a 1:1 ratio of protein-to-reservoir using the Mosquito robot (TTP LabTech). The 96-well plates were placed at 18°C and resulted in several positive conditions, of which some are shown in Figure 25. Besides spherulites (c in Figure 25), other crystal types were obtained in the commercial screens (Figure 25). Bi-pyramidal crystals (b in Figure 25) grew in Hampton I, No.14 (200mM $CaCl_2$, 100mM HEPES-Na, pH 7.5, 28% PEG400) and crystals of rhomboid shape (a-1 to a-3 in Figure 25) grew in Nextal II, No. 33 (100mM $MgCl_2$, 100mM NaCl, 100mM Na-acetate, pH 4.5, 30% PEG400). A correlation between crystal size and diffraction quality was apparent. Crystals of ~200 μm across diffracted to ~5Å whereas crystals of 400 μm across diffracted to 2.8Å. To achieve the best crystallisation conditions for each purification, which always differed in their lipid and/or detergent content, test plates around the optimal condition (Table 7) were pipetted. The plates were then checked after three days to determine whether the precipitant concentration needed to be adjusted and if NaCl or PEG400 had to be added to the reservoirs. Crystals f-1 to f-3 in Figure 25 resulted from purified SeMet-BetA 18-12-2006.

However, the growth of a single crystal in a clear drop could not always be achieved. Two different types of precipitation were mainly observed, as shown in Figure 26. The first type usually had a brownish colour (Figure 26 a and c) may imply denatured protein. The second precipitation type is a colourless phase separation (Figure 26 b). This type of precipitation usually does not contain denatured protein; moreover small crystals were detected at higher magnification (Figure 26 b).

Protein precipitation is a sign of excess precipitant concentration and in the case of BetP prevented further crystallisation. However, in the case of BetA or SeMet-BetA, the best and largest crystals were obtained from drops that showed no precipitation of either type.

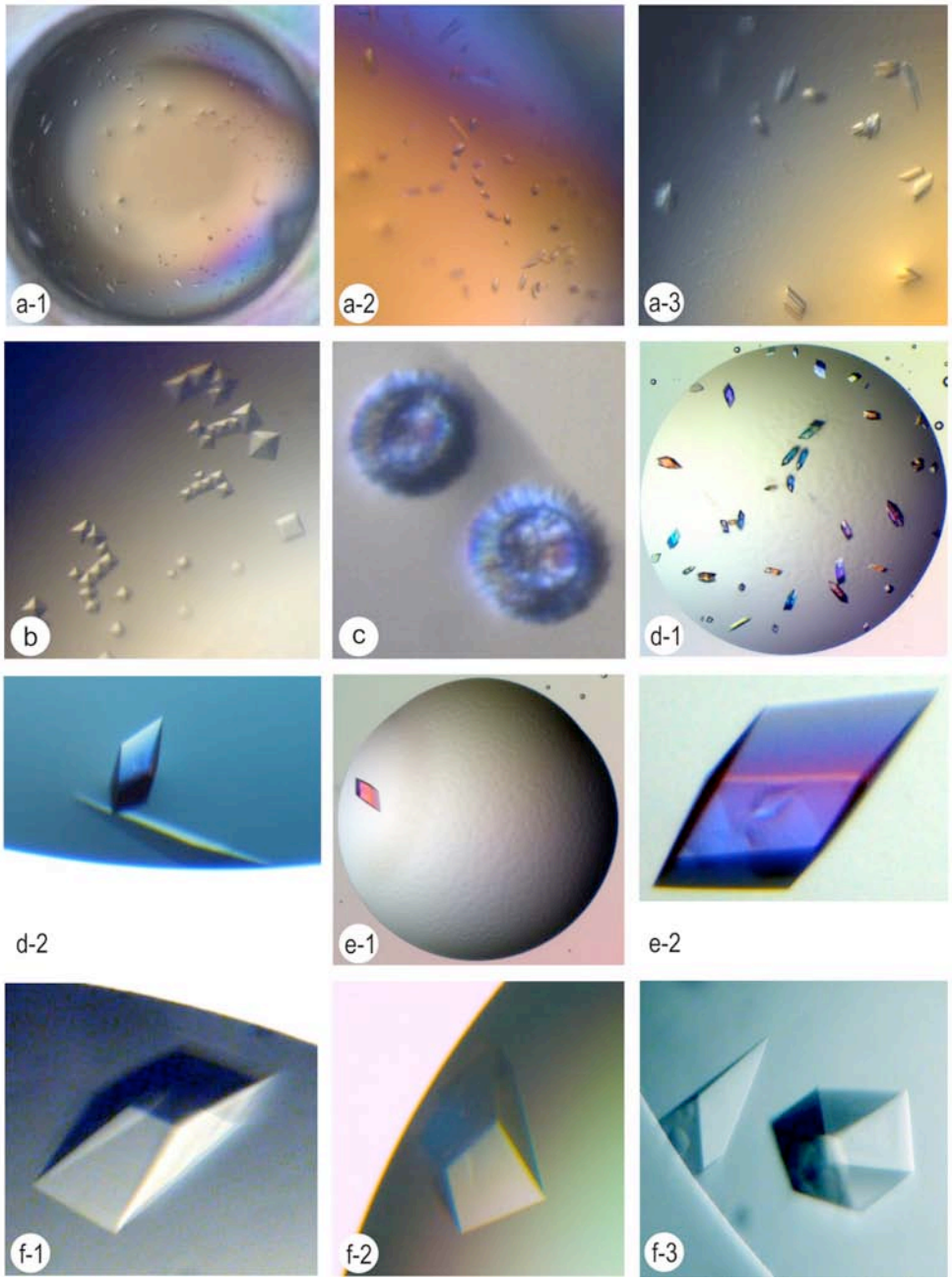


Figure 25| Crystal gallery of SeMet-BetA crystals. **a-1** to **c**, first hits from sparse matrix screens with a drop volume of 400 µl and a protein-to-reservoir ratio of 1:1. **a-1** Nextal Class II No. 33 (100 mM $MgCl_2$, 100mM NaCl, 100mM Na-acetate, pH 4.5, 30% PEG400); **a-2** and **a-3** are enlarged in pictures of drop **a-1**; **b**, Hampton I No. 14 (200 mM $CaCl_2$, 100 mM HEPES-Na, pH 7.5, 28% PEG400); **c**, spherulites growing in JBS 2 No. 29 (100 mM Zinc acetate, 50mM Tris pH 7.5, 10% PEG6000). Crystals shown in **a-1** to **c** derived from SeMet-BetA purification 08-08-2006 (~8 mg/ml). **d-1** and **d-2** show the crystals of SeMet 21-11-2007 (~12 mg/ml) grown in 14% PEG400, 100 mM Na_3-citrate buffer, pH 5.55, 150 mM NaCl; **e-1** and **e-2** show the crystal from purification SeMet-BetA-20-11-2007 (~10 mg/ml) grown in 16% PEG400, 100 mM Na_3-citrate buffer, pH 5.5, 100 mM NaCl; **f-1** to **f-3** from SeMet-BetA 18-12-2006 (~8 mg/ml) appeared in 19% PEG400, 100 mM Na_3-citrate buffer, pH 5.5, 100 mM NaCl. All crystals grew at 18°C.

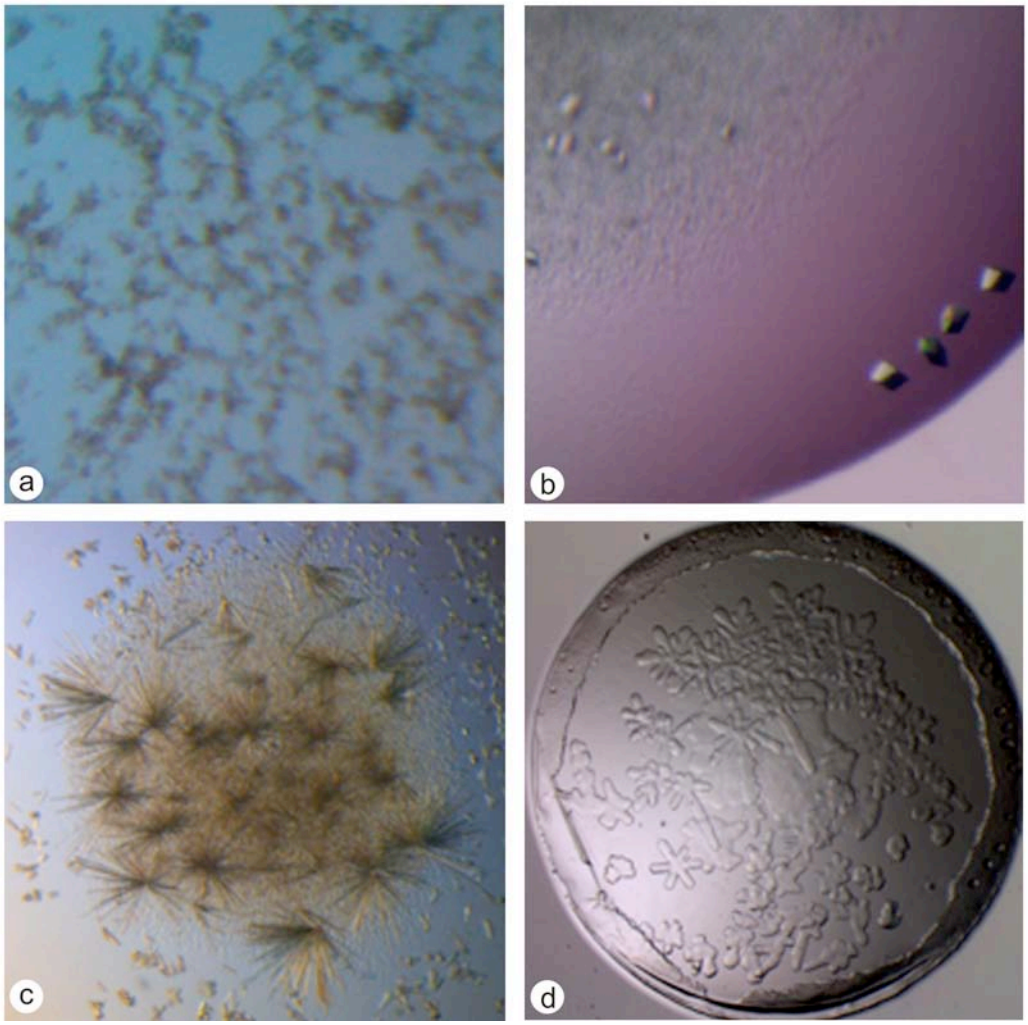


Figure 26| Precipitation types during crystallisation of SeMet-BetA. a brown precipitation**, b** colourless precipitation showing nuclei formation and crystal growth**, c** brown precipitation and needle like crystals**, d** colourless precipitation and oily phase separation.

4.6. Discussion

Heterogeneity of BetA and SeMet-BetA

The initial step and primary requirement for structure determination of a protein is to obtain it in sufficient and pure amounts. BetA and SeMet-BetA could be purified by recombinant technology using the Strep-II affinity tag, which led to a pure protein in sufficient amounts for structural studies. If the Strep-II affinity tag works for a certain system, it is preferable to the commonly used His-tag, because negatively charged cytoplasmic proteins or those containing histidine clusters are observed to bind non-specifically to the NiNTA column material. By comparison, interaction of the StrepII affinity tag with the Streptactin material is much more specific.

The purification of membrane proteins requires detergents for solubilisation, and subsequent detergent exchange is often necessary in order to obtain 3D crystals. Initially, a detergent was selected for its efficiency in the extraction process and in keeping the bioactivity of the membrane protein. Yet, this does not always lead to well-diffracting 3D crystals. It has been observed that changes in the alkyl chain of the detergent have an influence on 3D crystallisation (Michel, 2003; Palma et al., 1999; Shinzawa-Itoh et al., 1995). Extraction and further purification results in formation of diverse protein-detergent or protein-detergent-lipid complexes. However, the exact amount and composition of detergent and lipids being bound to the membrane protein is not easily assignable.

Variations in lipid and/or detergent content and different BetA and SeMet-BetA oligomers were detected by TLC, SEC and CN-/BN-PAGE (Figure 14, Figure 18, Figure 19, Figure 20, Figure 21 and Figure 22). The variable stoichiometry of either bound lipid, β-DDM or Cymal-5 might be one reason why the purification of a monodisperse sample was not feasible. Furthermore, a membrane protein that shows a tendency for 2D and 3D crystal growth might tend to form unspecific oligomers or aggregates during purification. Interestingly, the molecular weight ratios for higher BetP oligomers in the 4-12% CN-/BN-PAGE are all divisible by three. This implies that the different BetP oligomers must arise from trimer-trimer interactions. Another interesting feature observed in the 4-12% CN-/BN-gels is the double trimer band of BetA. Both trimers may have a qualitative or quantitative difference in lipid and detergent composition. Why this is only visible for the trimer fraction may be

explained by the high concentration of trimer fraction in the sample applied on the CN-/BN gels. Specific lipid mass-spectrometry (Ejsing et al., 2006; Tsai et al., 2007) analysis of the two different trimers, which can be cut out of the CN-/BN-gel, would give more information on the defined composition of bound detergent and lipid.

As determined by TLC (Figure 22) lipids remain bound in some BetA samples throughout the entire purification. In some cases, phospholipids have been shown to be important for stabilising solubilised membrane proteins (Hunte, 2005). Interestingly, the 2D crystal quality of BetP was observed to be very dependent on the lipid environment (Morbach and Kramer, 2005; Schiller et al., 2006; Tsai, 2008; Tsai et al., 2007; Tsai and Ziegler, 2005). However, addition of various lipids at the stage of affinity column purification did not result in better 3D BetP crystals (Ressl, 2006).

Lower temperature results in higher monodispersity?

Lower growth temperature during heterologous expression in *E. coli* cells positively influenced the monodispersity of BetA, as judged by SEC profiles in (Figure 14 and Figure 24). The profiles showed a single peak in the trimer range when BetA was purified from an overnight expression at 24°C. On the other hand, SEC profiles were very diverse when BetA derived from a three hour expression at 37°C. There was no difference in growth medium but in the induction point. The 24°C overnight BetA expression was induced below an OD_{600} of 1, whereas induction for the three hour expression at 37°C was chosen at an OD_{600} of ~1.8.

It is known that *E. coli* cells vary their membrane lipid composition (Cronan, 1968), which is dependent on growth temperature (de Siervo, 1969). The major phospholipid classes of *E. coli* (phosphatidyl ethanolamine (PE), cardiolipin (CL) and phosphatidyl glycerol (PG)) were quantified by phosphate assays at different stages of the growth cycle at 27° and 37°C (de Siervo, 1969). The major finding was a decrease of lipid phosphorus from cultures grown at 37°C and an accumulation of phospholipids in cultures grown at 27°C. Complex quantitative changes in patterns of the PE, PG and CL lipids were observed between these two temperatures as well (de Siervo, 1969). In the stationary growth phase at 27°C, when *E. coli* cells are usually harvested, PG is replaced by phosphatidyl glycerol-phosphate (PGP), which introduces additional negative charges to the membrane. Studies on chill activation of

BetP revealed a change in lipid composition of *C. glutamicum* cells upon changed growth temperatures as well (Ozcan et al., 2007). These studies showed that an increase of growth temperature resulted in a decrease of PG. As the activity of BetP is highly dependent on the lipid composition of the membrane (Rubenhagen et al., 2000), lower growth temperature during heterologous expression in *E. coli* cells might result in a optimal lipid composition and hence protect fold, stability and function of BetP during purification. In this work SEC profiles were monodisperse and crystal growths derived more readily from low temperature expression than of samples from high temperature expression. Phospholipids thus play an essential role in membrane protein folding and structure stabilisation.

BetP biosynthesis at 37°C may be reconsidered because longer expression time at lower temperature may promote better insertion and folding of BetP in the membrane. Furthermore, lowering the temperature to at least 27°C can result in an accumulation of important phospholipids in the membrane. Further careful analysis is necessary to find the effect of expression time, temperature and membrane lipid composition on the resulting BetP sample quality.

5. Crystallographic data collection, processing and statistics

Synchrotron radiation (Table 6) with its high brilliance and small beam size was necessary to collect suitable data sets for structure determination of BetP. Crystals were frozen prior to data collection as described in chapter 3.2.3. To reduce the effects of radiation damage, data were usually collected in several data sub-sets of one crystal by translation along one crystal axis (Figure 30). All data sub-sets were processed with XDS and scaled with either XSCALE or SCALA (3.3.6). Indexing was straightforward and revealed an orthorhombic crystal lattice. Screw axes along *a, b* and *c* were observed by systematic absences indicating the spacegroup $P2_12_12_1$. To collect a >90% complete data set for an orthorhombic crystal, a 90° rotation around one of the twofold axis was sufficient to cover the complete asymmetric unit cell, when the correct starting point is chosen. Data redundancy was accomplished by collecting several data sets from one crystal. BetA crystals were usually isomorphous so that scaling of data sets from different crystals was feasible. No variations over 4Å in each unit cell dimension were observed for BetA and SeMet-BetA crystals.

5.1. Native BetA

Two most promising native data sets were collected as described in 3.3.5 during the diploma work (Ressl, 2006). In the present thesis, these two native data sets were reprocessed and their statistics are presented below. Prior to data processing, correct beamstop masking of the images was necessary to prevent inclusion of overloads or other artefacts around the beamstop region, which influence data scaling. Therefore, beamstop masking improved data quality, especially when no separate low-resolution data were collected.

The two native data sets were reprocessed including all high resolution reflections with a mean $I/\sigma(I)$ value higher than 1. Although this resulted in high *R* values, especially in the high resolution shell, it was necessary to include all weak reflections from the higher resolution shell to compensate data loss from high data anisotropy. Each data set was analysed for its anisotropy by the programs *anisoscale* (3.3.7), PHASER (McCoy, 2007) or using the PHASER_MR script manually (script see 10.1). Using the *B* tensors from either the PHASER_MR or *anisoscale* output

(3.3.7) anisotropic data were initially corrected. The *B* tensor values were applied to the data using CAD (script see 10.3). Furthermore, an additional overall negative *B* factor was applied to the data using CAD (script see 10.4). To monitor the effect of anisotropy correction for each data set, the resulting profile from *anisoscale* before and after correction is shown in Figure 28 and Figure 32.

Table 8| Crystallisation conditions, data collection and statistics for crystal Native1.

Crystallisation condition:	19% PEG400, 100 mM NaCl, 100 mM Na_3-citrate buffer, pH 5.55 at 18°C
Cryo-condition:	28% PEG400, ~25% Glycerol, 100 mM NaCl, 100 mM Na_3-citrate buffer, pH 5.5
Data collection:	beamline ID 14:2 (ESRF), λ=0.933, exposure time=40 sec., oscillation range=0.5°, detector distance=285mm, data range 1-593 frames starting at 0°-0.5°, no data collection strategy used
Crystal parameter	a, b, c (Å) = 116.4, 128.7, 182.8 α, β, γ (°) = 90
Data statistics (Scala):	

Summary data for Project: BetA Crystal: nat1 Dataset: nat1

	Overall	InnerShell	OuterShell
Low resolution limit	39.47	39.47	3.56
High resolution limit	3.37	10.67	3.37
Rmerge	0.109	0.043	0.574
Total number of observations	355114	13581	6610
Total number unique	36061	1372	3061
Mean(I)/sd(I)	13.0	46.9	1.0
Completeness	88.9	97.0	53.0
Multiplicity	9.8	9.9	2.2

The data set Native1 was processed to a resolution of 3.37Å. The last resolution shell has high *R* factors (>50%) and low mean I/σ(I) <1. Nevertheless, 3061 unique reflections were observed in this resolution shell with a multiplicity >2. Weak and noisy data in the last resolution bin still contains valuable information. The large errors are taken into account during refinement using maximum likelihood algorithms. Weak data can be very valuable in the early stages of structure determination, *e.g.* phasing and heavy atom site search and during the use of the single isomorphous replacement/anomalous scattering (SIRAS).

The low to medium resolution bin of the native BetA data is acceptable as judged by statistics. When the native data sets Native1 and Native2 were collected in the diploma work (Ressl, 2006), no data collection strategy programs had been used and radiation sensitivity of the crystal had not been carefully evaluated. Later data collection in this work usually started with three test images per crystal that were initially processed to obtain the space group, unit cell parameters and crystal orientation with respect to the X-ray beam. From the latter, a specific starting and finitely angle for data collection was determined, defining the rotation wedge for the complete data range (Dauter, 1999; Dauter, 2005). Beam attenuation was tried out as well to diminish recorded overload reflections.

The directional dependence of diffraction was obtained in the diffraction image of the crystal (Figure 27). Whereas the upper left and lower right quadrant of the diffraction image in Figure 27 showed only a few diffraction spots, the upper right and lower left quadrant showed diffraction to a resolution of ~2.9Å. Analysis for data anisotropy of data set Native1 showed a high mean anisotropy *B* value of 67 $Å^2$ (Figure 28a).

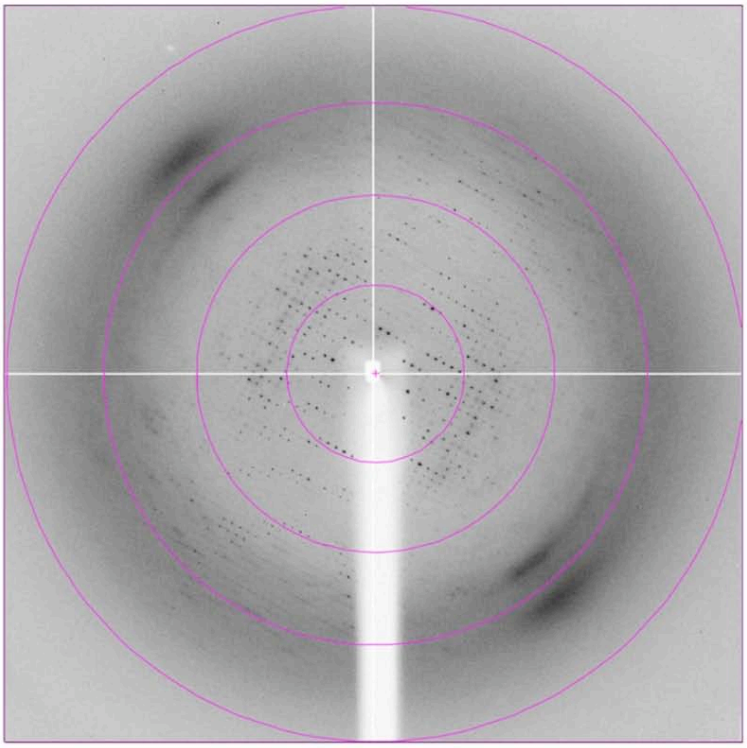


Figure 27| Anisotropic diffraction of the Native1 crystal. Different resolution bins from the inner to the outer purple circles: 11.8, 5.9, 3.9 and 2.9 Å.

Anisotropy correction for data set Native1 was achieved by application of the *B* tensors from SHARP: B11=33.6 B22=0 B33=-43.8 B12=0 B13=0 B23=10.2, using CAD (script see 10.3). The mean *B* factor decreased from a severe value of 67.28 Å^2 to lower anisotropic value of 49.15 Å^2 (Figure 28). The difference between the non-corrected and corrected data set Native1 is shown by the profiles in Figure 28a and b.

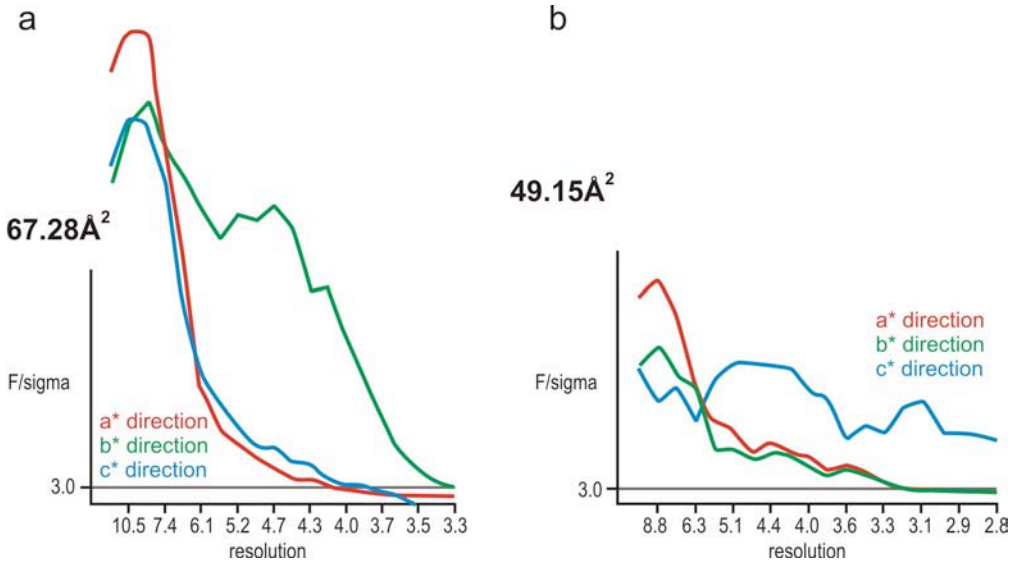


Figure 28| Correction for data anisotropy of Native1. a shows diffraction F/sigma for the non-corrected data set Native1 on the y axis for each direction *a** (red), *b** (green) and *c** (blue) against resolution in Å on the x axis for the non-corrected data set Arg5; **b** shows the same plot for the corrected data set.

Table 9| Crystallisation conditions, data collection and statistics for crystal Native2.

Crystallisation condition:	18% PEG400, 250 mM NaCl, 100 mM Na_3-citrate buffer, pH 5.55 at 18°C
Cryo-condition:	28% PEG400, ~25% Glycerol, 100 mM NaCl, 100 mM Na_3-citrate buffer, pH 5.5
Data collection:	beamline PXII (SLS), λ=0.98, exposure time=3 sec., oscillation range=0.5°, detector distance=300mm, data range 1-397 frames starting at 0°-0.5°, no data collection strategy used
Crystal parameter	a, b, c (Å) = 118.0, 130.1, 183.4 α, β, γ (°) = 90
Data statistics (Scala):	

Summary data for Project: BetA Crystal: Native2 Dataset: Native2

	Overall	InnerShell	OuterShell
Low resolution limit	39.47	39.47	3.35
High resolution limit	3.18	10.05	3.18
Rmerge	0.105	0.039	0.551
Total number of observations	279509	11609	13831
Total number unique	44496	1599	4750
Mean(I)/sd(I)	9.3	29.6	1.0
Completeness	92.3	95.1	69.2
Multiplicity	6.3	7.3	2.9

Data set Native2 was processed to a resolution of 3.18 Å and has a slightly lower *R* factor than Native1 but showed higher data anisotropy of 69.75 Å^2 (Figure 29). After correction, data set Native2 was usable to a resolution of 3.18 Å with a mean anisotropy *B* value of 51.24 Å^2.

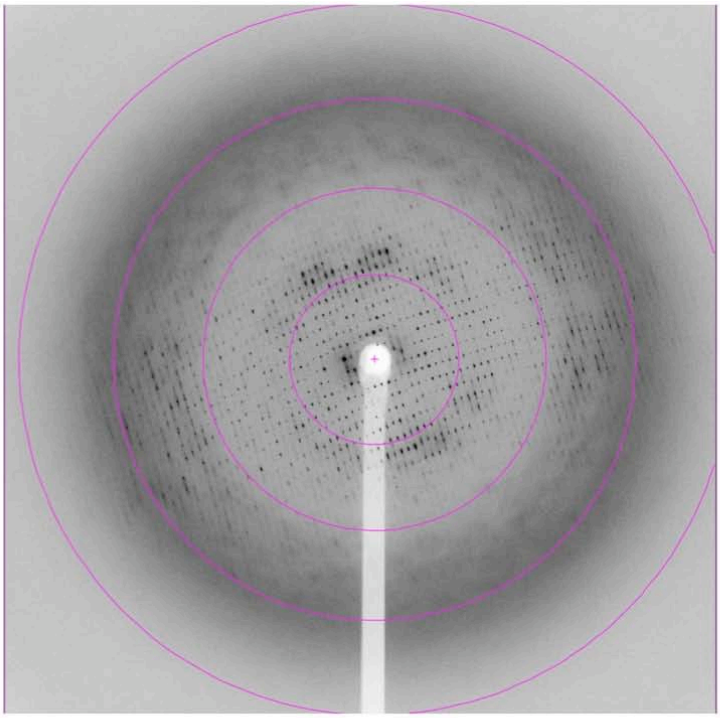


Figure 29| Anisotropic diffraction image of the Native2 crystal. Different resolution bins from the inner to the outer purple circles: 10.9, 5.5, 3.6 and 2.7 Å.

5.2. SeMet-BetA

A fluorescence scan was used to assess the incorporation of selenium and to determine the correct wavelength for MAD or SAD experiment (3.3.8) prior to data collection on SeMet-BetA crystals. Form factor values f' and f'' were determined from the fluorescence scan. The SAD data set SeMet1 was collected at the PXII beamline at the SLS (Table 6) using the peak wavelength of 0.9794 Å. Four peak data sub-sets were collected to achieve high anomalous data redundancy. To reduce the effects of radiation damage, the four data sets were collected from different positions on the crystal (Figure 30). After indexing and merging the four data sets with XDS, the Friedel mates were treated separately, and converted with POINTLESS to MTZ format. Each data batch was sorted and assigned in SORTMTZ, and scaled together using SCALA. Table 10 shows the statistics for the scaled data from all four data sets. The redundancy-weighted merging R factor R_{meas} lies between 10.4 and 11.3% for the merged datasets.

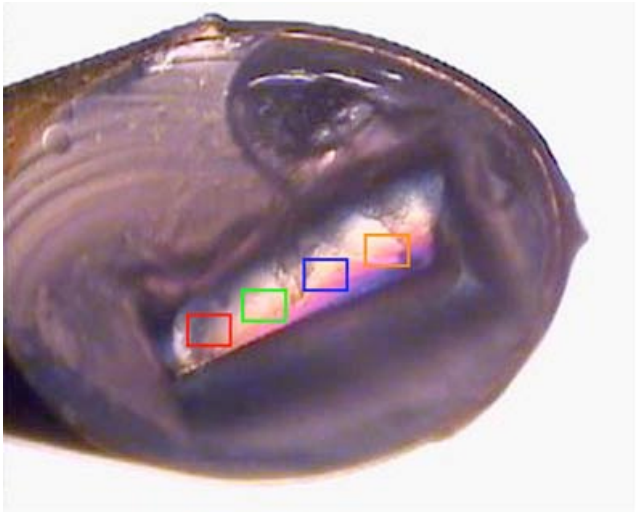


Figure 30| Crystal SeMet1 in a Litho loop. The red, green, blue and orange rectangles show the positions of the corresponding data of run1, run2, run3 and run4, respectively.

Data set SeMet1 was processed to a resolution of 3.35 Å. Although data quality statistics were better when compared with the native data, data anisotropy was very high, with a mean anisotropy *B* value of 92.37 Å^2 (Figure 32a). The first correction for data anisotropy of SeMet1 was accomplished by application of the *B* values from SHARP: B11=37.59 B22=0 B33=-43.95 B12=0 B13=0 B23=6, using CAD (script see 10.3). This initial correction together with the second correction during phasing and refinement as described in chapter 3.3.8, 3.3.12 and 7.2, resulted in a lower mean *B* value of 33.4 Å^2 (Figure 32b). The anomalous signal from the data set SeMet1 could be used for heavy atom site location and further structure determination (7.1, 7, 7.4, 8).

Table 10| Crystallisation conditions, data collection and statistics for crystal SeMet1

Crystallisation condition:	19% PEG400, 100 mM NaCl, 100 mM Na_3-citrate buffer, pH 5.5 at 18°C
Cryo-condition:	Paraffin
Data collection:	beamline PXII (SLS), λ=0.97944, exposure time=1sec., attenuator=Al3 (17% beam intensity (BI), oscillation range=0.5°, detector distance (dd)=440mm, data range run1=120°-210°, run2=30°-200° run3=120°-210°, attenuator=Al2+Al1 (29% BI) run4=120°-210°, attenuator=Al2+Al1 (29% BI), exposure time=2sec., dd=350mm
Crystal parameter	a, b, c (Å) = 118.3, 129.4, 183.6 α, β, γ (°) = 90

Data statistics (Scala):

Summary data for Project: Se Crystal: semet1 Dataset: semet1

	Overall	InnerShell	OuterShell
Low resolution limit	39.47	39.47	3.54
High resolution limit	3.36	10.62	3.36
Rmerge	0.099	0.044	0.648
Rmeas (within I+/I-)	0.104	0.046	0.870
Rmeas (all I+ & I-)	0.113	0.057	0.933
Rpim (within I+/I-)	0.032	0.014	0.574
Rpim (all I+ & I-)	0.025	0.013	0.533
Total number of observations	701985	27656	4374
Total number unique	37213	1393	2526
Mean(I)/sd(I)	23.2	89.3	1.0
Completeness	90.7	96.7	44.1
Multiplicity	18.9	19.9	1.7
Anomalous completeness	85.2	97.6	11.9
Anomalous multiplicity	10.4	11.5	1.6
DelAnom correlation between half-sets	0.709	0.747	-0.055
Mid-Slope of Anom Normal Probability	1.287		

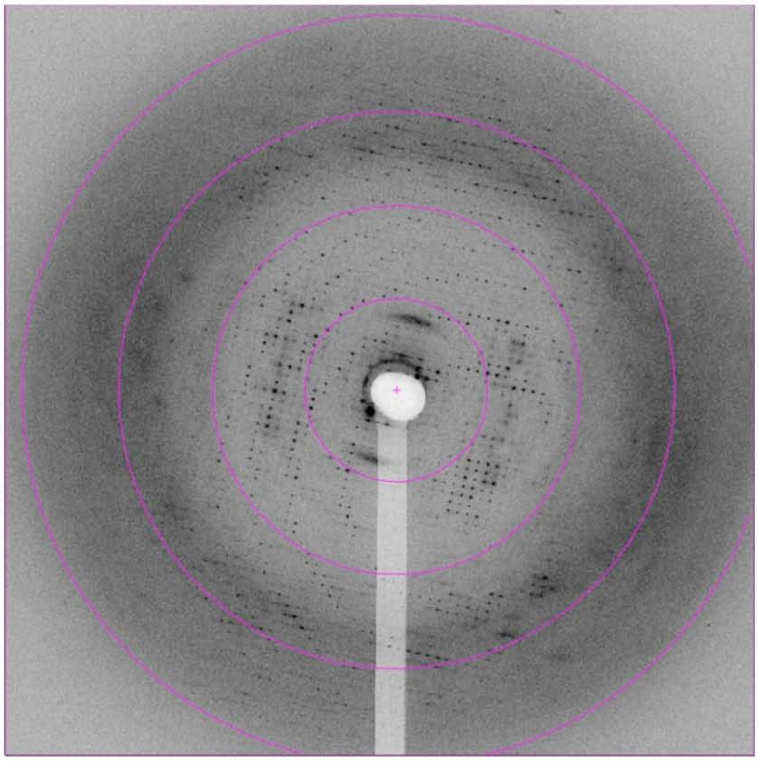


Figure 31| Anisotropy diffraction of the SeMet1 crystal. Different resolution bins from the inner to the outer purple circles: 15.5, 7.3, 5.2, 3.5 Å.

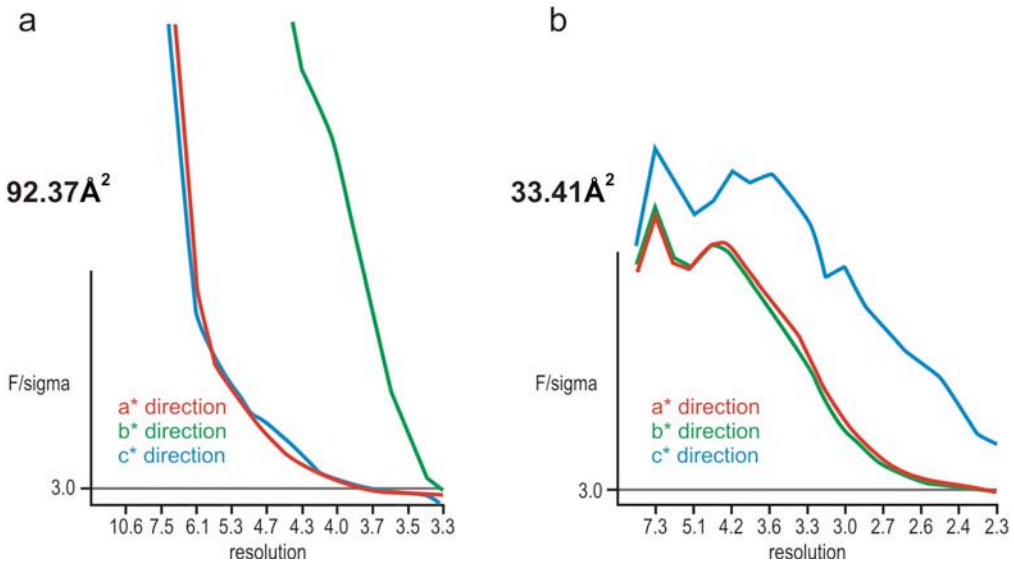


Figure 32| Correction for data anisotropy of SeMet1. a shows diffraction F/sigma on the y axis for each direction *a** (red), *b** (green) and *c** (blue) against resolution bins on the x axis in Å. **b**, shows the graph for data set SeMet1 after the final anisotropy correction during the refinement procedure (7).

5.3. Discussion

BetA and SeMet-BetA crystals rarely diffracted to resolutions above 3Å, were affected by data anisotropy, and due to their high solvent content, crystals were affected by radiation damage. These factors have an impact on downstream data processing and might determine the feasibility of structure determination. Therefore data collection and processing had been treated with special care. The following describes some important considerations for data collection, treatment of weak data and data anisotropy of BetA and SeMet-BetA crystals that might also be valid for other protein crystals of similar quality.

Medium to low-resolution data collection

Apart from knowing what to expect from a crystal and what one is aiming for in a specific project, one should always have a plan of how to collect the best data set possible (Dauter, 1999). Collection of a good data set depends on many variables, which are either due to the nature of the crystal, or arise from the specifications of the diffractometer or the beamline at a Synchrotron facility.

If medium-resolution data sets are expected, then the first aim in data collection should be a correct and completely measured low-resolution data set. Correct and complete low-resolution data is crucial for structure determination because it contains measured intensities that harbour important information concerning the overall shape of the molecule. Incomplete low-resolution data results in a loss of information about the overall shape and possible connecting features in the electron density map.

Low-resolution data collection should avoid recording overloads if possible. CCD detectors have a limit in the number of bits per pixel that can be recorded by the electronic read-out system. Overloads are pixels on a CCD detector, which are electronically "overloaded" beyond the bit depth of the detector and thus intensities that are too strong cannot be measured properly. All overloaded measurements get rejected in the scaling step and are then missing from the data set, resulting in incomplete information.

After initial structure determination, the model can subsequently be refined against an independently collected high-resolution data set. Special care should be taken in structure determination by molecular replacement. This method relies on the

Patterson function (3.3.4) calculated from the square of the amplitudes, and is thus very much depend on the strongest reflections. Therefore, all of the strongest reflections should be present in the data.

Macromolecular crystals often show a high degree of mosaicity. Mosaicity arises from slight disorientation of the crystal-building units, the mosaic blocks. Since data are collected by the rotation method, the data set is collected from successive exposures corresponding to the successive small rotation ranges of the crystal. Some reflections, which started diffraction during one exposure, will still diffract at the next exposure. As a consequence various fractions of the total intensity of such "partials" are recorded in more than one image. By contrast, the fully recorded reflections are those for which all intensity is recorded in a single diffraction image. Reflections on the reciprocal lattice are grouped in families of parallel planes and are grouped in ellipses. Partial reflections reside near the edges of the ellipse and if the rotation angle is too wide, those partial reflections appear in the next diffraction plane, causing them to overlap and thus making indexing of these data images difficult. Therefore data collection of highly mosaic crystals should be recorded in small rotation angles choosing a minimum rotation of half the mosaicity value, also known as fine-slicing data collection (Dauter, 1999). Crystal annealing can help to lower crystal mosaicity. This method involves warming the flash-cooled crystal to room temperature and flash-cooling it again (Harp et al., 1998). Annealing of BetP crystals did not result in improved diffraction.

Radiation damage during data collection

Radiation damage has to be taken into account especially when data are collected at undulator sources of third-generation synchrotrons. During primary radiation damage events, electrons in the outer atomic orbital absorb a highly energetic X-ray quantum and get released. As a direct result, free electrons lead to the breakage of interatomic bonds in protein crystals. Secondary events result from the propagation of radicals created by the absorption events. The secondary damage can be diminished by data collection at low temperature, even though the primary damage does not depend on the sample temperature. The global effect is a loss of diffraction by the protein crystal.

Since BetP crystals suffered from severe radiation damage, it was beneficial to collect a complete and redundant data set from more than one crystal or, if the crystal size allowed, more data sub-sets from different parts of one crystal and merge these data sub-sets together (Figure 30). This is also beneficial if the crystals are isomorphous to each other. If data redundancy is high enough, data suffering from radiation damage can be corrected by the interpolation of the intensities to zero-damage state (Diederichs et al., 2003).

Weak intensities

When maximum-likelihood based programs are used, the inclusion of weak intensities in all subsequent structure determination steps does not effect structure determination negatively. In former days the conventional lowest data cut-off based on the $I/\sigma(I)$ was 1.5 to 1.3. This meant that all data below this threshold were usually discarded. When dealing with medium to low-resolution data, this might decide whether structure determination is possible or not. With the emergence of maximum-likelihood based programs (Adams et al., 1997; Adams et al., 1999; Blanc et al., 2004; McCoy et al., 2004; Murshudov et al., 1997; Pannu et al., 1998; Read, 2001; Terwilliger, 2000; Terwilliger, 2001), it is now possible to include all weak ($I/\sigma(I)$ <1.3) diffraction data for structure determination. BetA and SeMet-BetA data contained weak intensities with large experimental error estimates. These were included because they were automatically down-weighted by the maximum-likelihood algorithm (3.3.12). In general, including all data into all subsequent steps of structure determination allows the use of all present information and results in improved electron density maps. All data should generally be included when dealing with medium-resolution data, especially if the data is anisotropic (Brunger et al., 2009b).

Processing strategy for suboptimal data

Suboptimal medium-resolution data for crystallographic structure determination is a data set with Bragg spacing below a resolution value necessary to resolve the mean distance of a covalent chemical bond (~1.3 Å). In addition, it has weak intensities is relatively noisy, reflected by high *R*-factors. Furthermore, it may be anisotropic, merohedrally twinned or contain a large $I/\sigma(I)$ falloff due to radiation damage at the latest stages of data collection.

There is no common rule for processing difficult data and each project has its own challenges. Nevertheless, for initial processing, XDS is favoured over MOSFLM by many crystallographers dealing with difficult medium to low resolution data (personal communication by various people in the field). One of the reasons is 3D profile fitting of the intensities in XDS compared to the 2D profile fitting in MOSFLM. XDS allows variation in data statistics to be observed over several sections of images. Hence, low quality data can be separated and treated differently; for example with other resolution cut-offs. Afterwards, specifically treated data can be merged and scaled together. The above mentioned zero-dose correction for radiation damage is implemented in XDS and can be turned on or off, depending on the severity of radiation damage.

Correction for data anisotropy

Correction of data anisotropy is crucial for all subsequent data handling steps in structure determination. The main problem in anisotropy correction is the convention between different programs: how they analyse data anisotropy, how they correct it and how their output files are formatted.

For initial anisotropy correction it is recommended to use PHASER with the script 10.1. This writes out a corrected MTZ file with additional corrected *F_iso* and *SIGF_iso* columns. The server program *anisoscale* uses PHASER for anisotropy correction but does an additional ellipsoid truncation of the data. This ellipsoid truncation should be used with care on medium resolution data. The truncation algorithm is primitive and might discard too many weak but well measured reflections and thus reduces the overall resolution.

Furthermore, PHASER records the anisotropic tensors *B11*, *B22* and *B33* in the output file, but not the remaining directional information *B12* for the *h/k*, *B13* for the *h/l* and *B23* for the *k/l* plane. All six values are necessary for correction of data anisotropy with CAD. However, the first corrected MTZ file from PHASER with its initial isotropic *F* and *SIGF* values can be further corrected in CAD using script 10.4. By applying a negative isotropic *B*-factor (B-factor sharpening) to the initially corrected data, its higher resolution terms get restored and revaluated, which results in an increase in detail for higher resolution features. Yet, the cost of more detail is an increased noise level in the resulting electron density map.

During phasing in SHARP, data is analysed for its anisotropy as well, but is not automatically corrected. SHARP writes all six *B*-factor values into the logfile that can be further applied in CAD. With a model available, the values deriving from SHARP lead to a further correction for data anisotropy. This is due to a better estimation of these values using F_{calc} from a (partial) model. Therefore, it is a very good idea to correct data anisotropy in all subsequent steps of structure determination. It leads to a nominally isotropic data set in which the slope of *F* versus resolution should be almost equal in all directions. It should be noted that $F/\sigma F$ does not change because both are scaled by the same multiplication constant.

As an experimental correction for crystal anisotropy, controlled dehydration might be considered (Kuo et al., 2003). Crystal dehydration was already carried out by Max Perutz in the mid-1940s (Perutz, 1946). Controlled dehydration of the crystal leads to reduction of solvent content and thus possibly to a closer packing and better ordered crystals. The correction for anisotropy by this method can be explained by the promotion of more crystal contacts by reduction of the solvent content. The controlled dehydration experiment requires access to the experimental set up at the Max-Planck-Institute of Biochemistry in Martinsried and it may be beneficial to improve diffraction quality of BetP crystals.

6. Phasing

The phase problem in X-ray crystallography is one of the main cruxes in structure determination. The single-wavelength anomalous dispersion (SAD) or multi-wavelength anomalous dispersion (MAD) approach allows analysis of recorded anomalous differences in the data and determination of phase information. Anomalous data collection (3.3.8) was carried out at the beamline PXII at the SLS and ID14-4, ID23-1, ID29 at the ESRF. After fluorescence scans around the specific edge of the derivative element (Table 11), the wavelength values for the inflection and peak region were obtained (3.3.8). For SAD experiments, data were collected at the peak wavelength. For MAD data collection the remote data was collected first, followed by the inflection data and finally the peak data. This is the usual practice in MAD data collection to minimise radiation damage.

Table 11| Properties for anomalous data collection of several elements. The elements used for the anomalous data collection are listed with the edges used given in keV and Å. The theoretical values for edges of the elements are taken from the http://skuld.bmsc.washington.edu/scatter/ website. Numbers marked with an asterisk diverge from these theoretical values and were used in the experimental setup.

Element	Edge	keV	Å
Se	K	12.6578	0.9795
Xe	L-I	5.4528	2.2738
		6.50065*	1.90726*
Ta	L-III	9.8811	1.2548
Br	K	13.4737	0.9292
Rb	K	15.1997	0.8157
Hg	L-III	12.2839	1.0093

The heavy atoms listed in Table 5 were screened in this work, either for crystal soaking or co-crystallisation. Most resulted in diffraction loss of the crystals. Phasing approaches using the noble gas Xe (6.1), the tantalum cluster $[Ta_6Br_{12}]^{2+}$ (6.2) and Selenium (7.1) resulted in collection of anomalous SAD or MAD data sets.

6.1. Phasing with Xenon

The inert noble gas xenon (Xe) is sometimes used for mild derivatisation of native crystals. It diffuses freely through the crystal and usually binds to small hydrophobic patches without causing structural rearrangements. This often results in highly isomorphous derivative crystals. The advantage of phasing with Xe is its weak interaction with the protein. Moreover, it binds to hydrophobic cavities, which makes it very suitable for membrane protein X-ray crystallography. Whereas "classical" heavy atoms, such as Hg or Pt (Table 5), tend to bind strongly to protein and often induce protein degradation or brakeage of crystal contacts, leading to loss of diffraction. Incorporation of Xe can be achieved by pressurising a native crystal with xenon gas in a specific pressure cell (Oxford Cryosystems) (Figure 33). By adjusting the gas pressure the number of potential binding sites and their occupancies can be changed.

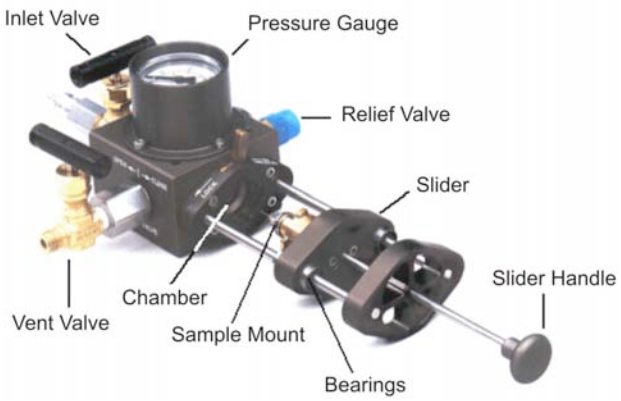


Figure 33| Pressure cell from Oxford Cryosystems used for Xe derivatisation. The crystal is mounted on the sample mount and is then transferred into the pressure chamber. Xe gas is applied through the inlet valve into the chamber and the pressure can be applied with a hand pump and monitored with the pressure gauge. After the required incubation time (min), Xe gas can be released slowly by opening the vent valve.

Crystals for Xe derivatisation derived from the expression BetA-03-11-2006 and were grown in 96-well plates with a drop volume of 1µl for ease of transportation (Figure 34).

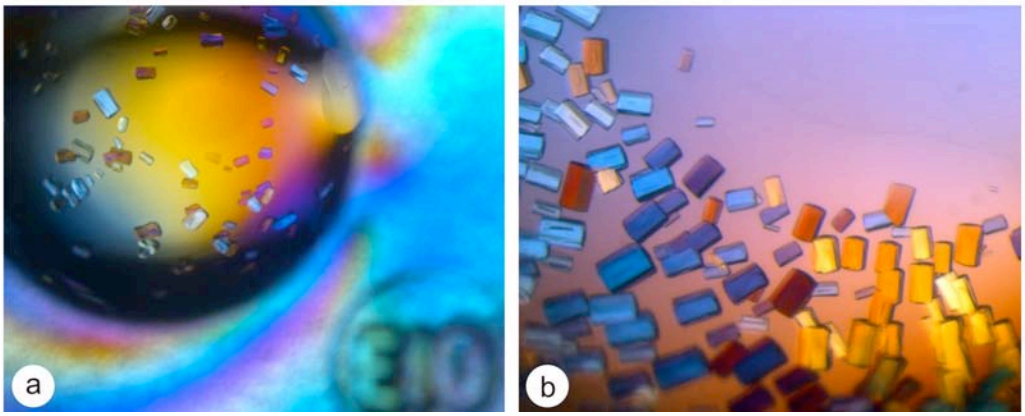


Figure 34| Native crystals of BetA grown in sitting drops for Xe derivatisation. **a** overview of a 1µl sitting drop with crystals, **b** enlarged picture on crystal used for Xe derivatisation.

Phasing experiments with Xe were performed at the PXII beamline at the SLS (Table 6). This beamline is equipped with a pressure cell, which allowed on-site experimental handling and parameter adjustments. The wavelength range of the PXII beamline ranges from 0.6199 – 1.9075Å, which did not allow a prior fluorescence scan for Xe nor peak data collection at the exact Xe L-I edge wavelength of 2.2738Å (Figure 35). Instead, Xe data were collected at a wavelength of 1.90719Å, which relates to an absorbance number of 10.22 electrons (*e*). As the peak wavelength of 2.2738 Å corresponds to absorption of 12.91 *e*, data collection at 1.90719Å matches more a high remote data than peak data (Figure 35). However, the anomalous signal can still be detected from such data.

Specific data collection conditions and statistics for data sets Xe1 and Xe3 are given below. Both were collected at the same wavelength but differed in Xe pressure applied and cryoprotection (Table 12). Xe data were collected in several sub-data sets, which were in turn collected from different positions on the crystals (Figure 30).

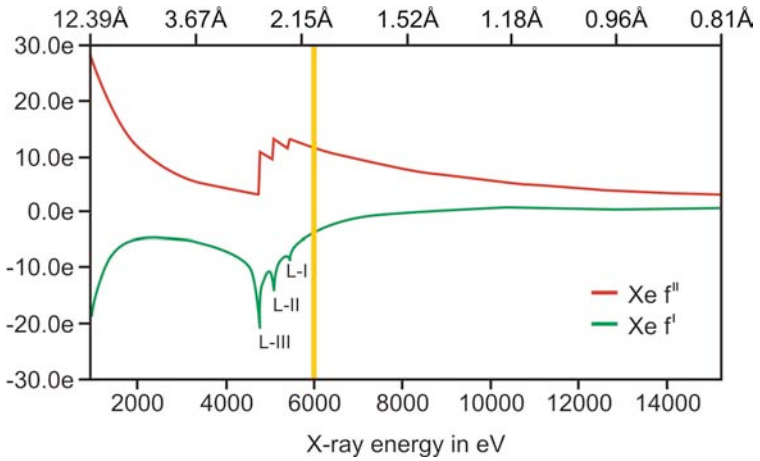


Figure 35| Theoretical Xe scan showing the L-III, L-II and L-I edge. The upper x-axis shows the wavelength in Å and the corresponding energy in keV is plotted on the lower x-axis. The y-axis shows the number of electrons (*e*) absorbing at the specific wavelength for the element Xe. The form factor *f*'' is plotted in green and *f*' in red. The yellow line indicates the position of the measured wavelength in this experiment. The three absorption edges are indicated by L-III, L-II and L-I.

Table 12| Experimental conditions for Xe derivatisation of native crystals.

conditions	Xe1	Xe3
pressure (bar)	10	8
time (min)	20	20
Cryoprotection	Paraffin	-

Crystal Xe1 was cryoprotected with paraffin oil prior to incubation in the pressure cell at 10 bar for 20 minutes in the Xe atmosphere. After release of the pressure the crystal was immediately flash frozen into liquid nitrogen. Crystal Xe3 was incubated at 8 bar for 20 minutes and was directly flash frozen without paraffin oil. Xe data was processed with XDS, treating the Friedel mates separately, converted to an MTZ file using POINTLESS and specific batch numbers were assigned for each run in SORTMTZ. Further data scaling was accomplished with SCALA and the anomalous signal was analysed with Xprep (Bruker Nonius) (Table 14, Table 17). Xe sites were searched using SHELXD and further phasing and refinement of initial sites was performed with SHARP/autoSHARP.

Table 13| Crystallisation conditions, data collection and statistics for crystal Xe1

Crystallisation condition:	19% PEG400, 150 mM NaCl, 100 mM Na_3-citrate buffer, pH 5.5 at 18°C
Cryo-condition:	Paraffin
Data collection:	beamline PXII (SLS), λ=1.90719, exposure time=0.3 sec., attenuator=none 100% BI, oscillation range=1°, detector distance (dd)=180 mm, run1, run2=100 frames starting at 148° run3, run4=100 frames starting at 148° run5, run6=100 frames starting at 223°
Crystal parameter	a, b, c (Å) = 118.38, 129.05, 183.34 α, β, γ (°) = 90

Data statistics (Scala):

Summary data for Project: Xe Crystal: Xe1 Dataset: Xe1-merged

	Overall	InnerShell	OuterShell
Low resolution limit	30.00	30.00	3.16
High resolution limit	3.00	9.49	3.00
Rmerge	0.213	0.056	2.092
Rmeas (within I+/I-)	0.223	0.058	2.233
Rmeas (all I+ & I-)	0.227	0.063	2.225
Rpim (within I+/I-)	0.065	0.016	0.763
Rpim (all I+ & I-)	0.049	0.013	0.608
Total number of observations	1083650	41657	62850
Total number unique	52124	1913	4982
Mean((I)/sd(I))	8.4	40.2	1.1
Completeness	91.6	96.7	61.1
Multiplicity	20.8	21.8	12.6
Anomalous completeness	83.2	97.6	35.3
Anomalous multiplicity	10.2	11.5	7.2
DelAnom correlation between half-sets	0.447	0.635	-0.040
Mid-Slope of Anom Normal Probability	1.084	-	-

Although data statistics for data set Xe1 show an $I/\sigma(I)$ of 1.1 up to a resolution of 3 Å, R factors higher than 100% indicated that higher resolution reflections were not reliable (Table 13). Therefore anomalous data analysis was carried out up to 3.5 Å and location searches for Xe sites in SHELXD were conducted using the data set Xe1 up to 4.5 Å, based on the cutoff of 1.3 $I/\sigma(I)$ for anomalous signal data analysis from Xprep (Table 14).

Table 14| Xprep anomalous signal statistics for Xe1.

resolution	Inf	8.0	6.0	5.3	5.1	4.9	4.7	4.5	4.3	4.1	3.9	3.7
bin Å	8.0	6.0	5.3	5.1	4.9	4.7	4.5	4.3	4.1	3.9	3.7	3.5
anomalous Signal[a)]	5.14	3.21	2.02	1.72	1.64	1.60	1.54	1.45	1.39	1.14	1.33	1.27
anomalous Signal[b)]	5.37	2.72	1.75	1.51	1.45	1.42	1.36	1.32	1.28	1.31	1.26	1.24

a) signal based on input sigmas
b) signal based on variances of *F*+ and *F*-

Six Xe sites with occupancies greater than 0.5 were found after several cycles running the program SHELXD and were further refined using SHARP/autoSHARP. The resulting electron density map from SHARP did not show interpretable density.

Table 15| Heavy atom site parameters for data set Xe1.

Site	**x**	**y**	**z**	**Occ**	**Site**	**x**	**y**	**z**	**Occ**
Xe1	24.355	22.944	7.316	1.000	**Xe4**	19.242	24.062	3.582	0.587
Xe2	30.24	21.577	5.729	0.898	**Xe5**	14.987	20.125	4.258	0.558
Xe3	27.505	26.353	7.753	0.740	**Xe6**	27.049	18.102	5.525	0.550

Data set Xe3 was collected in five data sub-sets (Table 16), which were processed, merged and scaled in the same way as Xe1. Data statistics for data set Xe3 showed comparable *R* factors and $I/\sigma(I)$ to data set Xe1. Interestingly, the anomalous signal obtained in Xe3 (Table 17) was a factor of ~ 2.5 lower than the signal obtained in Xe1 and site search with SHELXD did not result in Xe sites with occupancies higher than 0.5. Whether this difference in anomalous signal derives from the different treatment prior to pressurisation and flash freezing cannot be concluded by analysis of these two data sets. However, more experiments would be required to find out whether paraffin oil enhances binding of Xe gas in crystals that in turn might result in higher occupancies.

Table 16| Crystallisation conditions, data collection and statistics for crystal Xe3

Crystallisation condition:	19% PEG400, 150 mM NaCl, 100 mM Na_3-citrate buffer, pH 5.5 at 18°C
Cryo-condition:	none
Data collection:	beamline PXII (SLS), λ=1.90719, exposure time=0.3 sec., attenuator=none 100% BI, oscillation range=1°, detector distance (dd)=180 mm, data range run1, run2, run3 =100 frames starting at 253° run4 =100 frames starting at 340° run5 =100 frames starting at 0°
Crystal parameter	a, b, c (Å) = 118.94, 128.94, 185.24 α, β, γ (°) = 90
Data statistics (Scala):	

Summary data for Project: Xe Crystal: Xe3 Dataset: Xe3-merged

	Overall	InnerShell	OuterShell
Low resolution limit	30.00	30.00	3.16
High resolution limit	3.00	9.49	3.00
Rmerge	0.213	0.056	2.092
Rmeas (within I+/I-)	0.223	0.058	2.233
Rmeas (all I+ & I-)	0.227	0.063	2.225
Rpim (within I+/I-)	0.065	0.016	0.763
Rpim (all I+ & I-)	0.049	0.013	0.608
Total number of observations	1083650	41657	62850
Total number unique	52124	1913	4982
Mean((I)/sd(I))	8.4	40.2	1.1
Completeness	91.6	96.7	61.1
Multiplicity	20.8	21.8	12.6
Anomalous completeness	83.2	97.6	35.3
Anomalous multiplicity	10.2	11.5	7.2
DelAnom correlation between half-sets	0.447	0.635	-0.040
Mid-Slope of Anom Normal Probability	1.084	-	-

Table 17| Xprep anomalous signal statistics for Xe3.

resolution	Inf	8.0	6.0	5.3	5.1	4.9	4.7	4.5	4.3	4.1	3.9	3.7
bin Å	8.0	6.0	5.3	5.1	4.9	4.7	4.5	4.3	4.1	3.9	3.7	3.5
anomalous Signal[a)]	1.89	1.99	1.68	1.59	1.56	1.47	1.48	1.30	1.22	1.23	1.14	1.10
anomalous Signal[b)]	1.94	1.53	1.32	1.31	1.28	1.24	1.26	1.17	1.13	1.18	1.12	1.11

a) signal based on input sigmas
b) signal based on variances of *F*+ and *F*-

6.2. Phasing with Ta_6Br_{12}

Clusters represent a special class of heavy metal compounds that can be used for phase determination by isomorphous replacement or anomalous scattering methods. At low resolution, the individual scatterers in heavy atom clusters become indistinguishable and behave as one large "super-atom" (Banumathi et al., 2003).

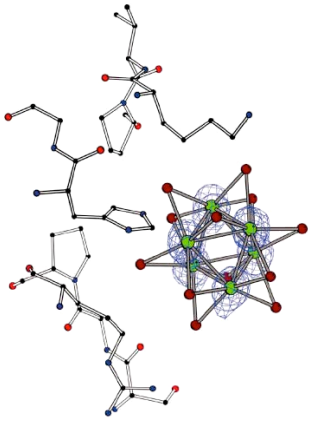


Figure 36| $[Ta_6Br_{12}]^{2+}$ cluster binding site in the crystal structure of glucose isomerase (Banumathi et al., 2003).

Because X-ray diffraction is proportional to the square of the number of electrons in a given scatterer, a cluster scattering at low resolution gives rise to a much stronger isomorphous and anomalous signal than the individual cluster atoms added together. Clusters are most commonly used for isomorphous replacement phasing of large macromolecular assemblies. The $[Ta_6Br_{12}]^{2+}$ cation consists of an octahedron of tantalum ions, surrounded by 12 bromine ions (Figure 36). Binding of one $[Ta_6Br_{12}]^{2+}$ cation add up 856 electrons to a protein, which results in a considerable contribution to the scattering power. The symmetric architecture of this cluster allows binding to the protein in a number of different but equivalent orientations (Schneider and Lindqvist, 1994).

In this work a solution of $[Ta_6Br_{12}]^{2+}$ was kindly provided by Dr. Thomas Meier (Frankfurt). $[Ta_6Br_{12}]^{2+}$ cluster was used for MAD and SAD phasing using the L-III edge at a wavelength of 1.2548 Å corresponding to 9.8811 keV. Derivatisation was performed by soaking native BetA crystals with the dark green solution of $[Ta_6Br_{12}]^{2+}$. 0.2 µl of the ~2mM $[Ta_6Br_{12}]^{2+}$ solution was added to one crystal drop with a volume of 3 µl. Green colouring of the crystals in the $[Ta_6Br_{12}]^{2+}$ solution was

observed under the stereo microscope (MZ 16, Leica, D-Burgsolms) and recorded with a digital camera (Leica, D-Burgsolms) over a maximum time of 26 hours (Figure 37). The over time increased intensity in green colour of the crystals shows a presumably time dependent incorporation of $[Ta_6Br_{12}]^{2+}$ (Figure 37).

As judged solely by optical observation, crystals did not suffer from $[Ta_6Br_{12}]^{2+}$ soaking. Crystals did not dissolve or crack, which is usually caused by chemical stress or by osmotic stress when applying a chemical solution different from the crystal drop.

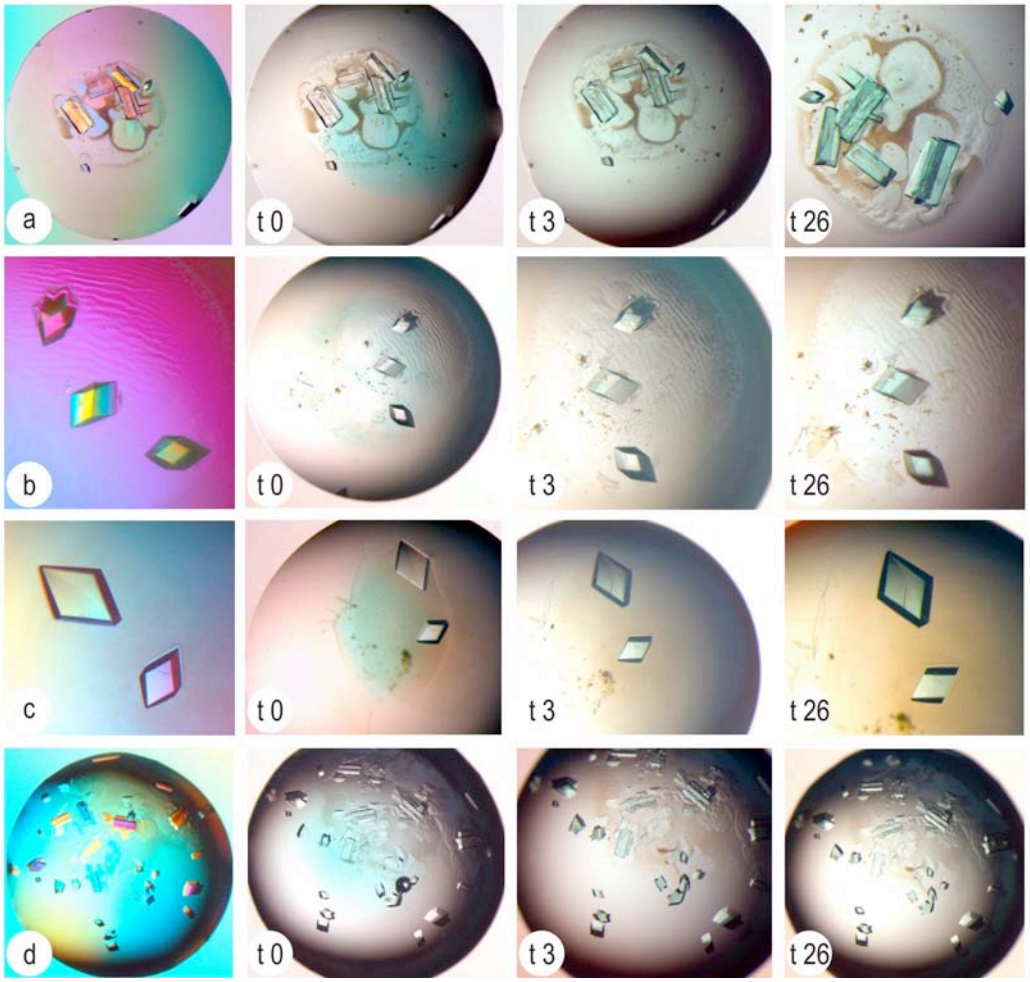


Figure 37| Soaking timeline of native BetA crystals with $[Ta_6Br_{12}]^{2+}$. Pictures **a** to **d** show the original crystal drop visualizing the crystals with a polarization filter prior to application of $[Ta_6Br_{12}]^{2+}$ solution. For each crystal drop the soaking timeline is documented for **t0**, $[Ta_6Br_{12}]^{2+}$ solution was just added; **t3**, after three hours soaking time; **t26**, after 26 hours soaking time.

Crystals were cryo-protected with paraffin oil (3.2.3) and flash frozen into liquid nitrogen. An X-ray fluorescence scan was performed prior to data collection of $[Ta6Br12]^{2+}$ BetA crystals (Figure 38).

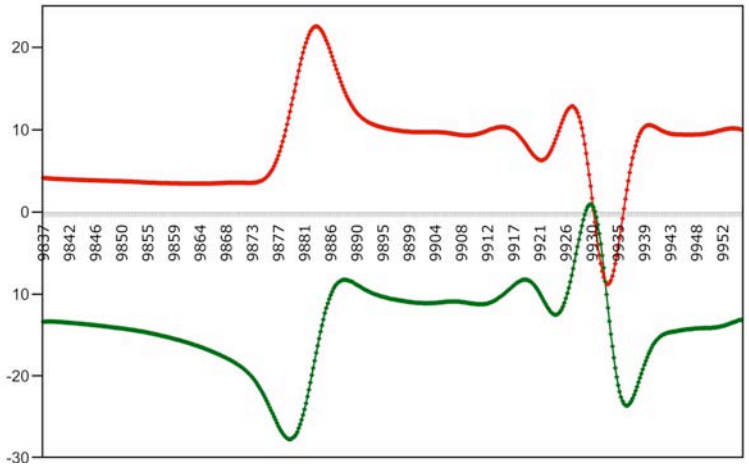


Figure 38|X-ray fluorescence scan of $[Ta_6Br_{12}]^{2+}$ BetA crystals using the L-III edge. The red line shows the trend for the imaginary part of the scattering from factor *f"*; the green line shows the course of the real part of the scattering form factor *f'* with respect to the X-ray energy in eV on the x-axis.

$[Ta6Br12]^{2+}$ BetA data was collected at the PXII beamline at the SLS at a wavelength of 1.25413 Å, which corresponds to a theoretical absorbance of 10.43 *e*. The data collection program STRATEGY was used to determine the optimal starting angle for data collection. $[Ta6Br12]^{2+}$ BetA crystals showed severe sensitivity to radiation damage. One MAD data set was collected but could not be used for phasing because of its poor statistics due to the severe radiation damage. However, two SAD data sets were collected and processed successfully. $[Ta6Br12]^{2+}$ BetA data was processed as the previous SeMet and Xe data sets. First analysis for anomalous signal was performed with Xprep (Bruker Nonius) (Table 19).

Table 18| Crystallisation conditions, data collection and statistics for crystal tabr_b3

Crystallisation condition:	21% PEG400, 100 mM NaCl, 100 mM Na_3-citrate buffer, pH 5.6 at 18°C
Soaking condition:	26h
Cryo-condition:	Paraffin
Data collection:	beamline ID29 (ESRF), λ=1.25413, exposure time=1 sec., attenuator=24.4%, oscillation range=0.5°, detector distance (dd)=448.34 mm, data range run1 = 180 frames starting at 33° run2 = 90 frames starting at 213° run3 = 90 frames starting at 303° run3 = 90 frames starting at 258°
Crystal parameter	a, b, c (Å) = 120.67, 129.94, 187.01 α, β, γ (°) = 90
Data statistics (Scala):	

Summary data for Project: tabr_b3 Crystal: tabr_b3 Dataset: tabr_b3

	Overall	InnerShell	OuterShell
Low resolution limit	**30.00**	**30.00**	**3.79**
High resolution limit	**3.60**	**11.38**	**3.60**
Rmerge	**0.233**	**0.165**	**1.053**
Rmeas (within I+/I-)	**0.277**	**0.190**	**1.444**
Rmeas (all I+ & I-)	**0.284**	**0.186**	**1.489**
Rpim (within I+/I-)	**0.146**	**0.092**	**0.982**
Rpim (all I+ & I-)	**0.113**	**0.068**	**0.822**
Total number of observations	**206980**	**9368**	**14264**
Total number unique	**34603**	**1158**	**4789**
Mean((I)/sd(I))	**4.9**	**22.9**	**0.9**
Completeness	**99.2**	**94.5**	**95.9**
Multiplicity	**6.0**	**8.1**	**3.0**
Anomalous completeness	**97.0**	**96.2**	**81.5**
Anomalous multiplicity	**3.1**	**4.7**	**1.6**
DelAnom correlation between half-sets	**0.098**	**0.147**	**0.009**
Mid-Slope of Anom Normal Probability	**0.985**		

Data statistics from SCALA show high *R* factors and a low anomalous signal for this data set. Further analysis of anomalous signal in Xprep resulted in a very low signal in the low-resolution shell (Table 19). This very low anomalous signal was not suitable for further site search with SHELXD.

Table 19| Xprep anomalous signal statistics for tabr_b3.

resolution	Inf	8.0	6.0	5.8	5.6	5.4	5.2	5.0	4.8	4.6
bin Å	8.0	6.0	5.8	5.6	5.4	5.2	5.0	4.8	4.6	4.4
anomalous Signal[a)]	2.03	1.32	1.21	1.26	1.38	1.32	1.29	1.26	1.21	1.2
anomalous Signal[b)]	1.75	1.15	1.04	1.06	1.17	1.11	1.08	1.06	1.02	1.04

a) signal based on input sigmas
b) signal based on variances of *F*+ and *F*-

6.3. Discussion

Phasing with Xenon

Xenon (Xe) has been used successfully for solving the phase problem in several X-ray crystallographic studies, for example for the photosystem I from *Synechoccus elongates* (Krauss et al., 1996). Xe is a noble gas with anaesthetic properties, which interacts with protein in a non-covalent way, involving just weak-energy van der Waals forces that make Xe binding reversible. A general feature of Xe is that it binds to hydrophobic regions in the protein, which should make it suitable for membrane protein X-ray crystallography. Nevertheless, due to its unspecific and reversible binding, it tends to result in low occupancies of several unspecific binding sites, which may be difficult to detect. Furthermore, the additional handling and pressurisation of the crystal can lead to loss of diffraction.

Xe phasing thus can be advantageous if crystals are very sensitive towards osmotic or chemical stress during the application of the specific phasing compounds. In this work Xe phasing was tried but not further emphasised because of the phasing success with SeMet. Nevertheless, it would be interesting to find out if varying pressure and pressure time could optimise Xe phasing results. The reproducibility of BetA crystallisation would make such a study possible. Once optimal pressure and time, as well as the effect of posterior oil application is determined for the localisation of the Xe sites, it would be interesting to investigate preferential binding locations of Xe in membrane proteins. Furthermore it would be valuable to find out whether variations in detergent or lipids affect Xe binding.

Phasing with $[Ta_6Br_{12}]^{2+}$

Evidently, the benefit of phasing with $[Ta_6Br_{12}]^{2+}$ clusters is the green colour, which indicates its successful absorption in native BetA crystals. Another advantage of $[Ta_6Br_{12}]^{2+}$ phasing compared to Xe phasing is the absorption edge at a wavelength of 1.254 Å, which is easily accessible on tunable synchrotron beamlines. Furthermore, due the cluster size it also provides strong phasing power even if it is just partially incorporated.

Nevertheless, the crystal solvent content is important for the success of the $[Ta_6Br_{12}]^{2+}$ phasing. Crystals with solvent content below 38% might not acquire green colour and hence might not specifically bind $[Ta_6Br_{12}]^{2+}$ (Banumathi et al., 2003). But, if the solvent channels in the crystal are of such nature that they allow diffusion of $[Ta_6Br_{12}]^{2+}$ clusters anyway, then a crystal with a low solvent content could be derivatised with $[Ta_6Br_{12}]^{2+}$. As membrane protein crystals usually have a high solvent content, $[Ta_6Br_{12}]^{2+}$ could be included into the list of suitable compounds for phasing.

The disadvantage of phasing with $[Ta_6Br_{12}]^{2+}$ is low site occupancy, which derives from the size of this cluster and its slow diffusion time. In the case of $[Ta_6Br_{12}]^{2+}$ BetA phasing, occupancies were not sufficient to locate sites. Two possible factors could have resulted in low incorporation or low occupancies. First, the soaking time of up to 26h may not have been long enough. Second, the binding sites may be too unspecific. On the other hand, radiation damage was a major issue using this compound. Due to the lower binding energy of the electrons in the outer orbital of heavy atoms, these are "kicked" out from these orbital much easier upon interaction with X-ray radiation and thus produce radicals. This effect can be minimised with back soaking of the crystal so that specifically bound heavy atom clusters remain in the crystal. Furthermore, one could use a smaller tantalum cluster that might migrate to more specific binding sites.

Choosing a SAD over MAD

The very first and obvious limitation in a SAD or MAD phasing is the requirement of a tunable synchrotron beamline. Furthermore, there are some advantages and disadvantages to take into account between SAD and MAD.

The clear disadvantage of SAD is phase ambiguity (3.3.8). Anomalous diffraction data collected solely from single peak wavelength do not produce unimodal phase-probability distributions (Figure 10). For this reason, electron density maps computed from SAD phases are generally difficult, and in some cases impossible, to interpret. Density modification such as solvent flattening and NCS averaging are used to discriminate between the two phase probabilities. Therefore the success of SAD phasing depends strongly on data quality, solvent content and the existence of NCS or a high symmetry spacegroup. Furthermore, if the derivative crystal is isomorphous to native crystals, the phase ambiguity can be solved by SIRAS.

In the MAD experiment, on the other hand, collecting additional data near the absorption edge solves the phase ambiguity. The differences between the diffraction amplitudes at different wavelength are then used to obtain a unique phase choice.
Although SAD is very challenging for *de-novo* structure determination at moderate to medium resolution, it was favoured over MAD because SeMet-BetA crystals were more sensitive towards radiation damage than native BetA crystals. Considering the change in wavelength for a MAD experiment and the amount of data taken, radiation damage throughout the experiment can cause non-isomorphism between the MAD data sets. The most important requirements for successful MAD phasing are a high degree of isomorphism between the various datasets, a high signal-to-noise ratio of the anomalous differences, and a good correlation of the anomalous differences at different wavelengths. MAD data collection requires a considerably longer collection time compared to SAD. Avoiding radiation damage in a MAD experiment would have the consequence of less redundant peak data that results in a more noisy and weaker anomalous signal.

Therefore the highest priority at this stage was a highly redundant SAD data set, providing the most accurate measurements of the anomalous signal to locate the selenium sites. The high number of Se sites in BetP and the relatively low data quality meant that data collection and site search was a delicate procedure. As SAD data collection time is shorter, the measurement of accurate anomalous differences in the steep region around the inflection point of the *f''* curve (Figure 9) is the main difficulty for successful SAD phasing. Being just a few eV off the edge and toward the inflection point might result in a non-detectable anomalous difference.

Radiation damage was found to be not as prominent as expected when SeMet was used as an anomalous scatterer. However, selenium positions show higher *B* factors, especially if determined from SAD data sub-sets measured later (Rice et al., 2000), which points towards increased radiation damage.

With SAD phasing it is generally more difficult to derive accurate phases than with MAD. Therefore SAD data has to be processed with special care. The SeMet1 data set was reprocessed in different data batches, each with individually assigned resolution cutoffs. In addition a careful reassignment of the beamstop region was necessary to improve data quality for all subsequent phasing processes.

Multi-crystal averaging using the native data set Native1 and Native2 together with the SeMet1 SAD data did not result in better phasing statistics and finally no improvement of the electron density. This was possibly due to the high non-isomorphism between the SeMet data set and the native data sets and their high data anisotropy.

The different success rate between SAD and MAD phasing is much smaller at medium to low-resolution (Rice et al., 2000) and the recording of anomalous differences is most important for the success of anomalous phasing. Therefore, it is best to concentrate on data collection at the *f''* maximum instead of measuring the dispersive differences (Gonzalez, 2007).

Hence, even though more challenging, using SAD was the right choice for solving the phase problem for BetP.

7. Structure determination of SeMet-BetA

7.1. Phase determination by SAD with SeMet

Protein derivatisation with selenium (Se) is an elegant way of substitution of sulphur with selenium. *E. coli* strains, auxotroph for methionine are used and selenomethionine is supplied during their cell growth. As a result, a high percentage of the protein has a high incorporation of selenomethionine in place of methionine.

SeMet-BetA data were collected after an X-ray fluorescence scan (Figure 39). The SAD data set SeMet1 was processed as described in 5.2. The anomalous signal was analysed with XPREP (Table 20) and showed a promising anomalous signal of 1.38, based on the input sigmas, up to a resolution of 4.5 Å. 1.3 is the proposed reasonable anomalous signal cutoff for anomalous data as analysed by (Dauter, 2006).

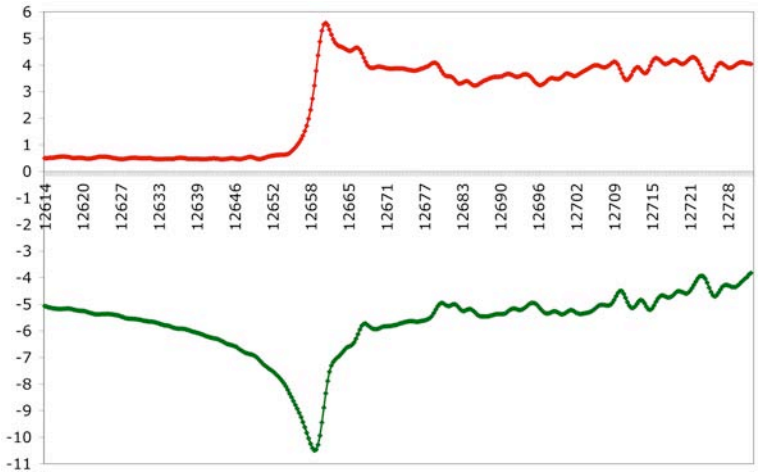


Figure 39| X-ray fluorescence scan of a SeMet-BetA crystal at the K-edge for Se. The red line shows the course for the imaginary part of the scattering factor *f''*; the green line shows the course of the real part of the scattering factor *f'* with respect to the X-ray energy in eV on the x-axis.

Table 20| Xprep anomalous signal statistics for SeMet1.

resolution	Inf	8.0	6.0	5.3	5.1	4.9	4.7	4.5	4.3	4.1	3.9	3.7
bin Å	8.0	6.0	5.3	5.1	4.9	4.7	4.5	4.3	4.1	3.9	3.7	3.5
anomalous Signal[a)]	5.87	3.49	2.09	1.79	1.64	1.56	1.46	1.38	1.38	1.29	1.27	1.16
anomalous Signal[b)]	4.17	2.90	1.83	1.57	1.43	1.37	1.30	1.25	1.26	1.21	1.23	1.15

a) signal based on input sigmas
b) signal based on variances of *F*+ and *F*-

SAD Δ*F* values were calculated from data set SeMet1 (5.2) and used in SHELXD to determine the selenium positions. Several runs with programs SHELXD and HySS were performed of more than 100 cycles each with different resolution cutoffs. Finally, a SHELXD run including data only to 4.5Å resulted in a successful substructure determination. The choice of resolution cutoff can be crucial for the success of substructure solution (Dall'Antonia et al., 2003). Data up to a resolution of 4.5Å had sufficient anomalous signal (Table 20) used by SHELXD and resulted in more than 50 Se sites. Yet, a clear solution was not detected by SHELXD as judged by the correlation coefficient profile for anomalous differences.

Table 21| Initial heavy atom site parameters for data set SeMet1 deriving from SHELXD.

Site	**x**	**y**	**z**	**Occ**	**Site**	**x**	**y**	**z**	**Occ**
Se1	111.191	34.548	38.674	1.000	Se13	82.741	27.792	22.768	0.731
Se2	120.647	25.461	54.007	0.931	Se14	151.198	21.407	66.685	0.698
Se3	124.932	24.641	63.829	0.923	Se15	127.994	3.727	54.247	0.690
Se4	145.927	10.693	42.919	0.915	Se16	94.021	6.750	88.397	0.688
Se5	104.813	30.799	46.298	0.912	Se17	142.303	16.298	74.929	0.639
Se6	94.952	30.035	63.992	0.888	Se18	122.070	27.318	72.104	0.581
Se7	90.973	36.792	59.263	0.867	Se19	90.971	8.493	68.293	0.578
Se8	121.848	12.898	37.214	0.864	Se20	89.401	16.561	44.331	0.567
Se9	119.993	37.543	39.139	0.843	Se21	113.577	37.789	16.520	0.532
Se10	105.208	25.874	64.109	0.836	Se22	74.097	28.828	71.851	0.479
Se11	102.310	44.148	16.025	0.781	Se23	100.822	15.193	28.223	0.449
Se12	109.919	5.737	68.775	0.775	Se24	79.537	27.965	83.598	0.417

SHELXD writes out coordinates of the Se sites with their occupancy values into a PDB file. The coordinates were analysed manually for a significant drop in the occupancy value (Table 21) as well as for their location in the unit cell. Based on prior knowledge from biochemistry and electron microscopy, a trimeric organisation of Se sites was expected. Therefore all coordinate files were checked for a trimeric distribution of the selenium sites. Only one out of these many files showed a trimeric Se site distribution and a meaningful solvent region as well (Figure 40). This coordinate file resulted in 13 manually chosen Se sites with reasonable occupancy that were related by 3-fold symmetry out of a total of 56 (Figure 40). This was the crucial result that allowed further structure determination of BetP.

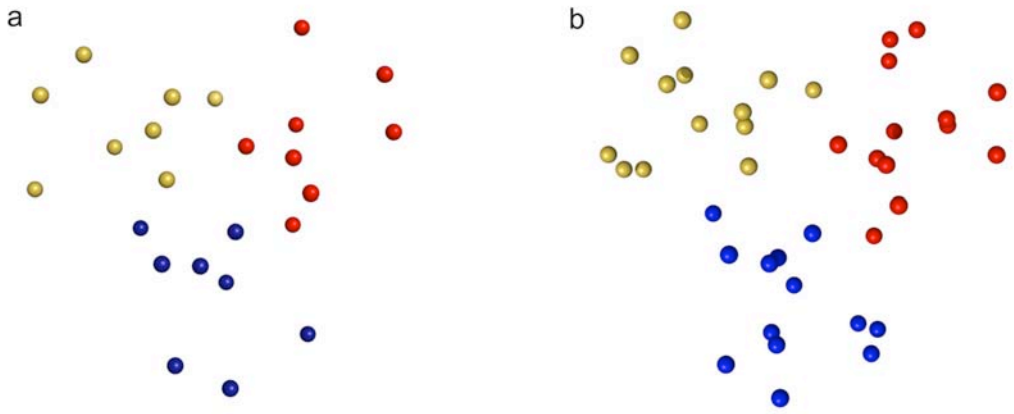


Figure 40| Selenium sites from SHELXD for data set SeMet1. a Initial 24 Se sites that are related by 3-fold symmetry with highest occupancies listed in Table 21; **b** initial 39 Se sites related by 3-fold symmetry including those that have low initial occupancies under 0.4.

For a better phase estimation from these first Se-sites, initial 39 Se-sites were refined and a substructure search was completed using SHARP. The almost complete incorporation of selenium during protein synthesis was indicated by the high occupancies and low *B* factors for most of the Se sites. SHARP phasing and refinement resulted in 53 Se sites that were refined for their occupancies and *B* factors, leading to initial phases from which protein electron density map was calculated (Table 22 and Figure 41).

Table 22| Initial SAD phasing statistics for SeMet1

Phasing statistics	
data set	SeMet1
resolution range (Å)	39.37 – 3.35
anomalous phasing power (overall)	1.094
and resolution bin 5.27 – 4.97	1.129
anomalous R_{cullis}	0.8
FOM (overall)	0.26
resolution bin 5.27 – 4.97	0.36

The quality of the resulting electron density map, shown in Figure 41, was sufficient to identify the SeMet-BetA trimer, and the phase problem was solved. Nevertheless, the phases were correct only to low resolution: helices were almost featureless, with poor side chain information and poor helix connectivity. Therefore, this initial map (Figure 41) did not allow model building of SeMet-BetA; either by hand or automatically.

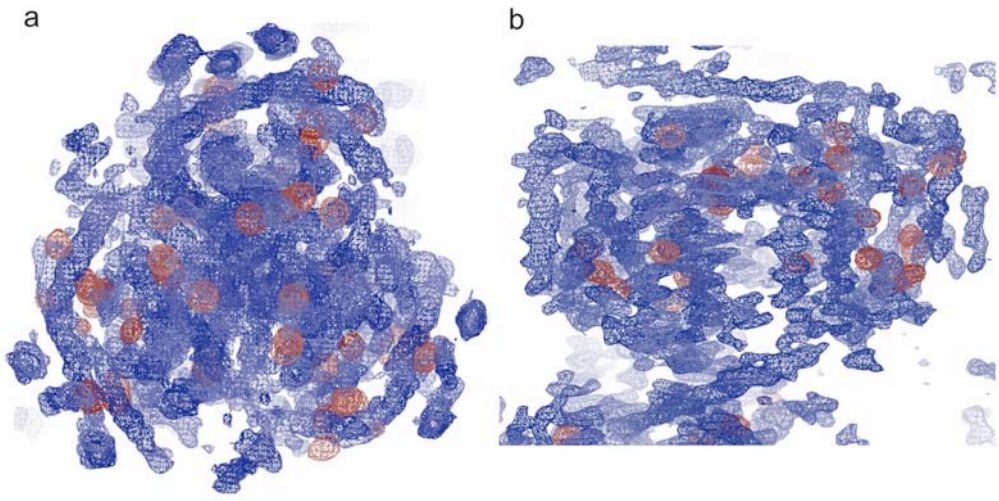


Figure 41| The initial SeMet-BetA electron density map. The initial observed electron density map shown in blue at 1.3σ after solvent flattening with 64% solvent content. The anomalous difference map for the selenium sites is shown in red at 3.5σ. **a** shows the top view and **b** the side view on the SeMet-BetA trimer.

39 selenium sites formed three groups from which the NCS operators the three protomers were determined using LSQKAB. The initial electron density map was then modified by solvent flattening using 64% solvent content in SOLOMON (Abrahams and Leslie, 1996), as implemented in SHARP, and three-fold NCS

averaging with DM. Nonetheless, these density modification attempts together with subsequent negative *B*-factor sharpening of the map did not result in an electron density map suitable for model building either. This demanded further correction of data anisotropy and improvement of the initial phase information.

7.2. Re-phasing, density modification and anisotropy correction.

SAD phasing with the SeMet-BetA data set SeMet1 solved the phase problem but did not result in an electron density map suitable for model building (7.1.). Reprocessing of the data set was necessary to increase data quality, including careful masking of the beamstop region and division of data sub-sets into resolution shells, which were rescaled for anisotropy as described in 3.3.7 and 5.2. SAD phasing in SHARP using this reprocessed data gave slightly better phasing statistics (Table 23).

For initial model building, the electron density map was improved by using only density modification in SOLOMON with a solvent content of 55%. A negative B-factor value of -106 Å^2 (3.3.10) (DeLaBarre and Brunger, 2006) was applied to the solvent flattened map, followed by NCS-averaging in DM. Since the resulting electron density map showed a lot of features suggesting helices, model building was initiated using the helix-building tool of ARP/wARP 6.1. This resulted in 1233 automatically built amino acids in 85 helix fragments, which were then included into a second round of density modification using solvent flattening with 55% solvent content and three-fold NCS averaging.

Table 23| SAD phasing statistics for SeMet1 including partial models

Phasing statistics	
data set	SeMet1
resolution range (Å)	39.37 – 3.35
anomalous phasing power (overall)	1.094
and resolution bin 5.27 – 4.97	1.129
anomalous R_{cullis}	0.9
FOM (overall)	0.26
resolution bin 5.27 – 4.97	0.36

The resulting phases from the second round of density modification were then used in BUCCANEER to initiate automatic model building. Three cycles of building

resulted in a model with 1785 residues in 82 chains, of which the longest segment consisted of 177 residues (Figure 42).

This initial model was refined using BUSTER-TNT and further anisotropy correction of the data set SeMet1 was achieved by scaling the anisotropic F_{obs} to the isotropic F_{calc} of the refined model (script 10.5). The scaled observed amplitudes were used for another round of phasing in SHARP, followed by density modification in SOLOMON and DM as well as a new round of automated model building in BUCCANEER (Figure 43). After these cycles of phasing, model building and data scaling with subsequent refinement cycle in BUSTER-TNT resulted in a model (buccaneer02) with R factors of R_{cryst}/R_{free} = 25.14/36.48 (Figure 43). Due to the incomplete model, mis-built side chains and not enforcing NCS restraints the difference between the R_{cryst} and R_{free} was large. At this point, the previously scaled F_{obs} were again scaled to the isotropic F_{calc} of the refined model and another round of SHARP phasing resulted in improved statistics shown in Table 24. Further density modification using 60% solvent content in SOLOMON and an additional NCS averaging in DM resulted in phases used in BUCCANEER for model building with the previous model as a starting point.

Table 24| SAD phasing statistics for SeMet1 after rescaling to F_{calc}.

Phasing statistics	
data set	SeMet1
resolution range (Å)	39.37 – 3.35
anomalous phasing power (overall)	1.091
and resolution bin 5.27 – 4.97	1.048
anomalous R_{cullis}	0.96
FOM (overall)	0.28
resolution bin 5.27 – 4.97	0.38

Specific sub-cycles were applied in BUCCANEER during automatic model building. The first four-cycles included a search for helices only, followed by a subsequent one-cycle round searching for other types of structures. This output was then given to a 10-cycle run, which resulted in a third initial model (buccaneer03 see Figure 42 and Table 25). This initial model contained 1818 residues in 51 chains, with the longest sequence having 227 residues. After pruning of the initial model, 1514 residues were fitted in sequence (10.6) (Figure 42).

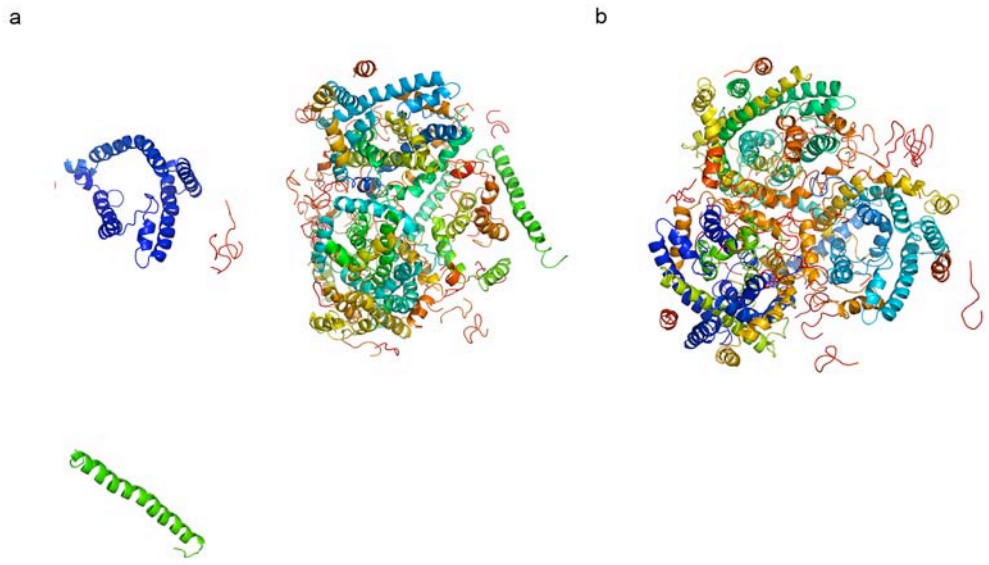


Figure 42| Built fragments from the second and third round of automated model building using BUCCANEER. Automatically built fragments are represented in the cartoon mode and rainbow colouring using PyMOL. **a**, the upper left fragment shows already a BetA trimer. The other fragments are part of the symmetry related BetA molecules. **b**, model fragments in buccaneer03, after 10 cycles of BUCCANEER building (symmetry related fragment not shown) are more complete compared to the model in **a**. Recurring cycles of scaling, density modification, model building and refinement resulted in a stepwise improvement of the phase quality and therefore an improved electron density map in which more and more parts of the model could be built.

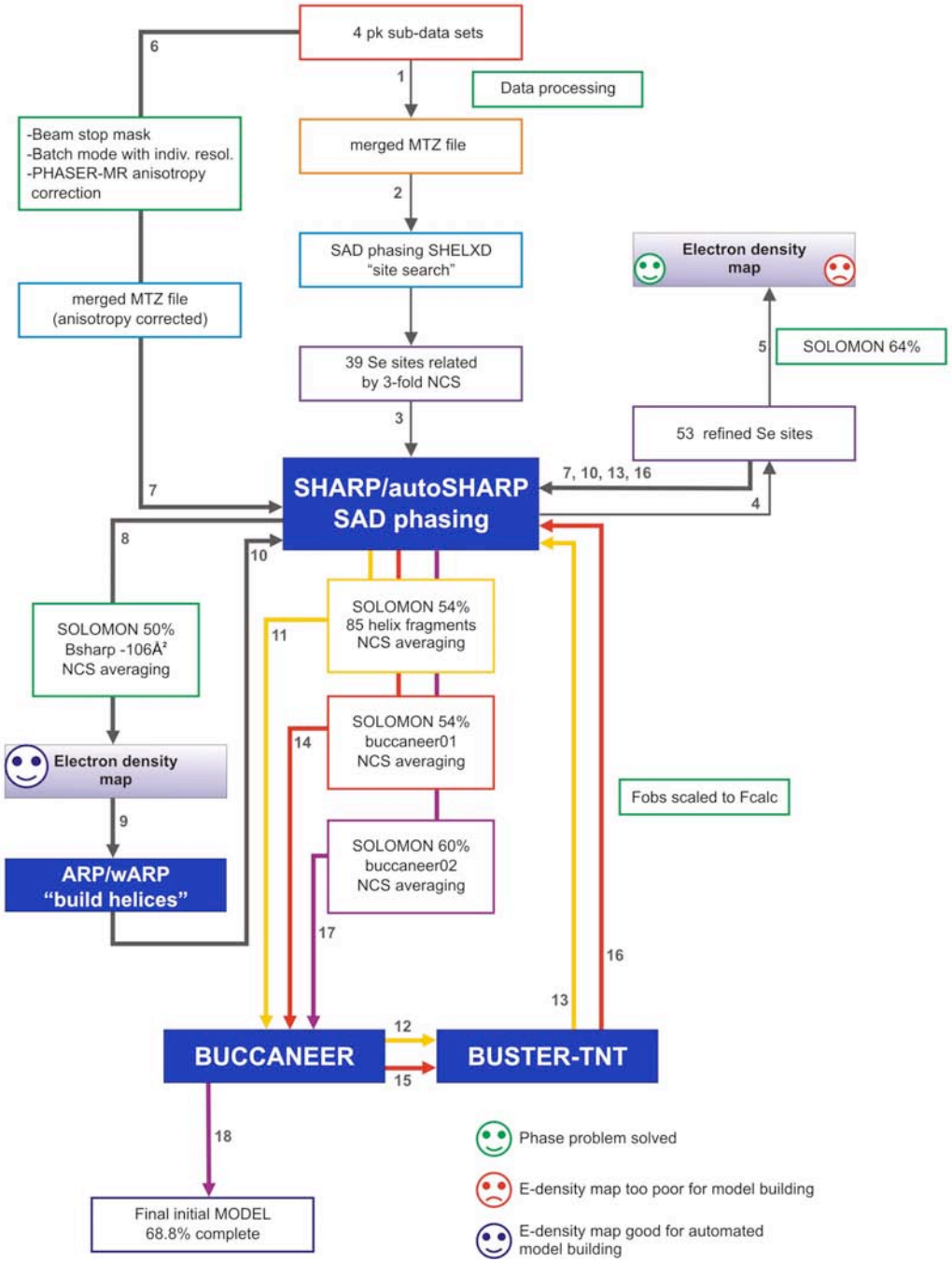


Figure 43| Flowchart of the processing steps towards an interpretable electron density map.

Altogether, around 20 different subsequent computational procedures summarised in the flowchart in Figure 43, resulted in an interpretable electron density map, with quality development illustrated in Figure 44.

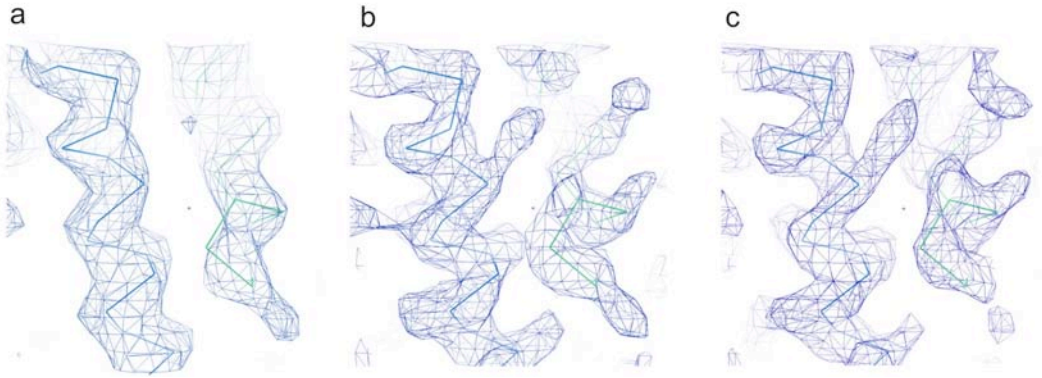


Figure 44| Quality progress of σ_A-*weighted* (mF_{obs}- DF_{calc}) ($2mF_{obs}$- DF_{calc}) electron density maps. Maps are shown in blue at 1.3σ and helices of the final model in green. **a,** initial density before data reprocessing and with no anisotropy correction nor NCS averaging. **b,** NCS averaged map after the second round of BUCCANEER building. **c,** resulting and not NCS averaged map of the initial trimer model_00.

A remarkable improvement in quality of the electron density map is evident (Figure 44). Side chains electron density greatly improved and the density for the helices allowed determining their directions.

7.3. Model building and refinement

Not all peptide stretches were built in sequence in the initial model. Therefore, unknown residues were removed and a refinement in BUSTER-TNT resulted in a model giving R_{cryst}/R_{free} of 24.78/32.36. Various chains in the third refined model (buccaneer03) were put into the right place manually using LSQKAB and Coot, which resulted in a trimer containing three chains (Figure 45). Chain A contained 409 residues, chain B 384 and chain C 392 out of 574 residues, indicating a 68% complete BetP trimer.

A sequence alignment (Clustal W 1.83) of chain A, B and C to the full BetP sequence showed the missing and incorrectly built parts (red letters in the original BetP sequence) and green stars indicate the sequence fit of chains A, B and C to the original BetP sequence:

```
BetP            MTTSDPNPKPIVEDAQPEQITATEELAGLLENPTNLEGKLADAEEEIILEGEDTQASLNW
A               ------------------------------------------------------------
B               ------------------------------------------------------------
C               ------------------------------------------------------------

BetP            SVIVPALVIVLATVVWGIGFKDSFTNFASSALSAVVDNLGWAFILFGTVFVFFIVVIAAS
A               ----PALVIVLATVVWGIGFKDSFTNFASSALSAVVDNLGWAFILFGTVFVFFIVVIAAS
B               --------IVLATVVWGIGFKDSFTNFASSALSAVVDNLGWAFILFGTVFVFFIVVIAAS
C               --------IVLATVVWGIGFKDSFTNFASSALSAVVDNLGWAFILFGTVFVFFIVVIAAS
                        ************************************************

BetP            KFGTIRLGRIDEAPEFRTVSWISMMFAAGMGIGLMFYGTTEPLTFYRNGVPGHDEHNVGV
A               KFGTIRLGRIDET-----VSWISMMFAAGMGIGLMFYGTTEPLTFYRNGVPGHDEHNVGV
B               KFGTIRLGRIDEA------SWISMMFAAGMGIGLMFYGTTEPLTFYRNGVPGHDEHNVGV
C               KFGTIR-------------SWISMMFAAGMGIGLMFYGTTEPLTFYRNGVPGHDEHNV--
                ******             ************************************

BetP            AMSTTMFHWTLHPWAIYAIVGLAIAYSTFRVGRKQLLSSAFVPLIGEKGAEGWLGKLIDI
A               AMSTTMFHWTLHPWAIYAIVGLAIAYSTFRVGRKQLLSSAFVPLIGEKGAEGWLGKLIDI
B               AMSTTMFHWTLHPWAIYAIVGLAIAYSTFRVGRKQLLSSAFVPLIGEKGAEGWLGKLIDI
C               AMSTTMFHWTLHPWAIYAIVGLAIAYSTFRVGRKQLLSSAFVPLIGEKGAEGWLGKLIDI
                ************************************************************

BetP            LAIIATVFGTACSLGLGALQIGAGLSAANIIEDPSDWTIVGIVSVLTLAFIFSAISGVGK
A               LAIIATVFGTACSLGLGALQIGAGLSAA--------------------------------
B               LAIIATVFGTACSLGLGALQIGAGLSAAN-------------------------------
C               LAIIATVFGTACSLGLGALQIGAGLSAA--------------------------------
                ****************************

BetP            GIQYLSNANMVLAALLAIFVFVVGPTVSILNLLPGSIGNYLSNFFQMAGRTAMSADGTAG
A               ------------------------------------------------------NIGTAG
B               -----------------------------------------------------ISADGTAG
C               ------------------------------------------------------ADGTAG
                                                                       ****

BetP            EWLGSWTIFYWAWWISWSPFVGMFLARISRGRSIREFILGVLLVPAGVSTVWFSIFGGTA
A               EWLGSWTIFYWAWWISWSPFVGMFLARISRGRSIREFILGVLLVPAGVSTVWFSIFGGTA
B               EWLGSWTIFYWAWWISWSPFVGMFLARISRGRSIREFILGVLLVPAGVSTVWFSIFGGTA
C               EWLGSWTIFYWAWWISWSPFVGMFLARISRGRSIREFILGVLLVPAGVSTVWFSIFGGTA
                ************************************************************

BetP            IVFEQNGESIWGDGAAEEQLFGLLHALPGGQIMGIIAMILLGTFFITSADSASTVMGTMS
A               IVFEQNGESIWGDGAAEEQLFGLLHALPGGQIMGIIAMILLGTFFITSADSASTVMGTMS
B               IVFEQNGESIWGDGAAEEQLFGLLHALPGGQIMGIIAMILLGTFFITSADSASTVMGTMS
C               IVFEQNGESIWGDGAAEEQLFGLLHALPGGQIMGIIAMILLGTFFITSADSASTVMGTMS
                ************************************************************

BetP            QHGQLEANKWVTAAWGVATAAIGLTLLLSGGDNALSNLQNVTIVAATPFLFVVIGLMFAL
A               QHGQLEANKWVTAAWGVATAAIGLTLLLSGGDNALSNLQNVTIVAATPFLFVVIGLMFAL
B               QHGQLEANKWVTAAWGVATAAIGLTLLLSGGDNALSNLQNVTIVAATPFLFVVIGLMFAL
C               QHGQLEANKWVTAAWGVATAAIGLTLLLSGGDNALSNLQNVTIVAATPFLFVVIGLMFAL
                ************************************************************

BetP            VKDLSNDVIYLEYREQQRFNARLARERRVHNEHRKRELAAKRRRERKASGAGKRR
A               VKDLSNDVIYLEYREQQRFNARLA-------------------------------
B               VKDLSNDVI----------------------------------------------
C               VKDLSNDVIYLEY------------------------------------------
                *********
```

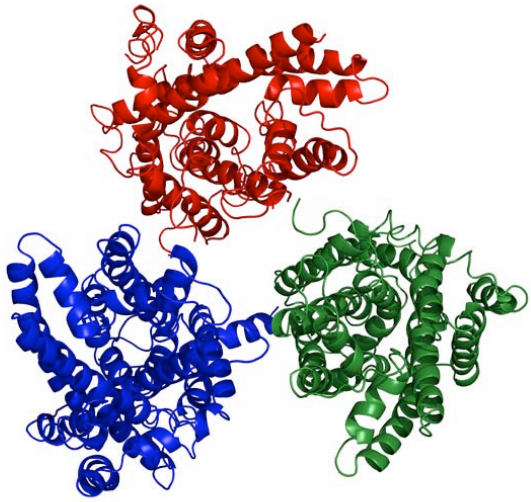


Figure 45| Trimer chains A, B and C after ten building cycles. Top view on the trimer with chain A (blue) containing 409 residues, chain B (red) containing 384 residues and chain C (green) containing 392 residues. Unknown residues within the trimer and symmetry related fragments were deleted.

This model was good starting point for further manual model building. Errors, such as mis-built residues or frame shifts in sequence had to be identified and corrected, and missing helices and loops had to be completed; this was an iterative process of refinement, quality control and rebuilding. Together with imposed stereo-chemical rules (ideal bond length and angles), refinement is an automatic adjustment of model parameters, the atomic coordinates, to improve their fit to the experimental data.

After building most of the missing parts with COOT and O, the removal of several frame shifts resulted in an improved model (BetP_TRIMER_ABC.pdb). Positive and negative difference densities from the ($\boldsymbol{F}_{obs}$- $\boldsymbol{F}_{calc}$)-map were contoured at 3σ and used as guidance for further manual building.

Several runs were performed using difference amplitudes and their standard deviations, with and without NCS restraints to find the correct refinement strategy. As a result, refinement was performed against the scaled amplitudes and their standard deviations derived from SHARP (FP and SIGFP) with strict NCS restraints, except for the C-terminal domain (see 10.8). Refinement of the model (BetP_TRIMER_ABC.pdb) under these conditions using BUSTER-TNT resulted in R_{cryst}/R_{free} of 26/30. In both programs experimental phase restraints and strict NCS

restraints were applied throughout the refinement. *Phenix.refine* included three translation-liberation-screw (TLS) groups per chain (for TLS see section 3.3.12 and for TLS group definition see 10.7) and used convential NCS restrain, whereas BUSTER-TNT used Local Structure Similarity Restraints (LSSR) as NCS restrain and ideal helix restraints. Around 90 cycles of building and refinement, using *phenix.refine* and BUSTER-TNT, were needed to achieve the final model of BetP.

In the first rounds of refinement used BUSTER-TNT and REFMAC-5 were used in parallel. Refinement in REFMAC-5 ended in a local minimum, indicated by an increase of R_{cryst}/R_{free} values, no improvement in electron density map and model geometry. This program was therefore not further used. After this decision, refinement was continued using *phenix.refine* and BUSTER-TNT alternately. All main refinement strategies and steps with their results are summerised in Table 25.

The combination of different refinement programs proved to be very useful: *phenix.refine* improved the stereochemistry of the model and allowed further correction using the phenix.geometry_minimization tool by solely optimising stereochemistry properties of the model without including any observed data (electron density); and BUSTER-TNT resulted in better guiding ($\boldsymbol{F}_{obs}$- $\boldsymbol{F}_{calc}$) electron densities, which allowed better judgment for model building. However, the qualities of the σ_A-weighted ($m\boldsymbol{F}_{obs}$- $D\boldsymbol{F}_{calc}$) ($2m\boldsymbol{F}_{obs}$- $D\boldsymbol{F}_{calc}$) electron density maps were of similar quality with both programs.

Table 25| Refinement and model building stages for SeMet-BetA.

STAGE OF MODEL	MODEL BUILDING, STRATEGY & PROGRESS	PROGRAM & STRATEGY	TLS groups[a)]	R_{cryst}	R_{free}	RAMACHANDRAN[b)]
buccaneer02	auto	BUSTER-TNT	-	25.14	36.48	-
buccaneer03	auto	BUSTER-TNT	-	24.78	32.36	-
BetP_TRIMER_ABC	manual -removed main sequence shifts -helix 7 (270 -357) build - C-term build up to 583	BUSTER-TNT NCS, MLhl, FPsha	-	26	30	60.46% / 15.49%
refmac-04	manual -loop 2, 5 and 6 remodelled in O using lego-Cα - several outliers removed	REFMAC-5 strict NCS -FPsha	-	32.54	35.67	88.09% / 2.31%
15_refine	manual -input file was remac-04	BUSTER-TNT NCS, MLhl, FPsha	-	25.83	28.8	80.39% / 5.24%
18_refine	manual -ligand glycine-betaine added - several loops remodelled - C-term up to -input model underwent several REFMAC-5 refinement cycles	BUSTER-TNT NCS, MLhl, FPsha	-	26.19	28.31	75.10% / 7.30%
phenix.refine01	manual -geometry_minimization (phenix.refine) - sequence shifts removed -NCS corrections	phenix.refine NCS, MLhl, TLS	18 TLS groups	27.45	31.44	81.48% / 5.76%
phenix.refine04	manual -geometry optimization -round with simulated annealing -outlier correction	phenix.refine NCS, MLhl, TLS	18 TLS groups	30.47	32.62	93.72% / 0.07%
Several rounds of manual building, correction of helices and loop regions, alternating refinement between phenix.refine and phenix.geometry_minimization with one round of simulated annealing, TLS refinement and BUSTER-TNT using LSSR NCS restraints.						
64_refine	manual	BUSTER-TNT -LSSR NCS, MLhl -target phenix.refine11	-	25.68	26.49	93% / 0.26%

a) TLS groups defined in 10.7 b) in most favoured regions/outliers as defined by MOLPROBITY

In the early stages of refinement, positive (F_{obs}- F_{calc}) electron densities appeared in all three monomers. Aromatic residues around the density were superposed on the substrate-binding pocket of the glycine-betaine binding protein ProX from *E. coli* (pdb ID 1r9l) (Schiefner et al., 2004a), confirming the binding pocket architecture and glycine-betaine position in BetP.

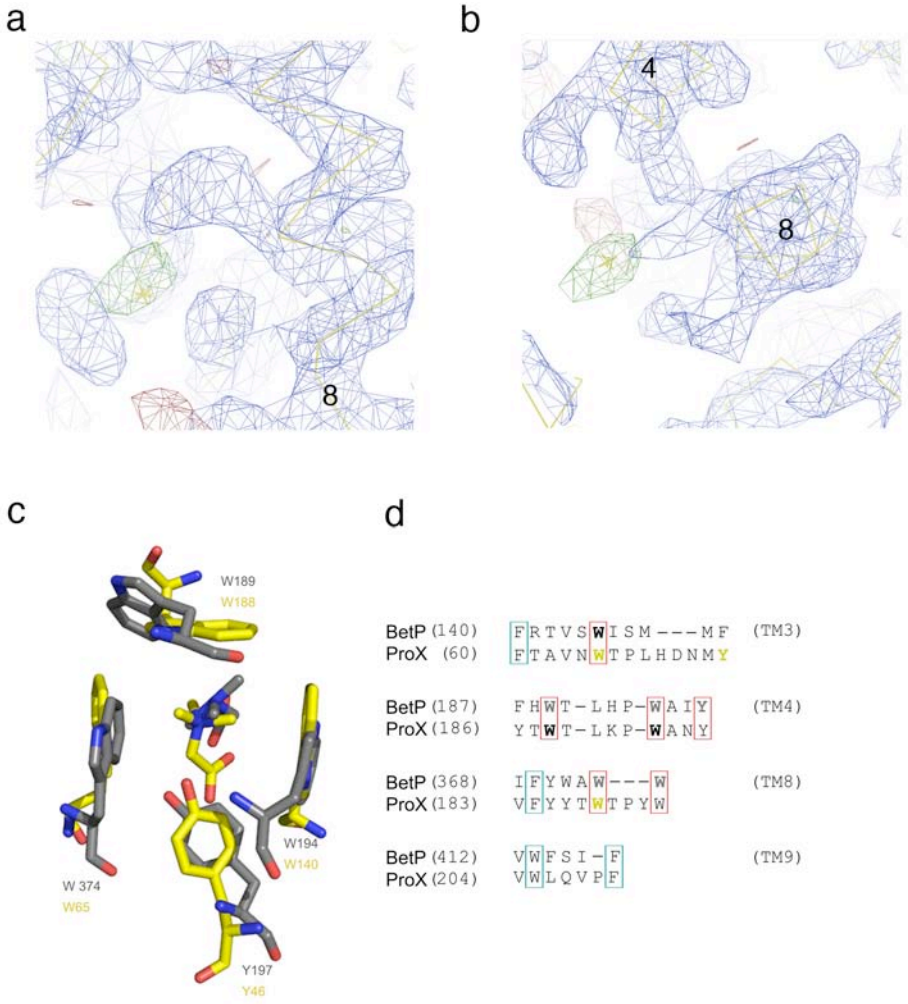


Figure 46|Positive Fo-Fc electron density for glycine-betaine and the strategy for its positioning in BetP. a, side view on transmembrane helix 8 in chain A with 2Fo-Fc electron density map (blue) at 1.3σ and the positive Fo-Fc density (green) countered at 2.4σ, indicating the glycine-betaine position; **b**, corresponding top view on transmembrane helix 8 and 4; **c**, overlay of the glycine-betaine binding pocket of BetP (grey) and the periplasmic glycine-betaine binding protein ProX from *E. coli* (yellow). **d**, The sequence homology between BetP and ProX in the 4-helix bundle TM3, TM4/TM8, TM9 of BetP is shown on the right hand side.

The density of the substrate did not improve dramatically during the whole refinement and was different in the three monomers. *B*-factor analysis (Figure 48) in

the final model suggests different substrate occupancies within the trimer. However, occupancy refinement was not feasible with this electron density map quality.

Table 26| Data collection and final refinement statistics.

	SeMet-BetA
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	118.09, 129.42, 182.94
α, β, γ (º)	90, 90, 90
	Peak
Wavelength	0.9794
Resolution (Å)	39.47 – 3.35 (3.55-3.35)
R_{merge}	9.9 (64.8)
$I/\sigma(I)$	23.2 (1.0)
Completeness (%)	90.7 (44.1)
Redundancy	18.9 (1.7)
Refinement	
Resolution (Å)	39.47 – 3.35 (3.55-3.35)
No. reflections	37151 (2967)
R_{cryst}/ R_{free}	25.68/26.49 (24.12/23.95)
No. atoms	11737
Protein	11353
Ligand/ion	384
average atomic B-factor	64.34
r.m.s deviations	
Bond lengths (Å)	0.002
Bond angles (º)	0.34

7.4. Structure validation of the BetP model

During BetP refinement (Table 25), values for R_{cryst}, R_{free} and the model geometry were monitored. Final statistics are given in Table 26. At this final stage all-atom contact analysis with MOLPROBITY (Davis et al., 2007; Lovell et al., 2003) showed no serious intra- or intermolecular clashes. The Ramachandran plot for the BetP model indicated that 93% of all torsion angles were in the favoured region (Figure 47). Most of the main chain torsion angles are in the favoured or allowed

region, except for residues Asp547 and Ile225 in all three chains, indicated by red circles in Figure 47(a). A correction for these outliers was not feasible because Asp547 forms a salt-bridge to Arg210, which stabilises this unfavourable position. Ile225 is situated in a region of increased flexibility, as indicated by the corresponding weak quality of the electron density map.

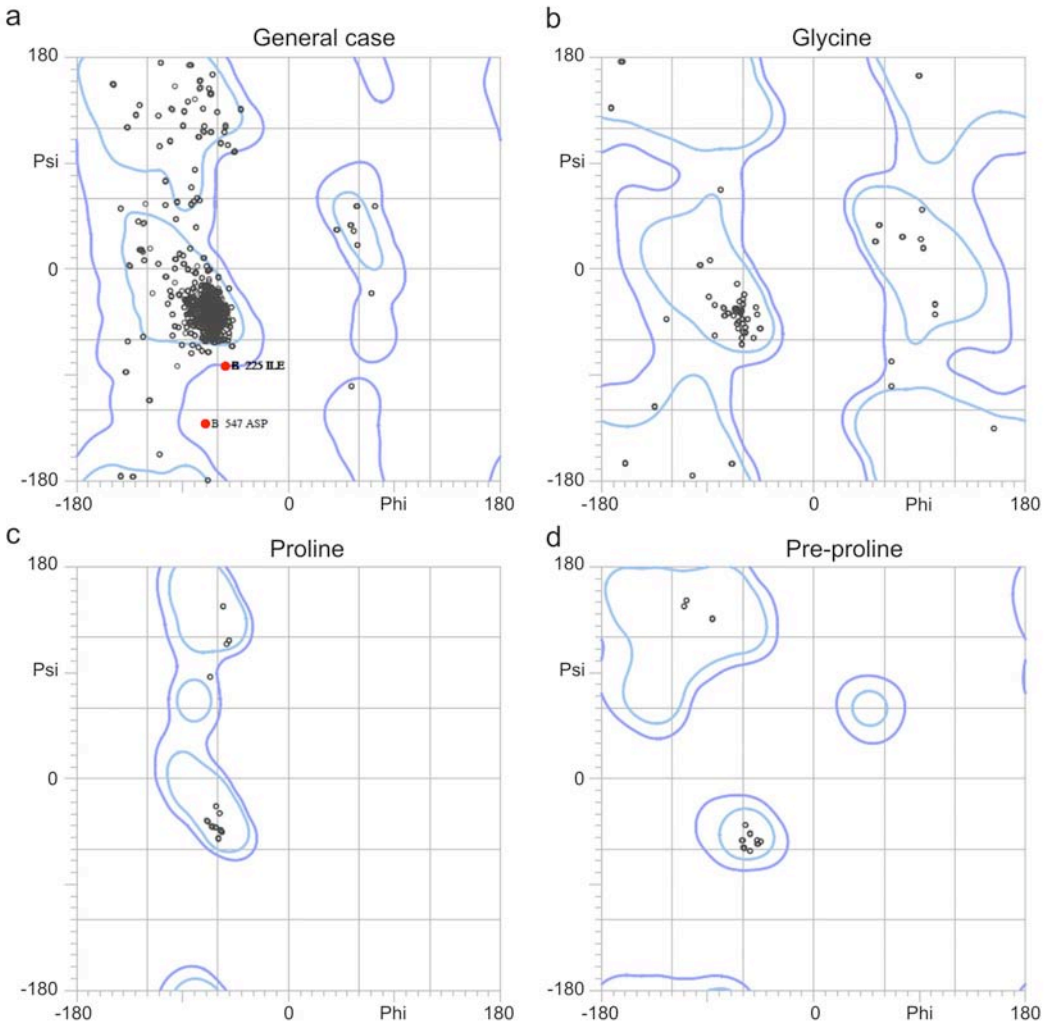


Figure 47| Ramachandran diagram for SeMet-BetA. Main chain torsion angles were analysed with MOLPROBITY and are indicated for each residue by open circles. Red circles indicate outliers. The inner region bordered by a light blue line indicates the energetically favourable regions and the dark blue line indicates the allowed region. **a**, general plot for all non-glycine residues. **b**, specific plot for torsion angles and energetically favoured regions of glycine residues. **c** and **d**, plots for proline and pre-proline residues.

Although NCS restraints were used throughout the refinement, differences between the chains A, B and C are observed in *B*-factor profiles (Figure 48). Chain C has the lowest average *B* factor and chain A the highest. Analysis of the trimer architecture reveal more stabilising interactions in chain C. Specific *B* factor variation between the three chains is most obvious in the region of transmembrane helix (TM) 3, TM6 and TM7 as well as in the region of TM11 and 12. These variations are promoted by the LSSR restraints used in BUSTER-TNT. Unlike conventional NCS constraints and very tight restraints, where domains are forced to be the same, LSSR uses a three-distance matrix in which the same residue in different chains is treated to have similar chemical environments. LSSR does not define domains and is therefore more dynamic than conventional NCS restraints and hence allows slight differences between the chains A, B and C of BetP.

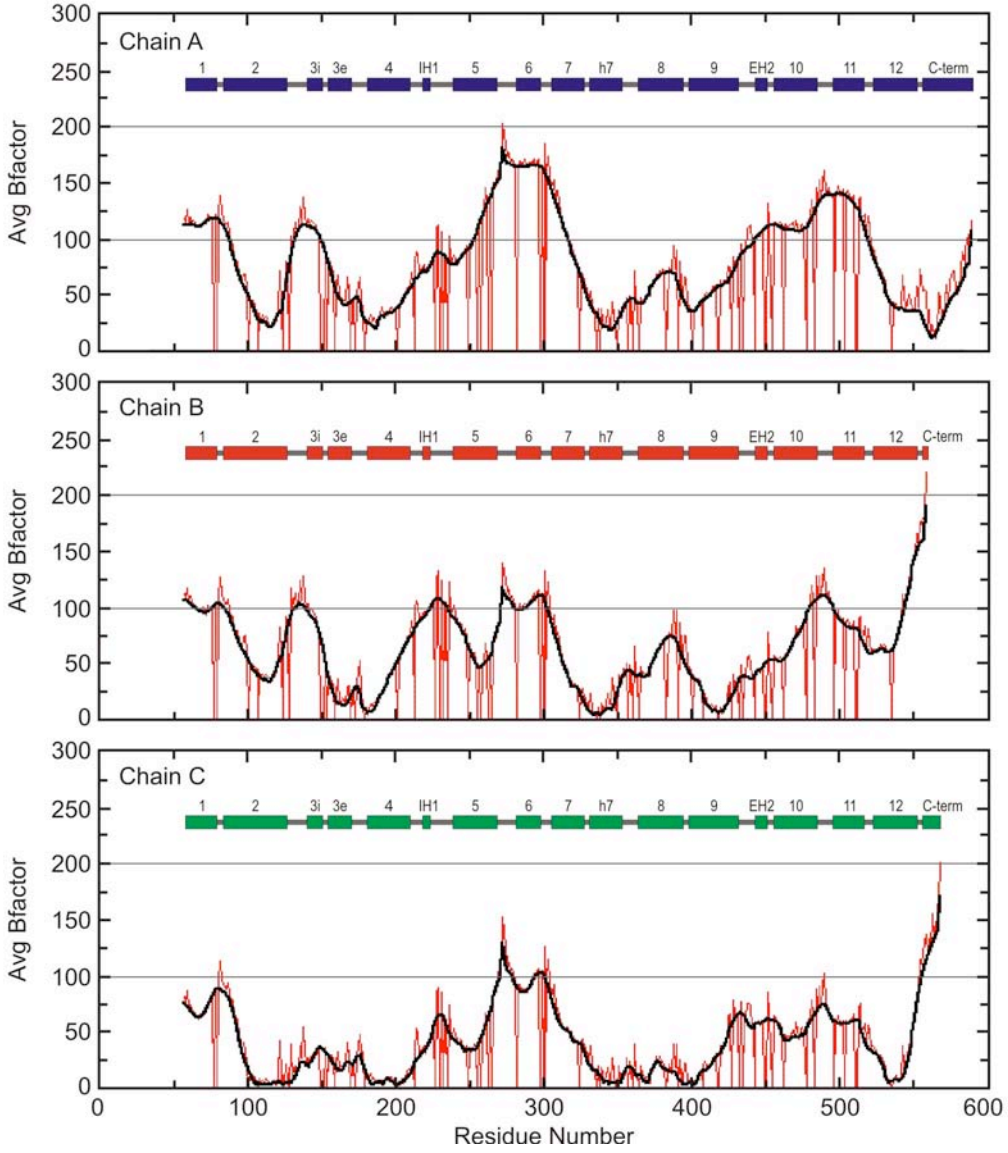


Figure 48| Plots of the main chain residual *B* factor for the final BetP model. From top to bottom, the *B* factor plot for chain A (blue), chain B (red) and chain C (green). Helices are indicated with bars and numbered corresponding to the residues sequence plotted on the x-axis. The average *B* factor is indicated by the black graph and plotted on the y-axis. The red graph shows the *B* factor per residue.

Regions with higher *B* factors in Figure 48 are also apparent in the real-space fit (Jones et al., 1991) of the main- and side-chain to the σ_A-weighted $(m\boldsymbol{F}_{obs}- D\boldsymbol{F}_{calc})(2m\boldsymbol{F}_{obs}- D\boldsymbol{F}_{calc})$ electron density map (Figure 49). Even though *B* factors were not used for the correlation calculations, the real-space fit expresses the correlation coefficient (y-axis in Figure 49) between the observed and calculated density from the model.

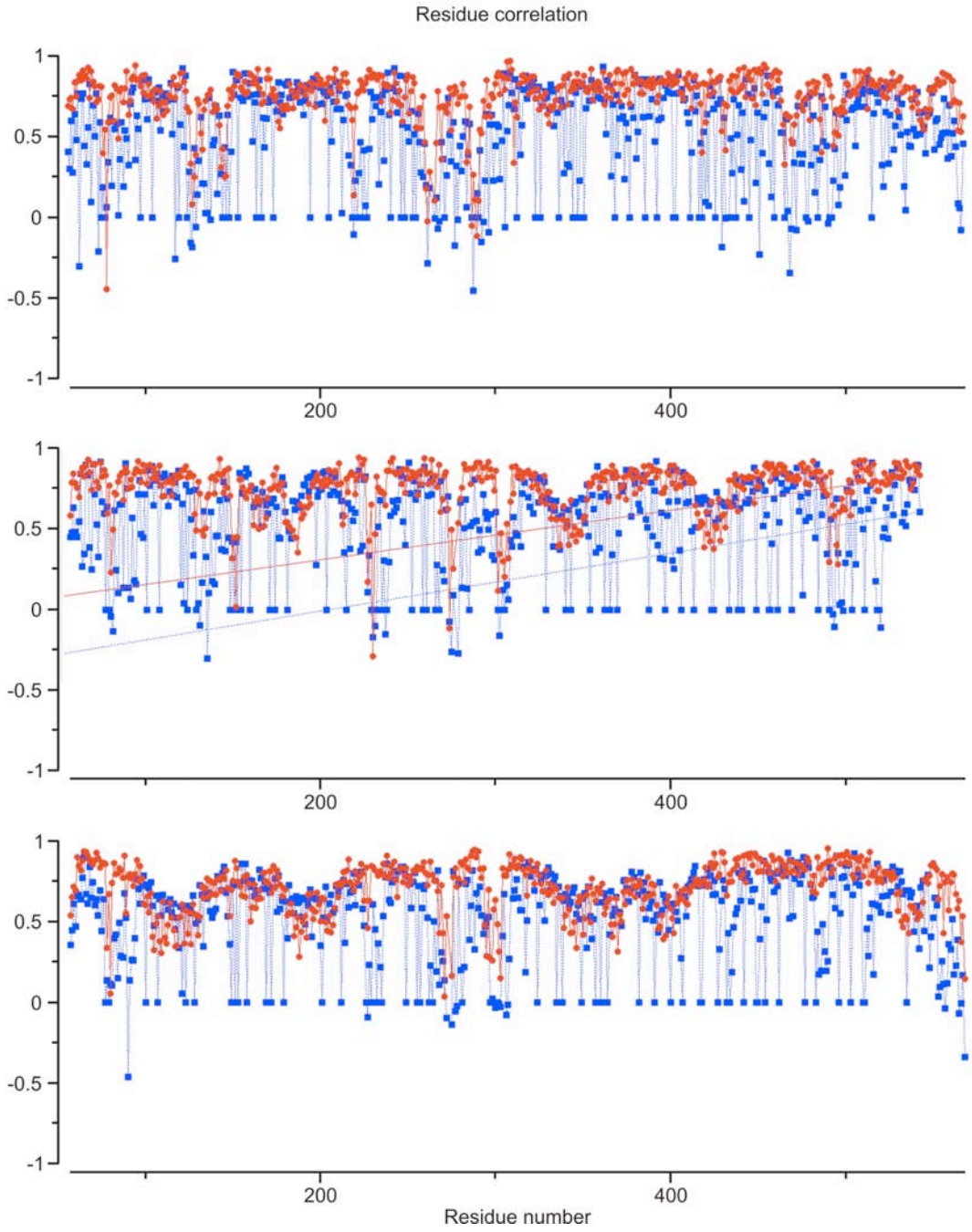


Figure 49| Real-space correlation for the final BetP model to its σ_A-weighted (mF_{obs}- DF_{calc}) ($2mF_{obs}$- DF_{calc}) electron density map. From top to bottom: the real-space fit for the main chain, red dots connected by a red line and the side chain, blue rectangles connected by a blue dotted line, for chain A, chain B and chain C, respectively. The residue numbers given at the x-axis and the correlation coefficient is given at the y-axis. The red and blue line crossing the plot for chain B are due to an unknown technical error and are not indicating any correlations.

The real-space correlation plot does not show very large differences between the three chains. Overall, chain C and A have the best real-space fit followed by chain B,

which shows most variations in the correlation curve. As already mentioned the real-space fit correlates with the *B* factor of the atoms. This means the higher the *B* factor of a given residue, the more mobile it is and the worse the quality of their electron density map and real-space fit. The regions with the highest *B* factor are located at the peripheral region of the trimer. Parts of the structure towards the center of the trimer show low *B* factor distribution, especially TM3, TM4, TM8 and TM9 (Figure 50). As the overall real-space fit for chain A is good, the region around TM6 shows the lowest fit. This correlates with the high *B* factor in this region of chain A (Figure 50).

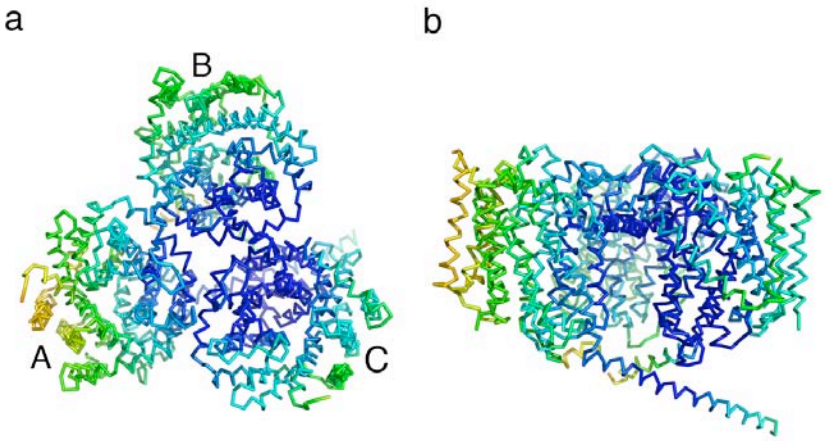


Figure 50| ***B*** **factor distribution of BetP. a**, top view on the trimer with chains marked with A, B and C. *B* factors are colour coded from low values in blue towards high values red. **a**, side view on chain A left hand side and chain C right hand side. Both views show lower average *B* factors localised towards the center of the trimer.

7.5. Discussion

Finding the right strategy for phase-improvement, building and refinement of the crystal structure

Using SeMet substituted protein for crystal structure determination is especially relevant when crystals diffract to only medium resolution. Knowledge of the protein sequence makes the SeMet positions into important landmarks during model building. It confirms the reliability and verifies the location of the amino acids sequence in the model.

Crystallisation conditions and cryo-protection for SeMet-BetA was similar to native BetA. The use of the expression protocol and media, provided by Molecular Dimensions resulted in an efficient incorporation of SeMet into BetA.

During anomalous data collection, radiation damage (5.3) can result in a loss of isomorphism of the data. Non-isomorphism of the data might prevent accurate measurement of the phasing signal. Taking radiation damage into account it might be preferable to collect SAD data instead of MAD data.

Solving a *de-novo* membrane protein structure of the size of BetP from SAD data is still a rare case in X-ray crystallography. Around 600 structures deposited in the protein data bank (PDB) (Berman et al., 2000) are solved by SAD and 38 are of medium resolution (up to 3Å) (Kirillova et al., 2007). From these reported 38 structures only five are membrane proteins: the cobalamin transporter BtuB at 3.1Å (PDB entry 1nqh) (Chimento et al., 2003), the bacterial nucleoside transporter Tsx at 3.1Å (PDB entry 1tlw) (Ye and van den Berg, 2004), the protein-conducting channel SecY solved at a resolution of 3.2Å (PDB entry 1rhz) (Van den Berg et al., 2004) and the Na^+/H^+ antiporter NhaA at 3.45Å (PDB entry 1zcd) (Hunte et al., 2005).

Besides the phase ambiguity, initial phase information deriving from a SAD experiment at medium resolution is usually not very precise, due the low resolution anomalous signal and low site occupancies. Density modification procedures, such as solvent flatting and NCS averaging are thus not just a tool to solve the phase ambiguity but also to improve the initial phase information. Usually these steps are used iteratively and phase improvement is achieved in each cycle (Terwilliger, 2000; Terwilliger et al., 2008). In this work phase improvement was carried out by density modification and additional anisotropy correction of the experimental data.

Furthermore, more than three cycles of subsequent phase refinement in SHARP were needed to obtain an electron density map that allowed further interpretation.

Non-crystallographic symmetry (NCS) was crucial in this work, not just for efficient density modification but also for reducing model parameters. The ratio between observations (number of reflections) and model parameters (x, y and z coordinates, occupancies and *B* factor) is usually low at medium resolution. In the case of SeMet-BetA, the final model contained 11737 atoms and with four parameters applied per atom, this results in 46948 total parameters (3.3.12). In the absence of NCS restraints, 46948 parameters would have to be refined with only 37151 observables from the experimental data. Therefore it was necessary to reduce the number of model parameters by applying isotropic *B* factor and NCS restraints. Nevertheless, larger anisotropic movements were taken into account with the use of TLS groups in *phenix.refine*.

Exploring the potentialities of different programs during model building and refinement was essential. However, if different programs are used care should be taken with data consistency. In particular, the set of reflections selected to calculate R_{free} must be unique in all used programs.

A general question is the expected *R* factors at a given resolution. As a quality criterium (Kleywegt and Jones, 2002), the *R* factors reported for BetP are comparable to those of other structures of similar resolution. Due to low weighting on the side of the experimental data and higher weighting on the stereochemical properties of the model, R_{free} factors tend to become lower. Analysis of the highest resolution shell in Table 26 shows that the R_{free} is lower compared to R_{cryst}. This might be explained by the NCS restraints, which were necessarily applied to run a successful refinement. NCS restraints should be taken into account when judging data statistics, not just because they can lower the *R* values (Kleywegt and Jones, 2002), but they also hinder the exploration of differences between domains especially at medium to low resolution. This problem was solved using the more dynamic LSSR treatment of NCS restraints in BUSTER-TNT, which allowed, to a certain extent, differences between equivalent residues in different chains of BetP (C. Vonrhein, personal communication).

Differences of R_{cryst} and R_{free} were observed during alternation between the two programs *phenix.refine* and BUSTER-TNT (Table 25). These differences might be due to variable applied weights in the programs and to their bulk-solvent correction. The version of BUSTER-TNT (Blanc et al., 2004) used, in this work has a sophisticated bulk-solvent correction. The solvent mask around the molecule can be blurred, which results in a smoother solvent border. Different blur factors can be applied and solvent structure factors can be calculated. This might be especially advantageous for membrane protein structure refinement because of their surrounding detergent micelle.

As already described in section 3.3.11, model building into a electron density map at medium resolution is challenging, and in many cases does not allow an exact positioning of a residue or even a helix. During model building of SeMet-BetA both, COOT and O building programs were used. The lego_cα function of O, which works with a library of high-resolution structures, was particularly useful for building in some poorly resolved parts of in the SeMet-BetA electron density map.

Some regions in the SeMet-BetA trimer were varyingly well resolved. Especially loop 2 (residues 125-138) showed quality differences in the electron density map between the three chains. Loop 2 was best resolved in chain C followed by chain B, and least density was observed in chain A. Usually building of such parts was carried out in a NCS averaged map. To judge whether a poorly resolved region would be included into the model or not, negative and positive densities of the (F_{obs}-F_{calc}) map were interpreted after refinement of the model. Stereochemical properties of the neighbouring atoms and the overall effect on the R_{cryst} and R_{free} values were monitored as well. Usually several rebuilding and refinement cycles were necessary to decide whether such a region would be included or not.

Model building and refinement is rather a subjective process based not just on the crystallographic data but also on the prior knowledge of the protein and the accompanying prejudices and expectations. In this study, it was of high priority to maintain the "philosophy" of not over-interpreting the given limitation of the data, such as the medium resolution as well as the impact of data anisotropy. However, crystal structures should be validated and interpreted not just by the given statistics but also, and more importantly, by the analysis of its actual experimental data, the electron density.

8. Structure of BetP

In the BetP crystal, contacts are mainly mediated by the C-terminal domain of chain A (Figure 51a, b) to the periplasmic side of the symmetry related trimer (Figure 51c). The contact is composed of four salt-bridges between arginines on the C-terminal domain and glutamic acids on the periplasmic side.

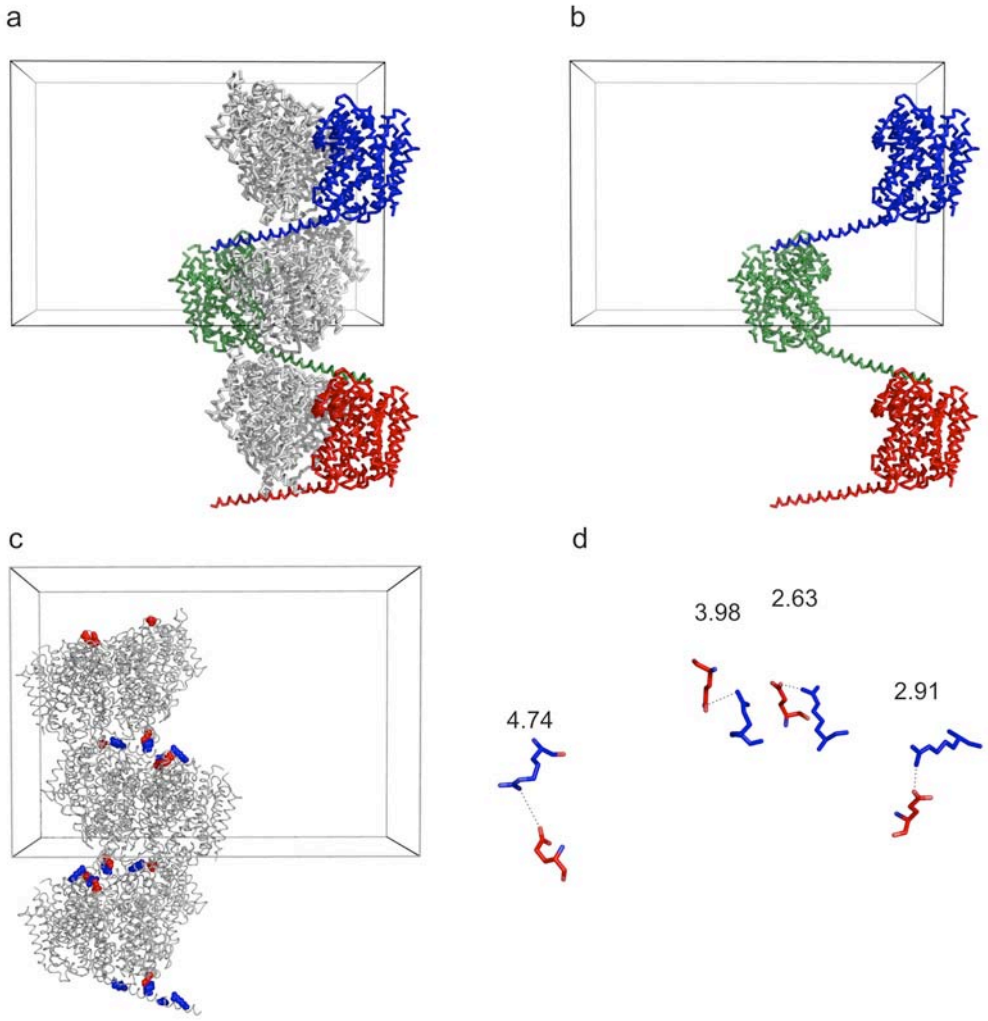


Figure 51| Crystal contacts in the BetP crystal. a, view along the b axis of the unit cell, chains A of the symmetry related trimer in blue, green and red, respectively; chains B and C in grey; **b,** same view as in **a** without chain B and C; **c,** view along the b-axis of the unit cell rotated 180° around the a axis; BetP trimer (grey) and residues involved in the crystal contact coloured in blue (arginines) and red (glutamic acids); **d** four main crystal contacts mediated by Glu428 (B) to Arg554 (A), Glu552 (C) to Arg568 (A), Glu175 (C) to Arg574 (A) and Glu175 (A) to Arg583 (A), the relative bond length between the residues are given in Å above the respective pair.

The four main crystal contacts are namely between Glu428 (B) to Arg554 (A), Glu552 (C) to Arg568 (A), Glu175 (C) to Arg574 (A) and Glu175 (A) to Arg583 (A) (Figure 51d). The mean bond distance is 3.5 Å with the largest distance of 4.74 Å between Glu428 (B) to Arg554 (A), which is off the maximum distance of 4 Å for a salt-bridge.

8.1. Protomer structure

The BetP protomer has a nearly cylindrical shape and contains twelve TM α-helices (Figure 52 and Figure 53). Whereas three helices at the perimeter of the protein (TM1, TM11, and TM6) run almost perpendicular to the membrane plane, the amphipatic α-helix 7 (α7) exhibits a nearly parallel orientation to the membrane plane. TM3, TM4, TM8 and TM9 form an iris-shaped helix bundle (4-helix bundle). The substrate, glycine-betaine is bound between the 4-helix bundle and the TM5, TM10 and TM12. TM5, TM10 and TM12 are three relatively long helices and tilted substantially by an angle of 39° to the membrane normal. Ten transmembrane helices are arranged as a five-plus-five transmembrane helix motif. TM3 to TM7 build the first repeat (repeat1) and TM8 to TM12 the second repeat (repeat2) (Figure 53, Figure 55). Repeat1 and repeat2 are intertwined and build the inverted repeat motif and the transporter core. TM2 and TM7 are long tilted helices located at the trimer interface and surrounding the inverted repeat motif. TM3 has a locally unwound segment approximately halfway across the membrane, which splits this helix into intracellular (TM3i) and extracellular (TM3e) halves. The discontinuous region between TM3i and TM3e goes from Phe146 to Ile152 and includes glycine residues Gly149 and Gly151, which show 87% conservation across the BCCT family (Figure 54). The helix pairs TM4-TM5 and TM9-TM10 are connected on the cytoplasmic or periplasmic side by extended loops referred to as intracellular helix (IH1) or extracellular helix (EH2), respectively. IH1 contains a 3_{10}-helix segment and EH2 an ordered α-helix segment.

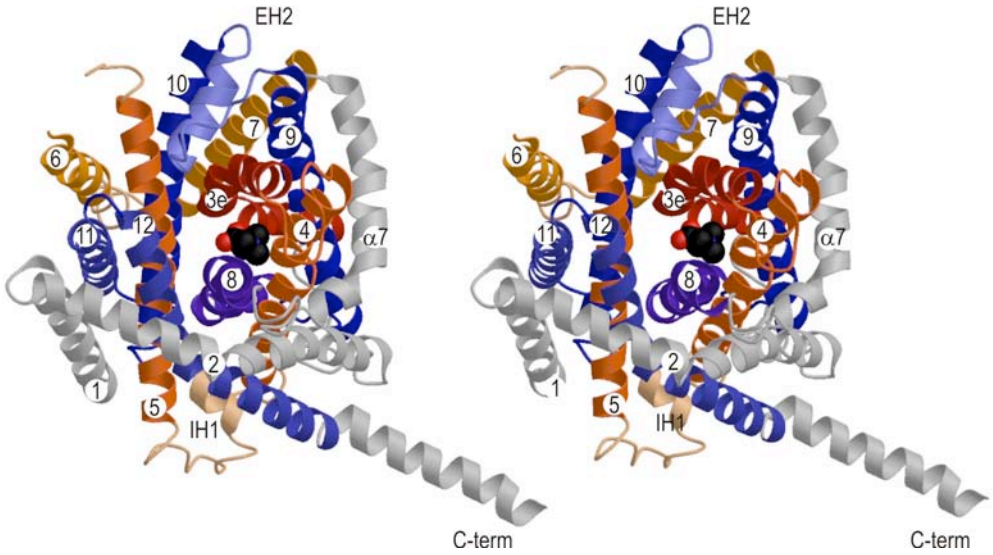


Figure 52| Stereo picture of protomer C. Top view of monomer C. Repeat 1 (TM3-TM7) colouring from red, N-terminus, to orange, C-terminus. The topologically related repeat 2 is coloured in blue (TM8-TM12) in the same way. TM1, TM2, the amphipathic helix 7 and the C-terminal domains are grey. The substrate glycine-betaine is in the middle of the protomer and is represented as black spheres for the carbon atoms, blue for nitrogen and red for oxygen atoms.

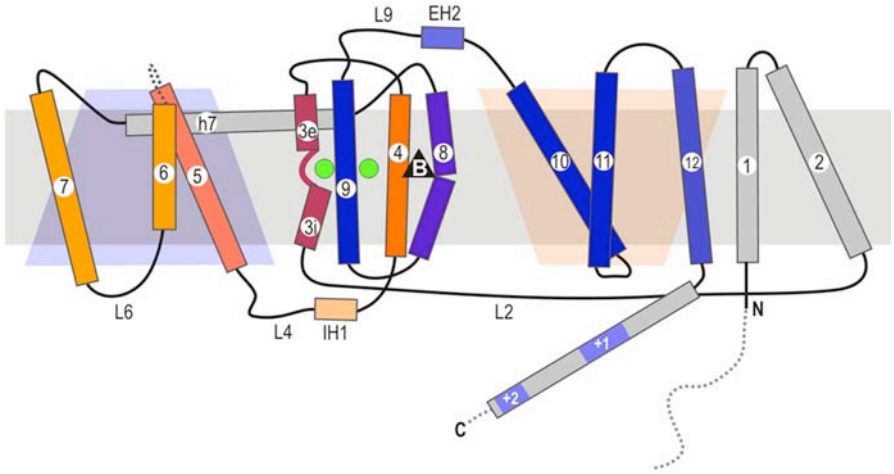


Figure 53| Inverted repeats in the BetP topology. Repeat 1 is coloured in orange-red and repeat 2 in blue (similar colouring as in Figure 52) and lighter coloured towards the C-terminus. The substrate is presented as a black triangle and the sodium ions as green circles. Unresolved residues are indicated by a dotted line. Important loops are indicated as L2, L4, L6 and L9. Two positively charged arginine clusters in the C-terminal domain are indicated by two blue regions.

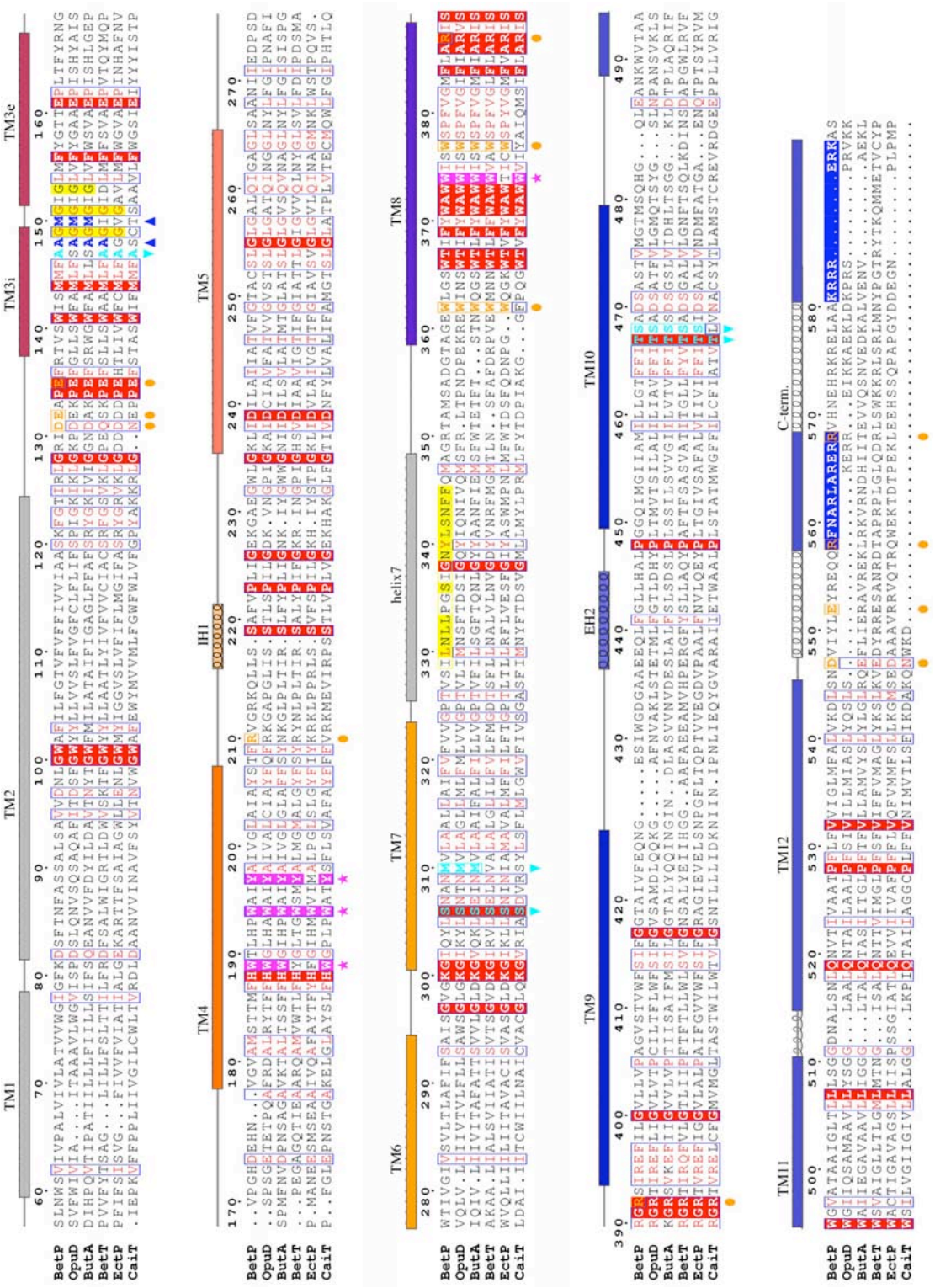


Figure 54| Amino acid sequence alignment of *C. glutamicum* BetP with five transporters of the BCCT family: Na^+/glycine-betaine transporter OpuD from *B. subtilis*, Na^+/glycine-betaine transporter ButA from *Tetragenococcus halophila*, H^+/choline transporter BetT from *E. coli*, Na^+/ectoine transporter EctP from *C. glutamicum*, and carnitine/γ-butyrobetaine antiporter CaiT from *E. coli* using ClustalW multiple sequence alignment displayed by ESPript 2.2 (http://espript.ibcp.fr/ESPript). Strictly conserved residues are displayed in red. α-

helices of the BetP structure are shown as cylinders above the BetP sequence. Residues potentially involved in coordinating sodium ions are marked with triangles shown in blue (Na1) or cyan (Na2). The conserved G-x-G-x-G motif of sodium-coupled transporters in the BCCT family in TM3 is yellow. The magenta stars indicate residues involved in glycine-betaine-binding. Orange coloured residues in TM8 and TM9 show additional substrate-binding sites, residues involved in trimer formation or regulatory interactions are labelled with an orange circle.

The ten transmembrane helices (Figure 53) contain an internal structural repeat, which is not detectable in the amino acid sequence of BetP. It relates TM3-TM4-TM5-TM6-TM7 to TM8-TM9-TM10-TM11-TM12 by a pseudo-two-fold axis in the plane of the membrane (Figure 55).

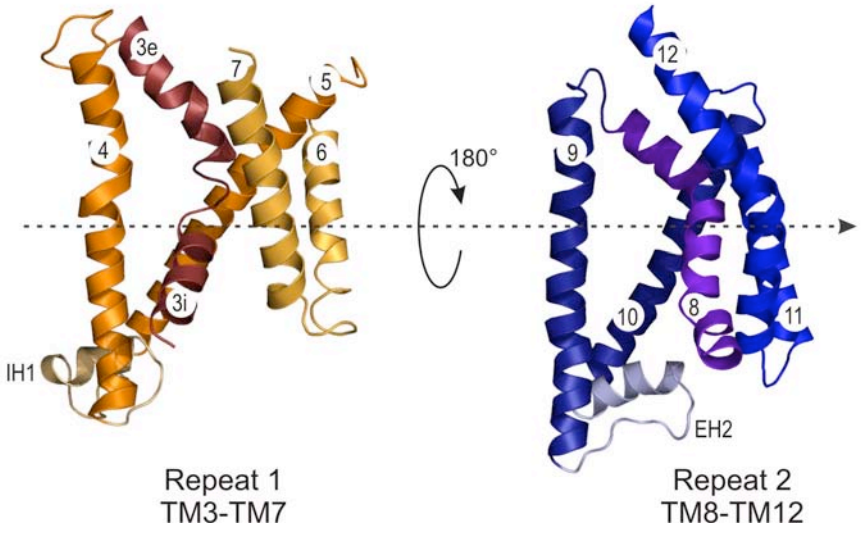


Figure 55| Inverted repeats in BetP. Repeat 2 (TM8-TM12) is rotated 180° around an axis in the middle of the membrane plane.

The Cα atoms of TM3 to TM7 can be superposed onto those from TM8 to TM12 with an r.m.s.d. of 3.5Å, and those of TM5 to TM7 superpose with those of TM10 to TM12 with an r.m.s.d. of 3.2Å. The V-shaped helix pair TM3 and TM4 is nested closely with the inverted V-shaped pair TM8 and TM9. The adjacent helices TM5-TM7 and TM10-TM12 surround TM3-4 and TM8-9 in the center of the BetP protomer and α-helix 7 (α7) and TM2 at the trimer interface appear to clamp the transporter core, comprising most of the lipid-exposed BetP surface. The inversion of the segments TM3 to TM7 and TM8 to TM12 allows for tight helix packing and gives BetP its symmetric cylindrical shape. The two long, curved helices α7 and TM2

appear to act as spacers at the trimer interface and surround the transporter core like a belt.

8.2. Trimer architecture

The three-fold axis of the BetP trimer runs perpendicular to the membrane. The three protomers are labelled A, B, and C in Figure 56a. Except for their three C-terminal domains (Figure 56b), the protomers obey NCS. The C-terminal domain is oriented differently in each protomer and thus breaks the NCS for this part of the structure. The C-terminal domain of protomer A is resolved to Arg589, protomer B to Arg558 and protomer C to Arg568 in the SeMet-BetA structure. Furthermore it contains two clusters of positively charged residues, from Arg558 to Arg568 and from Lys581 to Lys587 (Figure 53). The second cluster is only resolved in the C-terminus of monomer A where it mediates the main crystal contacts between symmetry related BetP trimers (Figure 51).

The trimer is triangular, measuring ~100 Å on edge, and ~63 Å in height (Figure 56d). On the cytoplasmic side, loop 2 and the α-helical C-terminal domains extend into the cell. The C-terminal helix of protomer A protrudes by about 30 Å into the cytoplasmic side, including an angle of 32° with the membrane normal.

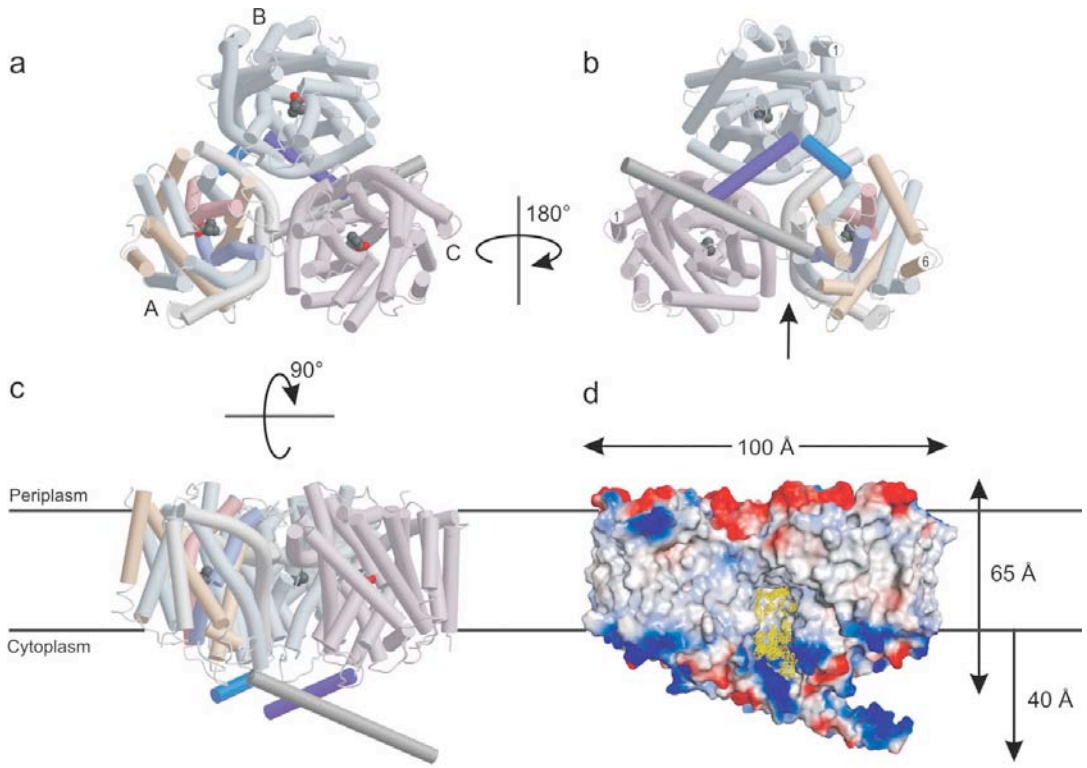


Figure 56| Trimer architecture of BetP. The BetP trimer as seen from the periplasmic **a** or cytoplasmic **b** side of the membrane. Protomer A (helices coloured as in Figure 52 but lighter colours, protomer B (grey) and C (light purple) are related by three-fold NCS with the exception of their C-terminal helices (darker colours). **b** In protomers A and C, the C-terminus faces the N-terminal TM1 (1) of protomer C and B, respectively, while in protomer B, the C-terminus points towards TM6 (6) in protomer A. The protomers are separated by a cleft (arrow). **c** View into the cleft between protomers A and C. The BetP trimer protrudes by about 10 Å from the periplasmic membrane surface. On the cytoplasmic side, the α-helical C-terminal domain of monomer A extends ~40 Å into the cell. **d** Electrostatic surface representation of the trimer oriented as in **c**. The electrostatic surface is coloured red (negative) to blue (positive). The F_{obs}-F_{calc} density map (contoured 2σ) is located in the hydrophobic cavity (yellow) that measures ~37 Å in height and ~30 Å in diameter. The boundaries of the lipid bilayer are indicated in **c** and **d**.

The trimer has a central conical cavity of 37 Å in height (Figure 56c and d) that ranges from 30 Å in diameter on the cytoplasmic side to 15 Å below the amphipatic helix α7 (Figure 57a). The cytoplasmic surface is positively charged, whereas the periplasmic surface is negatively charged (Figure 56d). The hydrophobic cavity is filled with non-protein density, most likely detergent or lipid (Figure 56d). The periplasmic side of the amphipathic α7 is rich in leucines and asparagines and show characteristics of a non-specific lipid interaction site (Hunte, 2005) and which may bind additional lipid or detergent in the central cavity to stabilise the trimer.

There are deep clefts between the protomers, which are accessible from the membrane (Figure 56d, Figure 57a). In the membrane these would be filled with lipids, which might contribute to a dynamic interaction of the protomers with each other. The main points of interaction between the protomers involve side chains of TM2, TM4, helix α7, TM9 and the C-terminal helices. Van der Waals contacts between TM2 and the horizontal amphipathic helix α7 of neighbouring protomer are made by Thr108, Phe112 in TM2 and Tyr340, Phe344, Phe345 in α-helix 7, His192 in TM4 interacting further with Thr108 in TM2 (Figure 57c). The main salt bridges between the neighbouring protomers are between Asn331 to Asp356 and Thr351 to Glu425 in TM9 of the neighbouring protomer. Trimer contacts on the cytoplasmic side mediated by the extended C-terminal helix are limited to two salt-bridges between Arg568 of protomer A and Glu552 of protomer C, and Arg568 of protomer C and Glu552 of protomer B, respectively.

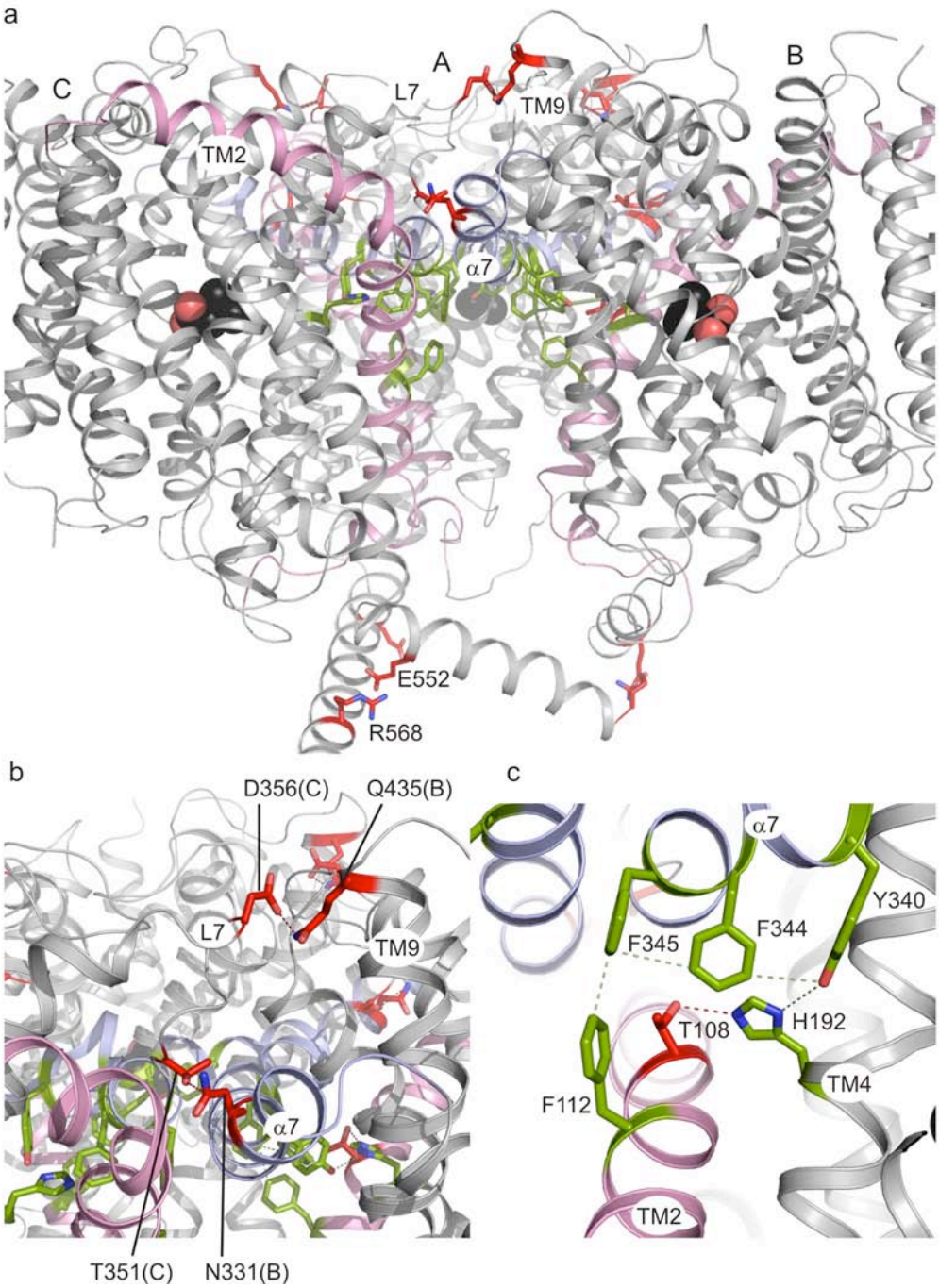


Figure 57| Interaction between the protomers. a, side view on the BetP trimer in grey with the central helix α7 (light blue) and TM2 (light pink), residues involved in salt bridges shown in red and aromatic residues shown in green; **b,** close-up on salt bridges between α7, TM9 and loop 7 **c,** close-up on aromatic interactions in chain B of helix α7 involving TM4 and TM2.

8.3. Substrate binding

Non-protein density was found in the core of BetP, approximately half way along TM4 and TM8, adjacent to the unwound region separating TM3i and TM3e. Side chains in TM4 and TM8 form the glycine-betaine-binding site. As in glycine-betaine-specific binding proteins (Schiefner et al., 2004b; Schiefner et al., 2004c), binding of the quaternary ammonium group involves an aromatic environment, with cation-π and van der Waals interactions (Figure 58a). Glycine-betaine is caged in an aromatic box formed by the indole groups of Trp374 (TM8), Trp189 and Trp194 (TM4). The indole groups of Trp194 and Trp374 are almost parallel, while Trp189 is nearly perpendicular to both and forms a lid on the periplasmic side. The side chain of Tyr197 makes up the fourth side of the aromatic substrate-binding box. The carboxyl group of glycine-betaine protrudes out of the box, forming a hydrogen bond with the main chain nitrogen atom of Gly151 in the unwound region of TM3. Trp194 is highly conserved in the BCCT family, while Trp189, Phe197 and Trp374 are only conserved in transporters of quaternary ammonium substrates.

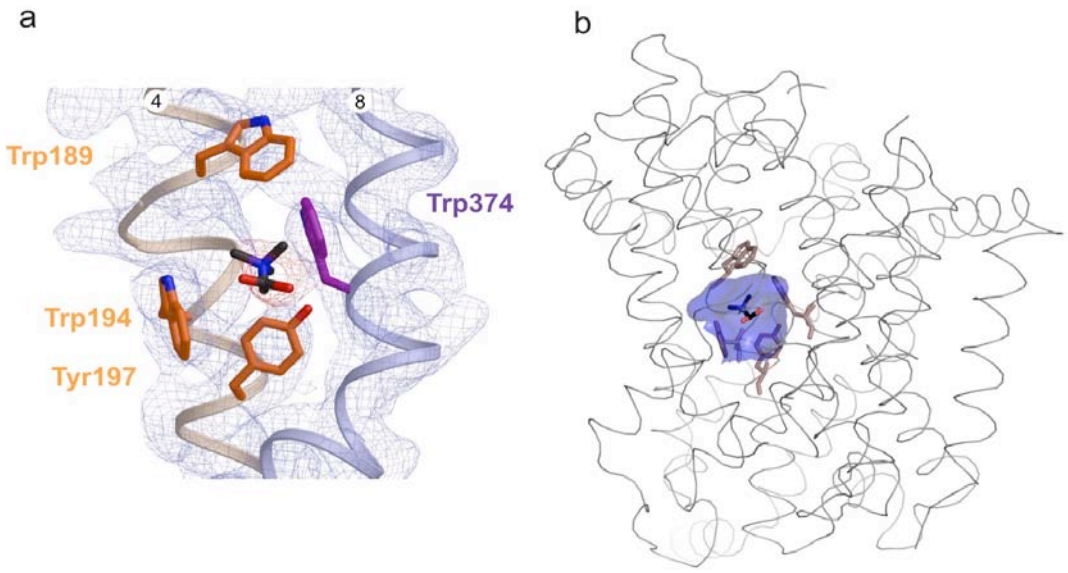


Figure 58| The betaine-binding site. a, σ_A-weighted $(mF_{obs}- DF_{calc})(2mF_{obs}- DF_{calc})$ map (blue), contoured at 1σ, showing density map for glycine-betaine in red, and the densities for TM4 (light brown) and TM8 (light blue) in blue. Side chains defining the tryptophan box of the betaine-binding motif are Trp189, Trp194, Trp374 and Tyr197. **b**, approximate same view on the betaine binding site. Residues of the Trp-box are in brown violet and glycine-betaine in black. The inner surface of the binding pocket is shown in transparent blue. Helices of protomer C are light grey.

Replacement of Trp189, Trp194, and Tyr197 in TM4 resulted in dramatically reduced betaine uptake rates (Figure 59, Figure 60). Mutation of Trp374 in TM 8 leaves the transport rate unaffected (Figure 60), suggesting that the substrate is still accommodated, probably at an altered position, in the modified binding pocket. Removal of Trp362 pointing out of the substrate pathway into the periplasm resulted in reduced activity (Figure 59, Figure 60) suggesting a role in substrate sequestration or its involvement in a periplasmic binding site. Mutation of Trp377 below the substrate-binding pocket inactivates BetP, while mutation of Trp366 and Trp371 do not affect activity (Figure 59, Figure 60). Trp377 might thus play a key role in substrate release.

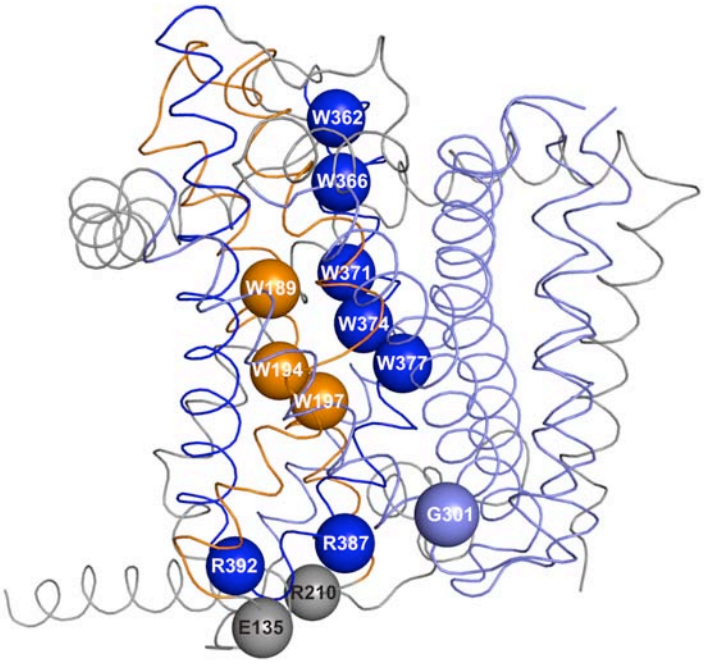


Figure 59| Location of characterised BetP mutants.

The inner surface of the binding site encloses a large volume compared to the size of glycine-betaine (Figure 58b). The binding site accommodates two sodium ions in addition to the glycine-betaine (8.4) and may also facilitate conformational changes of aromatic residues (Figure 61). The binding pocket might also harbour some specific water molecules, which cannot be detected at a resolution of 3.35Å. Specific water molecules might mediate residue interaction, or as discussed later (8.4), might mediate sodium coordination.

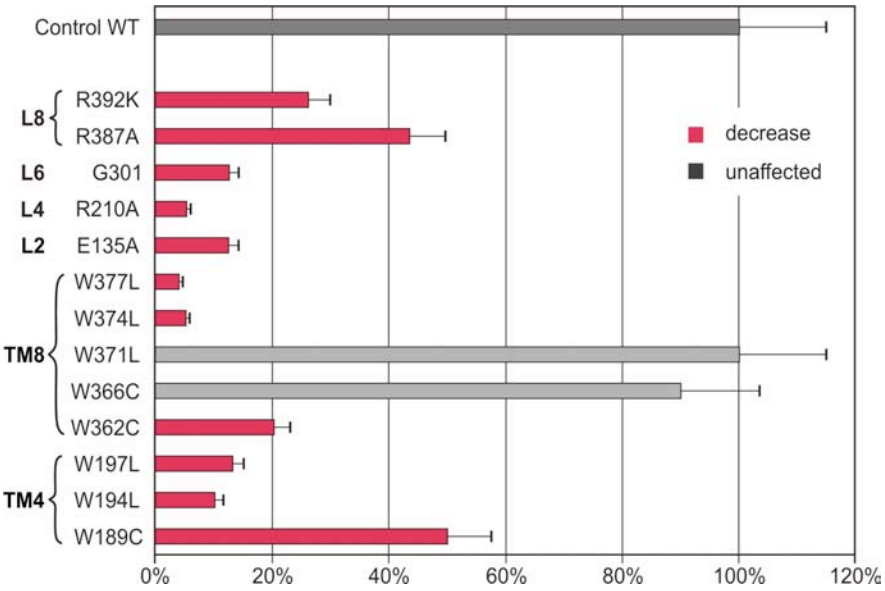


Figure 60| Activity measurement on selected BetP mutants. Transport activity determined from the initial slopes of [^{14}C]glycine-betaine uptake time curves of BetP mutants. Activity is expressed as percentage of WT control (100%) at optimal osmolarity. Immuno-blotting against the N-terminal StrepII tag assessed the expression levels of mutants and activity was normalized accordingly. Dr. Vera Ott, Department of Biochemistry, University of Cologne, made these mutants and measurement (Ressl et al., 2009).

There are 64 aromatic residues in each BetP protomer. Side chains of 23 of these residues line the possible pathway of the substrate along the 4-helix bundle (Figure 61). The assembly of aromatic residues in BCC transporters may be related to the broad spectrum of different osmolytes that can be transported. Assuming that BCC transporters exhibit the same overall fold with a rather narrow substrate pathway, a general solution for osmolyte transport would be to provide several aromatic boxes along the pathway. Clearly, such pathway composition creates surface properties ensuring that a wide variety of osmolytes can be transported without interaction with the protein backbone. Usually, osmolytes acts as co-solvents and are excluded from the first hydration shell of proteins by repulsive interaction with the protein backbone (Bolen and Rose, 2008). Coating the binding site with aromatic side chains would solve this problem, by providing a surface that would, according to Tanford's transfer model (Tanford, 1970), minimally repel osmolytes (Horn, 2006; Kuhlmann et al., ; Schiefner et al., 2004b; Schiefner et al., 2004d).

In the crystal structure, the bound substrate is not accessible from either side of the membrane. Tyr157, Phe156, Phe165, Trp366, Tyr370, Trp411, Phe412 and Phe416 lock the substrate binding site on the periplasmic side, and Glu135, Phe136, Trp141, Phe209, Phe384, Arg387, Arg390, Arg392, and Leu487 obstruct it on the cytoplasmic side.

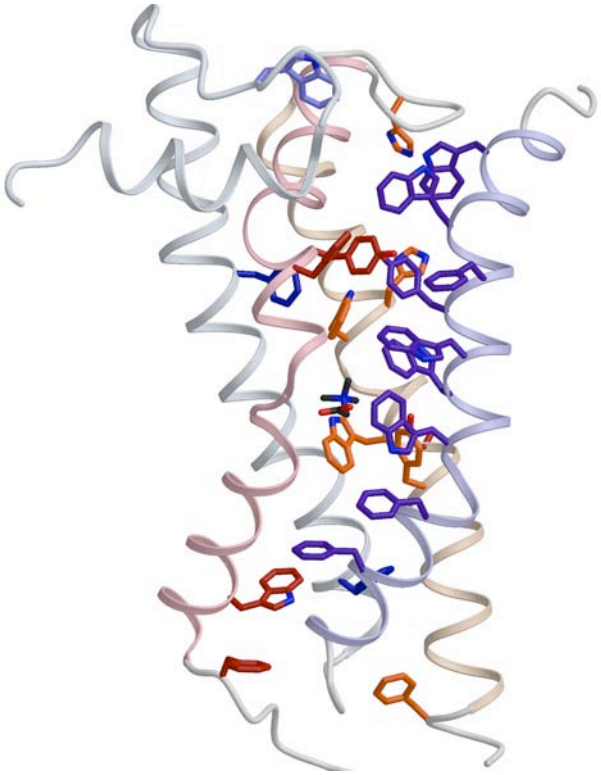


Figure 61| 23 aromatic side chains line a possible substrate pathway. 23 aromatic side chains in the central 4-helix bundle TM3, TM4/TM8, TM9 line a possible path of the substrate through the protomer.

8.3.1. The intracellular gate at Trp377

Transporters can have both intra- and extracellular gates that define the energy barriers for conformational switching between the inward- and outward-facing states. Although several point mutations along the putative substrate pathway render BetP inactive (Figure 59, Figure 60), Trp377 is of particular importance. It lies on the cytoplasmic side of the substrate binding site and, together with Phe380, comprises an obstruction to the release of glycine-betaine from the interior of the protein.

In the crystal structure Phe380 builds the only stabilising interaction to Trp377. Analysis of the stereochemical properties of Trp377 and alternative interaction

partners resulted in a proposed interaction to Thr250 in TM5. The crystal structure shows that to establish a hydrogen bond between Nε of Trp377 and Oγ of Thr250, major changes of the surrounding helices would be required.

A possible gating mechanism to facilitate the further course of substrate and ions towards the cytoplasm involving Trp377 could be described as follows:
In the closed conformation (Figure 62a) of the gate towards the cytoplasm (closed-gate), the position of Trp377 is stabilised by a hydrogen bond between Nε of Trp377 and Oγ of Thr250 and by a side-to-face aromatic interaction from Trp377 to Phe380. The by following their electrochemical potential towards the cytoplasm, sodium ions could break delicate aromatic interaction on their course. This allows Trp377 to flip into another rotamer conformation, resulting in a break of the hydrogen bond to Thr250. In general, sodium ions might break aromatic interactions and thus induce changes in their rotamer conformation to favour cation-π interaction to the ion. Upon loss of interaction with Trp377, Phe380 may rotate around the chi2 angle to build a cation-π interaction to a sodium ion (Figure 62). A break of these three bonds may result in a straightening of TM8, which in turn may induce further or simultaneous conformational changes in TM5 and TM3i. After the gate at Trp377 is opened, sodium ions plus substrate move towards the cytoplasm and along the aromatic residues in TM8 and TM4.

The crystal structure shows an intermediate conformation between these two hypothetical closed-gate and open-gate conformations (Figure 62c,d). The electron density map for Trp377 was not of equal quality in all three protomers, which points to its possible functional flexibility.

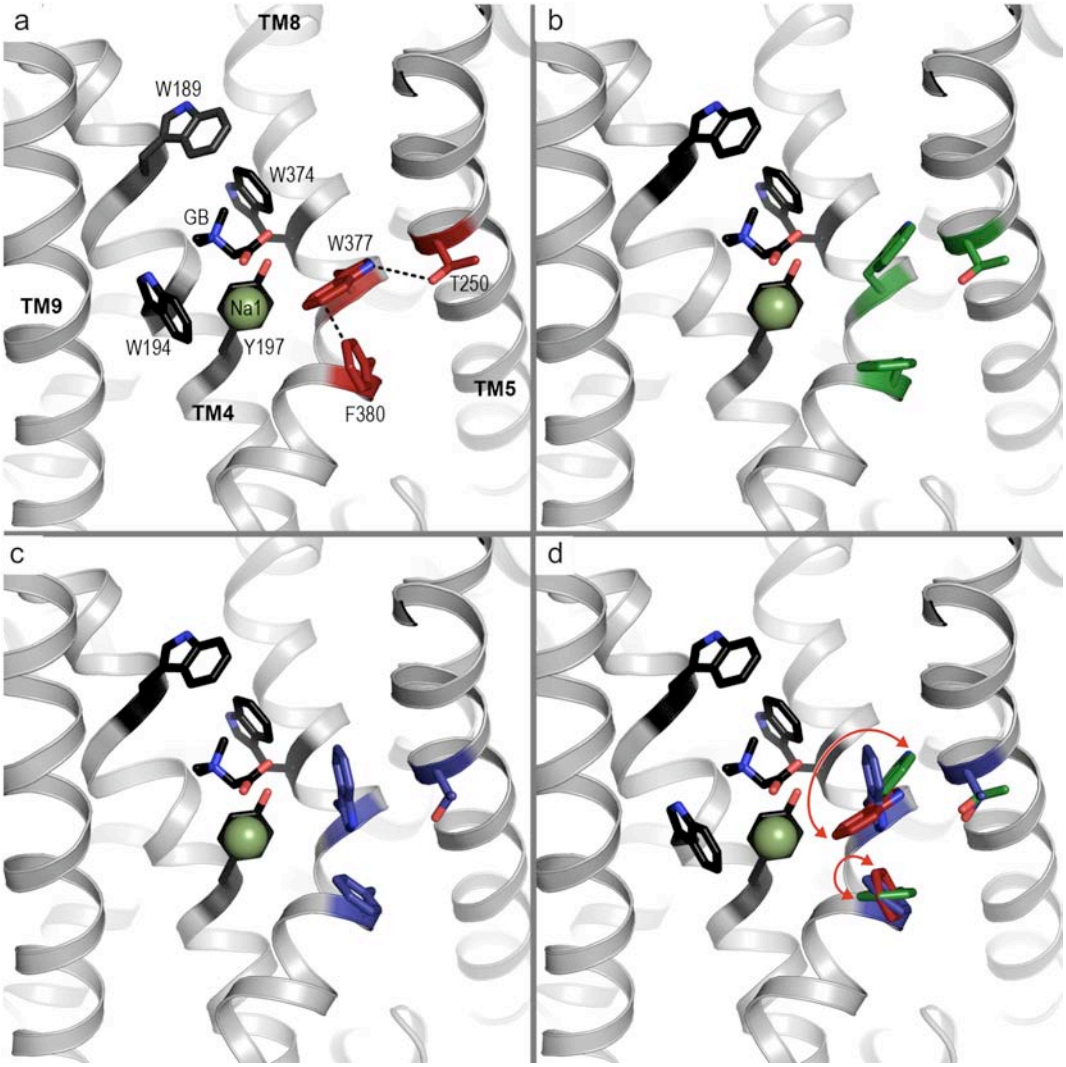


Figure 62| Possible gating by Trp377 and Phe380. View from the perspective of Na2 towards the binding pocket. Na1 in green, glycine-betaine (GB) and Trp374, 189, 194 and Tyr197 in black. Surrounding helices are shown in grey. **a**, putative closed gate is stabilised by an H-bond between the Nε of Trp377 and the Oγ of Thr250 (both in red) indicated with a black dashed line. Trp377 is further stabilised by an aromatic edge-to-front cation-π interaction to Phe380 (red) marked by a black dashed line. All residues for the closed gate are shown in pale orange. **b**, putative open gate of Trp377 and Phe380, both in green. Trp377 is in another rotamer conformation towards TM8 and Phe380 is rotated around the chi2 angle. Residues for the open gate are coloured in light green. **c**, side chain arrangement of the gate in the crystal structure of BetP, all Trp377, Phe380 and Thr250 coloured in blue. **d**, Superposition of all three gating states. Red arrows indicate the rotational gating movement of Trp377 and Phe380. The open and closed gating states were modelled by changing rotamer conformations and chi angles of the respective residues using COOT.

8.4. Putative sodium binding site

Sodium is required to energise glycine-betaine transport in BetP, which has a K_m for Na^+ of 4 mM (Farwick et al., 1995; Schiller et al., 2004a). The 3.35Å electron density map of the BetP crystal structure was not of enough quality to unambiguously assign the position of the sodium ions. Two putative Na^+ binding sites (Figure 63) were assigned into the BetP structure based on structural alignment of BetP on the $LeuT_{Aa}$ structure (Yamashita et al., 2005),

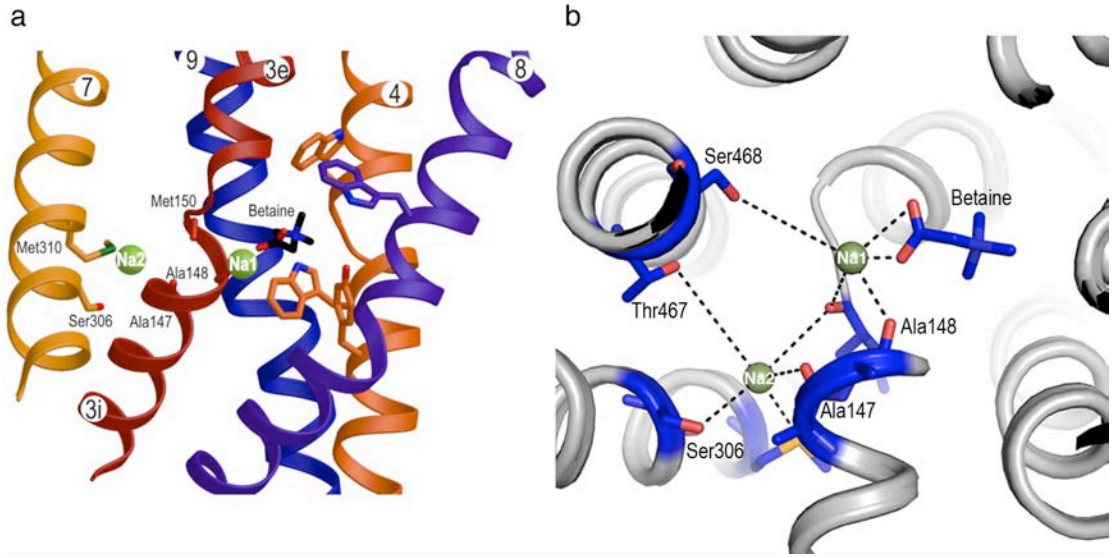


Figure 63| Putative sodium binding sites in BetP. a, Side view on sodium sites, sodium ions shown in green. Na1 is coordinated by the carboxyl group of betaine and residues in TM3i. Na2 is coordinated by residues in TM3i and TM7. **b,** View from the cytoplasm with sodium ions in green and black dotted lines indicate the coordination between the atoms of residues involved in Na^+ binding. The helices are displayed in light-grey.

Sodium binding in BetP seems to be facilitated by main chain interactions in the unwound glycine-rich stretch of TM3 with one Na^+ ion (Na1) coordinated by the carboxyl group of betaine, the carbonyl oxygen of Met150 in the unwound region of TM3 and the carbonyl oxygen of Ala148 in TM3i (Figure 63). The location of the second Na^+ ion (Na2) is assigned in the bent between TM3i and TM3e. The carbonyl oxygen of Ala147 in TM3i and the side chains of Ser306 and Met310 in TM7 coordinate Na2. Thr467 and Ser468 in TM10 are suggested to be additional ligands of Na2 with a probable interaction mediated by water molecules. Ser468 is conserved in the BCCT family, as are the corresponding Ser355 in the neurotransmitter sodium

symporter (NSS) family (Chang et al., 2004) and Ser364 or Ser 365 in the sodium solute symporter (SSS) family (Saier, 2000). However, the glycines flanking Met150 in the stretch of TM3 appear only to be conserved in transporters exhibiting a 2:1 sodium to substrate stoichiometry (Figure 64). Interestingly, sodium to substrate stoichiometries have been investigated and vary from 1:1, 2:1 to 3:1 among different transporters (Kavanaugh et al., 1992; Roux and Supplisson, 2000; Rudnick, 2002) and may be directly related to the nature of the substrate.

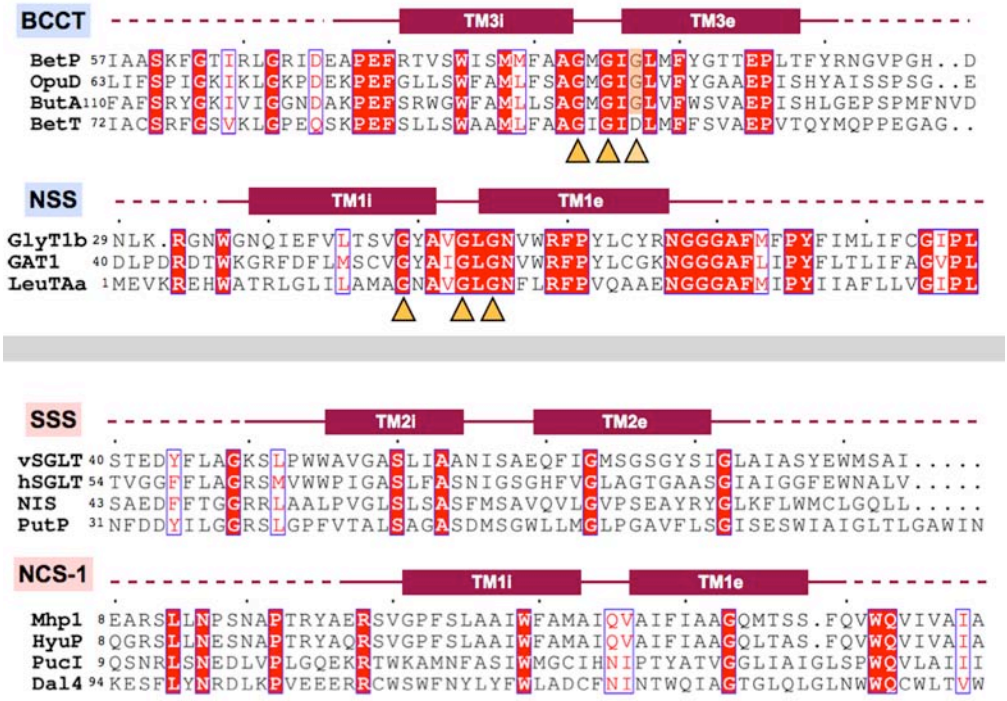


Figure 64| Sequence alignments of the first helix of the first repeat of transporters of the BCCT, NSS, SSS and NCS-1 family by ClustalW multiple sequence alignment displayed by ESPript 2.2 (http://espript.ibcp.fr/ESPript). Conserved glycines of the G-x-G-x-G motif are indicated by orange triangles.

In the BCCT family, high affinity betaine transporters (Figure 54) exclusively contain the conserved G-x-G-x-G motif in TM3, suggesting that the specificity to betaine is related to the structure of the unwound segment in BetP and may also depend on binding of Na1. The assigned position for the second sodium ion in BetP is in a slightly different location compared to the related sodium in the $LeuT_{Aa}$ structure. Molecular dynamics calculations showed the proposed position of Na1 to be very stable, compared to that of Na2 (Dr. Kamil Khafizov, personal

communication). Analysis of the electrostatic surface distribution around the sodium ions proposed the Na2 location in a negatively charged pocket that is formed by residues from TM7 and TM3i (Figure 65).

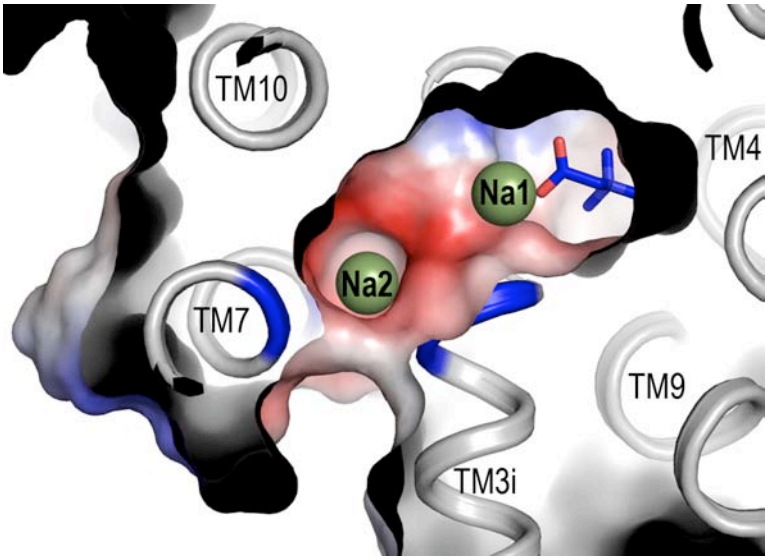


Figure 65| Electrostatic surface around the proposed sodium ions in the BetP structure. View from the cytoplasm with sodium ions in green. Helices are displayed in light grey and parts that are involved in sodium coordination are shown in blue. The electrostatic surface of the binding pocket was calculated using PyMOL. The carboxyl group of glycine-betaine coordinates Na1 and Na2 is located in a negatively charged pocket between TM3i, TM7 and TM10.

The substrate binding site and the assigned positions of the Na^+ ions are blocked from the periplasm in the BetP structure, but a narrow funnel between TM3i, TM7, TM8 and TM10 links the putative sodium sites to the cytoplasm (Figure 66a). The funnel has an elliptical constriction defined by Ala147 (TM3), Ile302 and Ser306 (TM7), Phe380 and Phe384 (TM8), and Ser471 (TM10), with a major axis of ~4.5 Å and a minor axis of ~2 Å (Figure 66b). If the BetP crystal structure would represent a cytoplasmic open conformation, the two sodium ions would have to pass through this narrow constriction towards a positively charged environment (Figure 66b). Furthermore, due to steric reasons and the limited side chain size of the residues involved in the constriction, the opening of the funnel likely requires a rearrangement of the polypeptide backbone, *i.e.* a considerable conformational change. Hence, from this and comparisons with other transporter structures, discussed in more detail in

chapter 8.9, it can be concluded that the BetP structure presents an intermediate state of the transporter.

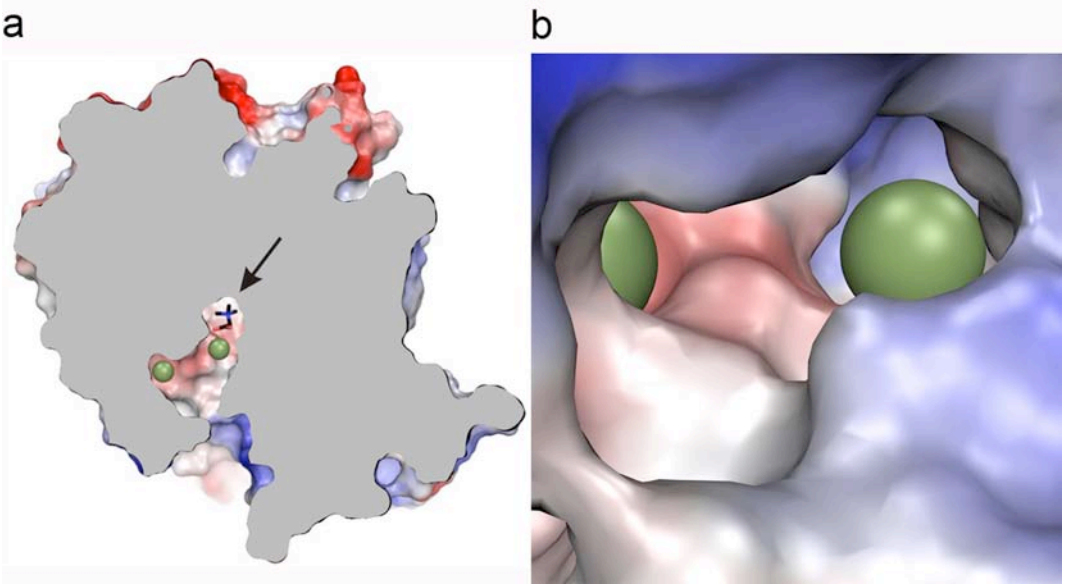


Figure 66| Cytoplasmic funnel towards the binding pocket. a, section through BetP protomer C in the plane of the substrate and the sodium ion (green) location. Accessibility of the two Na^+ ion positions and betaine (arrow) from the cytoplasm in a section through the protein volume. **b,** view from the cytoplasm on the elliptical constriction towards the binding pocket. Na1 is on the right hand side and Na2 is on the left hand side; both coloured green. Electrostatic surface potential is calculated and displayed in PyMOL.

8.5. Cation-π interactions in BetP

A delicate balance of interactions determines a protein structure. Hydrogen bonds, salt-bridges, and hydrophobic interactions play a role in folding and establishing the final structure of a protein. In addition, cation-π interactions are an important type of noncovalent binding interaction. Theoretical and experimental studies have shown that cation-π interactions are six-fold weaker than a salt-bridge in the gas phase. Calculation for an aqueous solution showed that the magnitude of salt-bridge interaction is reduced by over 50-fold, relative to the gas phase, whereas the cation-π interaction is reduced by less than 3-fold. This suggests that cation-π interactions will remain strong across a wide range of dielectric constants (Gallivan and Dougherty, 1999).

Cation-π interactions are stabilising interactions between a cation and π-orbital of an aromatic residue formed by sp^2 carbon electrons that results in an negative electrostatic potential in the middle of the aromatic ring. The cation-π interaction properties in aqueous media are known to be important molecular recognition sites and serve as general receptors for organic cations (Dougherty, 1996). Tryptophanes are preferred in these interactions and studies revealed that 25% of all tryptophanes are involved in cation-π interaction (Gallivan and Dougherty, 1999). Cation-π interactions are found in several important protein-ligand or drug receptor binding. For instance, the acetylcholine receptor shows a box-like architecture of tryptophans around acetylcholine (Xiu et al., 2009), similarly to the tryptophans in the binding side for glycine-betaine in BetP.

The overall distribution of aromatic residues in the BetP trimer shows that most cluster at the 4-helix bundle towards the central cavity of the trimer (Figure 67a). Together with the location of arginines or lysine residues, most of the cation-π interactions are located at the perimeter of the trimer (Figure 67b).

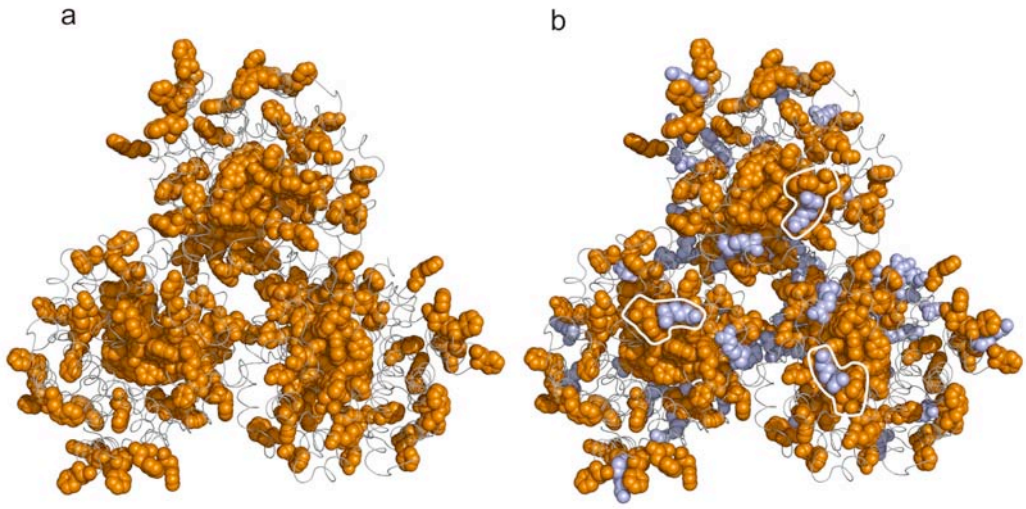


Figure 67| Distribution of aromatic and basic residues in the BetP trimer. a, top view on the trimer shows the overall distribution of aromatic residues in the structure. Most of the aromatic residues (orange) are located towards the middle of the trimer. **b,** same view as in **a** including the basic residues Lys and Arg (blue-grey) in the structure. The location of the cation-π between Trp431 and Arg167 is indicated by a white line.

Each protomer of BetP has five cation-π interactions: between Trp431 – Arg167, Phe83 – Lys81 and Phe187 – Arg350 on the periplasmic side and Phe209 – Arg390 and Trp141 – Arg392 on the cytoplasmic side. Two cation-π interactions between Trp431 - Arg167 and Trp141 – Arg392 are located at the 4-helix bundle on the periplasmic side and on the cytoplasmic side, respectively. Trp431 - Arg167 links the potentially important EH2 region with the upper part of TM3e and might be mediating conformational changes of EH2 during substrate acquisition (Figure 68). In this region, a secondary binding site was found for the Na^+/Cl^--coupled alanine transporter $LeuT_{Aa}$ from *Aquifex aeolicus*. $LeuT_{Aa}$ resembles BetP in the inverted repeat fold (Figure 55) and includes a structurally comparable region to EH2 of BetP. Furthermore, clomipramine bound to a similar located cation-π interaction site in $LeuT_{Aa}$ and blocked the transporter in its periplasmic open state (Singh et al., 2007). Trp141 – Arg392 links TM3i to loop8 and may play an important role in the activation of BetP and is further discussed in chapter 8.8.

Cross linking studies of these potentially important sites would be necessary to find out whether these cation-π interaction have an impact on transport and/or substrate acquisition.

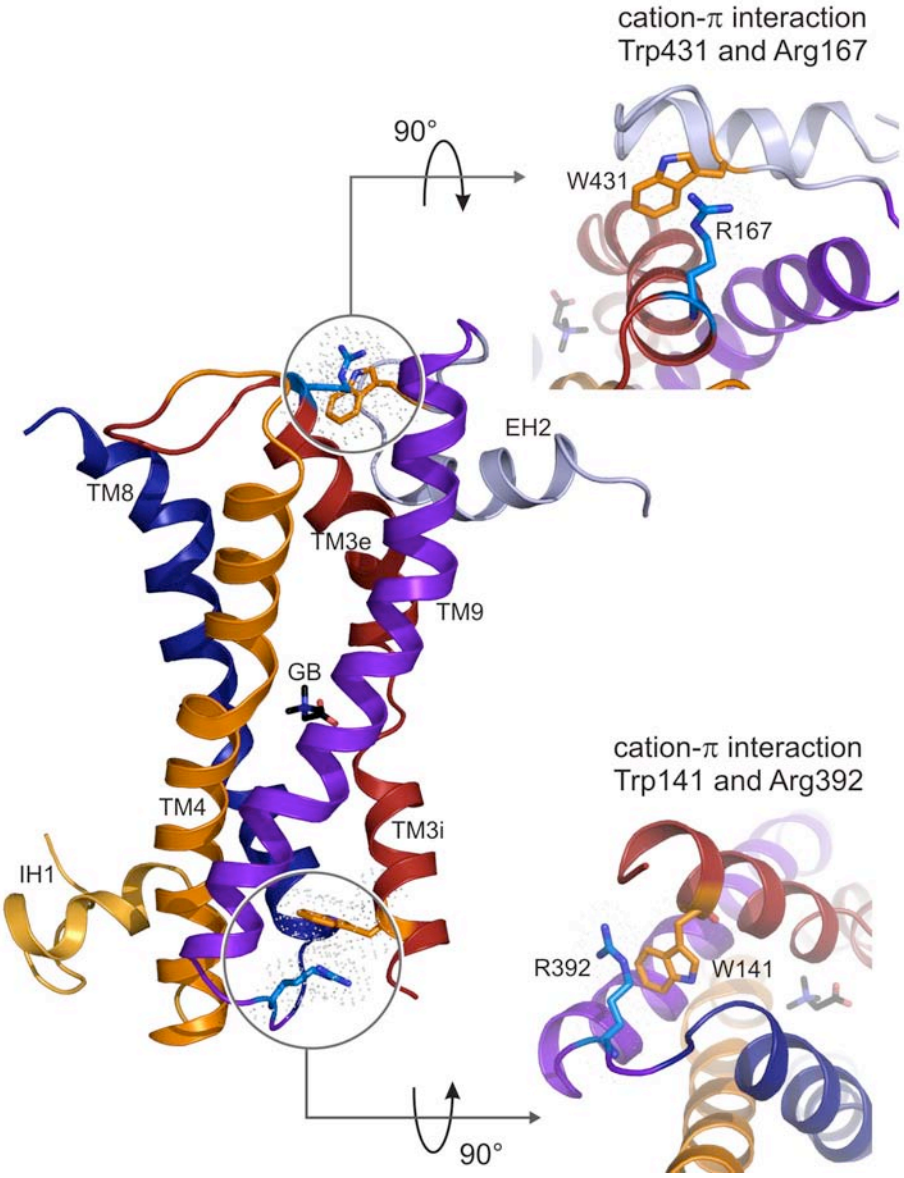


Figure 68| Cation-π interaction at the 4-helix bundle. Trp431 – Arg167 on the periplasmic side of the 4-helix bundle and Trp141 – Arg392 on the cytoplasmic side. A close up on both cation-π interactions is given on the right hand side, both show top views ~90 rotated from the side view of the 4-helix bundle.

8.6. Interactions of the C-terminus of protomers A, B and C

Besides the crystal contacts between the crystallographic related BetP trimers, interactions between protomers on the cytoplasmic side involve the C-terminal domain and towards the periplasmic side the amphipathic helix α7. Helix α7 makes contact with TM2, TM3, TM9 and to helix α7 of the other two protomers, while the C-terminal helix interacts with loop 2 and the C-termini of the neighbouring

protomers. The C-terminal domain contains two clusters of positively charged residues, from Arg558 to Arg568 and from Lys581 to Lys587 (Figure 69).

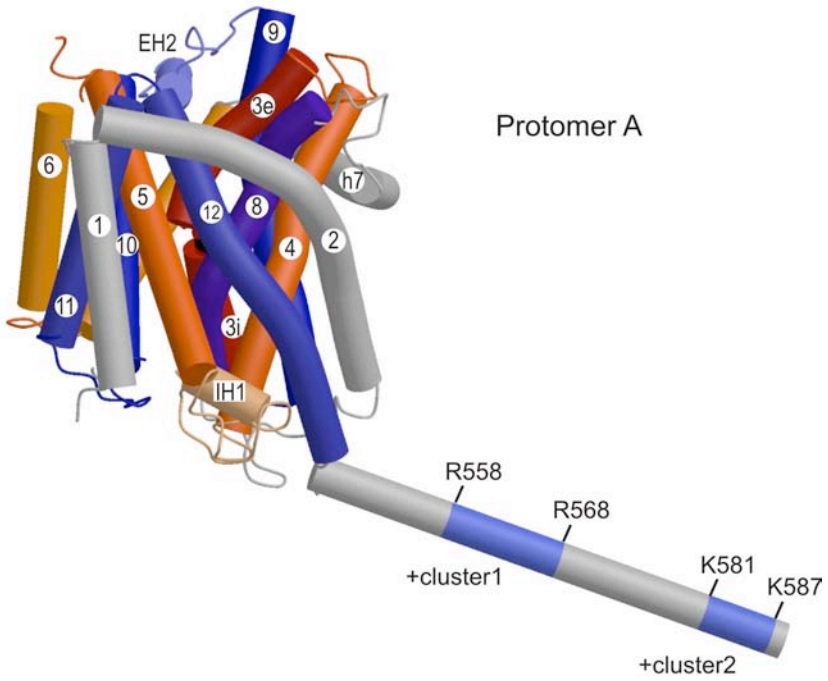


Figure 69| Positively charged clusters in the C-terminal domain of protomer A. The positively charged clusters in the C-terminal domain of protomer A are indicated as +cluster1 for Arg558 to Arg568 and as +cluster2 for Lys581 to Lys587 in blue.

In the crystal structure, all three helical C-terminal domains are different and do not obey NCS (Figure 56). They are resolved in protomer A to Arg589, in protomer B to Arg558 and protomer C to Arg568. The second cluster is only resolved in the C-terminus of protomer A where it is stabilised by the crystal contacts (Figure 51), which contribute to the break in NCS. However, other breaks of NCS are clearly not due to crystal contacts. Arg568 residues of protomers A and C form a salt-bridge to Glu552 of the adjacent protomer. This interaction is not found for Arg568 in protomer B (Figure 70). The structure of BetP also reveals an important role of loop 2, which was not observed by peptide array analysis (Ott et al., 2008). The role of loop2 could be supposed as a blocking region for Na^+ ions and substrate at the cytoplasmic funnel (Figure 66) in an outward-facing conformation of the transporter,

and might therefore undergo larger conformational changes. Loop2 contains the three negatively charged residues Asp131, Glu132 and Glu135. Asp131 interacts by a salt-bridge with Arg565 located in the C-terminal helix of the adjacent protomer C, and Glu132 with Arg390 in loop 8, which connects TM8 and TM9 in the 4-helix bundle (Figure 70). Loop 8 contains an additional positively charged patch of Arg387 and Arg392, and transport activity is sensitive to mutation of both arginines (Figure 60). Changes in orientation of the C-terminal helix might thus affect the orientation of Asp131 and in turn Glu132 and hence the conformation of loop 8, which gates the substrate pathway and thus might lead to a regulation of the transport cycle by ionic interactions between all three protomers. The described interactive cooperation suggests a connection between the activation cycle and the transport cycle in BetP.

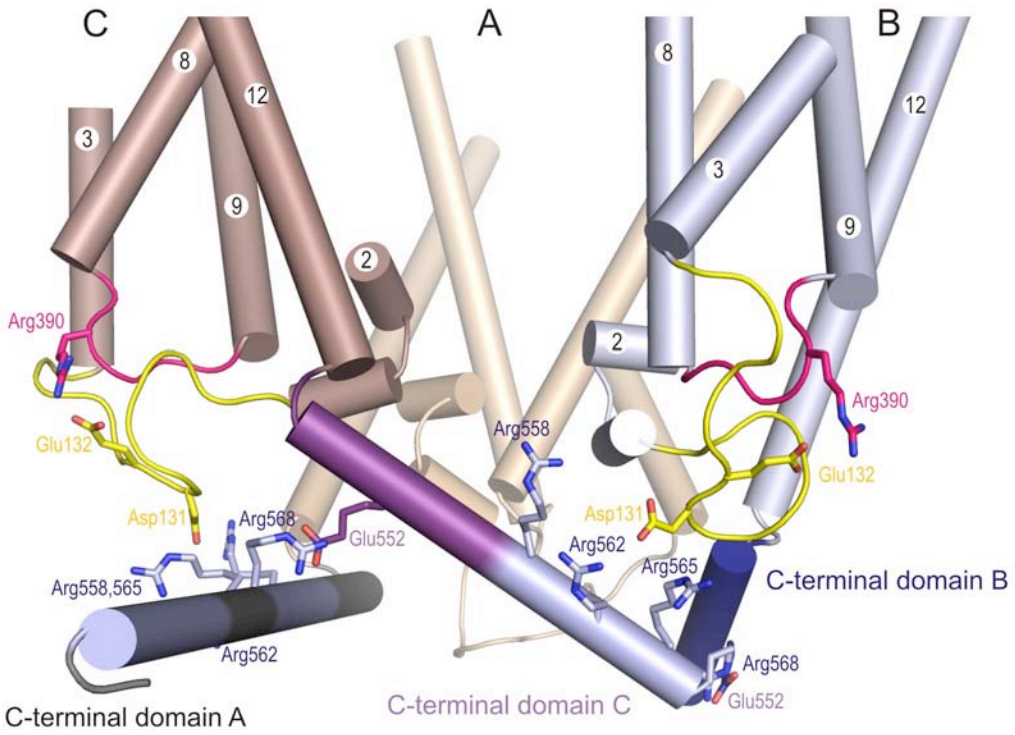


Figure 70| Ionic interactions of the C-terminal domains. Side view on protomer C (brown-purple) and B (blue) and protomer A (crème) with the involved helices. Cytoplasmic loop 2 (yellow) and loop 8 (pink) are shown with the respective residues involved in ionic interactions. The two positively charged clusters in each domain are shown in light blue. Residues forming salt-bridges are labelled separately.

In proteoliposomes, activation of BetP in response to osmotic stress seems to be specifically triggered by high internal K^+ concentration, although the activation optimum of 220mM (Schiller et al., 2004b; Schiller et al., 2004c) is high for a specific

K^+ binding region. In addition, the cation specificity depends on the membrane composition indicating concomitant or competing stimuli in osmoregulation. A direct interaction of the BetP C-terminus with charged lipids in the membrane was observed by surface plasmon resonance (Ott et al., 2008), and could influence the accessibility of a potential regulatory K^+-binding region. It might be possible that the different orientations of the C-terminal domains reflect functionally different conformations in the activation cycle. Regulation of transport might thus be a cooperative effect of interactions between the three BetP protomers. Interestingly, the number of conformationally essential residues, which do not tolerate a substitution for proline is larger when BetP is expressed in *E. coli* membranes compared to BetP expressed in *C. glutamicum* (Ott et al., 2008). Obviously, the impact of amino acid replacements on the regulatory behaviour of BetP depends on both the nature of the amino acid and the type of surrounding lipids.

It can be proposed that each protomer in the trimer activates its neighbour, taking advantage of the trimer architecture of BetP.

8.7. Resting state of BetP

During substrate translocation across a membrane, transporters undergo different conformational changes that allow either substrate binding or release, referred to as outward-facing or inward-facing states, respectively. The conversion from an outward-facing to an inward-facing state is thought to go through an occluded state of a transporter, in which the substrate is not accessible from either side of the membrane. The necessity of the occluded state was deduced from observations of transporter structures, in which the substrate was shown to be occluded at least to one side of the membrane (Gouaux, 2009). As active transport of a substrate against its concentration gradient cannot comprise an open channel conformation, it was required to add the occluded state to describe the transport mechanism. The lifetime and stability of such an occluded state may vary from transporter to transporter. The crystal structure of BetP shows that the substrate is occluded from the periplasmic side and, aside of the narrow funnel, the cytoplasmic side (Figure 66). Furthermore, the inverted repeat motif of BetP was analysed by structural comparison with other transporters sharing the inverted repeat motif. The

other transporters show either the outward- or inward-facing conformation. The structural superpositions of the two halves of the inverted repeat within and between each transporter, resulted in a symmetric arrangement of BetPs inverted repeat (see chapter 8.9) and is referred to as an intermediate state between the outward- and inward-facing conformations.

The intermediate and substrate bound state of BetP in the crystal structure may also present a functionally relevant resting state. In a resting state of BetP, the transporter is loaded with substrate and locked in a conformation that does not allow substrate release into the cytoplasm. This is facilitated by the interactions of the C-terminal domains within the trimer. Considering the high substrate specificity and affinity, the high accumulation ratios together with the rapidly reached maximum activity level of BetP, a resting state may serve as an efficient starting position for BetPs rescue response to hyperosmotic stress.

8.8. Atomic activation model of BetP

A mechanistic model of BetP activation by hyperosmotic stress was deduced by combining biochemical with structural data from this work. It involves a change of location of the C-terminal domain between a membrane-attached and a loop-attached state (Kramer, 2009). The state in which parts of the C-terminal domain are attached to the inner membrane surface is interpreted as the resting or inactive state of the transporter. At low hyperosmotic stress, the C-terminal domain is bound to a high density of negative charged lipids at the inner membrane surface.

In the activated state of the transporter the C-terminal domain is proposed to interact with parts of the cytoplasmic loops of the protomer and with the N-terminal domain. This conformational state will be favoured at high hyperosmotic stress, in which K^+ ions compete with the C-terminal domain for binding to the negative electrostatic potential at the inner membrane surface (Kramer, 2009).

The trimeric nature of native BetP was discovered previously by 2D electron crystallography (Ziegler et al., 2004), leading to the assumption that the trimeric state is directly connected with its activation and inactivation processes (Kramer, 2009). Furthermore, a protomer on its own is presumed to transport constitutively, perhaps comparable to a deregulated C-terminally truncated mutant of BetP. Therefore the

activation mechanism and the transport mechanism of BetP can be discussed separately.

The crystal structure clearly shows that the C-terminal domain of each BetP protomer is interacting primarily with the adjacent protomer. This leads to two main possibilities: (a) a functional interaction between the three BetP protomers in the trimer exists; (b) the interaction of the C-terminal domain might be directly linked to the activation cycle of BetP.

Alanine and proline scanning of the C-terminal helix indicated that the arginines in both clusters are important for regulation (Ott et al., 2008). Furthermore, peptide array analysis (Ott et al., 2008) suggests strong interactions of positively charged C-terminal domains with the negatively charged N-terminal domain. N-terminal truncation resulted in activation at higher osmolarity in *C. glutamicum* and in *E. coli* polar lipid proteoliposomes (Figure 2), indicating that the interaction between the N- and C-termini plays an important role in regulation. When not constrained by crystal contacts, the C-terminal helix appears to have freedom to move in a plane parallel to the membrane surface. In the BetP structure, the C-terminus of protomer C points towards TM1 of protomer B, suggesting a possible interaction with the N-terminal domain, while the C-terminus of protomer B is oriented towards TM6, pointing to a potential interaction site with residues in loop 4 and loop 8.

Assumptions and model

This model focuses mainly on the C-terminal domain of protomer A and residues in loop2, loop8 and Trp141 in TM3i of protomer C. Furthermore, it assumes possible interactions between the N-terminal domain and the C-terminal domain (Ott et al., 2008), and takes into account that the C-terminal domain interacts with the negatively charged inner membrane surface.

The atomic activation model starts from a resting state of the transporter, for which the specific location of interacting regions is described below (Figure 71).

The C-terminal domain (chain A) protrudes into the cytoplasm and passes under the neighbouring protomer (chain C). The first positively charged cluster (+cluster1) contains Arg558, Arg562 and Arg565 that build three possible salt-bridges with Asp131 in loop2 (Figure 71c). In the crystal structure, two different interaction

partners form a salt-bridge to Asp131: Asp131 (chain B) to Arg565 (chain C) and Asp131 (chain C) to Arg558 (chain A).

+cluster1 is subsequently followed by a short region (Figure 71), for which the electron density map of the side chains was not as well resolved as for side chains in +cluster1 and +cluster2. This suggests a higher flexibility for this short region between the two charged clusters. +cluster2 is located at the end of the C-terminal domain and contains Arg583, Arg584, Arg586 and Lys587 pointing towards the inner membrane surface. Lys581 and Arg582 point downwards into the cytoplasm (Figure 71a,c). This stretch of positive charges around the periphery of the C-terminal domain could be involved in simultaneous interactions with the membrane surface and a negatively charged cluster of the N-terminal domain, and thus serve to anchor the C-terminal domain to the membrane.

The N-terminal domain has the following sequence with two negatively charged regions, which are marked in red. Ser57, marked in green, is the first residue at the N-terminal side for which electron density was observed in the map and thus the crystal structure starts from:

MTTSDPNPKPIVEDAQPEQITATEELAGLLENPTNLEGKLADAEEEIILEGEDTQAS

The helix TM1 starts at Trp60 and Ser57 is thus already very close to the membrane embedded level of BetP, therefore the negatively charged cluster in the beginning (underlined above) of the N-terminal sequence interacts with the C-terminal domain and hence, seems to be an important trigger for the electrostatic interaction between the C-terminal domain and the membrane surface. Early transport measurement on N-terminal and C-terminal deletion mutants suggested a modulator function for N-terminal domain in the activation process of BetP (Peter et al., 1998). One of the main factors between the inactive/resting and active state of BetP is thought to be the increase of K^+ ions in the cell. Upon an increase of K^+ ions into the cell, it can be proposed that the N-terminus is thus modulating the strength of electrostatic interaction of the C-terminal domain to the lipid surface (Figure 72). I may be that the C-terminal and N-terminal domain are bound to each other at all time; throughout the activation cycle upon down regulation of transport.

Focusing now on the +cluster1, a conformational change could potentially be induced through the detachment of +cluster2 from the membrane. This might result

in a conformational change between the three possible interaction partners Arg558, Arg562 and Arg565 to Asp131 in loop2 (Figure 71, Figure 72). Loop2 contains Glu132, which forms a salt-bridge to Arg390 in loop8 in the inactive/resting state. A change of ionic interaction partners of Asp131 to one of the Arg residues in +cluster1 could therefore induce subsequent conformational change in loop2 and result in a switch in salt-bridges of Glu132 from Arg390 to Arg129. Loop8 contains a cation-π interaction between Arg392 to Trp141. Trp141 is located at the beginning of TM3i and both residues are conserved in the BCCT family (Figure 54, Figure 71). If conformational changes are induced in loop2 by changes in the C-terminal domain, this could in turn have an impact on conformational change in loop8 as well. The difference in length between these two loops is 13 residues of loop2 to 5 residues in loop8. Loop2 can therefore undergo larger conformational changes than the shorter loop8. Interestingly loop8 contains a glycine residue Gly391, which is located between Arg390 and Arg392, and is conserved in the BCCT family as well (Figure 54). Gly391 might therefore be an important pivot point for conformational changes occurring in loop8.

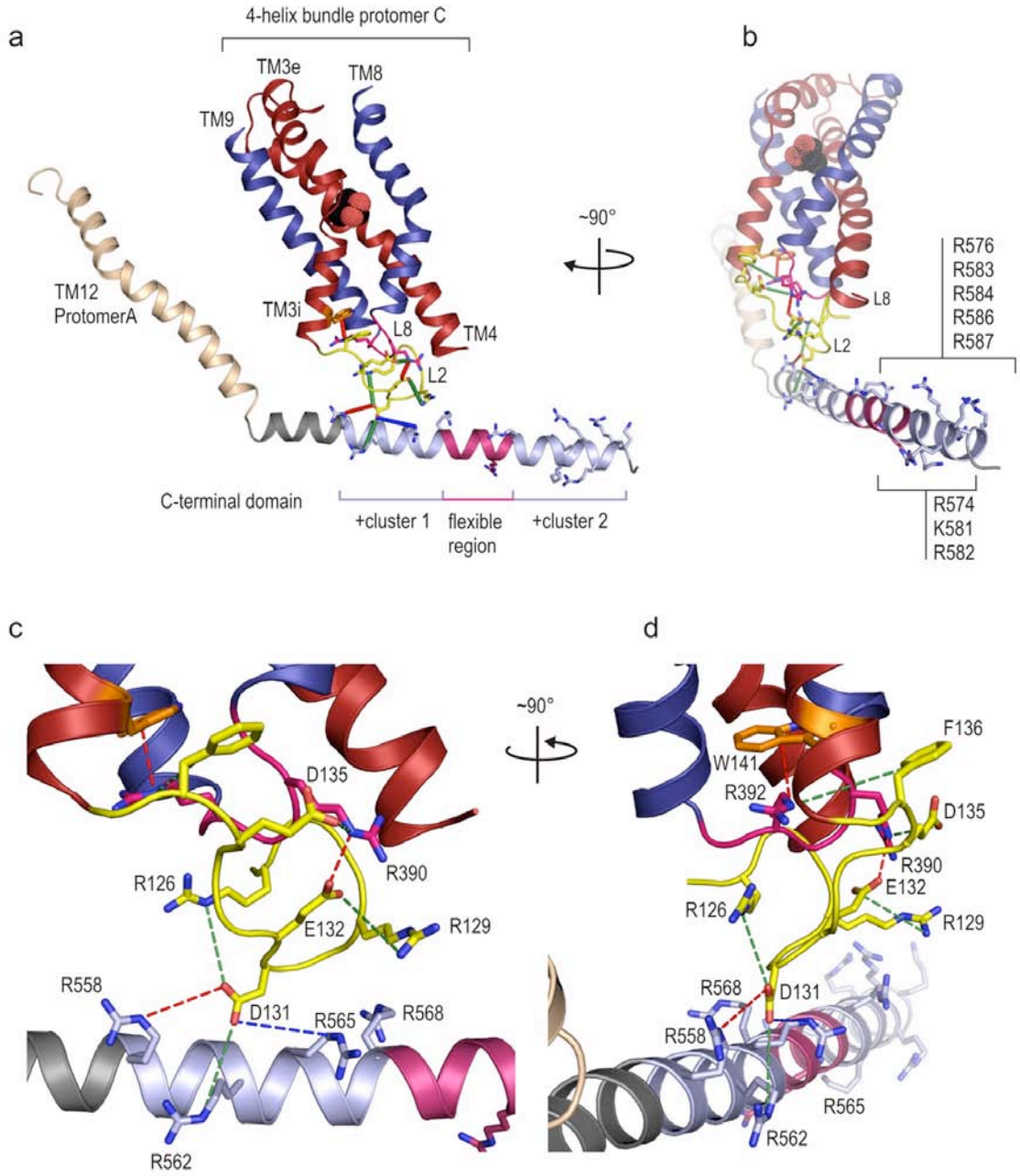


Figure 71| Region of interactions during activation of BetP. a, side view on the interaction site of the C-terminal domain of protomer A with the loop region of protomer C. TM3 and TM4 in red and TM8 and TM9 in blue of the 4 helix bundle of protomer C are the only helices shown directly involved in this interaction. **b**, view approximately 90° rotated **c**, closer up on the interaction network, view same in **a**. Possible binding partners are indicated with a dashed line; red and blue are observed interactions in the crystal structure and green indicates possible interactions up on conformational change in this region. **d**, closer view as in **c** approximately 90° rotated.

Conformational changes in loop2 could also open a pathway for K^+ ions towards the cation-π interaction between Trp141 and Arg392. Interestingly, in water the affinity of a cation-π interaction to ions has the following order: $K^+ > Rb^+ >> Na^+ > Li^+$ (Dougherty, 2007). Supposedly, a K^+ ion might be preferable to break this cation-π interaction, and therefore a first explanation for the K^+ specific activation of BetP. The break of this cation-π interaction might in turn allow TM3 and TM8 to straighten and induce other conformational changes in the surrounding helices, resulting in an opening of the cytoplasmic half of the transporter (Figure 72). Furthermore, TM3i and TM8 are both in the N-terminal part of the 4-helix bundle, for which a major conformational change is proposed during the alternating access transport mechanism (Forrest et al., 2008).

Cation-π interactions were proposed to establish K^+ selectivity in K^+ channels (Heginbotham and MacKinnon, 1992) and were later on confirmed in the K^+ channel structure (PDB ID 1JQ2, (Doyle et al., 1998). Whether the cation-π interaction between Trp141 and Arg392 can be specifically broken by a K^+ ion is speculative, but evidence for this could be the lower affinity of an NH_4^+ group to the π-orbital compared to the affinity of a K^+ ion (Ma and Dougherty, 1997).

During the activation event of BetP, the first change occurs at the region of electrostatic interactions between the inner membrane surface and the +cluster2 of the C-terminal domain, followed by changes at the region of ionic interaction sites between +cluster1 and loop2. Conformational changes of loop2 and loop8 may finally result in a break of the cation-π interaction between Trp141 and Arg392 that directly links the activation signal to the 4-helix bundle of the transporter. The electrostatic interactions at +cluster2 of the C-terminal domain involve about 13 possible residues, including five from the N-terminal domain. The ionic interaction, between +cluster1 and loop2 involves four residues and interactions between loop2 and loop8 are comprised of three to four residues. All possibly change their position and binding partner during conformational changes of the C-terminal domain, as indicated by green lines in Figure 71.

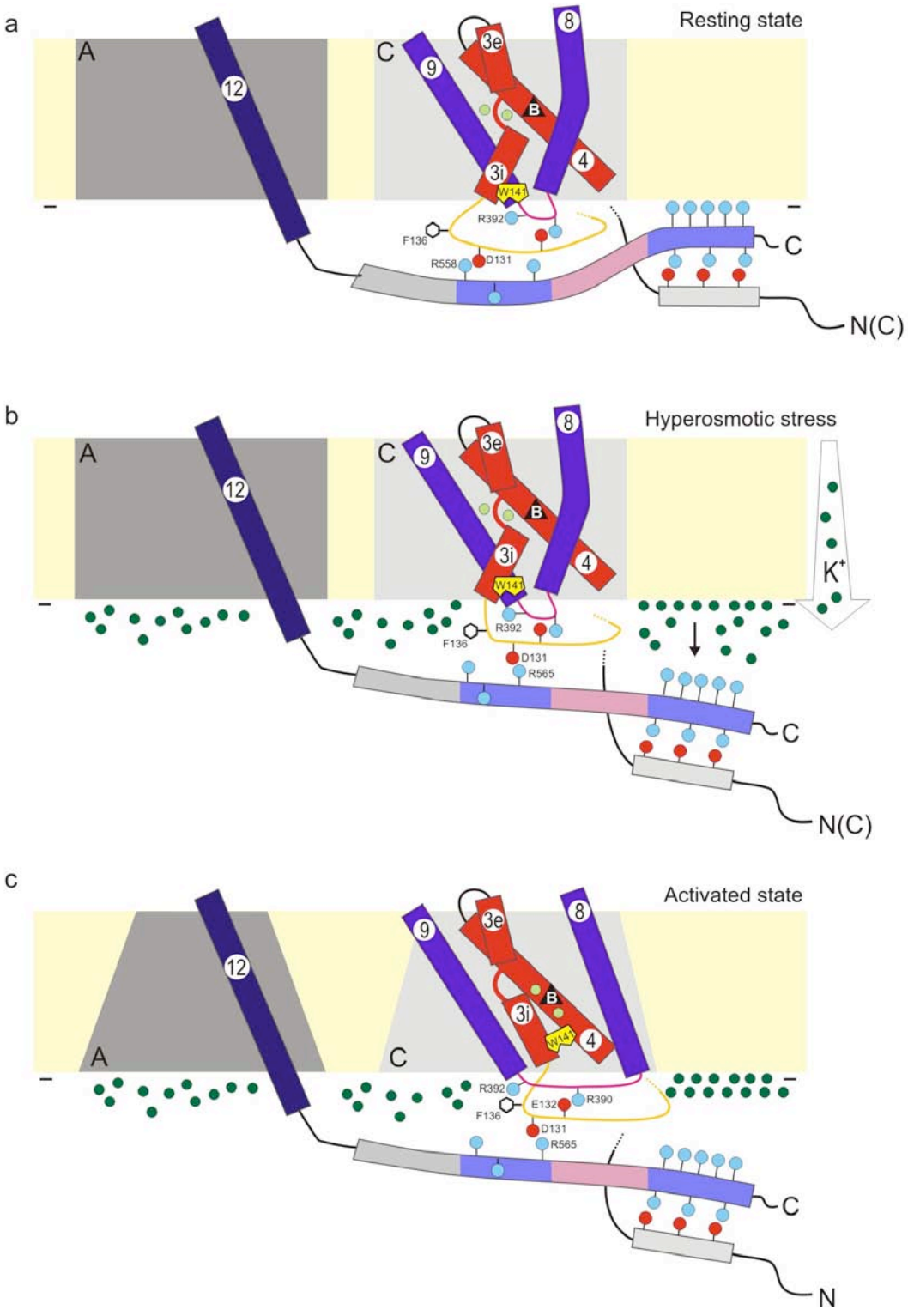


Figure 72| Atomic activation model. Negatively charged residues are shown as red circles and positively charged residues as light blue circles. K^+ ions are dark-green circles. **a,** The resting/inactive state reflects a state where substrate might be bound but no transport occurs. The cytoplasmic C-terminal domain is attached with the +cluster2 to the negatively charged lipid surface and is further stabilised by the N-terminal domain. Residue Asp131, pointing downwards from loop 2 (yellow line), contacts Arg558 in +cluster1 in the C-terminal domain (chain A). Further ionic

interactions between loop2 and loop8 (pink line) are indicated by the red and blue circles. Phe136 is shown as and open hexagon in loop2. The yellow object in TM3i indicates Trp141 building a cation-π to Arg392, light-blue circle underneath. **b,** hyperosmotic stress conditions followed by an influx of K^+ ions into the cell and result in the release of the +cluster2 from the inner membrane surface. This conformational change induces following changes in the region of ionic interaction partners between loop2 and the arginines in +cluster1. Further changes in loop2 and loop8 are transmitted towards the cation-π interaction of Trp141 to Arg392. Conformational changes of loop2 might open a putative pathway for K^+ ions towards this cation-π interaction. **c,** the brake of this cation-π interaction induces higher mobility of TM3i. Phe136 could here act as an alternative cation-π interaction partner for Arg392.

Questions and problems

The presented atomic activation model does not include large conformational motions of the C-terminal domain. The motion of +cluster2 away from the lipid surface and its interaction strength to the membrane surface may be triggered by bound N-terminal domain. The changes at +cluster2 can result in further rotational sideway motion of +cluster1 and to rearrangements of salt-bridges between +cluster1 and loop2.

Cross-linking studies of the clusters at the C-terminal domain to different cytoplasmic loops could show their specific relevance in this conformational signal transmission.

The osmosensing mechanism of the ATP-binding cassette (ABC) transporter OpuA from *Lactococcus lactis* has been elucidated in proteoliposome system (Biemans-Oldehinkel et al., 2006; Biemans-Oldehinkel and Poolman, 2003; van der Heide et al., 2001). Osmotic activation of the OpuA transporter is proposed to be mediated by variation of cytoplasmic ionic strength, which in turn affects lipid-protein interactions of its anionic C-terminal tail. In the case of ProP from *E. coli* (Racher et al., 2001), the C-terminal extension forms a homodimeric, antiparallel, α-helical coiled-coil structure. It is not essential for osmotic activation of the transporter but instead tunes the regulation over a specific osmolality range (Tsatskis et al., 2005). This would be similar to the role proposed for the anionic C-terminal tail of OpuA (Mahmood et al., 2006). ProP and OpuA respond similarly to increasing concentrations of Li^+, Na^+, K^+, $NH4^+$, Rb^+ and Cs^+ with phosphate, sulphate, or phosphate and Cl^- as anions. The signals sensed by OpuA and ProP may be similar, although the structural basis for ion sensing might be different. In contrast to OpuA and ProP, BetP displays

strong cation specificity for K^+ and smaller alkali ions such as Na^+ are much less effective (Schiller et al., 2004a).

It is questionable whether these osmosensing mechanisms elucidated for OpuA and ProP are comparable with the mechanism discussed for BetP, but the common theme is an ionic strength dependent electrostatic interaction underlying the sensing mechanism including a C-terminal domain.

8.9. Structural comparison with other transporters having the inverted-repeat motif

With determination of more and more membrane protein structures, common structural motifs have started to emerge. One of these is the inverted repeat, a tandem of five α-helical domains (five-plus-five) that have similar tertiary folds but opposite membrane orientations. Recent structures of $LeuT_{Aa}$ (Yamashita et al., 2005), vSGLT (Faham et al., 2008), Mhp1 (Weyand et al., 2008) and BetP, contain similar and unambiguous topologically inverted repeats (Table 27).

The structure of the monomeric $LeuT_{Aa}$ was solved to 1.65Å (Yamashita et al., 2005) from *Aquifex aeolicus* and is a member of the neurotransmitter sodium symporters (NSS) family, which use sodium and chloride electrochemical gradients to catalyse the thermodynamically uphill movement of a wide array of substrates. The $LeuT_{Aa}$ structure, in its outward-facing conformation, contains in total 12 TM helices of which the first ten are related by the inverted repeat motif (Figure 73a).

vSGLT (Faham et al., 2008), also a monomer, is a sodium/galactose symporter and member of the solute sodium symporters (SSS) from *Vibrio parahaemolyticus*. The ~3.0 Å structure contains 14 transmembrane (TM) helices in an inward-facing conformation with a core structure of inverted repeats relating helices TM2 to TM6 and TM7 to TM11 (Figure 73d).

Another monomer structure is the Nucleobase-Cation-Symport-1 (NCS1) transporter Mhp1, from *Microbacterium liquefaciens* (Weyand et al., 2008). The 2.85Å structure of this benzyl-hydantoin transporter contains 12 transmembrane helices, of which ten are arranged in the inverted repeat. The structure is of the outward-facing conformation (Figure 73b).

In addition to the inverted repeat motif, these transporters have in common that they are monomers, showing either the outward-facing or inward-facing substrate bound state. Otherwise they are from different families and not related by sequence similarity. BetP differs from the other transporters in at least two aspects: a) it is a native trimer and b) it shows an intermediate state in the crystal structure.

The structural superposition of repeat 1 on repeat 2 of BetP shows the smallest r.m.s.d. value, compared to the superposition of repeat 1 on repeat 2 of the other structures. This proposes BetP to be in a symmetric (intermediate) state (Table 27). The striking resemblance of the outward-facing conformation in $LeuT_{Aa}$ and Mhp1, which is also reflected by low r.m.s.d. values, suggests that the conformations observed in the different transporters are applicable to each other. Although all four structures are not related in sequence and are from different families, they share the same overall fold in the transporter core and their different appearance in conformational states, leads to a valuable comparison.

Table 27| Structure comparison with DaliLite (Holm L., 2000).

Structure comparison		R.M.S.D. (Å)
BetP repeat 1	BetP repeat 2	3.5
LeuT repeat 1	LeuT repeat 2	5.2
SGLT repeat 1	SGLT repeat 2	4.3
Mhp1 repeat 1	Mhp1 repeat 2	4.3
Scaffold		
BetP scaffold	LeuT scaffold	3.7
BetP scaffold	SGLT scaffold	3.9
BetP scaffold	Mhp1 scaffold	3.8
LeuT scaffold	SGLT scaffold	3.5
SGLT scaffold	Mhp1 scaffold	3.4
Mhp1 scaffold	LeuT scaffold	2.5
BetP TM10-TM12	LeuT TM8-TM10	3.1
BetP TM10-TM12	SGLT TM9-TM11	3.8
4-helix bundle		
BetP 4-helix bundle	LeuT 4-helix bundle	3.2
BetP 4-helix bundle	SGLT 4-helix bundle	3.1
BetP 4-helix bundle	Mhp1 4-helix bundle	3.4
LeuT 4-helix bundle	SGLT 4-helix bundle	3.5
SGLT 4-helix bundle	Mhp1 4-helix bundle	3.6
Mhp1 4-helix bundle	LeuT 4-helix bundle	3.2
Scaffold comparison used for superposition in Figure 75		
BetP TM6-7 TM11-12	LeuT TM4-5 TM9-10	2.9
BetP TM6-7 TM11-12	SGLT TM5-6 TM10-11	2.5

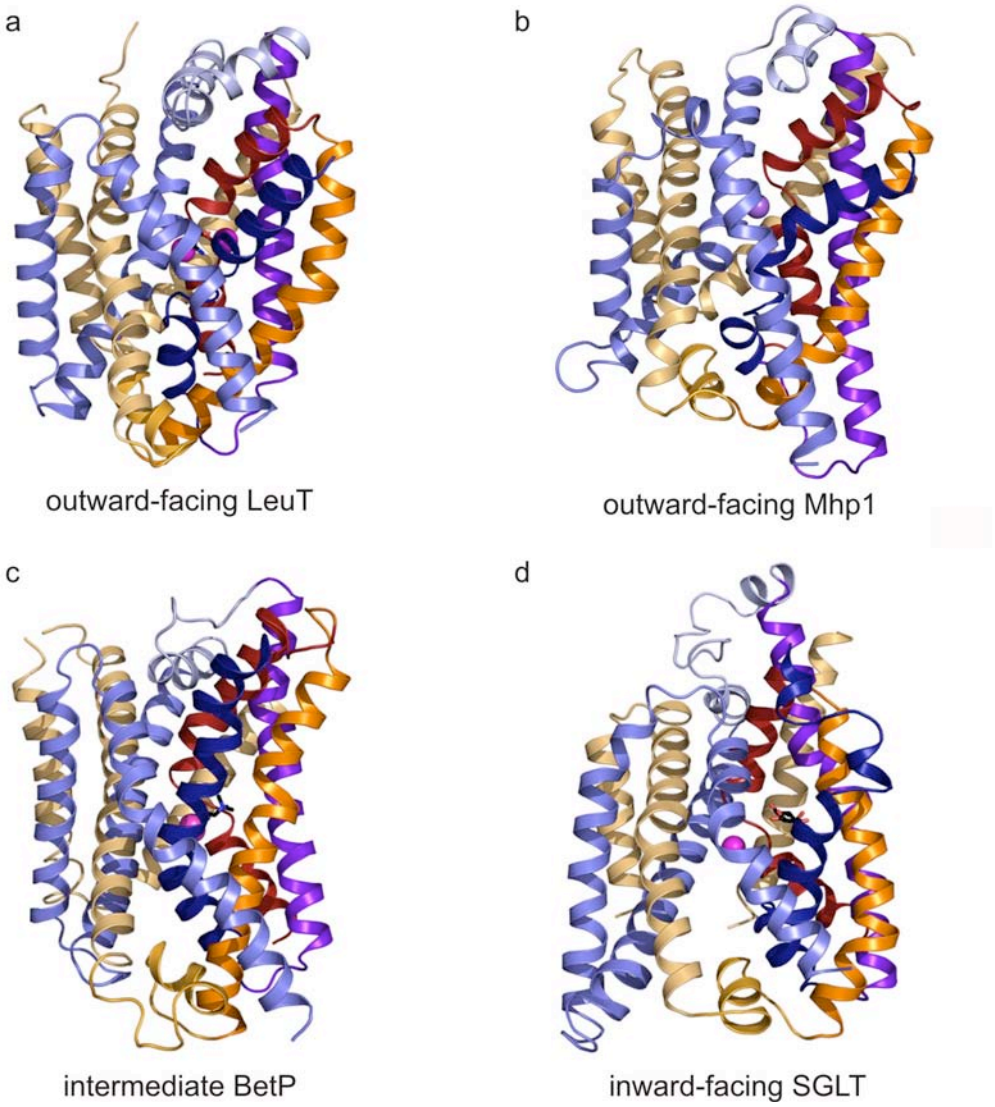


Figure 73| Four structures with the inverted repeat motif. Side view on all four structures with the periplasmic side pointing to the top and the cytoplasmic upwards. From each transporter, solely helices building the inverted repeat are shown. The first two helices of the 4-helix bundle in repeat1 are coloured red and orange and the scaffold region of repeat1 in light orange. The first two helices of the 4-helix bundle in repeat2 are coloured dark-blue and purple and the scaffold region of repeat2 in light blue. **a,** $LeuT_{Aa}$ outward facing conformation **b,** Mhp1 outward facing conformation **c,** BetP intermediate conformation **d,** vSGLT inward-facing conformation.

The different states of transporters may be observed because of their different residues that define gating from the outward- to the inward-facing conformation. The strength of the interactions between residues (salt-bridges, hydrophobic, H-bonds) at the intra- or extracellular gate dictates the energetically lower conformation of the transporters. The substrate binding site of BetP is completely obstructed towards the

periplasm and less obstructed towards the cytoplasm, which means that the energy barrier for BetP is much higher to the outward-facing conformation. Taking into account that the main purpose of the C-terminal domain is to keep the transporter inactive in the absence of hyperosmotic stress, the intermediate substrate bound state of BetP may be specifically facilitated by interactions of the C-terminal domain and thus enables to be observed in the crystal structure. An occluded intermediate state of a transporter can be described as a low energy conversion state between the outward-facing and the inward-facing state during substrate translocation. A functional relevance for an intermediate and occluded state would be a resting state of BetP, in which the transporter is loaded with substrate and ready for the inward directed substrate release.

8.10. Symmetry and asymmetry of the BetP trimer

The intermediate/resting state might be one of the three energetically favoured states during the transport cycle of BetP and the crystal contacts solely stabilise this state in all protomers, resulting in a symmetric BetP trimer. Unlike the symmetric BetP trimer of the crystal structure, the 7.5Å projection map of the wildtype and the 3D electron density map observed by electron cryo-microscopy (cryo-EM) data (Tsai, 2008; Ziegler et al., 2004) show BetP to be an asymmetric trimer. The structural results derived from cryo-EM indicate a possibility of alternating states between the protomers in the trimer, and hence suggest a coupling mechanism during substrate translocation. Coupling regarding the activation of BetP is clearly occurring between the C-terminal domain and the cytoplasmic loops of the neighbouring protomer, therefore activation is based on coupling events on the cytoplasmic side of the transporter. Conformational coupling events during transport may be mediated by interactions of the curved amphipatic helix α7 that maintains the main protomer-protomer contacts at the periplasmic membrane surface.

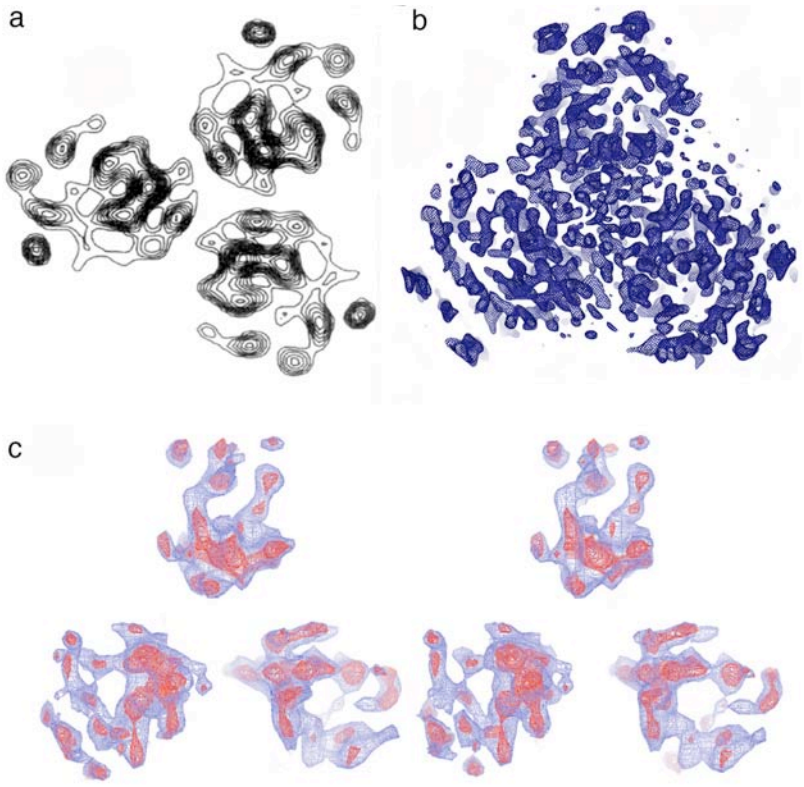


Figure 74| Electron densities for BetP from cryo-EM and X-ray data. a, averaged projection map at 7.5Å of the BetP wildtype (Ziegler et al., 2004). **b,** $2F_o$-F_c electron density map from the crystal structure of BetP, countered at 1.5σ and sliced towards the center of the trimer. **c,** stereoview of the electron density map calculated from tilted images series by cryo-EM, blue electron density map shown at 2σ with applied negative B-factor of -300Å² and red electron density map shown at 2.5σ with applied negative B-factor of -600Å² (Tsai, 2008).

8.11. Does transport in BetP include protomer coupling?

Two other open questions are whether BetP functions as a symmetric or asymmetric trimer during the transport cycle and whether coupling is necessary for transport. To date, there are only two other examples of membrane proteins known that show an asymmetric oligomer state: the multidrug efflux pumps EmrE dimer (Ubarretxena-Belandia et al., 2003) and the AcrB trimer (Murakami et al., 2002; Seeger et al., 2006; Sennhauser et al., 2007). The asymmetry of the EmrE dimer has not been attributed to two functionally distinct states because it must be dimeric to perform transport (Ubarretxena-Belandia et al., 2003). For the asymmetric AcrB trimer a coupling mechanism had been described and showed alternating access with

three different states, which was recently corroborated on the basis of a trimeric fusion construct (Takatsuka and Nikaido, 2009). AcrB, however, takes up the substrate from the hydrophobic interior of the membrane and extrudes it into the aqueous medium by peristaltic action (Seeger et al., 2006).

In BetP coupling during substrate transport may be transmitted over the amphipathic α-helix 7 together with conformational changes of the 4-helix bundle. The rotational movement of the 4-helix bundle, described below (8.12), could result in interactions between the neighbouring protomers. Furthermore, it can be suggested that the trimeric architecture of BetP supports two distinct coupling functions. Coupling events related to BetP activation and coupling associated to substrate translocation.

Tryptophan fluorescence studies of the transporter could elucidate whether cooperativeness between the protomers occurs during transport or not. These measurements use changes in the fluorescence properties of tryptophan residues during the change in environment upon substrate binding. If transport in BetP includes allosteric regulation, then this should be seen as a sigmoid curve in tryptophan fluorescence profile. A BetP monomer should consequently not show a sigmoid curve. This might be one possible experiment to answer this fundamental question.

8.12. Alternating access mechanism of Na^+-coupled transport

The mechanism of secondary transporters suggests that they alternate between at least two conformational states. In one state the transporters substrate-binding side is accessible to the extracellular environment (outward-facing conformation) and in the other the binding side is accessible to the cytoplasm (inward-facing conformation) (Mitchell, 1957). The alternating access model (Jardetzky, 1966) describes coupled transport by which the joint steps of substrate and ion binding are linked from the outward facing conformation followed by a transition/isomerisation of the transporter to the inward-facing conformation, allowing the release of substrate and ion.

The first crystal structure of a secondary transporter was the proton-driven multidrug efflux pump AcrB of the resistance nodulation cell division (RND) family

from *Escherichia coli* (Murakami et al., 2002). This was closely followed by two crystal structures from the major facilitator superfamily (MFS) of transporters, the glycerol-3-phosphate/phosphate antiporter GlpT (Huang et al., 2003) and the proton-coupled lactose symporter LacY (Abramson et al., 2003). Each of these structures revealed an internal two-fold structural pseudo symmetry that relates the N-terminal half of the transporter to the C-terminal half by an axis running through the center of the transporter. The crystal structures of GlpT and LacY show an outward-facing conformation that, suggested a switch-like movement, would result in the inward-facing conformation upon substrate translocation. Further insights into the mechanism derived from the structures of the aspartate transporter Glt_{Ph} of the dicarboxylate/amino-acid:cation symporter (DAACS) family (Yernool Dinesh, 2004), followed by the Na^+/H^+ antiporter NhaA from *E. coli* (Hunte et al., 2005) and the bacterial leucine transporter $LeuT_{Aa}$ and led to a proposed rocker-switch motion of the two symmetry-related halves (Forrest et al., 2008). The structures of Glt_{Ph}, NhaA and $LeuT_{Aa}$ not only revealed unique membrane protein folds but also supported the structural theme of internal two-fold symmetry and discontinuous transmembrane helices (Screpanti and Hunte, 2007).

The Glt_{Ph} homotrimer structure showed the relevance of the internal approximate two-fold symmetry to the transporter mechanism and suggested that the N-terminal and C-terminal half domains undergo an alternating, symmetry-related motion that opens and closes access to the substrate- and ion-binding sites (Yernool Dinesh, 2004). With the $LeuT_{Aa}$ structure, a different internal intertwined two-fold symmetry was observed. Ten transmembrane helices run parallel to the membrane plane through the center of the transporter and relate the first N-terminal five transmembrane helices to the second C-terminal five helices. Surprisingly, the inverted repeat seen in $LeuT_{Aa}$ (Yamashita et al., 2005) was also observed in the subsequently reported structure vSGLT (Faham et al., 2008), Mhp1 (Weyand et al., 2008) and BetP. The structures share the five-plus-five inverted repeat, which defines the architecture of their working transport machinery. Even though these four transporters do not share significant amino acid sequence identity or have the same number of transmembrane segments, their common inverted repeat motif further implies similarities in their transport mechanism.

Although the available transporter structures support the alternating access model of membrane transport (Forrest et al., 2008), so far none of them show the same transporter in more than one conformation. The central question of how the different conformations interconvert to bring about membrane transport has thus remained unanswered.

Structural differences between their repeats contribute to the asymmetric architecture of the outward- and inward-facing conformations of $LeuT_{Aa}$/Mhp1 and vSGLT, respectively (Figure 73). Modelling studies could only exploit the asymmetry between the outward- and inward-facing conformations and describe their transition as a switching conformational change within the two repeats (Forrest et al., 2008; Lolkema and Slotboom, 2008). However, the occluded intermediate state of BetP exhibits a symmetrical architecture (Figure 73, Table 27), which allows a more detailed description of the conformational changes by discussing the outward-facing, occluded-intermediate and inward-facing conformations observed in the different transporters in terms of a single, unified transport mechanism.

Structural alignments show a number of structurally related conserved glycines in BetP, $LeuT_{Aa}$, vSGLT and Mhp-1 in both repeats in their respective families. Recent hydropathy profile alignments (Smicun et al., 1999; Stephan et al., 1997) reveal a characteristic clustering of hydrophobic residues in TM helices in transporters of the unrelated NSS, SSS and the amino acid-polyamine-organocation (APC) family confirming their similar transporter core structure.

Upon changing from the outward-facing conformation ($LeuT_{Aa}$) over an intermediate state (BetP) to the inward-facing (vSGLT) conformation (Figure 75), a concerted, iris-like movement occurs in the 4-helix bundle. The periplasmic halves of TM3, TM4 and TM9 (BetP numbering) undergo a counter-clockwise rotation occluding the substrate-binding site on the periplasmic side (Figure 75). The concomitant movement of TM8 might be influenced by the different sizes and nature of the substrates in the three transporters. The rotation of the cytoplasmic half of the 4-helix bundle moves it out of the center of the transporter. In the scaffold surrounding the 4-helix bundle, TM10 straightens, while TM5 and TM12 approach TM3 and TM8 on the periplasmic side.

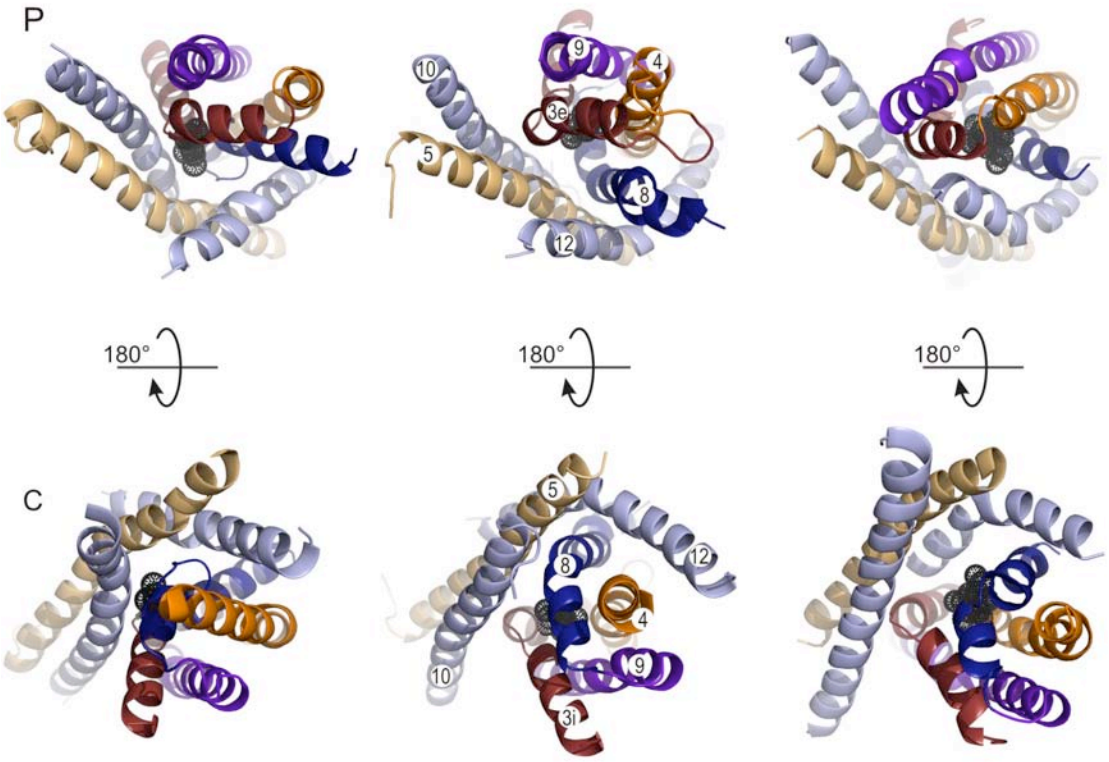


Figure 75| Comparison of bundle and scaffold regions of LeuT$_{Aa}$, BetP and vSGLT. From left to right, LeuT$_{Aa}$ outward-facing, BetP intermediate and vSGLT inward-facing conformation. Top view of the 4-helix bundles and a part of the scaffold (TM5, TM10 and TM12) from the periplasm (P) and cytoplasm (C). Grey spheres indicate the bound substrate. When observing the four helix bundle from left to right, a concerted counter clockwise rotational flip can be observed on the periplasmic side and clockwise rotational flip on the cytoplasmic side. Numbering of the helices is just shown for BetP.

On the cytoplasmic side, they move away from the 4-helix bundle, opening an access to the substrate-binding site between TM8 and TM12. The conformational changes associated with transport are due to concerted movements of the 4-helix bundle and the surrounding scaffold. The pivot points of these movements are located near the substrate and ion binding sites, close to the structurally conserved glycines as in the G-x-G-x-G motif in the stretch of TM3 and Gly301 in loop 6 (Figure 64). Notably, mutation of Gly301 renders BetP inactive (Figure 60). Gly301 is close to Ile302 and Ser306, both involved in the occlusion of the substrate from the cytoplasm. Therefore, this glycine might be important for the flexibility of loop 6 and thus for opening up this pathway to the cytoplasm.

Mutagenesis studies on the LeuT$_{Aa}$ homologue SERT reveal conformational changes during transport in the loop 9 connecting the 4-helix bundle with the C-

terminal part of the scaffold. A comparison of the three structures indicates major shift in the N-terminal part of this loop, moving out of the periplasmic vestibule in the transition to the inward-facing conformation (Figure 73).

In the intermediate state of the BetP structure, helices of the inverted repeat adopt intermediate positions between the outward and inward-facing conformations of $LeuT_{Aa}$ and vSGLT. Therefore, conformational changes during inward transport do not seem to be a mere switching motion between the two open states, as predicted by the alternating-access model, but to involve an intermediate state resulting from the concerted rotational movement of TM helices around the coupling ion binding sites.

8.13. Different mechanistic models for the alternating access mechanism

Over the last decade X-ray crystallography studies on transporters of different systems and families revealed that many integral membrane proteins consist of two domains of similar fold with five transmembrane helices each. They fold in an either parallel or opposite (antiparallel) orientation in the membrane. Furthermore, antiparallel five-plus-five domains can be arranged as one intertwined unit, which is also known as the inverted repeat fold. It can be hypothesised is that they evolved by gene duplication and gene fusion events from a dual topology ancestral membrane protein that adapted both orientations in the membrane and formed antiparallel homodimers.

A recent study of the DUF606 family of bacterial membrane proteins revealed the entire collection of intermediate states of such an evolutionary pathway: single genes that would code for dual topology homodimeric proteins, paired genes coding for homologous proteins with a fixed but antiparallel orientation in the membrane that would form heterodimers, and fused genes that encode antiparallel two-domain fusion proteins (Lolkema et al., 2008).

There might have been a selective advantage for fusion antiparallel two-domain membrane proteins observed in the above described crystal structures, compared to separate hetero- or homodimer membrane protein systems (Lolkema et al., 2008). The reason could be the very compact structure of this inverted repeat

motif, which in turn might facilitate a very effective transport movement. Furthermore, the latest atomic models of these functionally non-related but structurally related transporter families have provided insights into the principals of sodium-coupled transport, and are beginning to clarify an alternating-access mechanism that is distinctly different from that of the MFS family, possibly because of the different arrangement of the five-plus-five TM domain.

The two five TM helices domains of LacY, NhaA and EmrE are located side-by-side, either parallel or antiparallel, and form a homodimer transport unit, allowing a switch like motion during transport. The five TM helices of the inverted repeat of $LeuT_{Aa}$, vSGLT, Mhp1 and BetP are intertwined and form a closely packed five-plus-five TM transporter core, which allows the rotational component during the translocation process. Therefore LacY-fold like proteins might reflect an ancestral switch motion of the alternating access model and the LeuT-fold like transporter represent a further evolved iris like rotational motion for this type of transport.

The selective advantage for fusing the antiparallel and intertwined five-plus-five TM membrane domain could be: a) a smaller overall motion during transport enabled by the rotational component, b) its compact structure reduces the space for a transport unit needed in the membrane, c) the compact arrangement of the five-plus-five TM domain is a good basis to change substrate selectivity by only changing specific residues within a transporter family.

The necessity to facilitate fast translocation of high amounts of substrate, either involved in a rescue process (BetP) or in fast neuronal signal transition ($LeuT_{Aa}$), was very likely the evolutionary pressure to derive to this compact inverted repeat fold transport unit.

9. Conclusions and future perspectives

The Na^+-coupled glycine-betaine transporter BetP from *C. glutamicum* is the best-studied transporter in the field of osmotic stress response. The BetP crystal structure has greatly assisted the interpretation of the transporter-specific biochemical data, and has provided new insights into the global topic of sodium-coupled transport mechanism.

Limited resolution together with high data anisotropy made structure determination of BetP a challenging task. Initial phases were obtained using SAD, a still very exceptional practice in membrane protein crystallography. A specific combination of programs used subsequently and iteratively was needed to achieve an optimal correction for the high data anisotropy and optimisation of the low phase information.

The first atomic model from the BCC-transporter family is here presented. It reveals an substrate-bound, intermediate state of the transporter and demonstrated that the C-terminal domain of each BetP protomer is interacting with an adjacent neighbour.

A careful comparison between the inverted repeats of four independent structures of sodium-coupled symporters ($LeuT_{Aa}$, vSGLT, Mhp1 and BetP) extended the reach of the BetP structure into the alternating access mechanism field. In addition to the rocker-switch motion proposed until now, the structure of BetP in an intermediate state introduced an iris-like rotation component in the transport mechanism.

A functional resting state of BetP is proposed, based on the occluded structure determined in this work. It is suggested that BetP is loaded with substrate and kept inactive by the C-terminal domains until the activation signal is sensed. A cell membrane containing BetP is therefore not just pre-loaded with glycine-betaine but also enabled to respond very quickly and efficiently upon hyperosmotic stress.

An atomic activation model is described, suggesting changes in electrostatic interactions between the C-terminal domain and the inner membrane surface. A conformational change induced at the C-terminal domain results in a signal that is transmitted to a region of ionic interactions at cytoplasmic side of the 4-helix bundle of the transporter core, connecting activation with the transport process.

Two putative sodium sites were suggested and can now be addressed by new site-directed mutagenesis studies. It would be interesting to verify whether a mutation may drive BetP from its 2:1 to a 1:1 sodium to substrate stoichiometry.

With the structure in hand it was recently possible to create a mutant that does not form stable trimers. This provides a unique opportunity to measure the transport properties of a monomeric BetP system. Pilot experiments have already shown that the mutant is deregulated, similarly to C-terminal deletion mutants (Camilo Perez, personal communication). Furthermore, monomeric BetP may allow different crystal packing and possibly crystals diffracting to higher resolution. Another mutation in the G-x-G-x-G motif of BetP resulted in altered substrate specificity (Camilo Perez, personal communication). These new findings show the importance of atomic models for the design of site-directed mutational studies.

10. Appendix

10.1. PHASER_MR for data anisotropy analysis and correction

```
PHASER ‹‹ eof
MODE ANO
TITLe
HKLIn input.mtz
LABIn F=Fobs SIGF=Sigma
ROOT
eof
```

10.2. Using SFCHECK for data anisotropy analysis and correction

```
# sfcheck for data anisotropy correction and analysis
sfcheck -f file_sf_mtz
out a

# the out a command writes the corrected mtz file into a # cif file
which has to be converted again to a mtz file # with the program
CIF2MTZ
```

10.3. Using CAD for data anisotropy correction

```
# CAD for anisotropy correction

cad         \
  HKLIN1 name_input.mtz \
  HKLOUT name_output.mtz \
  << END-cad
TITL data anisotropy correction
LABI FILE 1 E1=F E2=SIGF
CTYP FILE 1 E1=F  E2=Q
SCALE file 1 1.0 B11 B22 B33 B21 B13 B23
# usually B21, B13 and B23 are set to 0
END
END-cad
#
```

10.4. Using CAD for negative B-factor sharpening

```
# CAD for negative B-factor sharpening

cad         \
  HKLIN1 name_input.mtz \
  HKLOUT name_output.mtz \
  << END-cad
TITL –B factor sharp
LABI FILE 1 E1=F E2=SIGF
CTYP FILE 1 E1=F  E2=Q
SCALE file 1 1.0 -100
# usually negative B values between -80 and -160 are good
END
END-cad
#
```

10.5. Using SCALEIT to scale F_{obs} to F_{calc}

```
#!/bin/sh (script Clemens Vonrhein)

f=FOSC
s=SIGFOSC
f=FPsha
s=SIGFPsha
f=FC
s=SIGFPsha

cad hklin1 data.mtz \
    hklin2 refined-data.mtz \
    hklout cad.mtz \
    <<EOF > /dev/null 2>&1
LABI FILE 1 ALL
LABI FILE 2 E1=$f E2=$s
EOF

scaleit hklin cad.mtz hklout scaleit.mtz <<EOF | awk '/ beta
matrix/,/Latent roots and vectors of/'
LABI FP=$f SIGFP=$s -
     FPH1=F SIGFPH1=SIGF DPH1=DANO SIGDPH1=SIGDANO
REFI ANISO
CONV NCYC 20 ABS 0.0001
EOF

#cv
exit
#cv

E1=FOSC E2=SIGFOS
E1=FPsha E2=SIGFPsha
E1=FC E2=SIGFPsha
```

10.6. BetA sequence

The theoretical sequence for BetA is given below. Green marks the StrepII tag and in red the exchanged glutamate cluster E44, E45 and E46 to A44, A45 and A46.

WSHPQFEKLENPTNLEGKLADAAAAIILEGEDTQASLNWSVIVPALVIVLATVVW
GIGFKDSFTNFASSALSAVVDNLGWAFILFGTVFVFFIVVIAASKFGTIRLGRID
EAPEFRTVSWISMMFAAGMGIGLMFYGTTEPLTFYRNGVPGHDEHNVGVAMSTTM
FHWTLHPWAIYAIVGLAIAYSTFRVGRKQLLSSAFVPLIGEKGAEGWLGKLIDIL
AIIATVFGTACSLGLGALQIGAGLSAANIIEDPSDWTIVGIVSVLTLAFIFSAIS
GVGKGIQYLSNANMVLAALLAIFVFVVGPTVSILNLLPGSIGNYLSNFFQMAGRT
AMSADGTAGEWLGSWTIFYWAWWISWSPFVGMFLARISRGRSIREFILGVLLVPA
GVSTVWFSIFGGTAIVFEQNGESIWGDGAAEEQLFGLLHALPGGQIMGIIAMILL
GTFFITSADSASTVMGTMSQHGQLEANKWVTAAWGVATAAIGLTLLLSGGDNALS
NLQNVTIVAATPFLFVVIGLMFALVKDLSNDVIYLEYREQQRFNARLARERRVHN
EHRKRELAAKRRRERKASGAGKRR

The working sequence contained a sequence error marked in orange:

- original sequence:

LENPTNLEGKLADA

- corresponding sequence with error:

IEGLETPTNLEGKLADA

10.7. TLS group definition in refinement with *phenix.refine*

```
tls = "chain A and (resid 57:120 or resid 551:589)"
tls = "chain A and (resid 139:210 or resid 430:548)"
tls = "chain A and (resid 214:325 or resid 360:426)"
tls = "chain B and (resid 55:120 or resid 551:585)"
tls = "chain B and (resid 139:210 or resid 430:548)"
tls = "chain B and (resid 214:325 or resid 360:426)"
tls = "chain C and (resid 57:120 or resid 551:585)"
tls = "chain C and (resid 139:210 or resid 430:548)"
tls = "chain C and (resid 214:325 or resid 360:426)"
```

10.8. NCS group definition in refinement with *phenix.refine*

```
reference = "chain A and (resseq 57:271 or resseq 278:545)"
selection = "chain B and (resseq 57:271 or resseq 278:545)"
selection = "chain C and (resseq 57:271 or resseq 278:545)"
```

10.9. Amino acid code and characteristics

Single Letter Code	Short Name	Name
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Characteristics
nonpolar (hydrophobic)
polar
basic
acidic

10.10. Structure analysis of arsenobetaine derivative BetP

Arsenobetaine [$(CH_3)_3As^+CH_2COO^-$] is the As analogue of glycine-betaine. Arsenobetaine (As-B) occurs naturally at high levels in most marine fish and other marine animals.

It would be possible to collect an anomalous data set for As-derivative crystals at the As K edge = 1.0448 Å equivalent to 11.8667 keV. If As-B binds to BetP then it should be possible to detect it specifically in the binding pocket.

First it should be tested if As-B has an impact on BetP stability. If it has no impact on its stability, then it should be possible to pre-incubate the BetP sample after SEC with As-B and co-crystallise it with additional As-B in the crystallisation buffer. This might be a straightforward procedure having the reproducible 3D protocols in hand. As-B might function as a basis for synthesis of specific inhibitors for BetP, which might lead to a more specific view on its transport mechanism.

Acknowledgements

Prof. Werner Kühlbrandt, danke ich sehr für die exzellente wissenschaftliche Betreuung und für manch moralische Unterstützung. Ich bin sehr dankbar und wertschätze es sehr in seiner Abteilung meine Doktorarbeit angefertigt haben zu dürfen. Es war eine grossartige Zeit.

Prof. Clemens Glaubitz, danke ich für die nette und wertvolle Unterstützung in letzter Minute und für die Übernahme des Gutachtens und der Doktorvaterschaft.

Dr. Christine Ziegler, danke ich sehr für die Überlassung dieses spannenden Projektes und ihre wissenschaftliche und motivierende Betreuung. Ich danke dir für dein Vertrauen in mich und meine Arbeit und die ein oder andere MIXA.

Dr. Anke Terwisscha van Scheltinga, turned out to be my second supervisor during my PhD and I owe her my greatest gratitude. Thank you for your patience in guiding me through the jungle of crystallographic programs and for always bringing me back on the track when I got lost. Thank you for being a supervisor, teacher and a friend. Actually, I do not remember who won the tomato growth competition in 2008?

Manchmal kommt der Berg (ich) eben zum Propheten, **Dr. Clemens Vonrhein**. Ich danke dir für die grossartige Wissensvermittlung in der Röntgenstrukturanalyse, ob im Kurs 2006 oder während unserer Zusammenarbeit. Danke für die Einblicke in deine "Philosophie" der Strukturaufklärung.

Prof. Reinhard Krämer danke ich sehr für die Einsicht in die Masse der biochemischen Daten und deren Interpretationen und habe grosse Ehrfurcht vor seiner Arbeit. Die biochemischen Ergebnisse seiner Gruppe waren eine faszinierende nicht wegzudenkende Bereicherung dieser Arbeit

Vera Ott & Sascha Nicklisch, bin ich unendlich dankbar für diese fruchtbare Kollaboration, den hervorragenden wissenschaftlichen Austausch und die unkomplizierte Zusammenarbeit. Ich glaube, wenn mich jemals jemand danach fragen sollte, wie eine perfekte Kollaboration in der Wissenschaft aussieht, dann nenne ich die unsere. Ich bin glücklich euch als Kollegen und Freunde zu kennen.

I thank **Prof. Werner Mäntele** and **Dr. Filiz Korkmaz** for introducing me in their FTIR world and the inspiring discussions on various BetP signals.

Although they might never read this acknowledgment, I would like to express my gratitude to all the **X-ray program developers** and to the great **X-ray community**, especially to people involved in CCP4, phenix and CNS.

Ich möchte mich sehr für die Ermöglichung der exzellenten wissenschaftlichen Arbeitsbedingungen bei der **Max-Planck Gesellschaft** bedanken. Des Weiteren bedanke ich mich für die Unterstützung dieser Arbeit bei der International Max

Planck Reseach School (**IMPReS**) und beim **SFB807** "Transport and Communication across Biological Membranes".

Remco Wouts, thank you very much for your continuous help related to various computational problems and questions, and for your persistence in forcing me to understand what this is all about. I learned a lot from you and due to your effort I don't call myself a "klicki-bunti" Windows generation anymore…☺.

Dr. Özkan Yildiz, danke ich für all die unkomplizierte und schnelle Hilfe in Computer- und Programmfragen. Danke für deinen grossartigen Einsatz die X-ray Programme "up-to-date" zu halten!

Prof. Claudia Büchel hat bei mir die Faszination an der Strukturbiologie geweckt und ich bedanke mich herzlich für ihr "mentoring" ☺.

I am glad that I got involved with **Dr. Lucy Forrest** and **Dr. Kamil Khafizov** during my PhD and that I had the chance to learn from their molecular dynamics studies. Thank you very much for the valuable discussions and the virtual world of molecular motions.

Helga Husmann bin ich zu tiefstem Dank verpflichtet. Sie hat nicht nur in vielen Abbildungen ihre "Zauber"-Hände mit im Spiel gehabt, sie hat mir auch in einigen stressigen Zeiten den Rücken freigehalten. Vielen Dank für die netten Gespräche rund um die, von uns beiden so geliebte Ornithologie, Lakritz und vieles andere.

Paolo Lastrico danke ich sehr für sein professionelles Angagement für das Titelbild. Es war sicherlich unser persönlicher Favorit und hat es bei Nature zumindest in das Inhaltsverzeichnis geschafft ☺. Danke für die hilfreichen Photoshop Tips!

Mike Strauss and **Rebecca Lam**, thank you very much for correcting some of my "denglish"!

I thank all the members of the **Structural Biology department** for the fantastic working atmosphere!
I would like to express my gratitude to **Deryck Mills** and **Janet Vonck** for their energy and angagement in keeping the social life in the Structural Biology department so valuable. If I could, I would give you a gold medal for you efforts! **Monika Hobrack** danke ich sehr für ihre vielen großartigen Hilfestellungen bei so manch bürokratischen Dingen des Doktorantenlebens. **Tea Pavkov** danke ich für schöne italienische Klettereien und nette Grazer Gespräche über Wissenschaft und darüber hinaus.

Ein grosser Dank geht an die Leute im **Ziegler Labor 2.090**, soweit ich sie noch kenne…: **Eva Schweikard**, **Andrea Neeb** danke ich für ihre Zeit einige schöne Kristallisationstropfen pipettiert zu haben. **Ulrike Geldmacher** danke ich für die unkomplizierte Probenweiterbearbeitung mit TLC und ihre Hilfestellungen im und 'ums Labor herum. **Katrin Rohde** danke ich, dass sie mich immer mal wieder ins Labor gebracht hat und mir damit zum Computeralltag eine schöne Abwechslung war! **Camilo Perez**, thank you for being so greatly involved and thank you very

much for the scientific and non-scientific discussions. **Caroline Koshi**, I wish you all the best on the BetP project and a lot of exiting results…as you already have ☺ I am more than happy for you!

Thanks to "The OFFICE"! Ich danke **Sabrina Schulze**, für etweige gemeinsame Lauf- und Rennrad-Kilometer und für alle wissenschaftlichen und nicht wissenschaftlichen Unterhaltungen. **Jonna Hakulinen** danke ich für die finnischen Eindrücke: "Finnische Lakritzschokolade ist der Himmel auf Erden!". Danke, dass du meine Fischies übernommen hast. **Sonja Kuhlmann** danke ich für die Arbeit "Rücken an Rücken" und den supernetten Austausch in fachlichen und nicht-fachlichen Dingen.
Danke euch dreien für die freundschaftliche und respektvolle Zeit im Büro.

Stephen Marino is so far one of the most thoughtful and precise scientists I have met. I am very grateful for all his constructive criticism and advice, and enjoyed the great times together outside the lab.

Ich möchte mich bei einem ganz besonderen Freund bedanken, **Ralph S.**. Danke, dass du mir meine Träume vor Augen führst und mich so weise in Geduld schulst. Danke für die Doktorandinnen Reha-Aufenthalte auf dem Lande.

Wahre Freundinnen, die braucht man um im Leben bestehen zu können. Ich bin von tiefsten Herzen, **Anita C.**, **Bianka B.**, **Sylvia K.**, **Rike H.**, **Simone V.** und **Lotte W.** zu Dank verpflichtet. Danke, dass ich wundervolle, wichtige und einzigartige Momente in meinem Leben mit euch erleben und teilen durfte. Danke für eure herzliche und unkomplizierte Freundschaft. "Die Zukunft bleibt dynamisch" ☺.

A cordial "Muito obrigada" to **Tiago B.**. I am grateful for sharing life&science with you. I thank you for all your great support and for being there for me.

Děkuji srdečně mojemu milymu **Dědečeku** pro všechny podpory a pro nádherně časy v Praze.

Honzo, jsi můj milovaný bratr, lhostejný co bylo…

Ich schätze mich jeden Tag glücklich solche **Eltern** zu haben. Ihr seit die Helden meines Lebens und danke euch herzlich für eure Liebe, Fürsorge, die Ermöglichung meines Studiums und meine phantastische Kindheit! *Já Vás moc miluju!*

CURRICULUM VITAE

Susanne Ressl
*29.04.1979
Vímperk (Winterberg), Czech Republic
German

EDUCATION AND TRAINING

1990-1999: Edertalschulgymnasium Frankenberg/Eder
Abitur (University entrance diploma)

04/2001 – 05/2006:
Academic studies in the faculty of Biology at the Johannes Gutenberg University Mainz. Zoological Physiology, Biophysics and Immunology as core areas.

11/2003 – 01/2004: Training at the Structural Biology department of the Max-Planck Institute of Biophysics Frankfurt. Subject: Preparation and characterisation of the membrane intrinsic protein fucoxanthin-chlorophyll protein for two-dimensional crystallisation. Supervisor: Prof. Claudia Büchel.

01/2005 – 02/2005: Training at Charité Berlin, Institute of Medical Physics and Biophysics, Prof. K.P. Hofman. Subject: Protein isolation from bovine retina and 3D crystallisation of macromolecules. Supervisor: Prof. Hui-Woog Choe

06/2005 – 02/2006: Diploma work at Max-Planck Institute of Biophysics Frankfurt in the Structural Biology department of Prof. W. Kühlbrandt. Subject: Röntgenstrukturelle und elektronenmikroskopische Untersuchungen zweier Mutanten des Glycinbetain Transportproteins BetP aus *Corynebacterium glutamicum*. Supervisor: Dr. C. Ziegler

05/2006: Diploma in Biology, with Zoological Physiology, Biophysics and Immunology as core areas, Faculty of Biology, Johannes Gutenberg University of Mainz.

DOCTORATE

05/2006 – 07/2009: Johan Wolfgang Goethe University Frankfurt am Main;
Max-Planck-Insitute of Biophysics, Department of Structural Biology, Prof. W. Kühlbrandt
Subject: X-ray structure determination of the Na^+-coupled glycine-betaine symporter BetP from *Corynebacterium glutamicum.*
Supervisor: Dr. C. Ziegler and Dr. A. Terwisscha van Scheltinga

FUNDING

This work was supported by the International Max Planck Reseach School (IMPReS) and funded by the SFB807 "Transport and Communication across Biological Membranes"

PUBLICATIONS

Ressl, S., Terwisscha van Scheltinga, A.C., Vonrhein, C., Ott, V. & Ziegler, C.: Molecular basis of transport and regulation in the Na+/betaine symporter BetP. (2009) Nature, 458, 47-52

Ressl, S., Terwisscha van Scheltinga, A.C., Vonrhein, C.: Membrane protein crystals with strong anisotropic diffraction: how to achieve interpretable maps using weak experimental phasing. (in preparation).

11. Bibliography

Abe, H., Akasaki, K., Seino, A. and Shirato, S. (1967) Production of primocarcin by Streptomyces IN-701. *J Antibiot (Tokyo)*, **20**, 167-173.

Abrahams, J.P. and Leslie, A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F_1 ATPase. *Acta Crystallogr.*, **D52**, 30-42.

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R. and Iwata, S. (2003) Structure and mechanism of the lactose permease of Escherichia coli. *Science*, **301**, 610-615.

Adams, P.D., Pannu, N.S., Read, R.J. and Brunger, A.T. (1997) Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc Natl Acad Sci U S A*, **94**, 5018-5023.

Adams, P.D., Pannu, N.S., Read, R.J. and Brunger, A.T. (1999) Extending the limits of molecular replacement through combined simulated annealing and maximum-likelihood refinement. *Acta Crystallogr D Biol Crystallogr*, **55**, 181-190.

Afonine, P.V., Grosse-Kunstleve, R.W. and Adams, P.D. (2005) phenix.refine. *CCP4 Newsl.* , **42** contribution 8.

Apell, H.J. (2004) How do P-type ATPases transport ions? *Bioelectrochemistry*, **63**, 149-156.

Arakawa, T. and Timasheff, S.N. (1985) The stabilization of proteins by osmolytes. *Biophys J*, **47**, 411-414.

Arndt, U.W., Champness, J.N., Phizackerley, R.P. and Wonacott, A.J. (1973) A ssingle-crstal oscillation camera for lage unit cells. *J. Appl. Cryst.*, **6**.

Banumathi, S., Dauter, M. and Dauter, Z. (2003) Phasing at high resolution using Ta6Br12 cluster. *Acta Crystallogr D Biol Crystallogr*, **59**, 492-498.

Bass, R.B., Strop, P., Barclay, M. and Rees, D.C. (2002) Crystal Structure of Escherichia coli MscS, a Voltage-Modulated and Mechanosensitive Channel *Science*, **298**, 1582-1587.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235-242.

Biemans-Oldehinkel, E., Mahmood, N.A. and Poolman, B. (2006) A sensor for intracellular ionic strength. *Proc Natl Acad Sci U S A*, **103**, 10624-10629.

Biemans-Oldehinkel, E. and Poolman, B. (2003) On the role of the two extracytoplasmic substrate-binding domains in the ABC transporter OpuA. *Embo J*, **22**, 5983-5993.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M. and Bricogne, G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr*, **60**, 2210-2221.

Blundell, T.L. and Johnson, L.N. (1976) Protein Crystallography. *Academic Press, New York*.

Bolen, D.W. and Rose, G.D. (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.*, **77**, 339-362.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**, 248-254.

Brunger, A.T. (1993) Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr D Biol Crystallogr*, **49**, 24-36.

Brunger, A.T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc*, **2**, 2728-2733.

Brunger, A.T. and Adams, P.D. (2002) Molecular dynamics applied to X-ray structure refinement. *Acc Chem Res*, **35**, 404-412.

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

Brunger, A.T., DeLaBarre, B., Davies, J.M. and Weis, W.I. (2009a) X-ray structure determination at low resolution. *Acta Cryst. D*, **65**, 128-133.

Brunger, A.T., DeLaBarre, B., Davies, J.M. and Weis, W.I. (2009b) X-ray structure determination at low resolution. *Acta Crystallogr D Biol Crystallogr*, **65**, 128-133.

Brunger, A.T. and Rice, L.M. (1997) Crystallographic refinement by simulated annealing: methods and applications. *Methods Enzymol*, **277**, 243-269.

Carter, C.W., Jr. and Carter, C.W. (1979) Protein crystallization using incomplete factorial experiments. *J Biol Chem*, **254**, 12219-12223.

CCP4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.

Chang, A.B., Lin, R., Keith, W., Tran, C.V. and Saier, M.H.J. (2004) Phylogeny as a guide to structure and function of membrane transport proteins. *Mol. Membr. Biol.*, **221**, 171-181.

Chimento, D.P., Mohanty, A.K., Kadner, R.J. and Wiener, M.C. (2003) Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. *Nat Struct Biol*, **10**, 394-401.

Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr*, **62**, 1002-1011.

Cronan, J.E. (1968) Phospholipid Alterations During Growth of Escherichia coli *Journal of Bacteriology* **95**, 2054-2061.

Culham, D.E., Henderson, J., Crane, R.A. and Wood, J.M. (2003) Osmosensor ProP of Escherichia coli responds to the concentration, chemistry, and molecular size of osmolytes in the proteoliposome lumen. *Biochemistry*, **42**, 410-420.

Cullis, A.F., Muirhead, H., Perutz, M.F., Rossmann, M.G. and North, A.T.C. (1962) A three-dimensional

Fourier synthesis at 5.5 A resolution. Determination of the phase angles. *Proc Roy Soc A*, **265**.

Dall'Antonia, F., Baker, P.J. and Schneider, T.R. (2003) Optimization of selenium substructures as obtained from SHELXD. *Acta Crystallogr D Biol Crystallogr*, **59**, 1987-1994.

Dauter, Z. (1999) Data-collection strategies. *Acta Crystallogr D Biol Crystallogr*, **55**, 1703-1717.

Dauter, Z. (2005) Efficient use of synchrotron radiation for macromolecular diffraction data collection. *Prog Biophys Mol Biol*, **89**, 153-172.

Dauter, Z. (2006) Estimation of anomalous signal in diffraction data. *Acta Crystallogr D Biol Crystallogr*, **62**, 867-876.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S. and Richardson, D.C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*, **35**, W375-383.

de La Fortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol*, **276**, 472-494.

de Siervo, A.J. (1969) Alterations in the Phospholipid Composition of Escherichia coli B During Growth at Different Temperatures. *Journal of Bacteriology*, **100**, 1342-1349.

Debeye, P. (1914) Interferenz von Röntgenstrahlen und Wärmebewegung. *Annal. Physiol.*, **43**, 49-95.

DeLaBarre, B. and Brunger, A.T. (2003) Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nature Structural Biology*, **10**, 856-863.

DeLaBarre, B. and Brunger, A.T. (2006) Considerations for the refinement of low-resolution crystal structures. *Acta Crystallogr D Biol Crystallogr*, **62**, 923-932.

Diederichs, K. (2006) Some aspects of quantitative analysis and correction of radiation damage. *Acta Crystallogr D Biol Crystallogr*, **62**, 96-101.

Diederichs, K. and Karplus, P.A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol*, **4**, 269-275.

Diederichs, K., McSweeney, S. and Ravelli, R.B. (2003) Zero-dose extrapolation as part of macromolecular synchrotron data reduction. *Acta Crystallogr D Biol Crystallogr*, **59**, 903-909.

Doublie, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol*, **276**, 523-530.

Dougherty, D.A. (1996) Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science*, **271**, 163-168.

Dougherty, D.A. (2007) Cation-pi interactions involving aromatic amino acids. *J Nutr*, **137**, 1504S-1508S; discussion 1516S-1517S.

Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T. and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science*, **280**, 69-77.

Drenth, J. (1994) Principles of protein x-ray crystallography. *New York, Springer-Verlag*.

Ejsing, C.S., Duchoslav, E., Sampaio, J., Simons, K., Bonner, R., Thiele, C., Ekroos, K. and Shevchenko, A. (2006) Automated Identification and Quantification of Glycerophospholipid Molecular Species by Multiple Precursor Ion Scanning. *Anal. Chem.*, **78**, 6202-6214.

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

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

Engh, R.A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr Sect A*, **47**, 392-400.

Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, **62**, 72-82.

Ewald, P.P. (1921) Das reziproke Gitter in der Strukturtheorie. *Z. Kristallogr.*, **56**.

Fabiola, F., Korostelev, A. and Chapman, M.S. (2006) Bias in cross-validated free R factors: mitigation of the effects of non-crystallographic symmetry. *Acta Crystallogr D Biol Crystallogr*, **62**, 227-238.

Faham, S., Watanabe, A., Besserer, G.M., Cascio, D., Specht, A., Hirayama, B.A., Wright, E.M. and Abramson, J. (2008) The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. *Science*, **321**, 5.

Farwick, M., Siewe, R.M. and Krämer, R. (1995) Glycine betaine uptake after hyperosmotic shift in *Corynebacterium glutamicum*. *J. Bacteriol.*, **177**, 4690-4695.

Forrest, L.R., Zhang, W.Y., Jacobs, M.T., Gesmonde, J., Xie, L., Honig, B. and Rudnick, G.A. (2008) Mechanism for alternating access in neurotransmitter transporters. *Proc. Natl. Acad. Sci. USA*, **105**, 10338-10343.

Gallivan, J.P. and Dougherty, D.A. (1999) Cation-p interactions in structural biology. *Proc. Natl. Acad. Sci. USA*, **96**, 9459-9464.

Garman, E. (1999) Cool data: quantity AND quality. *Acta Crystallogr D Biol Crystallogr*, **55**, 1641-1653.

Garman, E.F. and Owen, R.L. (2006) Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **62**, 32-47.

Gasser, W., O'Brien, J., Schwan, D., Wilcockson, D. and Bitter, P. (1977) Lecithin/spingomyelin ratio procedure by thin-layer chromatography. *Am J Med Technol*, **43**, 1155-1159.

Girard, E., Legrand, P., Roudenko, O., Roussier, L., Gourhant, P., Gibelin, J., Dalle, D., Ounsy, M., Thompson, A.W., Svensson, O., Cordier, M.O., Robin, S., Quiniou, R. and Steyer, J.P. (2006) Instrumentation for synchrotron-radiation macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **62**, 12-18.

Glykos, N.M. and Kokkinidis, M. (2000) On the distribution of the bulk-solvent correction parameters. *Acta Crystallogr D Biol Crystallogr*, **56**, 1070-1072.

Gonzalez, A. (2007) A comparison of SAD and two-wavelength MAD phasing for radiation-damaged Se-MET crystals. *J Synchrotron Radiat*, **14**, 43-50.

Gouaux, E. (2009) Review. The molecular logic of sodium-coupled neurotransmitter transporters. *Philos Trans R Soc Lond B Biol Sci*, **364**, 149-154.

Grasberger, B., Minton, A.P., DeLisi, C. and Metzger, H. (1986) Interaction between proteins localized in membranes. *Proc Natl Acad Sci U S A*, **83**, 6258-6262.

Haardt, M., Kempf, B., Faatz, E. and Bremer, E. (1995) The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system ProU of Escherichia coli K-12. *Mol Gen Genet*, **246**, 783-786.

Harp, J.M., Timm, D.E. and Bunick, G.J. (1998) Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. *Acta Crystallogr D Biol Crystallogr*, **54**, 622-628.

Heginbotham, L. and MacKinnon, R. (1992) The aromatic binding site for tetraethylammonium ion on potassium channels. *Neuron*, **8**, 483-491.

Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *Embo J*, **9**, 1665-1672.

Hendrickson, W.A. and Konnert, J.H. (1980) Incorporation of stereochemical information into crystallographic refinement. . *In Computing in Crystallography. (R. Diamond, S. Rameshan and K. Ventkateshan, ed.) ,Indian Academy of Sciences,*, 10.01-10.23.

Holm, L. and Park, J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**, 566-567.

Holm L., P.J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**, 566-567.

Horn, C.S.-B., L. Breed, J. Welte, W. Schmitt, L. Bremer, E. (2006) 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.

Huang, Y., Lemieux, M.J., Song, J., Auer, M. and Wang, D.N. (2003) Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science*, **301**, 616-620.

Hunte, C. (2005) Specific protein-lipid interactions in membrane proteins. *Biochem Soc Trans*, **33**, 938-942.

Hunte, C., Screpanti, E., Venturi, M., Rimon, A., Padan, E. and Michel, H. (2005) Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. *Nature*, **435**, 1197-1202.

Jancarik, J., Scott, W.G., Milligan, D.L., Koshland, D.E., Jr. and Kim, S.H. (1991) Crystallization and preliminary X-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-mediating aspartate receptor of Salmonella typhimurium. *J Mol Biol*, **221**, 31-34.

Jardetzky, O. (1966) Simple allosteric model for membrane pumps. *Nature*, **211**, 969-970.

Jones, T.A., Bergdoll, M. and Kjeldgaard, M. (1990) O: A macromolecular modeling environment. . *Crystallographic and Modeling Methods in Molecular Design. Eds.: C. Bugg & S. Ealick. Springer-Verlag*, 189-195.

Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110-119.

Kabsch, W. (1988) Evaluation of single-crystal X-ray diffraction from a position-sensitive detector. *JOURNAL OF APPLIED CRYSTALLOGRAPHY*, **21**, 916-924.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry

and cell constants. *JOURNAL OF APPLIED CRYSTALLOGRAPHY*, **26**, 795-800.

Kappes, R.M., Kempf, B. and E., B. (1996) Three transport systems for the osmoprotectant glycine betaine operate in Bacillus subtilis: characterization of OpuD. *J. Bacteriol.*, **178**, 5071-5079.

Kavanaugh, M.P., Arriza, J.L., North, R.A. and Amara, S.G. (1992) Electrogenic uptake of gamma-aminobutyric acid by a cloned transporter expressed in Xenopus oocytes. *J Biol Chem*, **267**, 22007-22009.

Killmann, H., Benz, R. and Braun, V. (1996) Properties of the FhuA channel in the Escherichia coli outer membrane after deletion of FhuA portions within and outside the predicted gating loop. *J Bacteriol*, **178**, 6913-6920.

Kirillova, O., Chruszcz, M., Shumilin, I.A., Skarina, T., Gorodichtchenskaia, E., Cymborowski, M., Savchenko, A., Edwards, A. and Minor, W. (2007) An extremely SAD case: structure of a putative redox-enzyme maturation protein from Archaeoglobus fulgidus at 3.4 A resolution. *Acta Crystallogr D Biol Crystallogr*, **63**, 348-354.

Kleywegt, G.J. (1996) Use of non-crystallographic symmetry in protein structure refinement. *Acta Cryst D*, **52**, 842-857.

Kleywegt, G.J. and Brunger, A.T. (1996) Checking your imagination: applications of the free R value. *Structure*, **4**, 897-904.

Kleywegt, G.J. and Jones, T.A. (1995) Where freedom is given, liberties are taken. *Structure*, **3**, 535-540.

Kleywegt, G.J. and Jones, T.A. (2002) Homo Crystallographicus-Quo Vadis? *Structure*, **10**, 465-472.

Kramer, R. (2009) Osmosensing and osmosignaling in Corynebacterium glutamicum. *Amino Acids*.

Kramer, R. and Morbach, S. (2004) BetP of Corynebacterium glutamicum, a transporter with three different functions: betaine transport, osmosensing, and osmoregulation. *Biochim Biophys Acta*, **1658**, 31-36.

Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* , **24**, 946-950.

Krauss, N., Schubert, W.D., Klukas, O., Fromme, P., Witt, H.T. and Saenger, W. (1996) Photosystem I at 4 A resolution represents the first structural model of a joint

photosynthetic reaction centre and core antenna system. *Nat Struct Biol*, **3**, 965-973.

Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774--797.

Kuhlmann, S.I., Terwisscha van Scheltinga, A.C., Bienert, R., Kunte, H.J. and Ziegler, C. Osmoregulated transport of compatible solutes in the halophilic bacterium *Halomonas elongata*: 1.55 Å high-resolution structure of the periplasmic ectoine-binding protein from TRAP-transporter TeaABC. *Biochemistry*, **47 (36)**, 9475-9485.

Kuo, A., Bowler, M.W., Zimmer, J., Antcliff, J.F. and Doyle, D.A. (2003) Increasing the diffraction limit and internal order of a membrane protein crystal by dehydration. *J Struct Biol*, **141**, 97-102.

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

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures *J Appl Cryst* **14.**, 379-400.

Leslie, A.G. (1992) Recent changes to the MOSFLM package for processing film and image plate data *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, **26**.

Leslie, A.G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr*, **62**, 48-57.

Leslie, A.G.W. (1987) A reciprocal space method for calculation of a molecular envelope using the algorithm of B. C. Wang. *Acta Crystallogr Sect A*, **43**, 134-136.

Lolkema, J.S., Dobrowolski, A. and Slotboom, D.J. (2008) Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family. *J Mol Biol*, **378**, 596-606.

Lolkema, J.S. and Slotboom, D.-J. (2008) The major amino acid transporter superfamily has a similar core structure as Na^+-galactose and Na^+-leucine transporters. *Mol. Membr. Biol.*, 1-4.

Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S. and Richardson, D.C. (2003) Structure validation by C-alpha geometry: phi, psi, and C-beta deviation *Proteins: Structure, Function, and Genetics* **50**, 437-450.

Lovell, S.C., Word, J.M., Richardson, J.S. and Richardson, D.C. (1999) Asparagine and glutamine rotamers: Bfactor

cutoff and correction of amide flips yield distinct clustering. *Proc Natl Acad Sci U S A* **96**, 400-405.

Lovell, S.C., Word, J.M., Richardson, J.S. and Richardson, D.C. (2000) The penultimate rotamer library *Proteins*

, **40** 389-408.

Ma, J.C. and Dougherty, D.A. (1997) The Cationminus signpi Interaction. *Chem Rev*, **97**, 1303-1324.

Mahmood, N.A., Biemans-Oldehinkel, E., Patzlaff, J.S., Schuurman-Wolters, G.K. and Poolman, B. (2006) Ion specificity and ionic strength dependence of the osmoregulatory ABC transporter OpuA. *J Biol Chem*, **281**, 29830-29839.

Matthews, B.W. and Czerwinski, E.W. (1975) Local scaling: a method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. *Acta Crystallogr Sect A*, **31**, 480-487.

McCoy, A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr*, **63**, 32-41.

McCoy, A.J., Storoni, L.C. and Read, R.J. (2004) Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr D Biol Crystallogr*, **60**, 1220-1228.

McPherson, A. (1990) Current approaches to macromolecular crystallization. *Eur J Biochem*, **189**, 1-23.

Messerschmidt, A. (2007) X-Ray Crystallography of Biomacromolecules A Practical guide. *WILEY-VCH Verlag*.

Michel, H. (2003) Crystallisation of membrane proteins. In: Macromolecular Crystallography, vol. F, . *Macromolecular Crystallography*, **F**.

Mitchell, P. (1957) A general theory of membrane transport from studies of bacteria. *Nature*, **180**, 134-136.

Morbach, S. and Kramer, R. (2002) Body shaping under water stress: osmosensing and osmoregulation of solute transport in bacteria. *Chembiochem*, **3**, 384-397.

Morbach, S. and Kramer, R. (2004) Osmosregulation and osmosensing by uptake carriers for compatible solutes in bactera. *Molecular mechansims controlling transmembrane resport (Boles, E. and Kramer, R.)*, **Springer Heidelberg**, 155-177.

Morbach, S. and Kramer, R. (2005) Structure and function of the betaine uptake system BetP of Corynebacterium glutamicum: strategies to sense osmotic and chill stress. *J Mol Microbiol Biotechnol*, **10**, 143-153.

Murakami, S., Nakashima, R., Yamashita, E. and Yamaguchi, A. (2002) Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*, **419**, 587-593.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.

Muthing, J. and Radloff, M. (1998) Nanogram detection of phospholipids on thin-layer chromatograms. *Anal Biochem*, **257**, 67-70.

Nicklisch, S. (2008) EPR-based structural and functional characterization of the C-terminal domain of the osmoregulated glycine betaine transporter BetP from Corynebacterium glutamicum. *Dissertation Universitaet Koeln, Germany*.

Ott, V. (2008) Der Regulationsmechanismus des Osmosensors BetP aus Corynebacterium glutamicum. *Dissertation Universitaet Koeln, Germany*.

Ott, V., Koch, J., Spate, K., Morbach, S. and Kramer, R. (2008) Regulatory properties and interaction of the C- and N-terminal domains of BetP, an osmoregulated betaine transporter from Corynebacterium glutamicum. *Biochemistry*, **47**, 12208-12218.

Ozcan, N., Ejsing, C.S., Shevchenko, A., Lipski, A., Morbach, S. and Kramer, R. (2007) Osmolality, temperature, and membrane lipid composition modulate the activity of betaine transporter BetP in Corynebacterium glutamicum. *J Bacteriol*, **189**, 7485-7496.

Ozcan, N., Kramer, R. and Morbach, S. (2005) Chill activation of compatible solute transporters in Corynebacterium glutamicum at the level of transport activity. *J Bacteriol*, **187**, 4752-4759.

Painter, J. and Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr Sect D*, **62** 439-450.

Palma, A.M., Thiyagarajan, P., Wagner, A.M. and Tiede, D.M. (1999) Effect of detergent alkyl chain length on crystallisation of a detergent-solubilized membrane protein: correlation of protein-detergent particle size and particle-particle interaction with crystallisation of the photosynthetic reaction center from Rhodobacter sphaeroides. . *J. Crystal Growth* **207**, 214–225. 239.

Pannu, N.S., Murshudov, G.N., Dodson, E.J. and Read, R.J. (1998) Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallogr D Biol Crystallogr*, **54**, 1285-1294.

Pape, T. and Schneider, T.R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.*, **37** , 843-844.

Perutz, M.F. (1946) The composition and swelling properties of haemoglobin crystals. *Trans. Faraday Soc.*, **42**, 187-195.

Peter, H., Burkovski, A. and Kramer, R. (1996) Isolation, characterization, and expression of the Corynebacterium glutamicum betP gene, encoding the transport system for the compatible solute glycine betaine. *J Bacteriol*, **178**, 5229-5234.

Peter, H., Burkovski, A. and Kramer, R. (1998) Osmo-sensing by N- and C-terminal extensions of the glycine betaine uptake system BetP of Corynebacterium glutamicum. *J Biol Chem*, **273**, 2567-2574.

Petsko, G.A. (1985) Preparation of isomorphous heavy-atom derivatives. *Methods Enzymol*, **114**, 147-156.

Poolman, B., Blount, P., Folgering, J.H., Friesen, R.H., Moe, P.C. and van der Heide, T. (2002) How do membrane proteins sense water stress? *Mol Microbiol*, **44**, 889-902.

Racher, K.I., Culham, D.E. and Wood, J.M. (2001) Requirements for osmosensing and osmotic activation of transporter ProP from Escherichia coli. *Biochemistry*, **40**, 7324-7333.

Ramachandran, G.N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv Protein Chem*, **23**, 283-438.

Read, R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Crystallogr Sect A*, **36**, 878-884.

Read, R.J. (1990) Structure factor probabilities for related structures. *Acta Crystallogr Sect A*, **46**, 900-912.

Read, R.J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D Biol Crystallogr*, **57**, 1373-1382.

Ressl, S. (2006) Roentgenstrukturelle und elektronenmikroskopische Untersuchungen zweier Mutanten des Glycinbetain Transportproteins BetP aus Corynebacterium glutamicum. *Diploma Thesis*.

Ressl, S., Terwisscha van Scheltinga, A.C., Vonrhein, C., Ott, V. and Ziegler, C. (2009) Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. *Nature*, **458**, 47-52.

Rice, L.M. and Brunger, A.T. (1994) Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins*, **19**, 277-290.

Rice, L.M., Earnest, T.N. and Brunger, A.T. (2000) Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr D Biol Crystallogr*, **56**, 1413-1420.

Roche, A.S. (2007) Lab FAQs Find a Quick Solution 3rd Edition.

Ronsch, H., Kramer, R. and Morbach, S. (2003) Impact of osmotic stress on volume regulation, cytoplasmic solute composition and lysine production in Corynebacterium glutamicum MH20-22B. *J Biotechnol*, **104**, 87-97.

Rosenbaum, G., Holmes, K.G. and Woitz, J. (1971) Synchrotron radiation as a source for X-ray diffraction. *Nature*, **230**, 43-437.

Rossmann, M.G. and Blow, D.M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr*, **15**, 24-31.

Roux, M.J. and Supplisson, S. (2000) Neuronal and glial glycine transporters have different stoichiometries. *Neuron*, **25**, 373-383.

Roversi, P., Blanc, E., Vonrhein, C., Evans, G. and Bricogne, G. (2000) Modelling prior distributions of atoms for macromolecular refinement and completion. *Acta Crystallogr D Biol Crystallogr*, **56**, 1316-1323.

Rubenhagen, R., Morbach, S. and Kramer, R. (2001) The osmoreactive betaine carrier BetP from Corynebacterium glutamicum is a sensor for cytoplasmic K+. *Embo J*, **20**, 5412-5420.

Rubenhagen, R., Ronsch, H., Jung, H., Kramer, R. and Morbach, S. (2000) Osmosensor and osmoregulator properties of the betaine carrier BetP from Corynebacterium glutamicum in proteoliposomes. *J Biol Chem*, **275**, 735-741.

Rudnick, G.A. (2002) Structure, Function, and Regulation. *Neurotransmotter Transporters (ed. Reith, E.A.)*, 25-52.

Saier, M.H.J. (2000) Families of transmembrane sugar transport proteins. *Mol. Microbiol.*, **35**, 699-710.

Sambrook, P.N., Champion, G.D., Browne, C.D., Cairns, D., Cohen, M.L., Day, R.O., Graham, S., Handel, M., Jaworski, R., Kempler, S. and et al. (1989) Corticosteroid injection for osteoarthritis of the knee: peripatellar compared to intra-articular route. *Clin Exp Rheumatol*, **7**, 609-613.

Schiefner, A., Breed, J., Bosser, L., Kneip, S., Gade, J., Holtmann, G., Diederichs, K., Welte, W. and Bremer, E. (2004a) 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.

Schiefner, A., Breed, J., Bösser, L., Kneip, S., Gade, J., Holtmann, G., Diederichs, K., Welte, W. and Bremer, E. (2004b) Cation-π 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.

Schiefner, A., Holtmann, G., Diederichs, K., Welte, W. and Bremer, E. (2004c) Structural basis for the binding of compatible solutes by ProX from the hyperthermophilic archaeon Archaeoglobus fulgidus. *J Biol Chem*, **279**, 48270-48281.

Schiefner, A., Holtmann, G., Diederichs, K., Welte, W. and Bremer, E. (2004d) Structural basis for the binding of compatible solutes by ProX from the hyperthermophilic archaeon *Archaeoglobus fulgidus*. *J. Biol. Chem.*, **279**, 48270-48281.

Schiller, D., Kramer, R. and Morbach, S. (2004a) Cation specificity of osmosensing by the betaine carrier BetP of Corynebacterium glutamicum. *FEBS Lett*, **563**, 108-112.

Schiller, D., Krämer, R. and Morbach, S. (2004b) Cation specificity of osmosensing by the betaine carrier BetP of *Corynebacterium glutamicum*. *FEBS Lett.*, **563**, 108-112.

Schiller, D., Ott, V., Kramer, R. and Morbach, S. (2006) Influence of membrane composition on osmosensing by the betaine carrier BetP from Corynebacterium glutamicum. *J Biol Chem*, **281**, 7737-7746.

Schiller, D., Rübenhagen, R., Krämer, R. and Morbach, S. (2004c) The C-terminal domain of the betaine carrier BetP of *Corynebacterium glutamicum* is directly involved in sensing K^+ as an osmotic stimulus. *Biochemistry*, **43**, 5583-5591.

Schmidt, T.G., Koepke, J., Frank, R. and Skerra, A. (1996) Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *J Mol Biol*, **255**, 753-766.

Schneider, G. and Lindqvist, Y. (1994) Ta6Br14 is a useful cluster compound for isomorphous replacement in protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 186-191.

Schneider, T.R. and Sheldrick, G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr*, **58**, 1772-1779.

Schomaker, V. and Trueblood, K.N. (1968) On the rigid-body motion of molecules in crystals. *Acta Crystallogr*
Sect B, **24**, 63-76.
Screpanti, E. and Hunte, C. (2007) Discontinuous membrane helices in transport proteins and their correlation with function. *J. Struct. Biol.*, **159**, 261-267.
Seeger, M.A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K. and Pos, K.M. (2006) Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science*, **313**, 1295-1298.
Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. and Grutter, M.G. (2007) Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol*, **5**, e7.
Shinzawa-Itoh, K., Ueda, H., Yoshikawa, S., Aoyama, H., Yamashita, E. and Tsukihara, T. (1995) Effects of ethyleneglycol chain length of dodecyl polyethyleneglycol monoether on the crystallisation of bovine heart cytochrome c oxidase *J. Mol. Biol.*, **246**, 572–575.
Singh, S.K., Yamashita, A. and Gouaux, E. (2007) Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature*, **448**, 952-956.
Smicun, Y., Campbell, S.D., Chen, M.A., Gu, H. and Rudnick, G. (1999) The role of external loop regions in serotonin transport. Loop scanning mutagenesis of the serotonin transporter external domain. *J. Biol. Chem* . 36058-36064.
Stephan, M.M., Chen, M.A., Penado, K.M. and Rudnick, G. (1997) An extracellular loop region of the serotonin transporter may be involved in the translocation mechanism. . *Biochemistry* 1322-1330.
Strauss, M., Hofhaus, G., Schroder, R.R. and Kuhlbrandt, W. (2008) Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. *Embo J*, **27**, 1154-1160.
Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D. and Eisenberg, D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*, **103**, 8060-8065.
Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, **41**, 207-234.
Takatsuka, Y. and Nikaido, H. (2009) Covalently linked trimer of the AcrB multidrug efflux pump provides support for the functional rotating mechanism. *J Bacteriol*, **191**, 1729-1737.
Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.*, **24**, 1-95.
Terwilliger, T.C. (1999) Reciprocal-space solvent flattening. *Acta Crystallogr D Biol Crystallogr*, **55**, 1863-1871.
Terwilliger, T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr*, **56**, 965-972.
Terwilliger, T.C. (2001) Maximum-likelihood density modification using pattern recognition of structural motifs. *Acta Crystallogr D Biol Crystallogr*, **57**, 1755-1762.
Terwilliger, T.C. (2003) Automated side-chain model building and sequence assignment by template matching. *Acta Crystallogr D Biol Crystallogr*, **59**, 45-49.
Terwilliger, T.C. and Berendzen, J. (1999) Discrimination of solvent from protein regions in native Fouriers as a means of evaluating heavy-atom solutions in the MIR and MAD methods. *Acta Crystallogr D Biol Crystallogr*, **55**, 501-505.
Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J. and Adams, P.D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr.*, **D64**, 61-69.

Tronrud, D.E. (2004) Introduction to macromolecular refinement. *Acta Crystallogr Sect D*, **60**, 2156-2168.

Tronrud, D.E., Ten Eyck, L.F. and Matthews, B.W. (1987) An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr Sect A*, **43**, 489-501.

Tsai, C.J. (2008) Three-dimensional structure of the glycine-betaine transporter BetP by cryo electron crystallography. *Dissertation Goethe Universitaet Frankfurt, Germany*.

Tsai, C.J., Ejsing, C.S., Shevchenko, A. and Ziegler, C. (2007) The role of lipids and salts in two-dimensional crystallization of the glycine-betaine transporter BetP from Corynebacterium glutamicum. *J Struct Biol*, **160**, 275-286.

Tsai, C.J. and Ziegler, C. (2005) Structure determination of secondary transport proteins by electron crystallography: two-dimensional crystallization of the betaine uptake system BetP. *J Mol Microbiol Biotechnol*, **10**, 197-207.

Tsatskis, Y., Khambati, J., Dobson, M., Bogdanov, M., Dowhan, W. and Wood, J.M. (2005) The osmotic activation of transporter ProP is tuned by both its C-terminal coiled-coil and osmotically induced changes in phospholipid composition. *J Biol Chem*, **280**, 41387-41394.

Ubarretxena-Belandia, I., Baldwin, J.M., Schuldiner, S. and Tate, C.G. (2003) Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J*, **22**, 6175-6181.

Van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C. and Rapoport, T.A. (2004) X-ray structure of a protein-conducting channel. *Nature*, **427**, 36-44.

van der Does, C. and Tampe, R. (2004) Changing orders--primary and secondary membrane transporters revised. *Chembiochem*, **5**, 1171-1175.

van der Heide, T., Stuart, M.C. and Poolman, B. (2001) On the osmotic signal and osmosensing mechanism of an ABC transport system for glycine betaine. *Embo J*, **20**, 7022-7032.

Van Dort, H.M., Knowles, D.W., Chasis, J.A., Lee, G., Mohandas, N. and Low, P.S. (2001) Analysis of integral membrane protein contributions to the deformability and stability of the human erythrocyte membrane. *J Biol Chem*, **276**, 46968-46974.

Vinothkumar, K.R., Raunser, S., Jung, H. and Kuhlbrandt, W. (2006) Oligomeric structure of the carnitine transporter CaiT from Escherichia coli. *J Biol Chem*, **281**, 4795-4801.

von Blohn, C., Kempf, B., Kappes, R.M. and Bremer, E. (1997) Osmostress response in Bacillus subtilis: characterization of a proline uptake system (OpuE) regulated by high osmolarity and the alternative transcription factor sigma B. *Mol Microbiol*, **25**, 175-187.

Vonrhein, C. and Schulz, G.E. (1999) Locating proper non-crystallographic symmetry in low-resolution electron-density maps with the program GETAX. *Acta Crystallogr D Biol Crystallogr*, **55**, 225-229.

Voss, S. and Skerra, A. (1997) Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng*, **10**, 975-982.

Wang, B.C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol*, **115**, 90-112.

Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E.P., Rutherford, N.G., Hadden, J.M., O'Reilly, J., Ma, P., Saidijam, M., Patching, S.G., Hope, R.J., Norbertczak, H.T., Roach, P.C.J., Iwata, S., Henderson, P.J.F. and Cameron, A.D. (2008) Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science*, **322**, 709-713.

Wilson, A.J.C. (1949a) The probability distribution of X-ray intensities. *Acta Crystallogr*, **2**, 318-321.

Wilson, A.J.C. (1949b) The probability distribution of X-ray intensities. *Acta Crystallogr*, **2**.

Winn, M.D., Isupov, M.N. and Murshudov, G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr Sect D* **57**, 122-133.

Winn, M.D., Murshudov, G.N. and Papiz, M.Z. (2004) Macromolecular TLS refinement in REFMAC at moderate

resolutions *Methods Enzymol*, **374**, 300-321.

Wittig, I., Karas, M. and Schagger, H. (2007) High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol Cell Proteomics*, **6**, 1215-1225.

Wolf, A., Kramer, R. and Morbach, S. (2003) Three pathways for trehalose metabolism in Corynebacterium glutamicum ATCC13032 and their significance in response to osmotic stress. *Mol Microbiol*, **49**, 1119-1134.

Wood, J.M. (1999) Osmosensing by bacteria: signals and membrane-based sensors. *Microbiol Mol Biol Rev*, **63**, 230-262.

Wood, W.B. (1966) Host specificity of DNA produced by Escherichia coli: bacterial mutations affecting the restriction and modification of DNA. *J Mol Biol*, **16**, 118-133.

Word, J.M., Lovell, S.C., LaBean, T.H., Taylor, H.C., Zalis, M.E., Presley, B.K., Richardson, J.S. and Richardson, D.C. (1999a) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms *J Mol Biol* **285**, 1711-1733.

Word, J.M., Lovell, S.C., Richardson, J.S. and Richardson, D.C. (1999b) Asparagine and glutamine: using

hydrogen atom contacts in the choice of side-chain amide orientation. . *J Mol Biol*, **285**, 1735-1747.

Xiu, X., Puskar, N.L., Shanata, J.A., Lester, H.A. and Dougherty, D.A. (2009) Nicotine binding to brain receptors requires a strong cation-pi interaction. *Nature*, **458**, 534-537.

Yamashita, A., Singh, S.K., Kawate, T., Jin, Y. and Gouaux, E. (2005) Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature*, **437**, 215-223.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103-119.

Ye, J. and van den Berg, B. (2004) Crystal structure of the bacterial nucleoside transporter Tsx. *Embo J*, **23**, 3187-3195.

Yernool Dinesh, B.O., Jin Yan, Gouaux Eric. (2004) Structure of a glutamate transporter homologue from Pyrococcus horikoshii. *Nature*, **431**, 811-818.

Ziegler, C., Morbach, S., Schiller, D., Kramer, R., Tziatzios, C., Schubert, D. and Kuhlbrandt, W. (2004) Projection structure and oligomeric state of the osmoregulated sodium/glycine betaine symporter BetP of Corynebacterium glutamicum. *J Mol Biol*, **337**, 1137-1147.