Structural and Functional Characterization of the Integral Membrane Proteins CitS and CCR5 by Electron Microscopy

Inauguraldissertation

Zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Fabian Kebbel

aus Ulm, Deutschland

Basel, Schweiz, 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel

auf Antrag von

Prof. Dr. Henning Stahlberg (Biozentrum, Universität Basel), Fakultätsverantwortlicher
Prof. Dr. Sebastian Hiller (Biozentrum, Universität Basel), Koreferent

Basel, den 21.5.2013

Prof. Dr. Jörg Schibler (Dekan)

Summary

Secondary transport proteins are integral membrane proteins found in every cell. They facilitate the transport of versatile substrates (e.g. nutrients, ions and drugs) across the hydrophobic membrane barrier. Independent on their mode of transport (symport/antiport) the uphill transport of the main substrate is driven by the coupled flux of a co-substrate downhill its electrochemical gradient. Malfunction of secondary transporter can cause severe physiological disorders like depression and obesity and therefore these transport proteins constitute attractive drug targets.

The main part of this PhD thesis is the structural and functional characterization of the secondary citrate/sodium symporter CitS from *Klebsiella pneumonia*, mainly by transmission electron microscopy (TEM). CitS is the best characterized member of the bacterial 2-hydroxycarboxylate transporter (2-HCT) family. It facilitates the secondary transport of bivalent citrate ions driven by a coupled flux of Na^+ across the inner membrane of the host. Hydropathy profiling and extensive biochemical experimentation prior to this study predicted CitS to represent a new structural fold as paradigm for numerous related proteins, so that it constitutes a highly attractive target for structural studies.

As a first step, two-dimensional (2D) crystals of recombinant CitS were produced by dialysis assisted reconstitution of pure detergent solubilized protein into bilayer forming phospholipids. Extensive screening of crystallization conditions led to highly ordered tubular 2D crystals suitable for structure determination by cryo-electron crystallography. Therefore, numerous sample preparation methods were evaluated, while plunge-freezing provided significantly better results compared to commonly used sugar embedding methods. As described in chapter 2, image processing of electron micrographs from plunge-frozen 2D crystals provided the projection structure of CitS at 6 Å resolution. The transporter appears as oval shaped dimer measuring 5*9 nm in the membrane plane. The dimer reveals three distinct structural domains being formed by two dense clusters of α-helices at each molecule's tip and a third, less dense domain in the center of the dimer. The domains are separated by solvent areas. Surprisingly, this architecture highly resembles that of the unrelated Na^+/H^+ antiporter NhaP1. In projection, each CitS monomer reveals eleven TMS that well match previous membrane topology predictions. Finally, we developed several models describing possible monomer-monomer interfaces and domain organizations.

In chapter 3, we describe the 3D structure of CitS at 6/15 Å resolution obtained by electron micrographs of tilted 2D crystal samples. Based on the 3D volume, we developed a molecular model that reveals eleven α-helices and two additional helical reentrant loops. The central dimerization domain is formed by seven partially tilted helices, while the distal cluster reveals 4 transmembrane segments surrounding the two reentrant loops. We also find internal structural symmetry for the strongly intertwined N- and C-terminal domains as prerequisite for substrate translocation by the 'alternating access' mechanism. Additional projection structures of CitS in various substrate environments (Na^+, K^+, acetate and citrate) allowed us to map the conformational space. The binding of citrate as main substrate induces a defined movement of α-helices spatially limited to the helix cluster in each monomer. This primarily occurs in the presence of Na^+, and much less with K^+ and highlights the high co-ion specificity. These findings also enable us to assign the dense helix cluster as substrate binding and translocation site.

In a second project, various biophysical techniques were used to characterize the recombinant G protein-coupled receptor (GPCR) CCR5. Besides its important role in immune responses, CCR5 also acts as co-receptor during HIV-1 target cell entry. In chapter 4, an innovative *E. coli* based expression platform is presented that enables the production of 10 mg purified protein from 1L cell culture. We could demonstrate ligand binding, structural integrity, homogeneity and stability of triply isotope labeled CCR5. This provides a promising starting point for ongoing structural studies, especially by nuclear magnetic resonance (NMR) spectroscopy.

Table of contents

Chapter 1 – Introduction

1.1 Biological membranes & membrane proteins

Biological membranes are essential for life since these versatile structures define and control the boarders of single organisms, cells or cellular organelles. Biomembranes are involved in a vast variety of biological processes: (1) formation and separation of biological compartments, (2) energy storage by maintaining electrochemical gradients, (3) uptake and secretion of nutrients and metabolites, (4) signal transduction, (5) control of enzymatic activities and (6) control of cell adhesion and mobility. Membranes consist of a lipid bilayer, usually 5-7 nm thick, and associated proteins while both can be directly linked to carbohydrates. The lipid content can be further subdivided into glycerophospholipids, sphingolipids and sterols[1].

The biological importance of membrane proteins is clearly reflected in two numbers. First, about 30% of a eukaryotic genome encode for membrane proteins[2] and second, membranes contain up to 80 % (w/w) of membrane associated proteins[1]. This abundance can be explained by their participation in nearly all membrane-involved processes. Peripheral membrane proteins are reversibly attached to the surface of the lipid bilayer mostly by weak electrostatic interactions (e.g. cytochrome *c*, phospholipase PLA_2, ankyrin and annexins) while they mainly contribute to enzymatic activities and structural aspects of membranes[1,3]. In contrast, integral membrane proteins partially or fully insert into the lipid bilayer as α-helical bundle or β-barrel. They are the key players in transport and signal transduction across the membrane[3].

1.2 Membrane transport proteins

Biological membranes constitute a perfectly designed hydrophobic barrier that is nearly impermeable for hydrophilic compounds such as ions and nutrients. However, their in- and efflux is essential to maintain all kinds of cellular processes. Therefore, the membranes of cells and organelles carry a vast variety of specific transport proteins that enable and regulate the substrate exchange across the lipid bilayer. These membrane transport proteins can be classified due to their functionality, as illustrated in figure 1.1. Channels and porins passively facilitate the selective diffusion of e.g. water and ions down their concentration gradient. Primary and secondary active transporters translocate their substrates under energy consumption against their electrochemical gradients[1,4].

1.2.1 Channels & Porins

Channels and porins enable and regulate the selective and fast flux of their polar substrates across the lipid bilayer downstream their electrochemical gradient. Typical substrates for channels are H_2O (aquaporins), K^+, Na^+, Ca^{2+}, H^+ and Cl^-, while each channel usually is highly specific for only one substrate species[1,5]. Their high selectivity (e.g. K^+ over Na^+ >1000-fold for potassium channels) and their typically high transport rates (10^7-10^8 molecules/s) are the results of some unique and striking structural features. In potassium channels, for instance, four

identical subunits form the pore in the center of the tetramer. The selectivity filter is formed by a set of carbonyl oxygens from four sequence motifs that allow K^+ ions to bind and translocate while imitating the hydration shell[5]. Furthermore, channels are often regulated by different stimuli, e.g. membrane potential, ions, biochemical ligands, mechano-sensing and even temperature[1,5].

Porins constitute another class of passive membrane transport proteins. These β-barrel shaped and water filled pores are predominantly found in mitochondria, chloroplasts and the outer membrane of Gram-negative bacteria[6]. Their main function is to facilitate the passive diffusion of a huge variety of solutes. Porins appear as both, highly specific (e.g. maltoporins) and unspecific transporter (outer membrane proteins, e.g. OmpF). The diffusion rate is regulated by the oligomeric state of the transporter, by a switchable loop within the cavity and by the electrochemical gradient of the substrate[6,7].

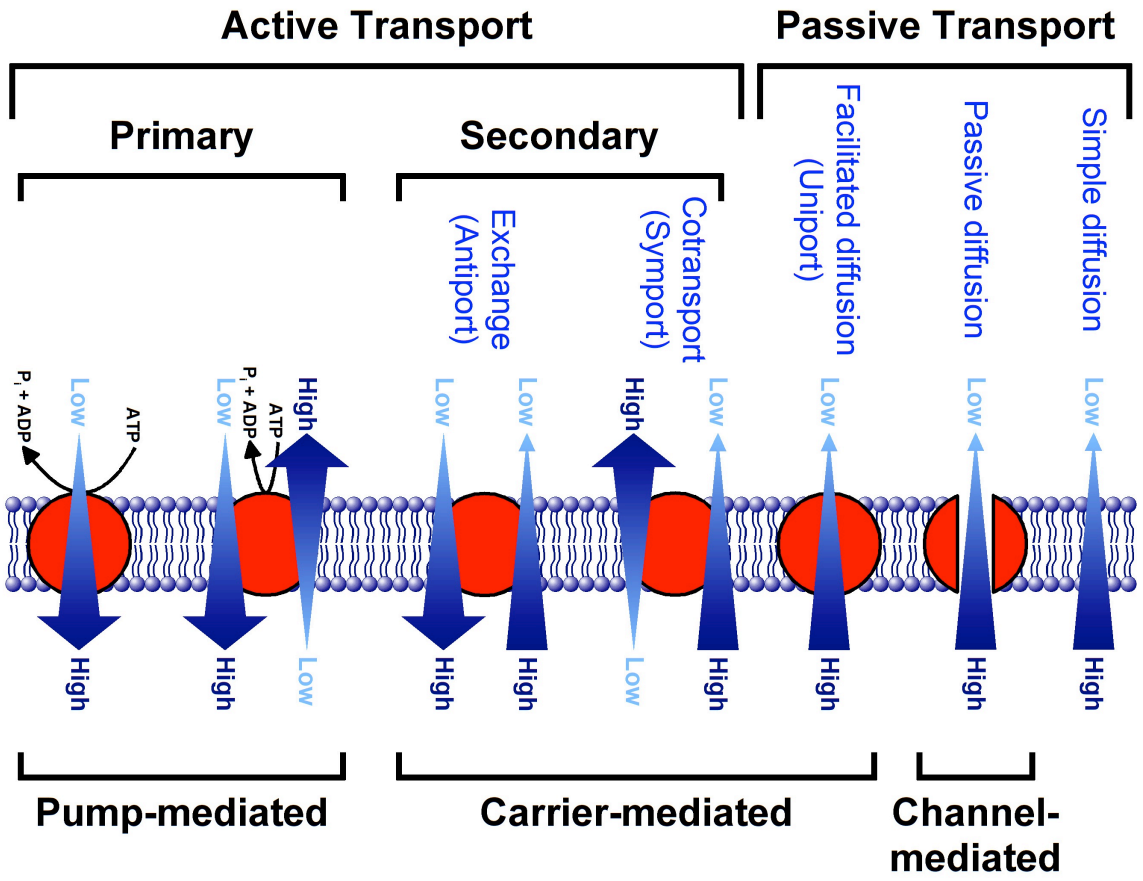


Figure 1.1 Overview of membrane transport processes. The translocation of substrates across lipid bilayers can be divided into active and passive processes. Active transport describes substrate translocation against concentration gradients. The energy input for most primary active transport proteins (pumps) is the hydrolysis of ATP. Secondary active transporters (carrier) use electrochemical gradients of co-substrates as energy source. Passive transport processes mediate substrate translocation downhill of electrochemical gradients and do not require a direct energy input. Passive translocation is based on simple diffusion, channel mediated passive diffusion or carrier mediated facilitated diffusion. Reprinted with permission from *PhysiologyWeb*, ©2000-2012.

1.2.2 Primary active transporters

Primary active transport proteins, also called molecular ion pumps, are integral membrane proteins that directly use an energy source to energize the cell membrane by establishing a transmembrane electrochemical potential. Primary active transport can be driven by (1) redox processes (e.g. complexes I-III as part of the respiratory chain), (2) light (e.g. photosynthetic reaction centers), and by (3) direct adenosine-triphosphate (ATP) hydrolysis (P-/V-/F-type ATPases and ATP-binding cassette (ABC) transporters)[8,9]. In all cases, the energy input is utilized to 'pump' substrates across the lipid bilayer against their electrochemical gradient. The resulting membrane potential then in turn can be used to drive other cellular processes, e.g. the formation of action potentials in neurons or the transport of nutrients and metabolites in and out of the cell by secondary active transporters[8].

1.2.3 Secondary active transporters

Secondary active transporters are found in every cell. They are involved in multiple biological processes such as nutrient uptake, efflux of metabolites and noxious substances and removal of neurotransmitters from the synaptic cleft. Malfunction of these processes can lead to severe physiological disorders such as epilepsy, depression and obesity. For this reason, secondary transporters are attractive drug targets[10]. As expected from their diverse function, there is only little sequence homology between the different classes and families[11].

During secondary transport, a substrate (main substrate) is translocated across the membrane upstream its electrochemical gradient. The energy source for this unfavorable reaction is the co-transport of a second substrate (co-substrate) downstream its own electrochemical gradient formed by primary active transporters, hence the term secondary. Based on their mode of transport, secondary active transporters can be grouped into symporters and antiporters[10,12] (Figure 1.1). During antiport, both substrates are translocated in opposite directions (bidirectional). A prominent example is given with the Na^+/Ca^{2+} exchanger NCX that shuttles Ca^{2+} out of the cell, driven by the influx of Na^+[13]. Symporters, on the other hand, use a unidirectional path with same directions of both substrates, as demonstrated for the proton dependent lactose importer LacY[14]. Since the direction of both transport modes is dictated only by the electrochemical gradient of the substrate(s), translocation can occur in both directions[15]. During the transport-cycle, usually 10^2–10^4 substrate molecules are moved across the membrane each second[1].

Secondary active transporters exhibit a huge diversity in terms of amino acid sequence, three-dimensional (3D) structure and the chemical nature of transported substrates. Based on their primary structure, more than 100 different families could be identified within the TC classification system, while 40 families can be found in humans[16]. This sequence diversity also leads to a vast variety of transported substrates, ranging from sugars, amino acids, ions, neurotransmitters, peptides, sterols, nucleosides/nucleotides and drugs to all kinds of biochemical metabolites, e.g. citric acid and glycerol-3-phosphate[15,16]. The co-substrate, however, is much less diverse; most secondary transporters use the electrochemical gradient of Na^+ or H^+ to drive the transport of the main substrate[10,12]. In a few cases, K^+ and/or Cl^- are additionally coupled to the sodium/proton flux, as reported for the serotonin transporter

SERT[16,17]. In most cases, the substrate affinity to the transporter is rather low which is reflected in a typical dissociation constant (K_D) in the μM range[15,18]. This enables high transport rates by substrate diffusion from the transporter. The structural basis of substrate specificity, ion coupling and transport activity will be discussed in section 1.2.3.4.

1.2.3.1 Structural features of secondary transporters

The very first 3D volume of a secondary transporter was presented in 2000, the Na^+/H^+ antiporter NhaA from *E. coli* at 7 Å resolution obtained by electron crystallography of two-dimensional (2D) crystals[19]. In 2003, the structure of AcrB represented the first transporter at atomic resolution, *i.e.* <4 Å[20]. Major advances and developments in biomolecular structure determination by x-ray diffraction (XRD) during the last decade led to a rapidly growing number of secondary transporter structures at atomic resolution. So far, 29 unique structures are reported[21], as summarized in table 1.1. Several of these atomic structures are complemented by lower resolved 3D structures from electron crystallography (see section 1.3.1.4 for details)[22,23]. Structures of representative transport proteins are additionally illustrated in figure 1.2.

All described secondary transporters are α-helical integral membrane proteins with 4-14 transmembrane helices while the major fraction carries 10-12 helices[15,22]. As shown in figure 1.2, most parts of the proteins are buried in the membrane, with no or only little protrusions out of the lipid bilayer. An exception here is AcrB as part of a tripartite complex spanning both membranes and the periplasmic space in gram-negative bacteria[24].

Numerous examples are available for different oligomeric states, including monomers (e.g. LacY[14]), dimers (e.g. NhaA[25]) or trimers (e.g. Glt_{Ph}[26]). Most transporters, however, are found in the dimeric form. The monomer-monomer interface can thereby be formed by β-sheets as exclusively found in NhaA[25] (Figure 1.2A), single helices[27] (LeuT, Figure 1.2B) up to seven helices[28] (*Vc*INDY, Figure 1.3C). Surprisingly, numerous phylogenetically and functionally unrelated transporters were found to share a common global structural fold with highly resembling 3D structures, e.g. the fold of LeuT (eight members, grey box, Table 1.1), the major facilitator superfamily (eight members, blue box) and NhaA (two members, green box). For all remaining folds so far only single structures are available, while more examples are expected to follow in prospective studies. Remarkably, each structural fold contains symporters and antiporters with highly resembling structures. This emphasizes the fact that the global protein architecture does not dictate the mode of transport[15].

Several transporters could be crystallized in the presence of native or artificial substrates, e.g. leucine in LeuT (Figure 1.2B) and bivalent citrate in *Vc*INDY (Figure 1.2C). In all cases, the substrates were found to bind in the center of the monomeric protein close to the middle of the membrane plane. More importantly, substrates can be found in every monomer, even if the native transporter exhibits higher oligomeric states (e.g. one citrate molecule in each *Vc*INDY monomer, Figure 1.2C). In combination with extensive biochemical experimentation, this feature allows the conclusion that the monomeric protein constitutes the functional unit of secondary transporters[15]. Oligomerization, however, was shown to play pivotal roles in structural stability and in regulatory aspects of transport[29,30]. The only known exception is the H^+/drug antiporter EmrE, where antiparallel dimerization is essential for functionality[31].

Table 1.1 Known 3D structures of secondary transporters by x-ray and electron crystallography

Protein	Transport activity	Fold	Resolution XRD [Å]	Resolution Electron Crystallography [Å]
LeuT *	Na^+/leucine symport	LeuT	1.6 [27]	-
vSGLT	Na^+/glucose symport	LeuT	2.7 [32]	-
Mhp1	Na^+/hydantoin symport	LeuT	2.8 [33]	-
BetP	Na^+/betaine symport	LeuT	3.3 [34]	8.0 [35]
AdiC	Arginine/agmatine antiport	LeuT	3.2 [36]	-
ApcT	H^+/amino acid symport	LeuT	2.3 [37]	-
CaiT	Carnithine/butyrobetaine antiport	LeuT	2.3 [38]	-
GadC	GABA/glutamate antiport	LeuT	3.1 [39]	-
EmrD	H^+/drug antiport	MFS	3.5 [40]	-
FucP	H^+/sugar symport	MFS	3.1 [41]	-
$PepT_{So}$	H^+/oligopeptide symport	MFS	3.6 [42]	-
$PepT_{St}$	H^+/oligopeptide symport	MFS	3.3 [43]	-
XylE	H^+/xylose symport	MFS	2.8 [44]	-
GlpT	Glycerol-3-phosphate/ PO_4^{3-} antiport	MFS	3.3 [45]	-
LacY *	H^+/sugar symport	MFS	3.6 [14]	-
PiPT	H^+/ PO_4^{3-} symport	MFS	2.9 [46]	-
NhaA *	H^+/Na^+ antiport	NHA	3.4 [25]	7.0 [19,47]
$ASBT_{NM}$	Na^+/taurocholate symport	NHA	2.2 [48]	-
AAC1	ADP/ATP antiport	AAC1	2.2 [49]	-
Glt_{Ph} *	Na^+/aspartate symport	Glt_{Ph}	3.5 [26]	-
CNT	Na^+/nucleoside symport	CNT	2.4 [50]	-
YiiP	Zn^{2+}/H^+ antiport	YiiP	2.9 [51]	13.0 [52]
AcrB	H^+/drug antiport	AcrB	2.9 [24]	-
***Ec*ClC**	H^+/Cl^- antiport	ClC	3.0 [53]	-
EmrE	H^+/drug antiport	SMR	3.8 [31]	7.0 [54,55]
UraA	H^+/uracil symport	UraA	2.8 [18]	-
NCX	Na^+/Ca^{2+} antiport	NCX	1.9 [13]	-
***Vc*INDY ***	Na^+/divalent anion symport	INDY	3.2 [28]	-
***Pf*MATE**	H^+/drug antiport	MATE	3.2 [56]	-

Abbreviations: ADP/ATP (adenosine di-/triphosphate), GABA (γ-amino butyric acid), MFS (major facilitator superfamily), RND (resistance nodulation cell division), SMR (small multidrug-resistance), XRD (x-ray diffraction). All listed atomic structures were solved using XRD. Proteins where only low resolution 3D data from electron crystallography is available are not listed (see section 1.3.1.4 for details). Members of a common fold are highlighted with colored boxes. Only one representative structure of each protein is listed. Proteins marked with an asterisk are illustrated in figure 1.2.

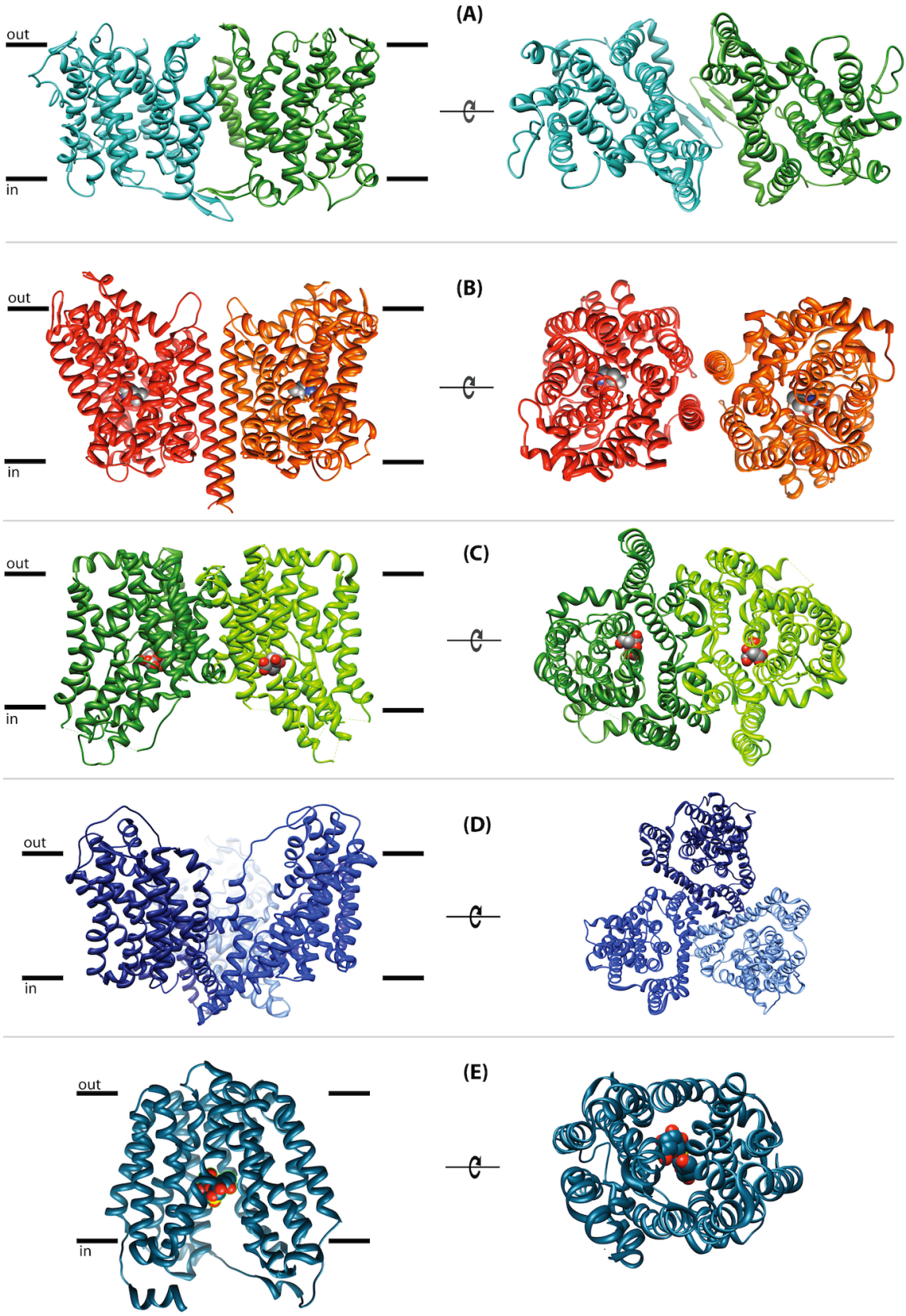


Figure 1.2 Structural diversity of secondary active transporters. Secondary transporters exhibit a wide range of 3D structures. Sideview (left) and topview (right) of **(A)** NhaA (pdb 3FI1), **(B)** LeuT (pdb 2AG5), **(C)** *Vc*INDY (pdb 4F35), **(D)** Glt_{Ph} (pdb 1XFH) and **(E)** LacY (pdb 1PV7). Monomers within higher oligomers are depicted in different colors. The membrane plane is shown as black lines. If available, bound substrates are displayed as spherical molecules.

Another prominent feature of most available structures is an occuring internal structural symmetry, where two defined domains in a single monomer are structurally related to each other[15]. This internal symmetry can be of different form and can have different origins. First, monomeric transporters can be composed of two or more defined structural repeats, *i.e.* helical domains with significant sequence homology as a result from an assumed ancient gene duplication event. This can be found for the mitochondrial ADP/ATP antiporter AAC1[49] and all known members of the MFS fold[57], e.g. the lactose permease LacY[14]. Here, the resulting symmetry axis runs through the center of the monomeric protein in between the two distinct and symmetry related N- and C-terminal halves of the monomer perpendicular to the membrane plane (Figure 1.3A-C). The high structural similarity between both domains is reflected in a low root-mean-square deviation (RMSD) of 2-3 Å[15]. Second, even without or only little sequence homology (<20 %), an uneven number of helices within each of several helical repeats (e.g. 2*5 helices) leads to a pseudo two-fold symmetry referred to as 'inverted topology'. This is found for *Vc*INDY and for all members of the LeuT fold (Table 1.1). Here, the two symmetry related domains are strongly intertwined with an inverted orientation in the membrane plane. The resulting apparent symmetry axis runs parallel to the membrane, again through the center of the molecule (Figure 1.3D-F). Superposition of symmetry related helical domains yields typical RMSD values of 3-5 Å[12].

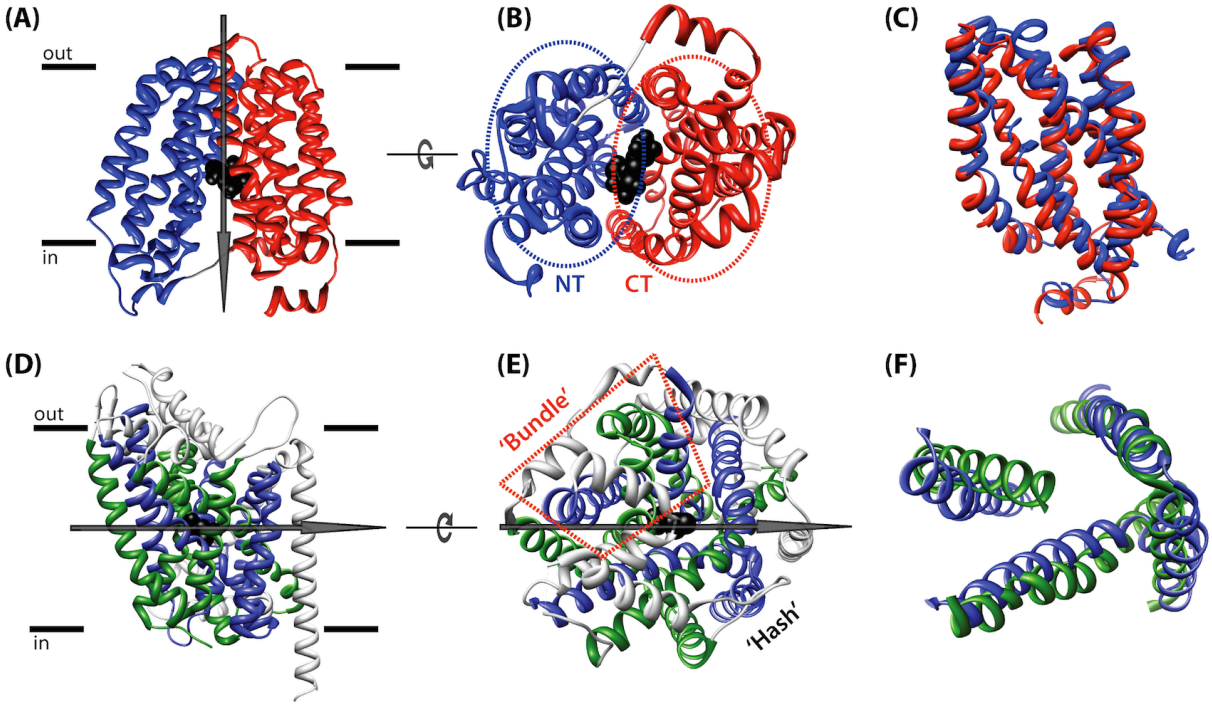


Figure 1.3 Internal structural symmetry within monomeric secondary transporters. LacY viewed from side **(A)** and intracellular space **(B)**. The six α-helices of the N- and C-terminal domain are depicted in blue and red, respectively. The symmetry axis (arrow) runs through the center of the molecule perpendicular to the membrane plane. The N- and C-domains can be well superimposed **(C)**. LeuT viewed from side **(D)** and top **(E)**. Five α-helices from each of both structural repeats are depicted in blue and green, respectively. Both domains exhibit strong intertwining. Symmetry unrelated helices are depicted in light grey. The characteristic 'bundle' domain with four helices is indicated (red line), the remaining molecule is referred to as 'hash'. The symmetry axis runs through the molecule's center parallel to the membrane. Helices 2-5 and 7-10 can be well superimposed **(F).** Thiogalactoside (LacY) and leucine (LeuT) as substrates are shown as black spheres in the center of the corresponding monomer.

Importantly, in both described cases each monomeric protein exhibits at least two distinct structural domains that can be of different origin. First, these domains can be formed by the two symmetry related parts itself as found for LacY (Figure 1.3B). Second, the structural distinct domains can contain helices of both symmetrical elements. The latter then leads to distinct domains that are not symmetry related itself, but contain parts of both symmetrical domains, e.g. the 'hash' and 'bundle' domains as found for transporters of the LeuT fold (Figure 1.3E). In both cases, however, the symmetry axes and the domain interfaces run through the central substrate binding site in the monomer as the functional unit. This already indicates an important functional role of structural symmetry for the substrate translocation mechanism, as discussed in the following section.

1.2.3.2 The transport model of 'alternating access' for secondary symport

In 1966, a first model was developed that described the structural basis of solute transport across membranes by secondary active transporters[58]. In this model of 'alternating access' the transporter protein cycles through a set of defined conformational states that provide a unique structural framework for efficient substrate transport (Figure 1.4A). All available atomic structures from secondary transporters significantly contributed to the understanding of the molecular details of substrate translocation. The availability of several structures within one common fold (e.g. LeuT and other members of the MFS fold, Table 1.1) and, especially, the availability of different conformations of single unique transporters significantly enhanced the knowledge on the conformational dynamics during the transport cycle[15]. On the background of this thesis, only the principle of secondary symport is described.

The current model of alternating access for an importing symporter is illustrated in figure 1.4. Here, the substrates first bind to the empty transporter in the outward facing conformation 'Ce' where the binding site is only accessible from the outer side (Figure 1.4A/B). This is followed by the closure of outer molecular gates to hinder substrate diffusion. The gate closure is thereby facilitated by the substrate induced rearrangement of single amino acid side chains or by the bending of single α-helices and/or helical hairpins, as found for LacY and LeuT, respectively[15,33]. The transport cycle then proceeds by a substantial conformational change from the closed outward facing to the closed inward facing conformation 'Ci'. During this structural switch the transporter passes through the closed occluded form 'Cc', where the substrates are inaccessibly buried within the protein (Figure 1.4A/C). This occluded state corresponds to the energetic 'transition state' of the transport cycle. Importantly, with the exception of Glt_{Ph}[26], the positions of all substrates at their corresponding binding sites remain unaffected and unchanged during this conformational change. This observation led to the model of a 'single binding center gated pore' (SBCGP) as an alternative mechanistic description for secondary transport[15].

The structural rearrangement from the outward to the inward facing state usually requires movements of whole characteristic domains within the protein. This can be achieved by the rocking movements of two symmetry related N- and C-domains against each other, as found for LacY and other proteins of the MFS fold (Figure 1.4B-D). This 'rocker switch' mechanism effectively opens and closes the central substrate binding sites alternating to either side of the membrane[15,45]. A variation of the described alternating rocker switch is available for the LeuT

fold. During this 'rocking bundle' mechanism, distinct helix clusters - the helical 'bundle' and 'hash' domains (Figure 1.3E) - rock against each other to control substrate accessibility[59]. A third but less common variation is described as 'gating mechanism' for Glt_{Ph}[60]. All models, however, are in good agreement with the overall idea of alternating access accomplished by the movement of helical domains against each other. Importantly, all described mechanisms involve the movements of distinct domains that arise from internal structural symmetry.

Finally, the opening of inner molecular gates (again on the level of side chains or helix bending) then enables the substrates to dissociate from the transporter protein into the cytosolic space. The transport cycle is then completed by switching from the empty internal form back to empty external conformation, where the protein is ready to start further cycles[15].

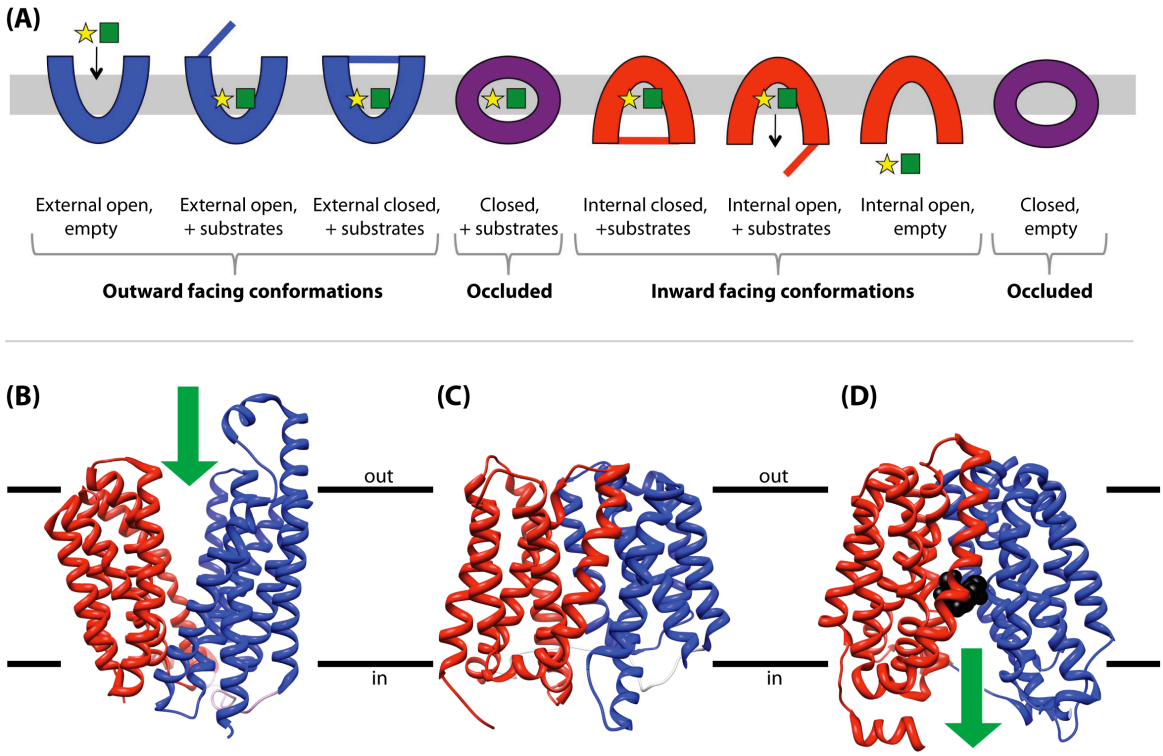


Figure 1.4 The principle of secondary symport by the 'alternating access' mechanism. (A) Secondary transporters exhibit three main conformations with different outward facing (blue), occluded (purple) and inward facing states (red). Substrate binding at one side of the membrane induces the closure of outer molecular gates, followed by a substantial conformational change leading to the inward facing conformation. Opening of inner molecular gates enables the substrates to be released. This cartoon illustrates secondary symport. Main- and co-substrates are shown as yellow stars and green rectangles. Adapted and modified from[15]. **(B)** V-shaped outward facing conformation of FucP (pdb 3O7Q). **(C)** Occluded state of EmrD (pdb 2GFP). **(D)** A-shaped inward phasing conformation of LacY (pdb 1PV7). The symmetry-related N- and C-terminal halves in blue and red, respectively, rock against each other. If available, substrate is shown as black spheres. Green arrows show substrate diffusion routes.

1.2.3.3 Driving forces and limiting steps in secondary transport

For a better understanding of the driving forces during secondary transport, the described cycle can also be treated as an 'enzymatic' process, while the chemical nature of the substrate is unaffected, but its location is changed. During translocation, the protein has to overcome several free energy barriers ΔG. Local energy minima are described for the empty protein and for the substrate bound forms with closed gates. Energy maxima were identified for the occluded states as well as for the substrate bound protein with open gates[15]. To ensure high transport rates, however, the energy profile has to be smoothened, which is achieved by the utilization of binding energy through binding of both, the main- and co-substrate to the empty symporter. Thereby, the substrate binding sites of the empty protein do not perfectly match the substrates, while they do in the closed occluded form as transition state. This 'induced transition fit' mechanism significantly lowers ΔG of the occluded transition state and enables the transporter to proceed in the translocation pathway[15,61].

Another driving force for substrate translocation is the electrochemical potential of the substrates, with the chemical potential $\Delta\mu$ (concentration gradient) and the electric potential ΔV (membrane potential) as the two components[12]. While the main substrate is transported against its chemical gradient, the co-substrate goes downhill its $\Delta\mu$. Lower chemical gradients of the co-substrate usually lead to a decreased transport rate reflected in higher K_M and K_D values of the main substrate[62,63]. The influence of the electric membrane potential is more complex. ΔV was found to have a substantial influence on K_M values and, especially, on the conformational change of the empty transporter from the inward to the outward open form[64]. More importantly, the release of the co-substrate (Na^+ in SGLT, H^+ in LacY) from the transporter is considerably accelerated by higher membrane potentials [15,65]. The structural and mechanistic details of these relations are, however, not fully understood.

Three possible parameters have been described to constitute the rate-limiting step during secondary transport. The intestinal Na^+/glucose symporter SGLT1 was found to be limited by the rate of the conformational change from the empty inward to the empty outward conformation and by the interdependent binding of both substrates[66]. A third possible limiting step is the dissociation rate of substrates from the protein, as described for the intracellular proton release from the lactose permease LacY[67]. Most probably, the rate limiting steps vary for different transport proteins and cannot be generalized[15].

1.2.3.4 Substrate binding and ion coupling

As a general rule, secondary transporters usually are specific for only one single or for several structural similar main substrates while translocation only occurs if the corresponding co-substrate is co-transported, either simultaneously as during symport or alternating as during antiport[15]. This has been experimentally proven for numerous transporters. Only one single main substrate is selectively transported e.g. by Glt_{Ph} (aspartate)[63], NCX (Ca^{2+})[13], GAT-2 (GABA)[68] and CitS (bivalent citrate)[69]. A less specific transport activity is found for e.g. *Vc*INDY transporting several di-/tricarboxylic intermediates of the Krebs cycle including citrate, succinate, malate and fumarate. $PepT_{So}$ and $PepT_{St}$ catalyze the uptake of a wide range of di- and tripeptides[43] and AcrB exports drugs of varying chemical nature[24]. In most cases, however, the

specificity for the co-ion is much higher. *Vc*INDY is effectively driven by Na^+ but to a much less extent by Li^+ and no activity is found for K^+[28]. A similar connection was found for numerous anti- and symporters including Glt_{Ph}[63], NCX[13] and CitS[70]. In summary, the overall specificity for the main substrate is usually lower compared to the co-ion. In fact, other co-ions may bind to the protein, but activity is abolished[71,72]. This behavior arises the following question: what is - besides energetic reason - the functional and structural implication of specifically co-transported co-ions? The molecular explanation can be found within atomic 3D structures of numerous transport proteins, e.g. *Vc*INDY and LeuT (Figure 1.5).

The crystal structure of *Vc*INDY could visualize one Na^+ ion bound to the protein. This high affinity sodium 'Na1' is mainly coordinated by the backbone oxygen of S150 and the side chain of N151 (Figure 1.5A). Separated by the inner helical hairpin, the bound citrate and Na1 do not directly interact with each other. However, both ligands partially share identical residues for coordination (S150/N151). The second putative sodium 'Na2' would sit at the corresponding binding site formed by the outer helical hairpin, while being closer to the cytoplasmic space[28].

In the Na^+ dependent leucine transporter LeuT both Na^+ sites 'Na1/2' were crystallographically identified (Figure 1.5B). In contrast to *Vc*INDY, the high affinity Na1 is not only coordinated by backbone and side chain hydrogen bonds but, remarkably, also directly interacts with the carbonyl oxygen of bound leucine as main substrate. Na2 is adjacent but separated from leucine by helix 1 and again closer to the cytoplasmic space as Na1[27,73].

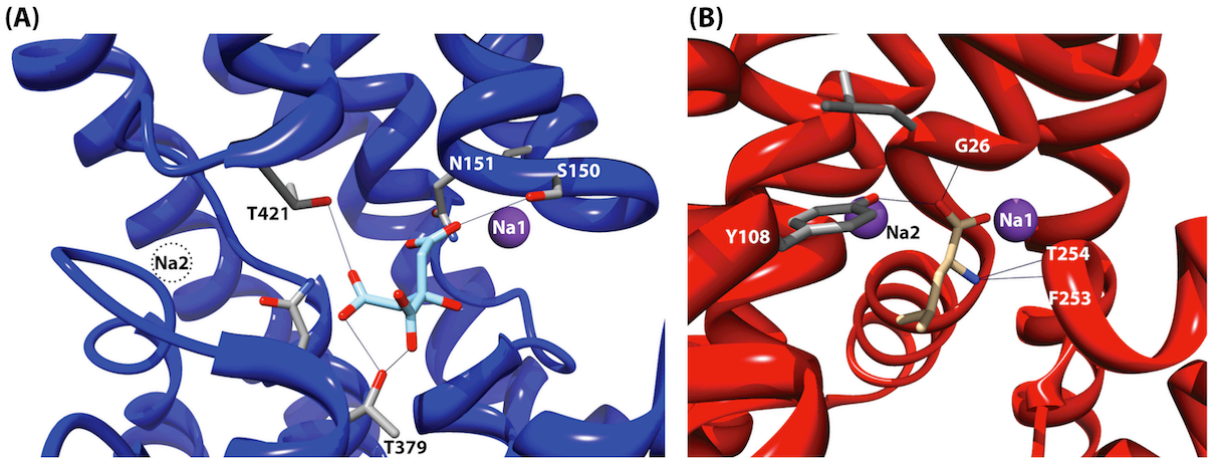


Figure 1.5 Molecular details of co-/substrate specificity and ion-coupling. (A) Binding pocket for citrate and Na^+ in *Vc*INDY (pdb 4F35). Citrate is coordinated by hydrogen bonds with three depicted residues, S150, T421 and T379. Na1 is adjacent but not in direct contact with citrate. Na2 is putative. **(B)** Binding pocket for leucine and Na^+ in LeuT (pdb 2AG5). Besides Y108, leucine is mainly coordinated by backbone hydrogen bonds and directly interacts with Na1. Na2 is separated from leucine/Na1 by helix 1. Important hydrogen bonds are depicted as black lines.

Among many others, these two examples provide powerful structural explanations for the highly specific coupled transport of Na^+, compared to e.g. K^+. In both cases Na1 is directly or at least indirectly involved in the binding of the main substrate. Na^+ thereby provides an additional framework for substrate coordination, mainly by electrostatic interactions and to a less extent by van der Waals coordination and desolvation effects[15]. In numerous reported cases Na^+ was found to first bind to the transporter, followed by the main substrate. All these aspects together suggest that Na^+ 'prepares' the binding pocket for the main substrate in a structural and

electrostatic way[15,27]. Interestingly, Na1 usually binds with higher affinity to the protein than Na2, underlining its importance in substrate binding. A closer inspection of the Na1 site in LeuT also explains the high specificity of Na^+ over K^+. Due to its larger ionic radius, K^+ simply would not fit into the binding site while resulting K^+/oxygen distances would be too small to effectively compensate ion dehydration[15]. Since K^+ does not bind, the binding site is not 'ready' for leucine and transport is abolished[27]. This analogon to ion channels is proposed to be a common feature of highly specific ion coupled secondary transport.

The reported cases also demonstrate functional implications of Na^+ co-transport. While Na1 mostly prepares the binding pocket for the main substrate, the role of Na2 is more complex. For numerous members of the LeuT fold, Na2 was found to bind exactly at the nexus of the two distinct protein domains, the bundle and hash (Figure 1.3E). As described, these domains move against each other to open and close the substrate binding site to either side of the membrane during alternating access. Importantly, Na2 is able to form electrostatic interactions with residues from both domains simultaneously. Based on these findings, Na2 is proposed to constitute the molecular linker that efficiently modulates these substantial conformational changes. Interestingly, the Na2 site is defect in distinct transitional conformational states, which also explains the lower affinity compared to Na1 and supports the model of lowered free energy states of particular conformations as the energetic basis of transport[15,27,73].

In summary, the requirement of specifically co-transported ions in secondary transporters is reflected in their in-/direct involvement in substrate coordination and their capability to mediate and control interactions between conformational related protein domains. Although the molecular details of these findings vary, the general concepts can be applied to most, if not all known secondary transporters.

1.2.4 CitS from *Klebsiella pneumoniae*

The main protein of interest for this thesis is the secondary citrate/Na^+ symporter CitS of *Klebsiella pneumoniae*. This transport protein will be introduced in detail in chapters 2 and 3.

1.3 Structural biology of membrane proteins

Despite their biological relevance, structural information on membrane proteins is still scarce - membrane proteins constitute < 1% among all proteins of known 3D structure. This is due to their amphipathic nature, which makes their handling difficult. Once a membrane protein is available in sufficient quality and amounts, there are several methods available for their structure determination, including X-ray diffraction (XRD), nuclear magnetic resonance spectroscopy (NMR) and (Cryo-)electron microscopy. XRD thereby provides the most powerful, sophisticated and productive technique. Every method has its very specific principles and requirements in point of sample preparation, data collection and data processing.

In X-ray diffraction, 3D crystals are produced by a strictly controlled precipitation of proteins in an aqueous environment. Usually, thousands of different crystallization conditions have to be screened to obtain 3D crystals with sufficient order and size. 3D crystals are then subjected to X-ray radiation that is diffracted by the protein lattice in the crystal, providing the X-ray diffraction pattern. However, diffraction patterns only contain the structure factor amplitudes while the corresponding phases are not readily accessible. Therefore, other strategies have to be applied, e.g. molecular replacement, anomalous X-ray scattering or heavy atom methods. Taken together, a complete dataset can be used to calculate the electron density map of a protein, which is then used to build an atomic protein model[74].

NMR provides another method for the structure determination of (membrane) proteins. Briefly, this spectroscopic technique uses the interaction of the magnetic dipole moment of nonzero spin nuclei with an applied electromagnetic field. NMR spectra are generated by first placing the protein sample in a strong magnetic field leading to an anti-/parallel alignment of the nonzero spin nuclei (e.g. ^{1}H, ^{13}C or ^{15}N). This is followed by a perturbation of the equilibrium (spin flip) by applying a radio frequency (RF) pulse. The RF that is absorbed by a particular nucleus strongly depends, among others, on the chemical environment, e.g. adjacent amino acids in a protein. The subsequent return to the thermal equilibrium state (relaxation) also contains useful information about the structure and dynamics of proteins[75].

Structural information about biological macromolecules can also be obtained by (Cryo) electron microscopy (EM). Cryo-EM can be divided into three sub-techniques: electron tomography, single particle analysis and electron crystallography. In tomography, the object of interest (e.g. whole cells) is imaged under several angles to provide 3D information with resolutions up to 2nm[76]. In single particle EM, a huge number of single protein particles are imaged in random orientations followed by a computational merging into a 3D reconstruction providing resolutions up to 2 Å[77]. The third EM related method, electron crystallography, is the main technique applied in this thesis. The basic principles will be discussed in detail in the following sections.

1.3.1 Electron crystallography

Electron crystallography is a powerful technique to determine the 2D (projection) and 3D structure of membrane proteins. The major benefit of studying 2D crystals by electron crystallography is that membrane proteins can be studied in a lipid bilayer as their native environment, usually keeping the protein in its functional form[78]. Thereby, structural information is extracted from 2D crystals by electron microscopic studies. Originally developed by Richard Henderson et al. in the mid 1970s, this technique provided the first 3D structure of an integral membrane protein by visualizing the seven α-helices of Bacteriorhodopsin[79]. Major innovations in crystallization, sample preparation, electron microscopy and image processing during the last decades led to a growing number of successfully solved structures of membrane proteins at or close to atomic resolution, e.g. several aquaporins and the acetylcholine receptor[78]. Additionally, there are numerous structures of soluble proteins reported from electron crystallographic studies[80].

The major steps of the electron crystallographic workflow include (1) protein expression and purification, (2) two-dimensional crystallization, (3) sample preparation, (4) data collection, (5) image processing and (6) model building. Besides the initial expression and purification, all these crucial steps are described in the following sections.

1.3.1.1 Two-dimensional (2D) crystallization of membrane proteins

In the context of membrane proteins, 2D crystals are defined as a highly ordered two-dimensional protein array embedded in a lipid bilayer[23]. Using a stricter definition, protein-protein interactions within the crystal only appear in the x- and y-dimension within the membrane plane, while z-interactions are missing. However, numerous examples are reported for multilayered crystals that are still defined as 2D crystals, e.g. double-layered Aquaporin0[81] and multilayered II^{mtl} crystals[82]. There are several naturally occurring 2D crystals available, e.g. the light-driven proton-pump Bacteriorhodopsin found in purple membrane patches of *Halobacterium salinarum*[83] and connexins that form gap junctions in mammals[84]. Nevertheless, most reported 2D crystals used for structural studies have to be produced artificially from recombinantly expressed proteins[23].

The production of 2D crystals basically involves three bio-/chemical components: (1) the membrane protein, (2) bilayer forming phospholipids and (3) detergent molecules (Figure 1.6). In the initial phase of crystallization, the detergent solubilized membrane protein is mixed with detergent solubilized bilayer forming lipids at a defined lipid/protein ratio (LPR). Depending on the chosen lipid/detergent ratio (LDR), the incubation of this ternary mixture leads to the formation of detergent micelles, mixed detergent/lipid micelles and/or bicelles that are all capable of stabilizing the membrane protein in an aqueous environment and, simultaneously, avoid protein aggregation[1,23].

The formation of 2D crystals is induced by specifically removing the detergent molecules from the ternary mixture. This can be achieved by dialysis[85], Biobeads[86], Cyclodextrin[87] and controlled dilution[88]. The most relevant events during crystal formation are the formation of phospholipid bilayers, protein insertion into these and crystallization. In principle, three different crystallization mechanisms are proposed: (1) one-step crystallization, (2) direct

stepwise and (3) indirect stepwise crystallization[89]. In a one-step process, formation of lipid bilayers, protein insertion and crystallization occur simultaneously during detergent removal[90]. During direct stepwise crystallization, concerted bilayer formation and protein insertion is followed by crystallization[91]. During indirect stepwise crystallization, all three processes are separate events: proteins insert into preformed lipid bilayers followed by crystallization[92]. Since evidence is available for all described models, it is assumed that the crystallization mechanism is mainly triggered by intrinsic features of the membrane protein itself and by experimental conditions and can therefore not be generalized[89].

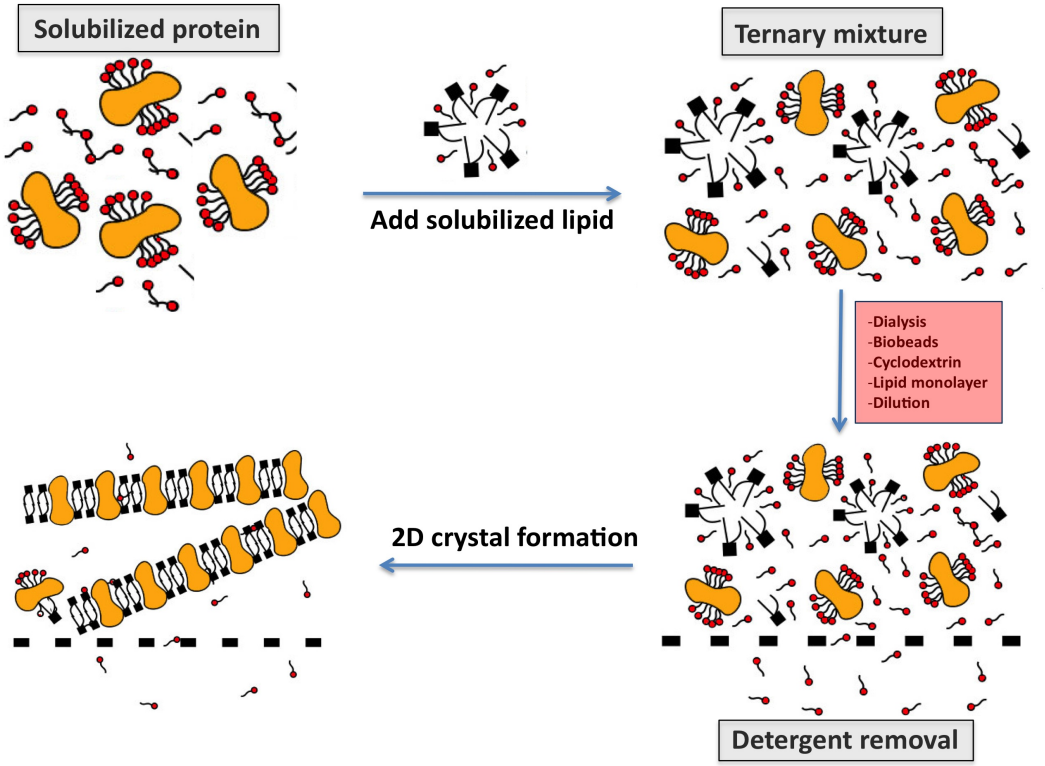


Figure 1.6 The basic principle of 2D crystallization of membrane proteins. The starting point is the pure target protein (orange) solubilized in detergent (red). A suitable detergent solubilized bilayer forming lipid (black) is added to the protein to yield the ternary mixture. The selective detergent removal from the mixture finally can induce the formation of a lipid bilayer that harbors a crystalline array of the membrane protein.

There are numerous parameters that effectively have great impact on the formation, shape, size and order of 2D crystals. These include protein parameters (stability, conformation, concentration and homogeneity), buffer conditions (pH, ionic strength, mono-/bivalent cations and buffer substance), phospholipids (head group, acyl chain length, saturation, lipid/protein ratio and cholesterol), detergents (non-/zwitter-/ionic, headgroup, acyl chain length, critical micellar concentration (CMC) and concentration), temperature (protein stability, diffusion rate, lipid configuration), detergent removal (technique and speed) and additives (inhibitors, ligands and reducing agents)[23]. In general, the effect of these parameters cannot be predicted, making a systematic screen of numerous different crystallization conditions inevitable. A promising strategy, however, is to find initial parameters that allow the reconstitution of a membrane protein into lipid bilayers, followed by the fine-tuning of these conditions to obtain highly ordered 2D crystals. Typical morphologies of the resulting 2D crystals include planar sheets, vesicles, tubular vesicles, tubes and helical tubes[23,93].

1.3.1.2 Cryo-EM: Sample preparation & Data collection

Once highly ordered 2D crystals of membrane proteins are available, these have to be subjected to a suitable sample preparation method that provides the structural and functional integrity of the membrane protein and that enables high-resolution data collection by electron microscopy. Therefore, two main hurdles have to be overcome: (1) dehydration of biological specimen in an electron microscope's high vacuum would lead to a collapse of its native structures and (2) biological specimen mainly consist of light atoms that are, compared to heavy atoms, much more prone to beam damage by inelastic electron scattering[94]. Strikingly, both resolution-limiting processes could be minimized by the development of cryo-EM.

In principle, two different sample preparations methods can be applied to biological specimen such as 2D crystals for cryo-EM: vitrification and sugar embedding. During vitrification, the fully hydrated sample on the EM grid is rapidly frozen in liquid ethane at liquid nitrogen temperature[95,96]. Thereby, the containing water vitrifies, i.e. solidifies without forming destructive ice-crystals. During sugar embedding, water molecules are replaced by less volatile compounds such as glucose[79], trehalose[97] and tannin[90] before the sample is frozen in liquid nitrogen. This technique was further improved by introducing back-injection and carbon sandwich methods[98].

For both cryo-EM variants, the subsequent data collection in the electron microscope is also performed at cryogenic conditions at liquid nitrogen or even liquid helium temperature[99]. This significantly reduces the impact of beam damage and preserves the sample in the vitrified state. Additionally, beam damage can be effectively minimized by data acquisition in the low electron-dose mode[100]. A reduction of beam-induced sample movements[101] and charging effects[102] by using the spot-scanning mode[103] can further improve the image quality. High-resolution data collection from 2D crystals can be performed in two different ways, direct imaging and electron diffraction[98]. Imaging even can be used for small crystalline patches (<1 μm) and readily provides amplitudes and phases. On the other hand, electron diffraction is restricted to large crystals and provides only amplitude information. Both, images and diffraction data can be recorded digitally or on photographic film[78].

1.3.1.3 Image processing: 2D & 3D data

The weak electron scattering propensity of light atoms in biological macromolecules only allows recording of electron micrographs with very low contrast and signal-to-noise ratios. Hence, intense image processing is inevitable to extract high-resolution structural data from these. The fundamental developments for processing electron crystallographic datasets were done by Henderson et al. at the MRC laboratories during their work on Bacteriorhodopsin[79,104]. An example for a more recent software solution is '2dx', which is still based on the MRC algorithms that were further optimized and integrated into a graphical user-interface[105,106]. The basic principle of processing single electron micrographs of 2D crystals in 2dx is illustrated in figure 1.7. The computed fast Fourier Transform (FFT) of the raw image is used to determine the repeating structure within the crystal, which spans a lattice in reciprocal space and provides the unit cell dimensions. In the power spectrum of the FFT the defocus value of each image can be identified by the detection of Thon rings. These initial steps are followed by crystal unbending, which is used to correct for crystal disorders introduced by crystal defects, imaging and

mechanical stress during sample preparation. Therefore, a small crystalline reference patch is used to localize crystal disorders by means of cross correlation, which can then be shifted, i.e. unbent, to the location as computationally predicted. In the resulting unbent image, high-resolution information is usually recovered and significantly enhanced[107]. In a next step, images have to be corrected for resolution dependent contrast reversals introduced by the contrast transfer function (CTF), a complex electron microscopic phenomenon[108]. During CTF correction, different resolution components have to be corrected for amplitude oscillations. The phases are then corrected according to their symmetry constraints. The final resulting amplitudes and phases can be back transformed to generate a 2D real-space projection structure of each processed image[109].

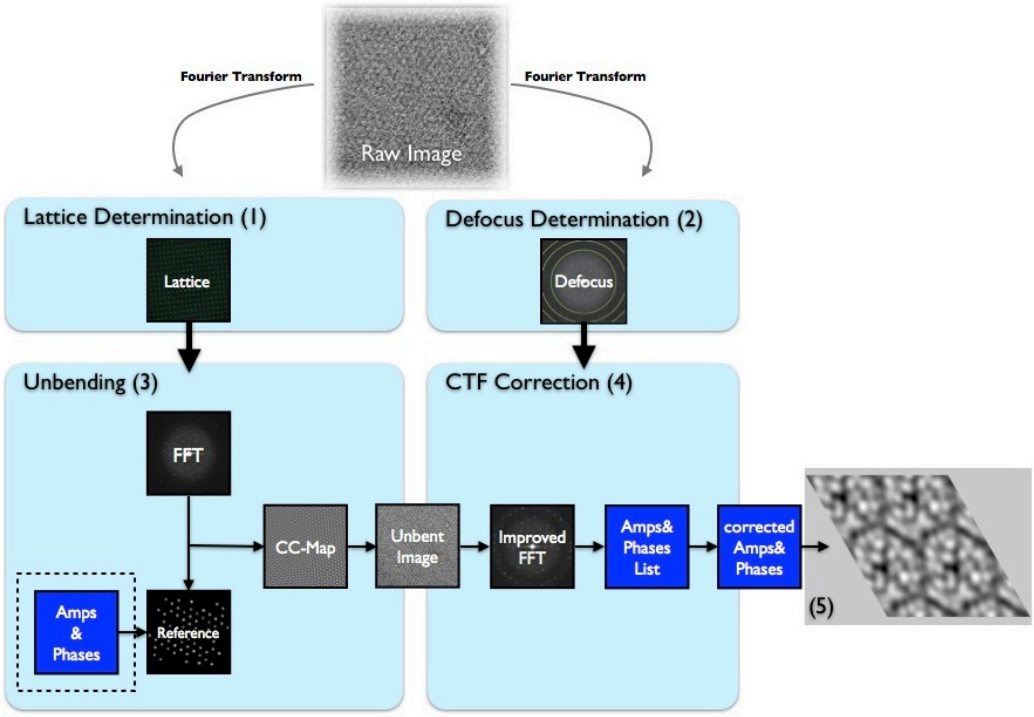


Figure 1.7 The basic principle of processing single images in 2dx. The main steps include the determination of (1) crystallographic lattice parameters and (2) the defocus value, both based on the Fourier Transform of the image. This is followed by (3) cross-correlative unbending to correct for crystal disorders. The final steps are (4) CTF correction and (5) map generation. Taken and modified from[109].

To obtain 3D information out of 2D crystals, projection structures at different tilt angles of the sample (up to 70°) in the microscope have to be recorded. Based on the central section theorem, these are then merged into a 3D dataset in Fourier space[110]. In principle, image processing of tilted samples is identical to untilted ones. As additional step the tilt geometry for each crystal, including tilt angle and axis, has to be determined. This can be achieved by measuring distorsions of the unit cell parameters, by quantifying defocus variations across the image, by the comparison of single images with the back-projection of a preliminary 3D volume and by the quantification of diffraction spot-splitting[111,112]. Amplitudes and phases in the z-dimension are continuously sampled along corresponding lattice lines[112]. The final resulting 3D volume usually is anisotropically resolved with worse data in the z-dimension perpendicular to the membrane plane. This is caused by uneven crystal samples, beam-induced specimen movement and mechanical limitations in sample tilting ('missing cone'). For these restraints, however, several promising solutions are available or under development.

1.3.1.4 Electron crystallography applied to secondary transporters

Electron crystallography is a proven technique for the structural characterization of secondary transporters. It has been successfully used to determine numerous projections and several 3D structures (see table 1.1)[22,23]. The most prominent examples include the Na^+/H^+ antiporters NhaA (7/15 Å)[19,47] and NhaP1 (6/15 Å)[113], the Na^+/betain symporter BetP (8 Å)[35], the formate/oxalate antiporter OxlT (6.5 Å)[114], the H^+/multidrug antiporter EmrE (7 Å)[54,55] and, recently, the Zn^{2+}/H^+ antiporter YiiP[52]. Besides these static structures, electron crystallography also provided detailed insights into conformational changes in secondary transporters. Substrate induced movements of single α-helices could be observed for EmrE[54], NhaA[47] and NhaP1[115]. In these studies, 2D crystals were extensively soaked in buffers of different biochemical composition prior to the sample preparation for cryo-EM. Due to missing protein-protein interactions in the z-dimension, 2D crystals usually tolerate slight conformational changes without breaking crystal contacts. This 'post-crystal modification' approach also proofs the accessibility of 2D-crystallized membrane proteins in the lipid bilayer and makes the challenging crystal growth at different conditions evitable[22].

1.4 Aims and structure of this thesis

Besides biochemical and biophysical data, structural information remains the main input source and bottleneck for the understanding of integral membrane proteins, e.g. secondary transporters and G protein-coupled receptors (GPCRs). Three-dimensional structures at atomic resolution are essential to answer the numerous remaining open questions in this field.

One final goal of this thesis was to determine the 3D structure of the secondary Na^+/citrate symporter CitS from *Klebsiella pneumoniae*. This transport protein constitutes an attractive target for structural studies since it is predicted to represent a novel, unique fold within the family of secondary transporters[116,117]. As a close collaboration between the groups of Prof. Henning Stahlberg (University of Basel) and Prof. Markus Grütter (University of Zürich), the 3D structure would be solved using a combination of electron crystallography and X-ray diffraction.

Since expression and purification of CitS have already been established prior to this thesis[118], the first step includes the production of high-quality 2D crystals of the transporter in an artificial lipid bilayer. This can be achieved by a strictly controlled detergent removal from highly pure, detergent solubilized CitS in the presence of lipids. This step usually requires the screening of numerous physico-chemical parameters that affect the formation, size and order of 2D crystals. Once high-quality 2D crystals of CitS are obtained, a suitable and reproducible sample preparation method has to be established. This method has to maintain the structural and functional integrity of CitS within the 2D crystals. Cryo-electron microscopy and subsequent image processing then could provide structural information at sub-nanometer resolution. A resulting 2D projection structure is expected to give first insights into the global architecture of membrane embedded CitS. This part of the thesis is described in chapter 2, which also includes a detailed introduction to CitS.

A further goal of this thesis is to obtain the first 3D structure of CitS by recording cryo-electron micrographs of tilted 2D crystal samples. Using this approach, the identification and localization of transmembrane segments and their assignment should be possible and would lead to a detailed molecular model. Additionally, the 3D data could be complemented by projection structures in different experimental conditions (presence/absence of Na^+ and citrate as substrates). This is expected to give a detailed insight into different conformations of the transporter as it is predicted from available models. This part of the thesis is described in chapter 3.

Finally, the moderately resolved 3D volume obtained by Cryo-EM could help to solve the 3D structure at atomic resolution. Therefore, the EM data could act as search model to perform molecular replacement with the existing X-ray diffraction data set, where no phases could be obtained so far. In the end, the resulting 3D structure would significantly contribute in understanding CitS, and secondary transporters in general, on a molecular level.

In chapter 4, the structural and biophysical characterization of the HIV related G protein-coupled receptor CCR5 is described. In collaboration with the group of Prof. Stephan Grzesiek (Biozentrum, University of Basel), the goal of this project was the bacterial overexpression of CCR5 for the subsequent structure determination by spectroscopic and/or crystallographic approaches. As side project of this thesis, the corresponding chapter 4 is handled independently regarding its introduction, applied techniques and results and is not referred to chapter 1.

1.5 References

1. Luckey, M. (2008). ***Membrane Structural Biology***, Cambridge University Press.
2. Engel, A. & Gaub, H. E. (2008). **Structure and mechanics of membrane proteins**. *Annu Rev Biochem* **77**, 127-148.
3. Hedin, L. E., Illergard, K. & Elofsson, A. (2011). **An introduction to membrane proteins**. *J Proteome Res* **10**, 3324-3331.
4. Saier, M. H., Jr. (2000). **A functional-phylogenetic classification system for transmembrane solute transporters**. *Microbiol Mol Biol Rev* **64**, 354-411.
5. Gouaux, E. & Mackinnon, R. (2005). **Principles of selective ion transport in channels and pumps**. *Science* **310**, 1461-1465.
6. Fairman, J. W., Noinaj, N. & Buchanan, S. K. (2011). **The structural biology of beta-barrel membrane proteins: a summary of recent reports**. *Curr Opin Struct Biol* **21**, 523-531.
7. Zeth, K. & Thein, M. (2010). **Porins in prokaryotes and eukaryotes: common themes and variations**. *Biochem J* **431**, 13-22.
8. Morth, J. P., Pedersen, B. P., Buch-Pedersen, M. J., Andersen, J. P., Vilsen, B., Palmgren, M. G. & Nissen, P. (2011). **A structural overview of the plasma membrane Na^+,K^+-ATPase and H^+-ATPase ion pumps**. *Nat Rev Mol Cell Biol* **12**, 60-70.
9. Hinz, A. & Tampe, R. (2012). **ABC transporters and immunity: mechanism of self-defense**. *Biochemistry* **51**, 4981-4989.
10. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. (2009). **Unlocking the molecular secrets of sodium-coupled transporters**. *Nature* **459**, 347-355.
11. Sobczak, I. & Lolkema, J. S. (2005). **Structural and mechanistic diversity of secondary transporters**. *Curr Opin Microbiol* **8**, 161-167.
12. Abramson, J. & Wright, E. M. (2009). **Structure and function of Na(+)-symporters with inverted repeats**. *Curr Opin Struct Biol* **19**, 425-432.
13. Liao, J., Li, H., Zeng, W., Sauer, D. B., Belmares, R. & Jiang, Y. (2012). **Structural insight into the ion-exchange mechanism of the sodium/calcium exchanger**. *Science* **335**, 686-690.
14. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S. (2003). **Structure and mechanism of the lactose permease of *Escherichia coli***. *Science* **301**, 610-615.
15. Forrest, L. R., Krämer, R. & Ziegler, C. (2011). **The structural basis of secondary active transport mechanisms**. *Biochim Biophys Acta* **1807**, 167-188.
16. Busch, W. & Saier, M. H., Jr. (2002). **The transporter classification (TC) system, 2002**. *Crit Rev Biochem Mol Biol* **37**, 287-337.
17. Torres, G. E., Gainetdinov, R. R. & Caron, M. G. (2003). **Plasma membrane monoamine transporters: structure, regulation and function**. *Nat Rev Neurosci* **4**, 13-25.
18. Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G., Deng, D., Dang, S., Zhang, X., Wang, J. & Yan, N. (2011). **Structure and mechanism of the uracil transporter UraA**. *Nature* **472**, 243-246.
19. Williams, K. A. (2000). **Three-dimensional structure of the ion-coupled transport protein NhaA**. *Nature* **403**, 112-115.
20. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. (2002). **Crystal structure of bacterial multidrug efflux transporter AcrB**. *Nature* **419**, 587-593.
21. Stephen White Laboratory, U. I. (2013). **Membrane proteins of known 3D structure** *http://blanco.biomol.uci.edu/mpstruc*.
22. Tsai, C. J. & Ziegler, C. (2010). **Coupling electron cryomicroscopy and X-ray crystallography to understand secondary active transport**. *Curr Opin Struct Biol* **20**, 448-455.
23. Abeyrathne, P., Arheit, M., Kebbel, F., Castano-Diez, D., Goldie, KN., Chami, M., Renault, L., Kühlbrandt, W. & Stahlberg, H. . (2012). **Electron Microscopy Analysis of 2D Crystals of Membrane Proteins**. 1 edit. Comprehensive Biophysics (Egelman, E., Ed.), 1.19, Academic Press.

24. Seeger, M. A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K. & Pos, K. M. (2006). **Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism**. *Science* **313**, 1295-1298.
25. Hunte, C., Screpanti, E., Venturi, M., Rimon, A., Padan, E. & Michel, H. (2005). **Structure of a Na^+/H^+ antiporter and insights into mechanism of action and regulation by pH**. *Nature* **435**, 1197-1202.
26. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. (2004). **Structure of a glutamate transporter homologue from *Pyrococcus horikoshii***. *Nature* **431**, 811-818.
27. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. (2005). **Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters**. *Nature* **437**, 215-223.
28. Mancusso, R., Gregorio, G. G., Liu, Q. & Wang, D. N. (2012). **Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter**. *Nature*.
29. Perez, C., Khafizov, K., Forrest, L. R., Krämer, R. & Ziegler, C. (2011). **The role of trimerization in the osmoregulated betaine transporter BetP**. *EMBO Rep* **12**, 804-810.
30. Herz, K., Rimon, A., Jeschke, G. & Padan, E. (2009). **Beta-sheet-dependent dimerization is essential for the stability of NhaA Na^+/H^+ antiporter**. *J Biol Chem* **284**, 6337-6347.
31. Chen, Y. J., Pornillos, O., Lieu, S., Ma, C., Chen, A. P. & Chang, G. (2007). **X-ray structure of EmrE supports dual topology model**. *Proc Natl Acad Sci U S A* **104**, 18999-19004.
32. Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., Wright, E. M. & Abramson, J. (2008). **The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport**. *Science* **321**, 810-814.
33. Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E. P., Rutherford, N. G., Hadden, J. M., O'Reilly, J., Ma, P., Saidijam, M., Patching, S. G., Hope, R. J., Norbertczak, H. T., Roach, P. C., Iwata, S., Henderson, P. J. & Cameron, A. D. (2008). **Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter**. *Science* **322**, 709-713.
34. Ressl, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. (2009). **Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP**. *Nature* **458**, 47-52.
35. Tsai, C. J., Khafizov, K., Hakulinen, J., Forrest, L. R., Krämer, R., Kühlbrandt, W. & Ziegler, C. (2011). **Structural asymmetry in a trimeric Na^+/betaine symporter, BetP, from *Corynebacterium glutamicum***. *J Mol Biol* **407**, 368-381.
36. Fang, Y., Jayaram, H., Shane, T., Kolmakova-Partensky, L., Wu, F., Williams, C., Xiong, Y. & Miller, C. (2009). **Structure of a prokaryotic virtual proton pump at 3.2 Å resolution**. *Nature* **460**, 1040-1043.
37. Shaffer, P. L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. (2009). **Structure and mechanism of a Na^+-independent amino acid transporter**. *Science* **325**, 1010-1014.
38. Schulze, S., Koster, S., Geldmacher, U., Terwisscha van Scheltinga, A. C. & Kühlbrandt, W. (2010). **Structural basis of Na(+)-independent and cooperative substrate/product antiport in CaiT**. *Nature* **467**, 233-236.
39. Ma, D., Lu, P., Yan, C., Fan, C., Yin, P., Wang, J. & Shi, Y. (2012). **Structure and mechanism of a glutamate-GABA antiporter**. *Nature* **483**, 632-636.
40. Yin, Y., He, X., Szewczyk, P., Nguyen, T. & Chang, G. (2006). **Structure of the multidrug transporter EmrD from *Escherichia coli***. *Science* **312**, 741-744.
41. Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., Wang, J. & Yan, N. (2010). **Structure of a fucose transporter in an outward-open conformation**. *Nature* **467**, 734-738.
42. Newstead, S., Drew, D., Cameron, A. D., Postis, V. L., Xia, X., Fowler, P. W., Ingram, J. C., Carpenter, E. P., Sansom, M. S., McPherson, M. J., Baldwin, S. A. & Iwata, S. (2011). **Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2**. *EMBO J* **30**, 417-426.
43. Solcan, N., Kwok, J., Fowler, P. W., Cameron, A. D., Drew, D., Iwata, S. & Newstead, S. (2012). **Alternating access mechanism in the POT family of oligopeptide transporters**. *EMBO J* **31**, 3411-3421.

44. Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y. & Yan, N. (2012). **Crystal structure of a bacterial homologue of glucose transporters GLUT1-4**. *Nature* **490**, 361-366.
45. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. (2003). **Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli***. *Science* **301**, 616-620.
46. Pedersen, B. P., Kumar, H., Waight, A. B., Risenmay, A. J., Roe-Zurz, Z., Chau, B. H., Schlessinger, A., Bonomi, M., Harries, W., Sali, A., Johri, A. K. & Stroud, R. M. (2013). **Crystal structure of a eukaryotic phosphate transporter**. *Nature* **496**, 533-536.
47. Appel, M., Hizlan, D., Vinothkumar, K. R., Ziegler, C. & Kühlbrandt, W. (2009). **Conformations of NhaA, the Na^+/H^+ exchanger from *Escherichia coli*, in the pH-activated and ion-translocating states**. *J Mol Biol* **388**, 659-672.
48. Hu, N. J., Iwata, S., Cameron, A. D. & Drew, D. (2011). **Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT**. *Nature* **478**, 408-411.
49. Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trezeguet, V., Lauquin, G. J. & Brandolin, G. (2003). **Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside**. *Nature* **426**, 39-44.
50. Johnson, Z. L., Cheong, C. G. & Lee, S. Y. (2012). **Crystal structure of a concentrative nucleoside transporter from *Vibrio cholerae* at 2.4 Å**. *Nature* **483**, 489-493.
51. Lu, M., Chai, J. & Fu, D. (2009). **Structural basis for autoregulation of the zinc transporter YiiP**. *Nat Struct Mol Biol* **16**, 1063-1067.
52. Coudray, N., Valvo, S., Hu, M., Lasala, R., Kim, C., Vink, M., Zhou, M., Provasi, D., Filizola, M., Tao, J., Fang, J., Penczek, P. A., Ubarretxena-Belandia, I. & Stokes, D. L. (2013). **Inward-facing conformation of the zinc transporter YiiP revealed by cryoelectron microscopy**. *Proc Natl Acad Sci U S A* **110**, 2140-2145.
53. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. (2002). **X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity**. *Nature* **415**, 287-294.
54. Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S. & Tate, C. G. (2003). **Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer**. *EMBO J* **22**, 6175-6181.
55. Fleishman, S. J., Harrington, S. E., Enosh, A., Halperin, D., Tate, C. G. & Ben-Tal, N. (2006). **Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain**. *J Mol Biol* **364**, 54-67.
56. Tanaka, Y., Hipolito, C. J., Maturana, A. D., Ito, K., Kuroda, T., Higuchi, T., Katoh, T., Kato, H. E., Hattori, M., Kumazaki, K., Tsukazaki, T., Ishitani, R., Suga, H. & Nureki, O. (2013). **Structural basis for the drug extrusion mechanism by a MATE multidrug transporter**. *Nature* **496**, 247-251.
57. Pao, S. S., Paulsen, I. T. & Saier, M. H., Jr. (1998). **Major facilitator superfamily**. *Microbiol Mol Biol Rev* **62**, 1-34.
58. Jardetzky, O. (1966). **Simple allosteric model for membrane pumps**. *Nature* **211**, 969-970.
59. Forrest, L. R. & Rudnick, G. (2009). **The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters**. *Physiology (Bethesda)* **24**, 377-386.
60. Reyes, N., Ginter, C. & Boudker, O. (2009). **Transport mechanism of a bacterial homologue of glutamate transporters**. *Nature* **462**, 880-885.
61. Jencks, W. P. (1989). **Utilization of binding energy and coupling rules for active transport and other coupled vectorial processes**. *Methods Enzymol* **171**, 145-164.
62. Robertson, D. E., Kaczorowski, G. J., Garcia, M. L. & Kaback, H. R. (1980). **Active transport in membrane vesicles from *Escherichia coli*: the electrochemical proton gradient alters the distribution of the lac carrier between two different kinetic states**. *Biochemistry* **19**, 5692-5702.
63. Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K. & Gouaux, E. (2007). **Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter**. *Nature* **445**, 387-393.

64. Parent, L., Supplisson, S., Loo, D. D. & Wright, E. M. (1992). **Electrogenic properties of the cloned Na+/glucose cotransporter: I. Voltage-clamp studies**. *J Membr Biol* **125**, 49-62.
65. Garcia-Celma, J. J., Ploch, J., Smirnova, I., Kaback, H. R. & Fendler, K. (2010). **Delineating electrogenic reactions during lactose/H+ symport**. *Biochemistry* **49**, 6115-6121.
66. Wright, E. M. (1993). **The intestinal Na+/glucose cotransporter**. *Annu Rev Physiol* **55**, 575-589.
67. King, S. C. & Wilson, T. H. (1990). **Towards an understanding of the structural basis of 'forbidden' transport pathways in the *Escherichia coli* lactose carrier: mutations probing the energy barriers to uncoupled transport**. *Mol Microbiol* **4**, 1433-1438.
68. Schlessinger, A., Wittwer, M. B., Dahlin, A., Khuri, N., Bonomi, M., Fan, H., Giacomini, K. M. & Sali, A. (2012). **High Selectivity of the gamma-Aminobutyric Acid Transporter 2 (GAT-2, SLC6A13) Revealed by Structure-based Approach**. *J Biol Chem* **287**, 37745-37756.
69. Pos, K. M. & Dimroth, P. (1996). **Functional properties of the purified Na(+)-dependent citrate carrier of *Klebsiella pneumoniae*: evidence for asymmetric orientation of the carrier protein in proteoliposomes**. *Biochemistry* **35**, 1018-1026.
70. Lolkema, J. S., Enequist, H. & van der Rest, M. E. (1994). **Transport of citrate catalyzed by the sodium-dependent citrate carrier of *Klebsiella pneumoniae* is obligatorily coupled to the transport of two sodium ions**. *Eur J Biochem* **220**, 469-475.
71. Tao, Z., Gameiro, A. & Grewer, C. (2008). **Thallium ions can replace both sodium and potassium ions in the glutamate transporter excitatory amino acid carrier 1**. *Biochemistry* **47**, 12923-12930.
72. Perez, C., Koshy, C., Ressl, S., Nicklisch, S., Krämer, R. & Ziegler, C. (2011). **Substrate specificity and ion coupling in the Na+/betaine symporter BetP**. *EMBO J* **30**, 1221-1229.
73. Krishnamurthy, H. & Gouaux, E. (2012). **X-ray structures of LeuT in substrate-free outward-open and apo inward-open states**. *Nature* **481**, 469-474.
74. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. (2008). **Overcoming the challenges of membrane protein crystallography**. *Curr Opin Struct Biol* **18**, 581-586.
75. Hong, M., Zhang, Y. & Hu, F. (2012). **Membrane protein structure and dynamics from NMR spectroscopy**. *Annu Rev Phys Chem* **63**, 1-24.
76. Kudryashev, M., Castano-Diez, D.,Stahlberg, H. (2012). **Limiting Factors In Single Particle Cryo Electron Tomography**. *Comput Struct Biotechnol J* **1**, e201207002.
77. Grigorieff, N. & Harrison, S. C. (2011). **Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy**. *Curr Opin Struct Biol* **21**, 265-273.
78. Wisedchaisri, G., Reichow, S. L. & Gonen, T. (2011). **Advances in structural and functional analysis of membrane proteins by electron crystallography**. *Structure* **19**, 1381-1393.
79. Henderson, R. & Unwin, P. N. (1975). **Three-dimensional model of purple membrane obtained by electron microscopy**. *Nature* **257**, 28-32.
80. Ellis, M. J. & Hebert, H. (2001). **Structure analysis of soluble proteins using electron crystallography**. *Micron* **32**, 541-550.
81. Gonen, T., Sliz, P., Kistler, J., Cheng, Y. & Walz, T. (2004). **Aquaporin-0 membrane junctions reveal the structure of a closed water pore**. *Nature* **429**, 193-197.
82. Stuart, M. C., Koning, R. I., Oostergetel, G. T. & Brisson, A. (2004). **Mechanism of formation of multilayered 2D crystals of the enzyme IIC-mannitol transporter**. *Biochim Biophys Acta* **1663**, 108-116.
83. Gualtieri, E. J., Guo, F., Kissick, D. J., Jose, J., Kuhn, R. J., Jiang, W. & Simpson, G. J. (2011). **Detection of membrane protein two-dimensional crystals in living cells**. *Biophys J* **100**, 207-214.
84. Unger, V. M., Kumar, N. M., Gilula, N. B. & Yeager, M. (1997). **Projection structure of a gap junction membrane channel at 7 Å resolution**. *Nat Struct Biol* **4**, 39-43.
85. Jap, B. K., Zulauf, M., Scheybani, T., Hefti, A., Baumeister, W., Aebi, U. & Engel, A. (1992). **2D crystallization: from art to science**. *Ultramicrosc* **46**, 45-84.

86. Rigaud, J. L., Mosser, G., Lacapere, J. J., Olofsson, A., Levy, D. & Ranck, J. L. (1997). **Bio-Beads: an efficient strategy for two-dimensional crystallization of membrane proteins**. *J Struct Biol* **118**, 226-235.
87. Signorell, G. A., Kaufmann, T. C., Kukulski, W., Engel, A. & Remigy, H. W. (2007). **Controlled 2D crystallization of membrane proteins using methyl-beta-cyclodextrin**. *J Struct Biol* **157**, 321-328.
88. Remigy, H. W., Caujolle-Bert, D., Suda, K., Schenk, A., Chami, M. & Engel, A. (2003). **Membrane protein reconstitution and crystallization by controlled dilution**. *FEBS Lett* **555**, 160-169.
89. Rigaud, J., Chami, M., Lambert, O., Levy, D. & Ranck, J. (2000). **Use of detergents in two-dimensional crystallization of membrane proteins**. *Biochim Biophys Acta* **1508**, 112-128.
90. Kühlbrandt, W. & Wang, D. N. (1991). **Three-dimensional structure of plant light-harvesting complex determined by electron crystallography**. *Nature* **350**, 130-134.
91. Dolder, M., Engel, A. & Zulauf, M. (1996). **The micelle to vesicle transition of lipids and detergents in the presence of a membrane protein: towards a rationale for 2D crystallization**. *FEBS Lett* **382**, 203-208.
92. Schmidt-Krey, I., Lundqvist, G., Morgenstern, R. & Hebert, H. (1998). **Parameters for the two-dimensional crystallization of the membrane protein microsomal glutathione transferase**. *J Struct Biol* **123**, 87-96.
93. Unwin, N. (2005). **Refined structure of the nicotinic acetylcholine receptor at 4Å resolution**. *J Mol Biol* **346**, 967-989.
94. Reimer L, R.-M. M. (1990). **Contrast in the electron spectroscopic imaging mode of a TEM.II.Z-ratio, structure sensitive and phase-contrast**. *J Microscopy* **159**, 143-160.
95. Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowall, A. W. & Schultz, P. (1988). **Cryo-electron microscopy of vitrified specimens**. *Q Rev Biophys* **21**, 129-228.
96. Dubochet J, C. J., Freeman R, Lepault J, McDowall AW. (1982). **Frozen aqueous suspensions**. *Ultramicroscopy* **10**, 55-61.
97. Chiu, P. L., Kelly, D. F. & Walz, T. (2011). **The use of trehalose in the preparation of specimens for molecular electron microscopy**. *Micron* **42**, 762-772.
98. Hite, R. K., Schenk, A. D., Li, Z., Cheng, Y. & Walz, T. (2010). **Collecting electron crystallographic data of two-dimensional protein crystals**. *Methods Enzymol* **481**, 251-282.
99. Fujiyoshi, Y., Mizusaki T, Morikawa K, Yamagishi H, Aoki Y, Kihara H, Harada Y. (1991). **Development of a superfluid helium stage for high-resolution electron microscopy**. *Ultramicroscopy* **38**, 241-251.
100. Kuo AM & RM, G. (1975). **Development of methodology for low exposure high resolution electron microscopy of biological specimen**. *Ultramicroscopy* **1**, 53.
101. Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., Henderson, R. & Grigorieff, N. (2012). **Beam-induced motion of vitrified specimen on holey carbon film**. *J Struct Biol* **177**, 630-637.
102. Hirai, T., Mitsuoka, K., Kidera, A. & Fujiyoshi, Y. (2007). **Simulation of charge effects on density maps obtained by high-resolution electron crystallography**. *J Electron Microsc (Tokyo)* **56**, 131-140.
103. Henderson, R. & Glaeser, R. M. (1985). **Quantitaive analysis of image contrast in electron-micrographs of beam-sensitive crystals**. *Ultramicroscopy* **16**, 139-150.
104. Crowther, R. A., Henderson, R. & Smith, J. M. (1996). **MRC image processing programs**. *J Struct Biol* **116**, 9-16.
105. Gipson, B., Zeng, X., Zhang, Z. & Stahlberg, H. (2007). **2dx - User-friendly image processing for 2D crystals**. *J Struct Biol* **157**, 64-72.
106. Gipson, B., Zeng, X. & Stahlberg, H. (2008). **2dx - Automated 3D structure reconstruction from 2D crystal data**. *Microsc Microanal* **14**, 1290-1291.
107. Gil, D., Carazo, J. M. & Marabini, R. (2006). **On the nature of 2D crystal unbending**. *J Struct Biol* **156**, 546-555.

108. Zanetti, G., Riches, J. D., Fuller, S. D. & Briggs, J. A. (2009). **Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging**. *J Struct Biol* **168**, 305-312.
109. Arheit, M., Castano-Diez, D., Thierry, R., Gipson, B. R., Zeng, X. & Stahlberg, H. (2013). **Image Processing of 2D Crystal Images**. *Methods Mol Biol* **955**, 171-194.
110. Penczek, P. A. (2010). **Fundamentals of three-dimensional reconstruction from projections**. *Methods Enzymol* **482**, 1-33.
111. Schenk, A. D., Castano-Diez, D., Gipson, B., Arheit, M., Zeng, X. & Stahlberg, H. (2010). **3D reconstruction from 2D crystal image and diffraction data**. *Methods Enzymol* **482**, 101-129.
112. Arheit, M., Castano-Diez, D., Thierry, R., Abeyrathne, P., Gipson, B. R. & Stahlberg, H. (2013). **Merging of image data in electron crystallography**. *Methods Mol Biol* **955**, 195-209.
113. Goswami, P., Paulino, C., Hizlan, D., Vonck, J., Yildiz, O. & Kühlbrandt, W. (2011). **Structure of the archaeal Na^+/H^+ antiporter NhaP1 and functional role of transmembrane helix 1**. *EMBO J* **30**, 439-449.
114. Hirai, T., Heymann, J. A., Shi, D., Sarker, R., Maloney, P. C. & Subramaniam, S. (2002). **Three-dimensional structure of a bacterial oxalate transporter**. *Nat Struct Biol* **9**, 597-600.
115. Vinothkumar, K. R., Smits, S. H. & Kühlbrandt, W. (2005). **pH-induced structural change in a sodium/proton antiporter from *Methanococcus jannaschii***. *EMBO J* **24**, 2720-2729.
116. Ter Horst, R. & Lolkema, J. S. (2012). **Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity**. *Biochim Biophys Acta* **1818**, 72-81.
117. Sobczak, I. & Lolkema, J. S. (2005). **The 2-hydroxycarboxylate transporter family: physiology, structure, and mechanism**. *Microbiol Mol Biol Rev* **69**, 665-695.
118. Huber, T., Steiner, D., Röthlisberger, D. & Plückthun, A. (2007). **In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na(+)-citrate symporter CitS as an example**. *J Struct Biol* **159**, 206-221.

Chapter 2 - The projection structure of CitS

The following section has been published in:

Journal of Molecular Biology (2012), 418, 117-126 (doi:10.1016/j.jmb.2012.02.016)

Projection structure of the secondary citrate/sodium symporter CitS at 6 Å resolution by electron crystallography

Fabian Kebbel[1,+], Mareike Kurz[2,+], Markus G. Grütter[2,*], and Henning Stahlberg[1,*]

[1] Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland
[2] Department of Biochemistry, University Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

*Corresponding authors
[+]These authors contributed equally

My contribution to this study was the 2D crystallization of CitS, sample preparation for EM, image recording, data processing, model building and data interpretation (see section 2.6 for a detailed description of author contributions).

2.1 Abstract

CitS from *Klebsiella pneumoniae* acts as a secondary symporter of citrate and sodium ions across the inner membrane of the host. The protein is the best characterized member of the 2-hydroxycarboxylate transporter (2-HCT) family, while no experimental structural information at sub-nanometer resolution is available on this class of membrane proteins. Here, we applied electron crystallography to two-dimensional (2D) crystals of CitS. Tubular 2D crystals were studied by cryo-electron microscopy (EM), producing the 6 Å resolution projection structure of the membrane embedded protein. In the $p22_12_1$-symmetrized projection map, the predicted dimeric structure is clearly visible. Each monomeric unit can tentatively be interpreted as being composed of 11 transmembrane α-helices. In projection, CitS shows a high degree of structural similarity to NhaP1, the Na^+/H^+ antiporter of *Methanococcus jannaschii*. We discuss possible locations for the dimer interface and models for the helical arrangements and domain organizations of the symporter based on existing models.

2.2 Introduction

Two different membrane protein classification schemes were established during the last decades. The classical transporter classification (TC) system[1] is based on the function (mode of transport) and molecular phylogeny of transport proteins. According to the TC system more than 250 putative transport protein families have been identified[1-3]. A more recent classification system (MemGen), developed by Lolkema & Slotboom (1998)[4,5], groups membrane proteins into structural classes based on their hydropathy profile that is proposed to represent a specific fold. According to the MemGen classification system, structural class ST[3] contains thousands of different secondary transporters from 32 families[6]. In both systems, secondary transporters represent one of the largest functional categories. These transporters exploit energy stored in ion and/or solute gradients across the membrane to drive substrate transport and are generally classed in three groups regarding their mode of energy coupling, i.e. symporters, antiporters, and uniporters[7]. Secondary transporters are ubiquitously spread across all kingdoms of life and their abundance is reflected in the vast diversity of encoded sequences.

Major efforts in biomolecular structure determination over the last decades led to an increasing number of three-dimensional (3D) crystal structures of secondary transporters. Among others, these are AcrB[8], LacY[9], GlpT[10], NhaA[11], ClC[12], Glt_{Ph}[13], LeuT[14], AAC1[15], SGLT[16], UCP2[17], EmrD[18], EmrE[19-22], FucP[23], Mhp1[24], BetP[25], AdiC[26], ApcT[27], CaiT[28], $PepT_{So}$[29], UraA[30] and ASBT[31]. However, none of those belongs to the structural class ST[3] of the MemGen classification system. Interestingly, the solved structures reveal several different folds and hence several different substrate translocation mechanisms[7,32]. The structures allow a first understanding of the translocation mechanism in the many different families of secondary transporters.

One family of secondary transporters is represented by the 2-hydroxycarboxylate transporters (2-HCTs). 2-HCTs are found in the ST[3] class in the MemGen system and, being the biochemically best-studied family in this class so far, serve as a paradigm for the 31 other families[6]. A characteristic feature of the exclusively bacterial 2-HCT family is the transport of substrates containing the 2-hydroxycarboxylate motif ($HO\text{-}CR_2\text{-}COO^-$) as found in citrate, malate or lactate[32,33]. Well studied 2-HCT members are proton symporters (CimH of *B. subtilis*, MaeP of *S. bovis*), precursor/product exchangers (CitP of *L. mesenteroides*, MleP of *L. lactis*) and sodium symporter (CitS of *K. pneumoniae*, MaeN of *B. subtilis*).

Functionally and structurally, the Na^+/citrate symporter CitS from *Klebsiella pneumoniae* is the best-characterized 2-HCT member. Studies on CitS have resulted in a detailed topological model of this transporter as a representative for all 2-HCT members (Figure 2.1A). The core structure for 2-HCTs consists of two homologous domains (N-/C-domain) connected by a large cytoplasmic loop. Each domain carries five transmembrane segments (TMSs). The two domains share a similar fold, but due to the odd number of TMSs they have opposite orientations in the membrane (inverted topology)[32,34,35]. CitS possesses one additional TMS at the N-terminus leading to a total of 11 TMS with the NH_2- and COOH-termini at the cytoplasmic and periplasmic side of the membrane, respectively[36] (Figure 2.1A). Furthermore, TMS VIII and IX are connected *via* an intracellular amphipathic surface helix (AH)[37]. Structural and biophysical studies suggested detergent solubilized and purified CitS to exist as an elliptical shaped homo-dimer[38,39]. Krupnik *et al* (2011) assigned the interface between the two monomers to the short

axis of the elongated particle, leaving the long axis for the interface between the N- and C-terminal domains within one monomer (Figure 2.1B)[40]. This study also suggested the N-terminal TMS to be located close to the dimer interface, while the long cytoplasmic loop was positioned to the end of the long axis. The Lolkema group identified a highly conserved arginine residue, Arg428, at the cytoplasmic end of TMS 11. It was proposed that Arg428 might be directly involved in the high-affinity binding of one carboxyl group of the bivalent citrate anion[41]. Additionally, CitS exhibits two reentrant loops Vb and Xa, which fold back in between the TMSs from opposite sides of the membrane. The tips of the loop are formed by the highly conserved GGxG sequence motif, which can be found in most of the 2-HCT members. It is hypothesized that these reentrant loops in the N- and C-domains might be in close proximity in the 3D structure at the interface of the two domains, thereby forming the translocation pathway for citrate and sodium ions[42,43].

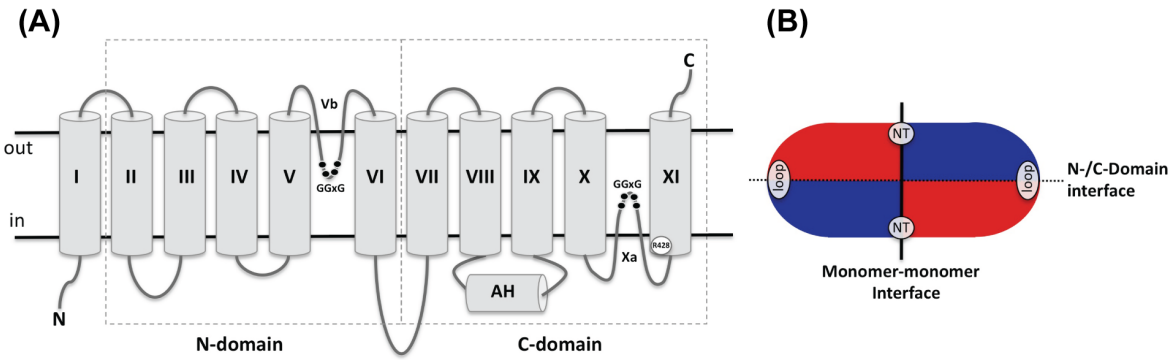


Figure 2.1 Topology model and domain organization of CitS. (**A**) Topology model of CitS. The protein consists of 10+1 transmembrane segments (TMS I-XI) organized as N-terminal and C-terminal domains (5 TMS each) with inverted topology, plus one single TMS. A large cytoplasmic loop connects both domains. Between the 4th and 5th TMS in each domain, a reentrant loop (Vb and Xa) folds into the membrane with the conserved GGXG motif at its tip. TMS XI carries the highly conserved Arginine428, which is involved in substrate binding. (**B**) Model of dimer-interface and domain organization. The monomer-monomer interface is formed by the short axis of the elliptical CitS dimer. The N- and C-domains (depicted in blue/red or red/blue, respectively) are separated by the long axis. The ends of both axes also house the N-terminus (NT) and the cytoplasmic loop.

Functionally it has been postulated that CitS co-transports citrate as a divalent citrate anion ($HCit^{2-}$) coupled with two Na^+-ions and one H^+ using the electrochemical gradient of Na^+[44,45]. During substrate translocation, CitS most likely exhibits two main conformational states, in which the substrate binding pocket is either exposed to the extracellular medium or the cytosol. This mechanism of 'alternating access' seems to be a common feature for secondary transporters as confirmed by high-resolution 3D structures of several transporters[46]. However, the exact transport mechanisms and substrate stochiometries in the 2-HCT transporter family are still not fully determined.

In this study we present the first sub-nanometer resolved structural data of CitS revealed by electron crystallography of two-dimensional (2D) crystals of recombinant CitS embedded in an artificial lipid bilayer. The projection structure of the membrane embedded protein at 6 Å resolution clearly shows a homodimeric structure with each monomer exhibiting eleven electron dense regions likely corresponding to transmembrane α-helices. Based on this projection structure, we discuss different models for possible dimer-interfaces as well as possible arrangements of the N- and C-terminal domains.

2.3 Results

2.3.1 Electron crystallography

Highly pure and homogeneous recombinant CitS expressed in *E. coli* was used for 2D crystallization experiments. The purity was confirmed by SDS-PAGE analysis showing one major band at ~35 kDa corresponding to monomeric CitS (Figure 2.2A). A second faint band at 65 kDa indicated the presence of small amounts of dimeric CitS in SDS, documenting a weak interaction between the two monomers. Numerous crystallization parameters had to be varied over a wide range to find the most suitable conditions for highly ordered 2D crystals (Table 2.1).

Table 2.1: Summary of tested 2D crystallization conditions for CitS

Parameter	Range tested	Best condition
pH	4 - 9	4.5
Protein concentration [mg/ml]	0.2 - 2	1.4
Lipid	*E. coli* polar lipid, POPC, POPE, DMPC, DOPC, DOPG, POPS, POPA	POPE:POPC (7:3 and 3:7)
Lipid-protein-ratio (LPR, w/w)	0.1 – 1.5	0.32 - 0.35
Detergent	DDM, DM	DM (0.2 %)
Temperature [°C]	4-40	32-34
Crystallization technique	Dialysis, Biobeads	Dialysis
Buffer (for pH 4.5)	citrate, acetate	acetate
NaCl [mM]	25-600	500
$MgCl_2$ [mM]	0-50	15
KCl [mM]	0-200	0
Glycerol [%] (v/v)	0-20	0
DTT [mM]	0-5	2

For high-resolution data collection, the crystals were plunge-frozen in liquid ethane without the addition of common cryo-protectants like trehalose, glucose or tannic acid. Other freezing techniques[47] did not yield high-resolution information. Image processing produced two lattices as a result of the two layers formed by the flattened tubes, which were processed independently. A summary of the crystallographic data can be found in table 2.2. The unit cell dimensions for type A crystals (in acetate buffer) were 96.0 Å x 106.0 Å with an included angle of 90.0°. The symmetry assignment was $p22_12_1$. Type A crystals yielded reliable structure factor phases up to a resolution of 4.5 Å, while the merged dataset from five micrographs was limited to 6 Å (Figure 2.2D). For the crystals of type B (in citrate buffer) we found unit cell dimensions of 70.9 Å x 68.6 Å with an included angle of 94.3° and $p2$ symmetry assignment. For this crystal form, only one micrograph could be processed to 9 Å resolution.

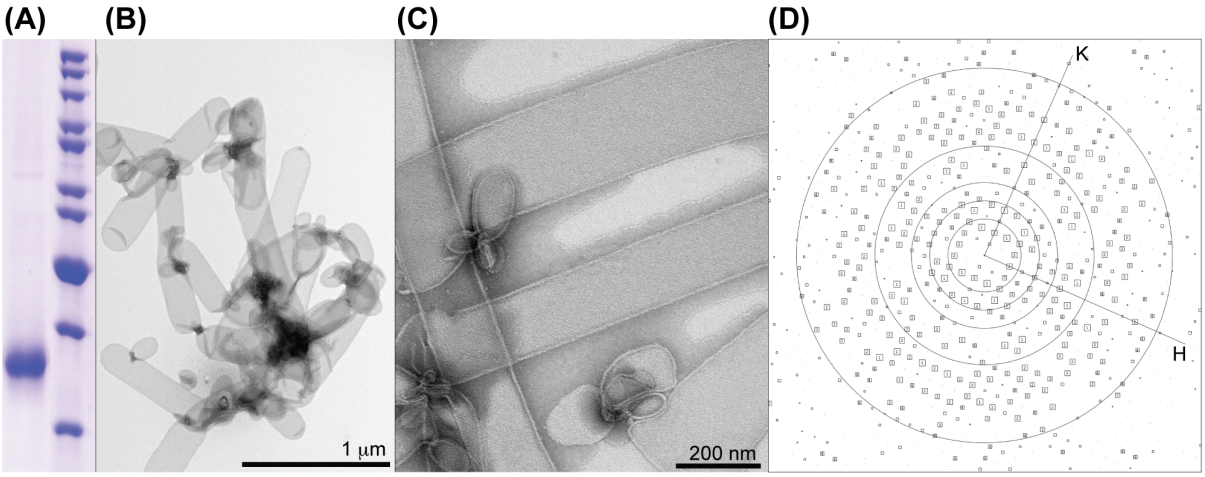


Figure 2.2 Purification and 2D crystallization of CitS. (**A**) SDS-PAGE of purified CitS. The prominent band at 35 kDa corresponds to monomeric CitS, while a faint band at 65 kDa represents the dimer. Molecular weight marker bands are 200, 150, 120 100, 85, 70, 60, 50, 40 and 30 kDa. (**B,C**) Electron micrographs of planar-tubular 2D crystals of CitS (type A, grown in acetate buffer) at different magnifications. Negative staining was performed with 2% uranyl acetate. (**D**) The computed powerspectrum of one single cryo-EM image of one CitS 2D crystal, shown here as IQ plot[48]. The lattice vectors are indicated as H and K. Resolution rings are 36, 24 18, 12 and 7 Å.

Table 2.2: Electron crystallographic data and statistics

	type A crystals (acetate)	**type B crystals (citrate)**
Plane group symmetry	$p22_12_1$	$p2$
Unit cell dimensions	a = 96.0 Å ; b = 106.0 Å γ = 90°	a = 70.9 Å : b = 68.6 Å γ = 94.3°
Number of processed images	5	1
Number of reflections (IQ≤4) in resolution range	∞ - 9.5 Å → 350 9.5 - 6.7 Å → 161 6.7 - 5.5 Å → 39 Σ reflections = 550	∞ - 18 Å → 26 18 - 12 Å → 36 12 - 9 Å → 32 Σ reflections = 94
IQ-weighted phase residuals in resolution range	∞ - 9.5 Å → 25.0° 9.5 - 6.7 Å → 34.6° 6.7 - 5.5 Å → 36.8°	∞ - 18 Å → 23° 18 - 12 Å → 29.6° 12 - 9 Å → 35.4°

2.3.2 Projection Structure

Figure 2.3 illustrates the resulting projection structure (A) and contour plot (B) of CitS at 6 Å resolution in acetate buffer (type A crystals). The crystallographic unit cell accommodates two elliptical shaped molecules with dimensions of 5.2 nm (short axis, a) by 9.6 nm (long axis, b). The same overall shape and dimensions for CitS can also be found in citrate buffer, where the unit cell is formed by one molecule (type B crystals, Figure 2.3C). The observed elliptical shape of the asymmetric unit is consistent with the low-resolution projection structure derived from single particle analysis of detergent solubilized dimeric CitS[38]. As expected, the crystal projection map shows a remarkably smaller outer dimension of the molecule (9.6 nm x 5.2 nm) than seen for the detergent-surrounded particles previously observed (16 nm x 8.4 nm)[38]. The obtained dimensions of the dimeric CitS symporter are similar to those of other secondary transporters like the bacterial chloride channel ClC[49] or the Na^+/H^+ exchangers NhaA[50], and NhAP1[51].

The 2D crystal arrangement shows CitS in a dimeric form, which corroborates the dimerization findings from Blue Native-PAGE, single particle electron microscopy, and fluorescence spectroscopy[38,39]. However, it is still unclear, whether the monomeric or the dimeric CitS forms the functional unit. In comparison, most available structures from secondary transporters so far exhibit higher oligomeric states like dimers (e.g. NhaP1) and trimers (e.g. the H^+/galactose symporter GalP). In those cases, the protomer constitutes the functional unit and oligomerization may predominantly enhance structural stability[7]. An exception is BetP, where trimerization was found to be important for function and regulation[52].

The projection map of the dimeric CitS shows a circularly arranged group of electron-dense regions at both ends of the long axis, while the central part of dimeric CitS is formed by a more rectangular arrangement. These clusters are separated by areas of low density. Surprisingly, a similar global architecture was previously found for the bacterial sodium/proton antiporter NhaP1, which exhibits 13 transmembrane α-helices organized in two homologues domains connected by helix 7 (Figure 2.3E, reproduced from Goswami *et al*., (2011)[51]).

The projection map of the putative CitS dimer (Figure 2.3D) with an assumed monomer-monomer interface formed by the short axis allows the identification of eleven stronger (blue circles) and four weaker densities (dashed blue circles) for each monomer. The stronger densities likely correspond to projections of α-helical transmembrane segments. Two of them (blue circles in figure 2.3B) are strongly contrasted and of limited extension, suggesting them to be in nearly perpendicular orientation relative to the membrane plane. These seem to be involved in defining the crystal contacts between adjacent dimers (Figure 2.3B). The number of strong densities in the CitS projection map is in agreement with the current model, predicting eleven membrane spanning α-helices (Figure 2.1A)[36]. The elongation of most TMS densities suggest those to be slightly tilted in the membrane plane, in line with other secondary transporters of known structure like Glt_{Ph}[13] or LeuT[14].

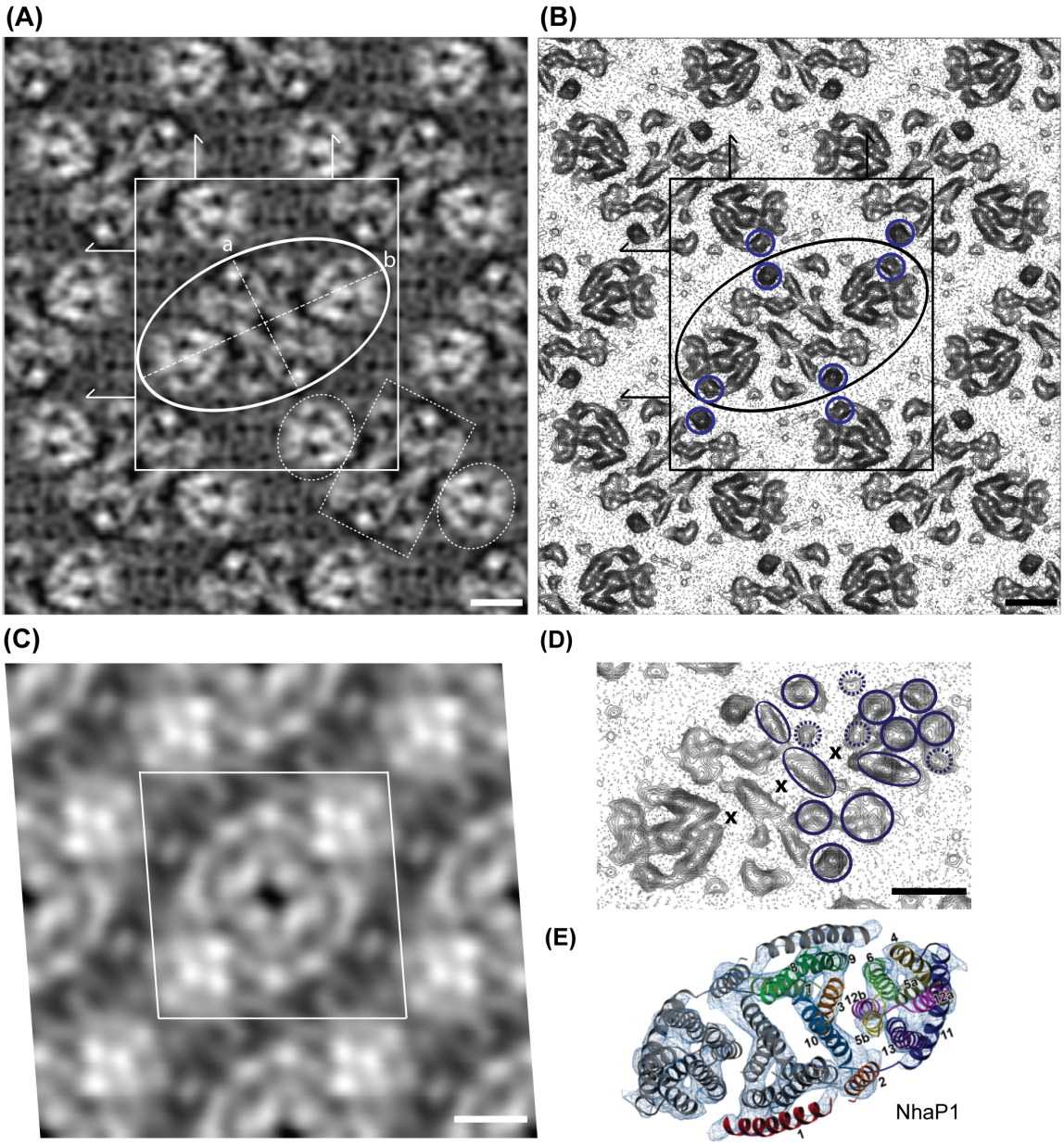


Figure 2.3 Projection structure and contour plot of CitS from type A crystals. (**A**) The merged projection map of CitS from type A crystals at 6 Å resolution. Each unit cell (96.0 Å x 106.0 Å and γ = 90°) contains two CitS dimers. The screw axes of the $p22_12_1$ plane group are marked by arrows and indicate alternating up and down orientation of adjacent dimers relative to the membrane plane. One CitS dimer is highlighted by a white ellipsis with axes of a = 5.2 nm and b = 9.6 nm. In the dimer in the lower right corner three characteristic structural areas are highlighted by dashed white circles and one rectangle. High and low electron densities are depicted in white and black, respectively. No temperature factor was applied. Scalebar is 2 nm. (**B**) Contour plot of the map in (A). One CitS dimer is highlighted by a black ellipsis. Blue circles indicate putative electron densities involved in crystal contacts. Scalebar is 2 nm. (**C**) Contour plot of CitS from type B crystals at 9 Å resolution. The unit cell (70.9 Å x 68.6 Å and γ = 94.3°) with applied *p*2 symmetry is marked by a white rhomboid and contains one CitS dimer. No temperature factor was applied. Scalebar is 2 nm. (**D**) Contour plot of one CitS dimer at 6 Å resolution. Prominent electron densities in one hypothetical monomer are marked by blue circles. Black crosses indicate regions of low density. Scalebar is 2 nm. (**E**) Model of NhaP1 viewed from top. Arabic numbers refer to the 13 helices in the monomer. Corresponding helices in the two domains are in the same color. Reproduced from[51] with kind permission.

2.3.3 Monomer-monomer interface, N- and C- terminal domain arrangement

According to crosslinking studies of BAD-tagged CitS, the monomer-monomer interface of CitS was proposed to be located at the short axis of the elliptical dimer[40]. Potential models that fulfill this criterion are presented in figures 2.4A and B. Both interfaces differ only in the monomer affiliation of the prominent perpendicular helix at both ends of the putative interfaces. Figures 2.4C/D present two alternative hypothetical interface locations. However, only the models in figures 2.4A and B are compatible with interfaces found in other dimeric secondary transporters such as NhaP1[53] and ClC[12]. The precise localization of the dimer interface will have to await the availability of a higher-resolution 3D structure.

Possible arrangements of the N- and C-terminal 5-helix domains within each CitS monomer are indicated in figures 2.4E-G. Krupnik *et al.*[40] argued that (1) the N-terminus of CitS might be located at the end of the short axis, and (2) the large cytoplasmic loop might be at the tip of the dimer's long axis. We also note that (3) an internal symmetry can roughly be discerned in the projection map of each CitS monomer along the long axis of the dimer, likely originating from a similar fold of the N- and C-terminal domains of CitS. Finally (4), Dobrowolski & Lolkema[43] assigned the substrate translocation site to the interface of both domains, formed by helices 5/6 and 10/11 with the associated reentrant loops. Based on these four assumptions, we propose three different feasible models for the assignment of N- and C-terminal halves of each monomer (Figures 2.4E-G).

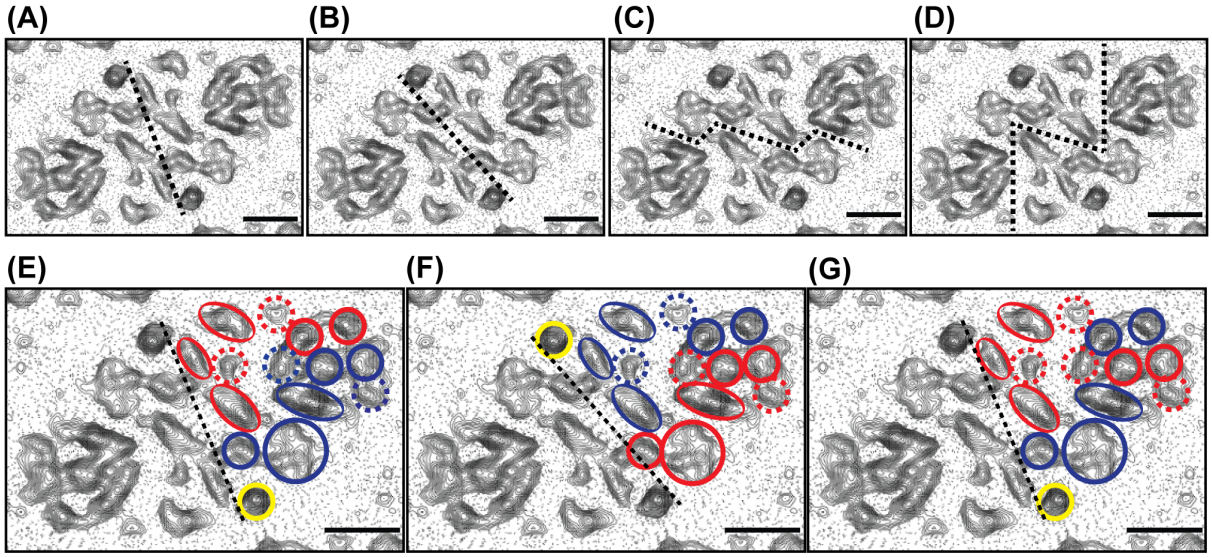


Figure 2.4 Possible monomer-monomer-interfaces and domain orientations in the CitS dimer. Each panel shows one CitS dimer projection contour plot at 6 Å resolution. (**A-D**) Hypothetical dimer interfaces are marked by dashed black lines. (**E-G**) Models of domain organization. The N-terminal helix (yellow) sits at the end of the short axis. The N-domain is highlighted in blue, while helices belonging to the C-domain are depicted in red. Scalebars are 2 nm.

In all models, the prominent perpendicular α-helix at the assumed protomer-interface is defined as the N-terminal helix 1 (compare figure 2.1A). Adjacent to this, we defined five helices as N-domain (blue) and C-domain (red). Models 1 and 2 differ in the position of the interface between two monomers, which leads to an altered position of helix 1 and with it, a flipping of the two domains. In both models, the two domains appear as structurally independent from each other

with only little helix intertwining. Similar domain organizations (in terms of limited domain intertwining) were previously found for secondary transporters of the major facilitator family, e.g. LacY, GlpT and EmrD[7]. In a third model (Figure 2.4G) we flipped the four helices found in the cluster at the tip of the dimer, which would lead to a higher degree of helix intertwining between the N- and C-terminal domains. Here, CitS would rather resemble secondary transporters of the LeuT fold (e.g. LeuT, BetP, CaiT) and, again, NhAP1. For all three models an approximate internal symmetry can be applied on the N- and C-terminal domains by rotating one of the domains by 180° in the membrane plane around the long axis of the dimer. This symmetry does not apply for the single N-terminal helix I, which speaks for a valid assignment of that yellow density in figures 2.4E-G as helix I. However, compared to other secondary transporters the internal symmetry we find in our models is rather weak, especially in the central part of the molecule. This might reflect a distinct conformation of CitS as it is trapped in the presented crystals. On the other hand, the weak internal symmetry could also be a structural feature of CitS and other members of the ST[3] class.

According to functional studies by Dobrowolski & Lolkema[42], the helical TMS 5/6 and 10/11 and the reentrant loops Va and Xb are involved in substrate translocation. In addition, the translocation site might be relatively distant from the dimer interface, as it is also found for most other secondary transporters[7,46,54]. Based on these findings, a possible substrate translocation site in CitS could be formed by the circular helix cluster at each tip of the dimeric molecule. However, a high-resolution 3D map of CitS is needed to clearly assign the helix model and translocation site.

2.4 Conclusion

In this study we present the first sub-nanometer projection map of the secondary citrate/Na^+ symporter CitS as a representative member of the 2-Hydroxycarboxylate-transporter family. The projection structure at 6 Å resolution exhibits at least eleven densities to which α-helical transmembrane segments can be assigned. These are organized in three major clusters in the dimeric molecule. The CitS projection map shows a high similarity to that of the unrelated Na^+/H^+ antiporter NhaP1. We propose a hypothetical model for the monomer-monomer interface in the CitS dimer, and discuss possible orientations of the two N- and C-terminal sub-domains of CitS. According to these models, helix I would be located at the end of the short axis of the elliptical dimer. The N- and C-terminal domains would exhibit an approximate internal symmetry, which can be recognized in the approximate mirror symmetry along the long axis of the molecule. We speculate the substrate translocation site to be formed by at least four helices at the distant tip of each monomeric molecule.

2.5 Materials & Methods

2.5.1 Protein expression and purification

CitS was expressed and purified to homogeneity as described before[55,56] with modifications. Briefly, the N-terminally His-tagged CitS was overexpressed in *E. coli* C43(DE3) by fermentation. Purification was performed by membrane solubilization in *n*-Dodecyl-β-D-maltoside (DDM) and ion metal-affinity chromatography (IMAC, Ni^{2+}-NTA, Quiagen). Detergent exchange to *n*-Decyl-β-D-maltopyranoside (DM) was performed during IMAC. The protein was further polished by passing over a size exclusion chromatography column (Superdex 200 10/300 GL, GE Healthcare) and concentrated with microcon (Amicon), molecular weight cutoff (MWCO) of 100 kDa.

2.5.2 2D crystallization

2D crystals of CitS were grown by microdialysis in 70 µl buttons sealed with a 14 kDa MWCO membrane. Lipids solubilized in 2% DM were added to the membrane protein solution (1.4 mg/ml in 0.2 % DM) and incubated on ice over night. 70 µl of this ternary mixture were dialyzed against 2 liters of buffer containing 20 mM sodium acetate, pH 4.5, 500 mM NaCl, 15 mM $MgCl_2$, 2 mM DTT and 2 mM NaN_3 at temperatures of 32 °C (2 days) and 34 °C (3 days). The quality of the harvested 2D crystals was evaluated by negative stain electron microscopy. This was done by adsorbing 3 µl crystal solution to 200 mesh carbon-coated copper grids that were rendered hydrophilic by glow-discharging in air for 20 s. Grids were washed in double-distilled water and stained with 2 % uranyl acetate. Pictures were taken on a Philips CM10 equipped with a LaB_6 filament and operated at 80 kV accelerating voltage.

2.5.3 Electron microscopy and image processing

For cryo-electron microscopy (cryo-EM), crystal solution on glow-discharged carbon coated holey carbon grids (Quantifoil R2/2, Quantifoil Micro Tools, Jena, Germany) was blotted and rapidly plunge frozen in liquid nitrogen cooled liquid ethane, using a MarkII Vitrobot (FEI, Eindhoven, Netherlands). The frozen grids were transferred to a Gatan-626 cryo holder and analyzed in a Philips CM200 transmission electron microscope, equipped with a field-emission gun (FEG) and operated at 200 kV. Pictures were taken at a nominal magnification of 50kx using low-dose imaging techniques with an electron dose of approx. 5 $e^-/Å^2$ and defocus values ranging from 0.5 to 0.9 µm. Images were recorded on Kodak SO-163 film, which was developed for 7 min in full strength Kodak D19 developer solution. Image quality was assessed by optical diffraction on a home-built laser diffractometer. The best images were digitized using a Heidelberg Primescan D 7100 scanner with a step size of 1 Å/pixel at the specimen level. Digital images were processed using the 2dx software suite[57,58], which is based on the MRC programs[59]. Images were corrected for crystal disorders by three rounds of unbending. This was followed by a correction for the contrast transfer function (CTF) and astigmatism. Symmetry was determined using the allspace program[60] within 2dx.

2.6 Acknowledgements and author contributions

We thank Mohamed Chami for providing continuous support for cryo-EM. We also thank Werner Kühlbrandt for permission to reproduce the NhaP1 map in Figure 2.3E. This work was in part supported by the Swiss National Science Foundation (SNF 315230_127545, NCCRs Structural Biology and TransCure), and the Swiss Initiative for Systems Biology (SystemsX.ch).

MK expressed and purified the protein. FK performed 2D crystallization, electron microscopy and model building. HS and FK carried out image processing. All authors interpreted the results and wrote the manuscript.

2.7 References

1. Saier, M. H., Jr. (2000). **A functional-phylogenetic classification system for transmembrane solute transporters**. *Microbiol Mol Biol Rev* **64**, 354-411.
2. Busch, W. & Saier, M. H., Jr. (2004). **The IUBMB-endorsed transporter classification system**. *Mol Biotechnol* **27**, 253-262.
3. Paulsen, I. T., Sliwinski, M. K. & Saier, M. H., Jr. (1998). **Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities**. *J Mol Biol* **277**, 573-592.
4. Lolkema, J. S. & Slotboom, D. J. (1998). **Estimation of structural similarity of membrane proteins by hydropathy profile alignment**. *Mol Membr Biol* **15**, 33-42.
5. Lolkema, J. S. & Slotboom, D. J. (2003). **Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis**. *J Mol Biol* **327**, 901-909.
6. Ter Horst, R. & Lolkema, J. S. (2012). **Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity**. *Biochim Biophys Acta* **1818**, 72-81.
7. Forrest, L. R., Krämer, R. & Ziegler, C. (2011). **The structural basis of secondary active transport mechanisms**. *Biochim Biophys Acta* **1807**, 167-188.
8. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. (2002). **Crystal structure of bacterial multidrug efflux transporter AcrB**. *Nature* **419**, 587-593.
9. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S. (2003). **Structure and mechanism of the lactose permease of *Escherichia coli***. *Science* **301**, 610-615.
10. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. (2003). **Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli***. *Science* **301**, 616-620.
11. Hunte, C., Screpanti, E., Venturi, M., Rimon, A., Padan, E. & Michel, H. (2005). **Structure of a Na^+/H^+ antiporter and insights into mechanism of action and regulation by pH**. *Nature* **435**, 1197-1202.
12. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. (2002). **X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity**. *Nature* **415**, 287-294.
13. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. (2004). **Structure of a glutamate transporter homologue from *Pyrococcus horikoshii***. *Nature* **431**, 811-818.
14. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. (2005). **Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters**. *Nature* **437**, 215-223.

15. Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trezeguet, V., Lauquin, G. J. & Brandolin, G. (2003). **Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside**. *Nature* **426**, 39-44.
16. Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., Wright, E. M. & Abramson, J. (2008). **The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport**. *Science* **321**, 810-814.
17. Berardi, M. J., Shih, W. M., Harrison, S. C. & Chou, J. J. (2011). **Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching**. *Nature* **476**, 109-113.
18. Yin, Y., He, X., Szewczyk, P., Nguyen, T. & Chang, G. (2006). **Structure of the multidrug transporter EmrD from *Escherichia coli***. *Science* **312**, 741-744.
19. Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S. & Tate, C. G. (2003). **Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer**. *EMBO J* **22**, 6175-6181.
20. Tate, C. G. (2006). **Comparison of three structures of the multidrug transporter EmrE**. *Curr Opin Struct Biol* **16**, 457-464.
21. Chen, Y. J., Pornillos, O., Lieu, S., Ma, C., Chen, A. P. & Chang, G. (2007). **X-ray structure of EmrE supports dual topology model**. *Proc Natl Acad Sci U S A* **104**, 18999-19004.
22. Korkhov, V. M. & Tate, C. G. (2009). **An emerging consensus for the structure of EmrE**. *Acta Crystallogr D Biol Crystallogr* **65**, 186-192.
23. Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., Wang, J. & Yan, N. (2010). **Structure of a fucose transporter in an outward-open conformation**. *Nature* **467**, 734-738.
24. Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E. P., Rutherford, N. G., Hadden, J. M., O'Reilly, J., Ma, P., Saidijam, M., Patching, S. G., Hope, R. J., Norbertczak, H. T., Roach, P. C., Iwata, S., Henderson, P. J. & Cameron, A. D. (2008). **Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter**. *Science* **322**, 709-713.
25. Ressl, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. (2009). **Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP**. *Nature* **458**, 47-52.
26. Fang, Y., Jayaram, H., Shane, T., Kolmakova-Partensky, L., Wu, F., Williams, C., Xiong, Y. & Miller, C. (2009). **Structure of a prokaryotic virtual proton pump at 3.2 Å resolution**. *Nature* **460**, 1040-1043.
27. Shaffer, P. L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. (2009). **Structure and mechanism of a Na^+-independent amino acid transporter**. *Science* **325**, 1010-1014.
28. Schulze, S., Köster, S., Geldmacher, U., Terwisscha van Scheltinga, A. C. & Kühlbrandt, W. (2010). **Structural basis of Na(+)-independent and cooperative substrate/product antiport in CaiT**. *Nature* **467**, 233-236.
29. Newstead, S., Drew, D., Cameron, A. D., Postis, V. L., Xia, X., Fowler, P. W., Ingram, J. C., Carpenter, E. P., Sansom, M. S., McPherson, M. J., Baldwin, S. A. & Iwata, S. (2011). **Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2**. *EMBO J* **30**, 417-426.
30. Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G., Deng, D., Dang, S., Zhang, X., Wang, J. & Yan, N. (2011). **Structure and mechanism of the uracil transporter UraA**. *Nature* **472**, 243-246.
31. Hu, N. J., Iwata, S., Cameron, A. D. & Drew, D. (2011). **Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT**. *Nature* **478**, 408-411.
32. Sobczak, I. & Lolkema, J. S. (2005). **Structural and mechanistic diversity of secondary transporters**. *Curr Opin Microbiol* **8**, 161-167.
33. Bandell, M., Ansanay, V., Rachidi, N., Dequin, S. & Lolkema, J. S. (1997). **Membrane potential-generating malate (MleP) and citrate (CitP) transporters of lactic acid bacteria are homologous proteins. Substrate specificity of the 2-hydroxycarboxylate transporter family**. *J Biol Chem* **272**, 18140-18146.

34. Lolkema, J. S. (2006). **Domain structure and pore loops in the 2-hydroxycarboxylate transporter family**. *J Mol Microbiol Biotechnol* **11**, 318-325.
35. Lolkema, J. S., Sobczak, I. & Slotboom, D. J. (2005). **Secondary transporters of the 2HCT family contain two homologous domains with inverted membrane topology and trans re-entrant loops**. *FEBS J* **272**, 2334-2344.
36. van Geest, M. & Lolkema, J. S. (2000). **Membrane topology of the Na(+)/citrate transporter CitS of *Klebsiella pneumoniae* by insertion mutagenesis**. *Biochim Biophys Acta* **1466**, 328-338.
37. Sobczak, I. & Lolkema, J. S. (2005). **Loop VIII/IX of the Na^+-citrate transporter CitS of *Klebsiella pneumoniae* folds into an amphipathic surface helix**. *Biochemistry* **44**, 5461-5470.
38. Moscicka, K. B., Krupnik, T., Boekema, E. J. & Lolkema, J. S. (2009). **Projection structure by single-particle electron microscopy of secondary transport proteins GltT, CitS, and GltS**. *Biochemistry* **48**, 6618-6623.
39. Kästner, C. N., Prummer, M., Sick, B., Renn, A., Wild, U. P. & Dimroth, P. (2003). **The citrate carrier CitS probed by single-molecule fluorescence spectroscopy**. *Biophys J* **84**, 1651-1659.
40. Krupnik, T., Dobrowolski, A. & Lolkema, J. S. (2011). **Cross-linking of dimeric CitS and GltS transport proteins**. *Mol Membr Biol* **28**, 243-253.
41. Sobczak, I. & Lolkema, J. S. (2004). **Alternating access and a pore-loop structure in the Na^+-citrate transporter CitS of *Klebsiella pneumoniae***. *J Biol Chem* **279**, 31113-31120.
42. Dobrowolski, A. & Lolkema, J. S. (2009). **Functional importance of GGXG sequence motifs in putative reentrant loops of 2HCT and ESS transport proteins**. *Biochemistry* **48**, 7448-7456.
43. Dobrowolski, A., Fusetti, F. & Lolkema, J. S. (2010). **Cross-linking of trans reentrant loops in the Na(+)-citrate transporter CitS of *Klebsiella pneumoniae***. *Biochemistry* **49**, 4509-4515.
44. Lolkema, J. S., Enequist, H. & van der Rest, M. E. (1994). **Transport of citrate catalyzed by the sodium-dependent citrate carrier of *Klebsiella pneumoniae* is obligatorily coupled to the transport of two sodium ions**. *Eur J Biochem* **220**, 469-475.
45. Pos, K. M. & Dimroth, P. (1996). **Functional properties of the purified Na(+)-dependent citrate carrier of *Klebsiella pneumoniae*: evidence for asymmetric orientation of the carrier protein in proteoliposomes**. *Biochemistry* **35**, 1018-1026.
46. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. (2009). **Unlocking the molecular secrets of sodium-coupled transporters**. *Nature* **459**, 347-355.
47. Gyobu, N., Tani, K., Hiroaki, Y., Kamegawa, A., Mitsuoka, K. & Fujiyoshi, Y. (2004). **Improved specimen preparation for cryo-electron microscopy using a symmetric carbon sandwich technique**. *J Struct Biol* **146**, 325-333.
48. Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). **Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy**. *J Mol Biol* **213**, 899-929.
49. Mindell, J. A., Maduke, M., Miller, C. & Grigorieff, N. (2001). **Projection structure of a ClC-type chloride channel at 6.5 Å resolution**. *Nature* **409**, 219-223.
50. Appel, M., Hizlan, D., Vinothkumar, K. R., Ziegler, C. & Kühlbrandt, W. (2009). **Conformations of NhaA, the Na^+/H^+ exchanger from *Escherichia coli*, in the pH-activated and ion-translocating states**. *J Mol Biol* **388**, 659-672.
51. Goswami, P., Paulino, C., Hizlan, D., Vonck, J., Yildiz, O. & Kühlbrandt, W. (2011). **Structure of the archaeal Na^+/H^+ antiporter NhaP1 and functional role of transmembrane helix 1**. *EMBO J* **30**, 439-449.
52. Perez, C., Khafizov, K., Forrest, L. R., Krämer, R. & Ziegler, C. (2011). **The role of trimerization in the osmoregulated betaine transporter BetP**. *EMBO Rep* **12**, 804-810.

53. Vinothkumar, K. R., Smits, S. H. & Kühlbrandt, W. (2005). **pH-induced structural change in a sodium/proton antiporter from *Methanococcus jannaschii***. *EMBO J* **24**, 2720-2729.
54. Abramson, J. & Wright, E. M. (2009). **Structure and function of Na(+)-symporters with inverted repeats**. *Curr Opin Struct Biol* **19**, 425-432.
55. Huber, T., Steiner, D., Röthlisberger, D. & Plückthun, A. (2007). **In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na(+)-citrate symporter CitS as an example**. *J Struct Biol* **159**, 206-221.
56. Kästner, C. N., Dimroth, P. & Pos, K. M. (2000). **The Na^+-dependent citrate carrier of *Klebsiella pneumoniae*: high-level expression and site-directed mutagenesis of asparagine-185 and glutamate-194**. *Arch Microbiol* **174**, 67-73.
57. Gipson, B., Zeng, X. & Stahlberg, H. (2008). **2dx - Automated 3D structure reconstruction from 2D crystal data**. *Microsc Microanal* **14**, 1290-1291.
58. Gipson, B., Zeng, X., Zhang, Z. & Stahlberg, H. (2007). **2dx - User-friendly image processing for 2D crystals**. *J Struct Biol* **157**, 64-72.
59. Crowther, R. A., Henderson, R. & Smith, J. M. (1996). **MRC image processing programs**. *J Struct Biol* **116**, 9-16.
60. Valpuesta, J. M., Carrascosa, J. L. & Henderson, R. (1994). **Analysis of electron microscope images and electron diffraction patterns of thin crystals of phi 29 connectors in ice**. *J. Mol. Biol.* **240**, 281-287.

Chapter 3 – Three-dimensional structure and conformations of CitS

The following section has been accepted for publication in *Structure*.

Structure and substrate induced conformational changes of the secondary citrate/sodium symporter CitS revealed by electron crystallography

Fabian Kebbel[1,+], Mareike Kurz[2,+], Marcel Arheit, Markus G. Grütter[2,*], and Henning Stahlberg[1,*]

[1] Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland
[2] Department of Biochemistry, University Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

*Corresponding authors
[+]These authors contributed equally

My contribution to this study was the 2D crystallization of CitS, sample preparation for EM, image recording, data processing, model building and data interpretation (see section 3.6 for a detailed description of author contributions).

3.1 Abstract

The secondary Na^+/citrate symporter CitS of *Klebsiella pneumoniae* is the best-characterized member of the 2-hydroxycarboxylate transporter family. The recent projection structure gave first insights into its overall structural organization. Here, we present the three-dimensional map of dimeric CitS obtained by electron crystallography. Each monomer has 13 α-helical transmembrane segments, six are organized in a distal helix cluster and seven in the central dimer interface domain. Based on structural analyses and comparison to *Vc*INDY we propose a molecular model for CitS, assign the helices and demonstrate the internal structural symmetry. We also present projections of CitS in several conformational states induced by the presence and absence of sodium and citrate as substrates. Citrate binding induces a defined movement of α-helices within the distal helical cluster. Based on this we propose a substrate translocation site and conformational changes that are in agreement with the transport model of 'alternating access'.

3.2 Introduction

Two available classification systems group secondary transport proteins according to their functionality and sequence homology (TC classification[1]) or their hydropathy profiles (ST[1-4], MemGen classification[2-4]). Both systems underline the enormous phylogenetic, functional and structural diversity among secondary transporters. The number of available high-resolution 3D structures for these proteins is growing rapidly, providing unexpected structural insights. So far, all structures reveal 4-14 α-helical transmembrane segments (TMS), the majority of protein monomers being comprised of 11-13[5]. Surprisingly, numerous unrelated secondary transporters seem to share common global structural folds, e.g., the fold of LeuT[6,7] and the fold of the major facilitator superfamily (MFS)[8]. Most secondary transporters occur as dimers or trimers[9] and most of their structures reveal internal structural symmetry within the single monomers based on (inverted) repeats of a defined number of helices. More importantly, the growing number of atomic structures within a common fold provides an unprecedented insight into the molecular mechanism of secondary active transport[9-11].

In the original model proposing 'alternating access', the secondary transporter alternately exposes its substrate binding sites to both sides of the membrane, which facilitates a unique framework for substrate translocation[12]. This model has been refined and extended by numerous crystallographic and biochemical breakthroughs[9]. According to current knowledge, secondary transporters cycle through defined structural states. Initial substrate binding, e.g., to the empty 'outward open' transporter, induces the closure of outer molecular gates, resulting in the closed 'occluded' conformation as a transition state. A further conformational switch opens the inner gates, leading to the 'inward open' state, where the substrate(s) can be released. The free energy barriers of these sometimes substantial conformational changes are overcome by utilizing the binding energy of both substrates (main- and co-substrate) to the transporter[9]. Different crystallographically-captured conformational states have allowed the transport cycle to be studied and visualized in detail, leading to three mechanistic models referred to as 'rocker-switch', 'gating' and 'rocking bundle'[9,13]. All of these models accentuate the necessity of internal structural symmetry and each is in good agreement with the original 'alternating access' model.

The secondary citrate/Na^+ symporter CitS of *Klebsiella pneumoniae* is the best characterized member of the 2-hydroxycarboxylate transporters (2-HCTs), a subclass of bacterial transport proteins within ST[3] of the MemGen system for which three-dimensional (3D) structural information is still not available. CitS is postulated to couple the import of two Na^+ ions and one bivalent citrate ion for anaerobic metabolism[14,15]. Models predict an inverted topology of 2*5 helices organized in two domains plus one N-terminal helix[16-19]. In confirmation, the two-dimensional (2D) projection structure of CitS at 6 Å resolution published recently[20] reveals 11 α-helical TMS organized in a distal helix cluster and a central dimerization interface. CitS thereby closely resembles the Na^+/H^+ antiporters NhaA and NhaP1[21,22]. A detailed insight into the structural and functional properties of a member of the ST[3] family, although not of a 2-HCT, was delivered by the recent crystal structure of the divalent anion/Na^+ symporter (DASS) *Vc*INDY[4,23].

Here, we present the three-dimensional (3D) map of the 2-HCT CitS at a resolution of 6 Å obtained by electron crystallography of 2D crystals. Each monomer of the dimer reveals 13 rod-shaped densities, representing 11 single α-helices and two putative helical reentrant loops. Resemblance to *Vc*INDY enables us to refine our model with respect to the membrane

orientation and dimer interface. Helices are assigned and the internal structural symmetry of the N- and C-terminal domain is documented. In addition, projection structures of CitS in different substrate combinations indicate a rearrangement of α-helices within the distal helix clusters after citrate exposure, particularly in the presence of Na^+. The data highlight the co-dependence of these two substrates, support our 3D model and demonstrate that the substrate binding site is part of the distal helix cluster. The observed helix movements are in agreement with those expected for molecular gates.

3.3 Results & Discussion

3.3.1 Electron crystallography

2D crystals of CitS were grown as described previously[20]; use of a temperature controlled dialysis machine with a sodium acetate buffer allowed the size and quality of the crystals to be improved. Tubular/vesicular 2D crystals with diameters up to 600 nm were obtained (Figure S3.1). Calculated Fourier transforms of cryo-electron microscopy (cryo-EM) images of the flattened tubes usually showed two lattices resulting from the two crystalline layers. On image processing in 2dx[24-28], these lattices were treated as two independent datasets. The unit cell parameters (96.0 Å x 106.0 Å with an angle of 90.0°) and the $p22_12_1$ plane group are the same as reported previously[20], and were unchanged for crystals that had been soaked in various substrates (Table 3.1). To extract 3D information, the sample was tilted up to 45° in the microscope for data collection. 3D merging of all 79 lattices enabled us to continuously sample amplitudes and phases along the lattice lines up to a vertical resolution of 15 Å (Table 3.1, Figure S3.2).

Table 3.1. Electron crystallographic data

	Na-Acetate (3D)	Na-Acetate (2D)	Na-Citrate (2D)	K-Acetate (2D)	K-Citrate (2D)
Unit cell dimensions	a = 96 Å, b = 106 Å, γ = 90°	a = 96 Å, b = 106 Å, γ = 90°	a = 96 Å, b = 106 Å, γ = 90°	a = 96 Å, b = 106 Å, γ = 90°	a = 96 Å, b = 106 Å, γ = 90°
Plane group	$p22_12_1$	$p22_12_1$	$p22_12_1$	$p22_12_1$	$p22_12_1$
Number of images	79[a]	11	9	13	11
Resolution in plane	6 Å	6 Å	6 Å	6 Å	6 Å
Resolution in z	15 Å	-	-	-	-
Defocus range (μm)	0.3-2.2	0.4-1.1	0.3-1.2	0.5-1.3	0.4-1.3
Total number of reflections	33441	1095	1034	1993	1806
Number of unique reflections	11480	217	212	219	221
Completeness[b]	79.3 %	-	-	-	-
Overall weighted R-factor	31.2 %	-	-	-	-
Overall weighted phase error	36°	30.5°	31.2°	29.3°	31.7°

[a] Tilt angle distribution: 10 (0°), 13 (15°), 29 (30°), 22 (40°), 5 (45°)

[b] calculated by number of unique reflections to 45° tilt with a figure of merit >50%

3.3.2 Three-dimensional map and structural model of CitS

Figure 3.1 shows the 3D map of CitS in the presence of Na^+ acetate at a resolution of 6 Å in the membrane plane and 15 Å in the vertical direction. As found in the earlier projection structure, viewed from the top dimeric CitS is oval measuring 52*96 Å[20]. In the z-dimension, the CitS dimer spans 40-60 Å through the lipid bilayer. The central part of the dimer is mostly buried in the membrane, while densities towards the end of the long axis extend 10-20 Å out of the bilayer. The resulting 'M-shape' of the side-view (Figure 3.1B) confirms the previous low-resolution single particle structure[29]. Dimeric CitS has three cavities, i.e., one at the center of the dimer and one at the center of each monomer. These cavities are probably filled with lipids or water, respectively.

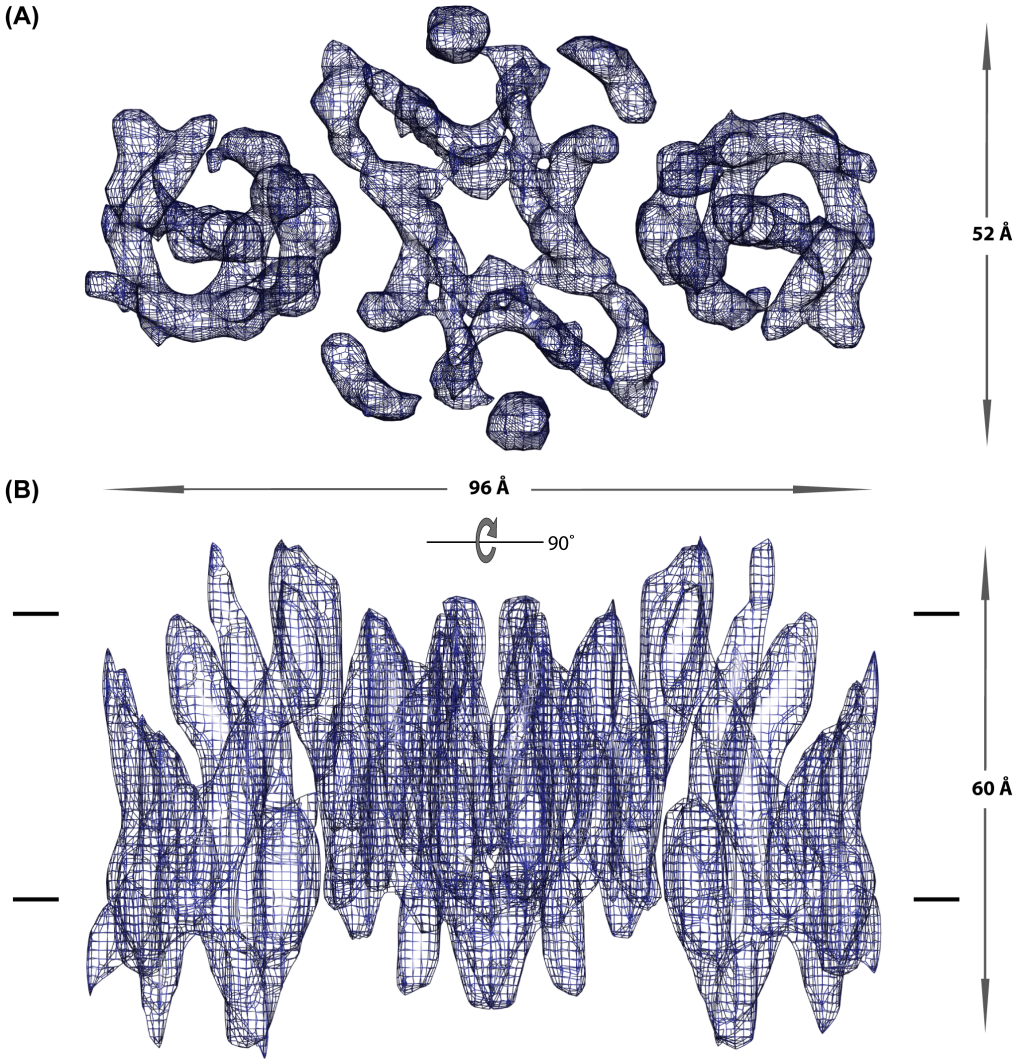


Figure 3.1 Three-dimensional map of CitS. The 3D map of CitS viewed from the top **(A)** and the side **(B)**. One dimer measures 52*96 Å within the membrane plane, spans 40-60 Å in the vertical direction and has a central dimer interface domain and two distal helix clusters.

Manual placement of α-helical poly-A chains into the 3D map of CitS led to the model shown in figure 3.2A/B. The two identical monomers of the dimer are highlighted in red and blue. The assignment of helices to each monomer is based on their location and proximity to other helices, and on the comparison to the structurally related *Vc*INDY (see below). Each CitS monomer has

13 α-helical TMS organized in two domains. The domain involved in dimerization, the interface domain, is comprised of seven helices that are partially tilted or kinked up to 45° relative to the membrane plane; contact between the two monomers is mainly provided by four helices. The second characteristic domain of each monomer, the distal domain, is formed by a dense cluster of six helical elements and is located at the distal tips of the dimer. Besides one kinked helix, most of these elements are nearly perpendicular in the membrane. Two central helices within the cluster are split into two shorter parts (cyan, Figure 3.2G).

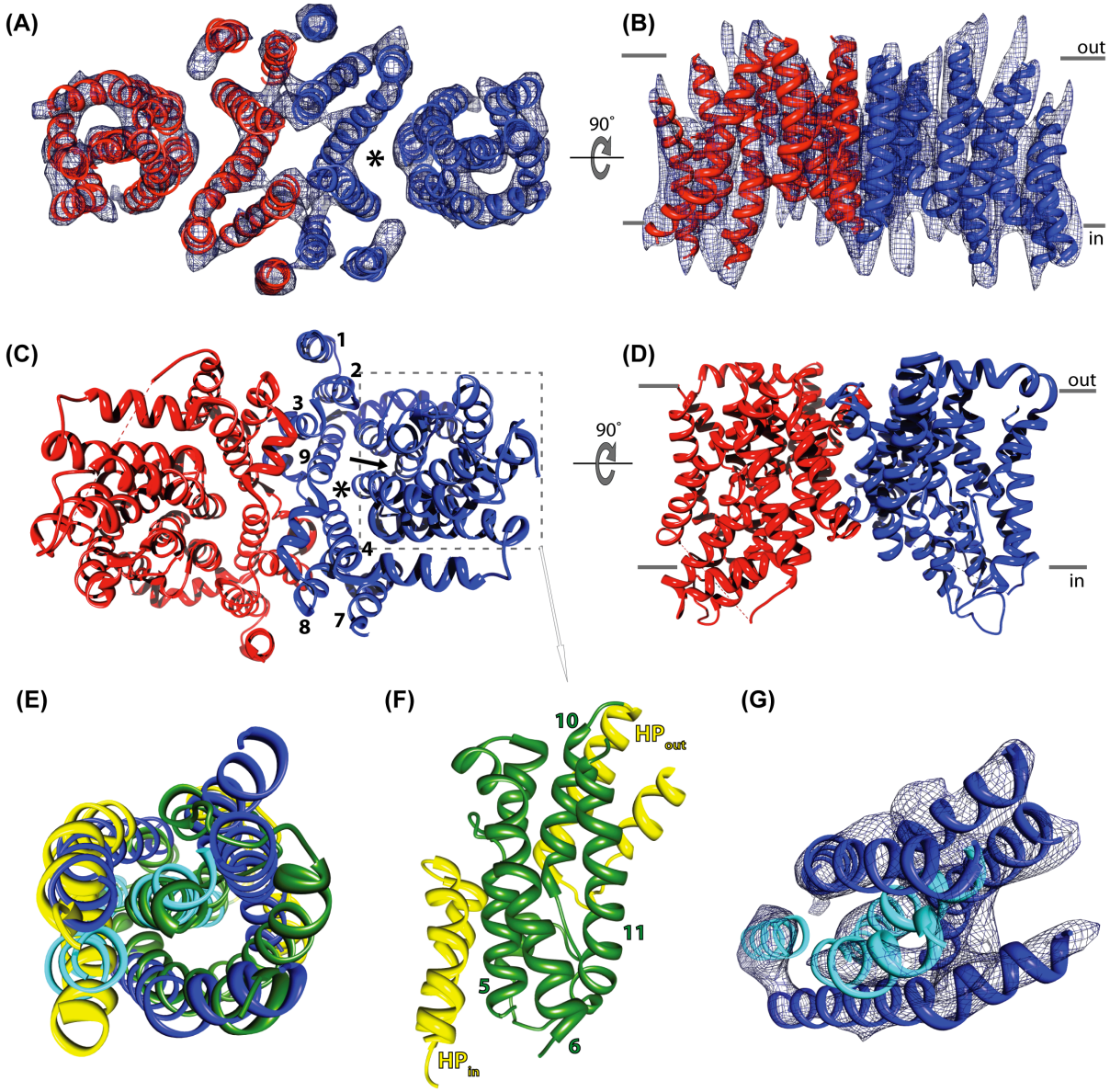


Figure 3.2 Structural model of CitS and comparison to *Vc*INDY. Dimeric CitS viewed from the cytosol (**A**) and the side (**B**). *Vc*INDY (pdb 4F35) viewed from the cytosol (**C**) and the side (**D**). Single monomers are colored in red/blue. The two proteins have a similar global architecture. The central dimerization domain and the distal helix clusters are separated by an aqueous basin (∗). The substrate binding sites in *Vc*INDY (→) are in the center of a monomer. **(E)** Superposition of the helix clusters from CitS (blue/cyan) and *Vc*INDY viewed from the cytosol. **(F)** Distal helix cluster of *Vc*INDY. TMS 5/6/10/11 (green) and $HP_{in/out}$ (yellow) are shown. The surface helices 4c/9c have been removed. **(G)** Distal helix cluster of CitS. Broken TMSs (cyan) may represent helical reentrant loops.

Of the 13 α-helices of each monomeric CitS molecule, 11 are single membrane-spanning helices while two are probably helical re-entrant loops. This extends and refines the model based on our previous projection structure[20], which exhibited 11 strong densities and four less dense regions. In the present model (Figure 3.2A), two of the light regions within the helix bundle and the central dimerization interface, respectively, appear to arise from single α-helical segments. Furthermore, the new model confirms that the monomer-monomer interface is formed by the short axis of the dimer[20]. Due to the lower vertical resolution of 15 Å, our dataset did not allow the visualization of amphipathic surface helices; these are predicted for CitS and present in other secondary transporters[23,30,31].

Hydropathy profiles of 2-HCTs and DASSs predict CitS and *Vc*INDY to share a very similar 3D structure with 10+1 TMSs plus two α-helical reentrant loops[4]. However, the low sequence homology of 14 % (Figure S3.3) does not allow homology modeling. Comparison of the CitS and *Vc*INDY (pdb 4F35) structures reveals numerous common features (Figures 3.2A-D). Both dimeric transporters have the same overall shape and architecture; the interface domain of each monomer contains seven partially tilted helices and there is a second distal helical cluster. Further, in both cases, the interface and distal domains of the monomers are separated by an aqueous basin (asterisks, Figures 3.2A/C). Viewed from the side, the CitS and *Vc*INDY dimers have a characteristic M-shape, and the position of the dimerization interface is almost identical. Assuming the same orientation of both proteins, CitS would protrude into the cytoplasmic space (Figure 3.2B). Another common salient feature is a vertically oriented helix at both ends of the dimer's short axis (TMS1 on the *Vc*INDY structure in Figure 3.2C). In *Vc*INDY, however, this is further away from the protein's main body.

Although the global structures of *Vc*INDY and CitS dimers look similar, there are significant differences in both the interface and distal helix clusters, and the individual monomers superimpose poorly (not shown). In particular, there are major differences in the helix positions and orientations at the dimer interfaces (TMSs 1-4 and 7-9; Figure 3.2A-D). Separate superposition of corresponding helix clusters reveals major structural matches (Figure 3.2E), but there are still substantial differences in the helical architecture. In *Vc*INDY the distal helix cluster is composed of four partially unwound TMSs (green) and four shorter helical reentrant loops, $HP_{in/out}$, each spanning half of the membrane (yellow; Figure 3.2F). This is also true for CitS (Figures 3.2G/3.3A), but rather than flanking the TMSs as in *Vc*INDY, the four short helices are adjacent to each other and at the very center of the cluster (cyan). Nevertheless, their length and proximity allow us to speculate that these four short helical elements represent the reentrant loops Vb/Xa. The significantly different positions of Vb/Xa and $HP_{in/out}$ in CitS and *Vc*INDY was to be expected since *Vc*INDY's $HP_{in/out}$ are found between TMSs 4/5 and 9/10, CitS's are predicted to be between helices 5/6 and 10/11[32]. Furthermore, in *Vc*INDY helix 11 sits at the outer border of the helical bundle and does not contribute to substrate binding[23], while TMS11 of CitS is postulated to be directly involved in citrate binding via R428[33]. Together, these differences underline the different molecular details of CitS and *Vc*INDY although both proteins are found within the same subclass of ST[3]. As a consequence, further structural analysis is essential to reliably assign the helices of CitS (see below).

3.3.3 Molecular model and internal symmetry of CitS

Cross-linking studies on CitS showed that helices 5/6 and 10/11 plus the reentrant loops Vb/Xa form the translocation site[32]. The number of TMSs corresponds well to our model of the distal helical cluster, in which case the remaining seven TMSs 1-4 and 7-9 constitute the dimer interface as in *Vc*INDY[4,23]. Further biochemical studies[19] and the *Vc*INDY structure[23] allow us to assign the perpendicular helix at the outer end of the dimerization interface as TMS1. Additional consideration of inter-helical distances led to the detailed molecular model of CitS shown in figure 3.3A. Helix 1 is depicted in yellow, TMSs belonging to the N-terminal domains in green (2-6) and TMSs belonging to the C-terminal in blue (7-11). Interestingly, helices of the C- and N-terminal domains intertwine much more than proposed in previous models[19,20]. In our new model, helix 11 and the helical reentrant loops Vb/Xa are adjacent to each other at the inner edge of the distal cluster. Thus, all known functionally important and highly conserved elements are positioned at the inner edge and center of the distal helical cluster, close to the aqueous basin. This includes Arg428 of TMS11 and the GGxG motifs at the tips of Vb/Xa, which are directly involved in substrate binding[32,33].

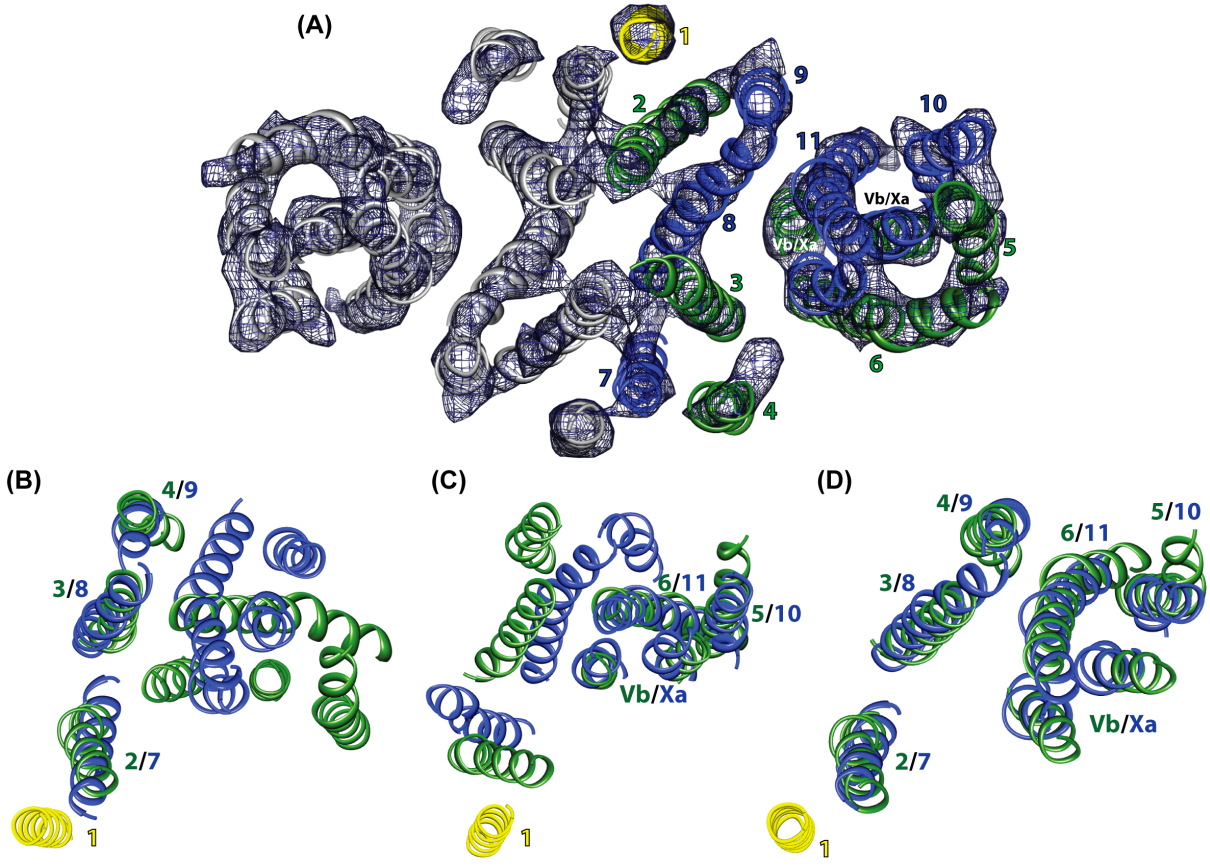


Figure 3.3 Molecular model and internal structural symmetry of CitS. (**A**) Molecular model of CitS. Helices belonging to one monomer are depicted in yellow (TMS1), green (TMS2-6) and blue (TMS7-11). Superposition of the N- and C-domains focusing on (**B**) the dimer interface and (**C**) the distal helix cluster. (**D**) Independent alignment of both domains.

Analysis for the expected internal structural symmetry of each monomer validated the presented molecular model and the helix assignment. Rotating helices 1-4 by 180° along the dimer's long axis gives a good match to TMSs 7/8/9 (Figure 3.3B), but the putative symmetry-related elements of the distal helix cluster do not fit each other. Similarly, a superposition focusing on the distal cluster, with TMSs 5/6 and Vb corresponding to 10/11 and Xa (Figure 3.3C), leads to a bad fit within the interface domain. Alignment of the two domains individually emphasizes the symmetry observed for each (Figure 3.3D). Other helix assignments do not fulfill the symmetry correlations expected for a CitS dimer and none result in a symmetry mate for TMS1, which again supports our model.

The transport cycle of *Vc*INDY is thought to be accomplished by a defined movement of the N- and C-terminal halves of the distal helix cluster relative to each other using the dimer interface as a static anchor point[23]. This would lead to the internal symmetry relationship being valid either for the central dimerization domain or for the helix cluster but not for both simultaneously, in agreement with our CitS model. Thus, the transport mechanisms employed by CitS and *Vc*INDY, and possibly by 2-HCTs and DASSs in general, are probably very similar. Overall, the presented molecular model confirms, refines and extends most previous findings for CitS.

3.3.4 Substrate induced conformational changes

2D crystals of CitS grown in buffer containing Na^+ acetate, were soaked in selected substrate combinations before cryo-EM sample preparation. This led to four different projection structures in (1) Na^+ acetate, (2) Na^+ citrate, (3) K^+ acetate and (4) K^+ citrate, each at 6 Å resolution. All four projections have the shape and dimensions of unsoaked dimeric CitS, and all datasets exhibit low phase residuals (Table 3.1) and reliable calculated diffraction spots (Figures S3.4-3.5). Difference maps were calculated from these four maps to examine the influence of the different substrates on the conformation of CitS.

The first difference map (Figure 3.4A) was calculated from the projections obtained in Na^+-/ and K^+-acetate. Structural differences are negligible across the whole dimer as indicated by bluish and reddish areas with a maximum intensity of +/- 0.2. In the presence of K^+ and absence of Na^+, citrate induced slight density changes (red peaks) towards the center of the distal helix cluster (Figure 3.4B). The largest structural changes were found when the Na^+-citrate projection was compared to its citrate free counterparts (K^+-/Na^+-acetate). Both difference maps (Figures 3.4C/D) exhibit strong peaks (±0.3/±0.5) at central and inner regions of the distal helix cluster, while the dimer interface domain is almost unaffected. The positions of the observed citrate-induced density changes are the same in both cases although the intensities are slightly different, and confirm the weaker peaks found in the absence of Na^+ (Figure 3.4B). Two additional difference maps shown in figure S3.5E/F confirm these results. In all difference maps, the background (lipid bilayer) is without noteworthy changes.

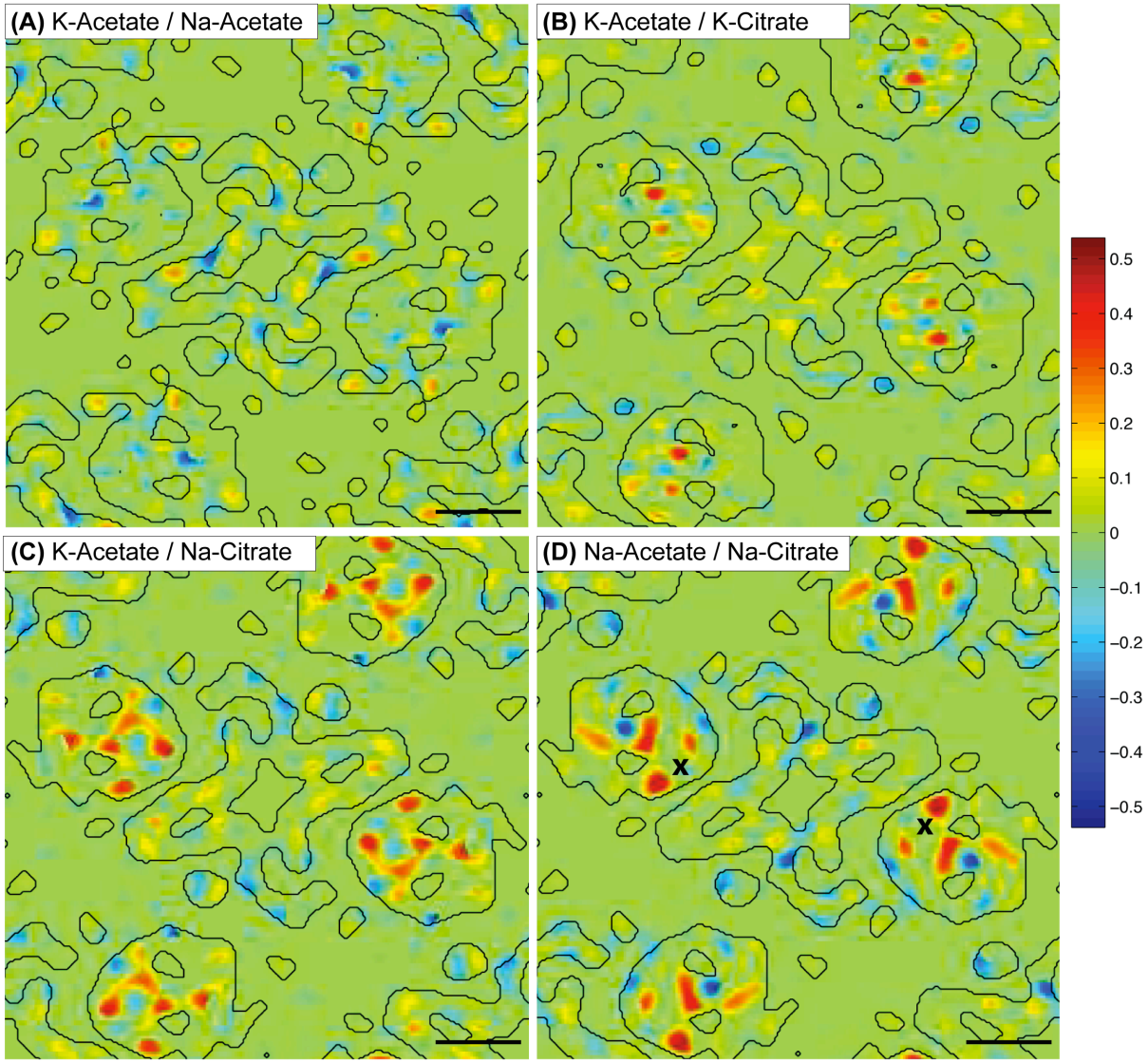


Figure 3.4 Substrate induced conformational changes. Difference maps of CitS in the presence of different substrates: (**A**) K-Acetate/Na-Acetate, (**B**) K-Citrate/K-Acetate, (**C**) K-Acetate/Na-Citrate, (**D**) Na-Acetate/Na-Citrate. Cation exchange in acetate (Na^+/K^+) only causes minor structural changes. Binding of citrate causes density shifts within the distal helix cluster, particularly in the presence of Na^+. The proposed substrate translocation site is indicated (x). The contour of the minuend is plotted. Scalebar, 2 nm.

The density changes observed demonstrate the rearrangement of α-helices within the distal helix cluster of each CitS monomer induced by the binding of citrate. This primarily occurs in the presence of Na^+; K^+ supports minor changes at similar locations. The central dimerization interface remains unaffected by substrate exchanges. These findings suggest (1) that substrate binding occurs within the distal helix cluster of each monomer, (2) that the central helices primarily provide the dimer interface and are not involved in substrate binding, and (3) that citrate induced conformational changes require Na^+ ions as co-substrate. Overall, these findings are in agreement with available models for secondary symport. With the exception of EmrE[34], the monomeric protein is the functional unit of the secondary transporter, oligomerization may regulate transport activity and enhance stability[35,36]. This also applies to CitS. None of the 14 helices at the center of the dimer, seven from each monomer, respond to substrate exchange, emphasizing their primary role as a static anchor point with little or no functional role. From the

citrate-induced α-helix movements observed at highly reproducible positions, the substrate binding site is located close to the inner edge of the distal helix cluster of each monomer (Figure 3.4D, black cross). In agreement, according to our model (Figure 3.3A) this site is formed by helices 5, Vb, Xa and 11, which harbor all known functionally important residues. Citrate would bind close to the aqueous basin at the center of the monomer, providing a structural framework for effective substrate translocation. The proposed binding site closely resembles that of the structurally related transporter *Vc*INDY (arrow, Figure 3.2C[23]).

As there was Na^+ acetate but no citrate in the buffer, our model probably corresponds to an empty 'inward open' or 'outward open' CitS conformation, 'Ci' or 'Ce' respectively. The conformational change occurring in the presence of citrate probably relates to the Na^+ and citrate induced closure of inner/outer molecular gates at the start of the transport cycle[9]. In the resulting inward or outward facing occluded state 'CSic' or 'CSec', the substrates are inaccessibly buried within the membrane. The movements of helices or hairpins during such gate closures are usually relatively small[37]. In good agreement, the observed shifts were in the range of 6 Å.

Like numerous other Na^+ coupled transporters, CitS was previously shown to be inactive in the presence of K^+, and less active in the presence of Li^+[14,23,38]. Na^+ usually binds to the protein first providing a suitable structural and electrostatic framework for the main substrate by direct or indirect interaction [6,23]. This is also valid for CitS[39]. Consequently, the weak citrate induced density shifts observed in the presence of K^+ (Figure 3.4B) were unexpected. Although the 2D crystals were extensively soaked in K^+ buffer, CitS might have retained minute amounts of previously bound Na^+, which then enabled the slight conformational change observed. Additional biochemical and higher resolution structural data are required to further elucidate ion coupling in CitS.

3.4 Conclusion

In this study we present the 3D model of the dimeric citrate/sodium symporter CitS of *Klebsiella pneumoniae* based on electron crystallography of two-dimensional crystals. Each CitS monomer is comprised of 13 helices. These are organized in two characteristic domains, seven being in a central cluster forming the dimerization interface and six in a dense distal cluster. Considering previous models, we developed a detailed molecular model of CitS in which we assigned 11 transmembrane helices, two helical reentrant loops and a substrate binding site. The global architecture of CitS resembles that of *Vc*INDY with substantial differences in the helix orientations and positions of reentrant loops. The helical assignments proposed in our model are validated by the internal structural symmetry within each CitS monomer. Additional structural analyses revealed conformational changes induced by the binding of Na^+ and citrate. The observed helix shifts are spatially limited to the distal helix cluster, and in agreement with gate movements predicted to take place during the transport cycle by the 'alternating access' mechanism.

3.5 Materials & Methods

3.5.1 2D Crystallization

CitS was expressed and purified to homogeneity as described previously[20]. 2D crystals of CitS were grown by dialysis using a home-built dialysis machine with a 14 kDa MWCO membrane[40]. Membrane protein solution (1.4 mg/ml in 0.2 % DM) and detergent-solubilized lipids (POPE:POPC 7:3 (w/w), 10 mg/ml in 2% DM) were mixed at lipid-protein ratios of 0.3-0.35 (w/w) and incubated on ice for 12 h. 100 µl of the ternary mixture were dialyzed against continuously exchanged buffer containing 20 mM sodium acetate, pH 4.5, 500 mM NaCl, 15 mM $MgCl_2$, 2 mM DTT and 2 mM NaN_3 at a flow rate of 0.2 ml/min. A specific temperature profile was applied during four days of dialysis (12 h 10 °C, increase to 34 °C within 12 h, two days at 34 °C, and decrease to 10 °C within 24 h). The resulting vesicular tubular crystals were stable at 4 °C for several months. Crystals were evaluated by negative stain transmission electron microscopy. This was done by adsorbing 3 µl of the crystal solution for 45 sec to 200 mesh carbon-coated copper grids that had been rendered hydrophilic by glow-discharge in air. Grids were washed in double-distilled water and stained with 2 % uranyl acetate. Images for crystal screening were taken at an accelerating voltage of 80 kV using a Philips CM10 equipped with a LaB_6 filament .

3.5.2 Sample preparation, electron microscopy and image processing

Data collection was performed by cryo-electron microscopy (cryo-EM). The crystal solution was adsorbed to carbon coated and glow-discharged holey carbon grids (Quantifoil R3/2, Quantifoil Micro Tools, Jena, Germany). Substrate exchange (Na^+/K^+, acetate/citrate) was achieved by extensive soaking of the adsorbed 2D crystals in 10 drops of the corresponding buffer, 5 seconds each. The buffers were (1) 20 mM sodium acetate, pH 4.5, 200 mM NaCl, 15 mM $MgCl_2$ and 2 mM DTT; (2) 20 mM sodium citrate, pH 4.5, 200 mM NaCl, 15 mM $MgCl_2$, 2 mM DTT; (3) 20 mM potassium acetate, pH 4.5, 200 mM KCl, 15 mM $MgCl_2$ and 2 mM DTT and (4) 20 mM potassium citrate, pH 4.5, 200 mM KCl, 15 mM $MgCl_2$ and 2 mM DTT. After excess soaking solution had been removed by blotting for 4.5 sec at 95 % relative humidity, vitrification was achieved by rapid plunge-freezing in liquid ethane at liquid nitrogen temperature using a MarkII Vitrobot (FEI, Eindhoven, Netherlands). Data acquisition at -180 °C sample temperature was performed at an acceleration voltage of 200 kV using a Philips CM200 microscope with a field-emission gun. Low-dose images with approx. 10 $e^-/Å^2$ at a nominal magnification of 50kx and defocus values of 0.3 to 1.2 µm (untilted images) or 0.5 to 2.2 µm (tilted images) were recorded on Kodak SO-163 film and developed for 7 min in full strength Kodak D19 developer. A home built diffractometer was used to evaluate the quality of the individual images, followed by digitization using a Heidelberg Primescan D 7100 scanner with a step size of 1 Å/pixel at the specimen level. Tilted and untilted images were processed using the 2dx software suite[24,26-28]. The 3D volume was calculated using the CCP4 software package[41], and visualized using Chimera[42].

3.5.3 Model building and difference maps

Difference maps were calculated in real space by subtracting two corresponding merged projection maps limited to 6 Å resolution. The individual projection maps were aligned onto each other within 2dx by the MRC program ORIGTITLT[43] and subsequently scaled to [0,1]. The raw difference maps were then obtained by subtracting one map from the other in real space. To determine the significance of the differences, we determined the variations within the individual projection maps using a similar approach to Appel et al.[21]. However, instead of calculating the difference between images of crystals incubated under identical conditions, we empirically determined the variations across the two compared conformations. This was done by combining images of both conformations and splitting each dataset into two equally Q-value weighted sets[24]. Merging these subsets results in two projection maps each containing both conformations. The difference between these two maps represents the variations within the data and yields an inferior threshold for conformational variations. We then only considered differences between conformations to be significant if they exceed this threshold by a factor of three. We visualized the final result as a heat map with Matlab and plotted the contour of the minuend. The color scale of our maps was adapted to the overall maximum difference of ±0.54.

To build the 3D model of CitS, helices composed of poly-A chains were manually placed into the obtained 3D volume in Chimera[42]. Kinks were introduced by adjusting the corresponding Φ and Ψ angles of the peptide backbone.

3.6 Acknowledgement and author contributions

We thank Mohamed Chami, Kenneth N. Goldie and Bill Anderson for providing excellent support for the cryo-EM. We also thank Shirley A. Müller for her helpful comments on the manuscript. This work was supported by the Swiss National Science Foundation (SNF 315230_127545, National Centers for Competence in Research (NCCR) Structural Biology and TransCure), and the Swiss Initiative for Systems Biology (SystemsX.ch). FK and MK contributed equally to this work. MG and HS inspired and designed the research. MK expressed and purified the protein. FK performed 2D crystallization, electron microscopy, and model building. FK, MA and HS carried out image processing. All authors wrote the manuscript and declare no conflict of interest.

3.7 Supplemental figures

Figure S3.1 Two-dimensional crystals of CitS. Tubular crystals from dialysis buttons (**A**) and a temperature controlled dialysis machine[40] (**B**). The crystal diameter increased from 250 nm up to 600 nm when the latter was used, which was highly beneficial for structural analysis by electron crystallography.

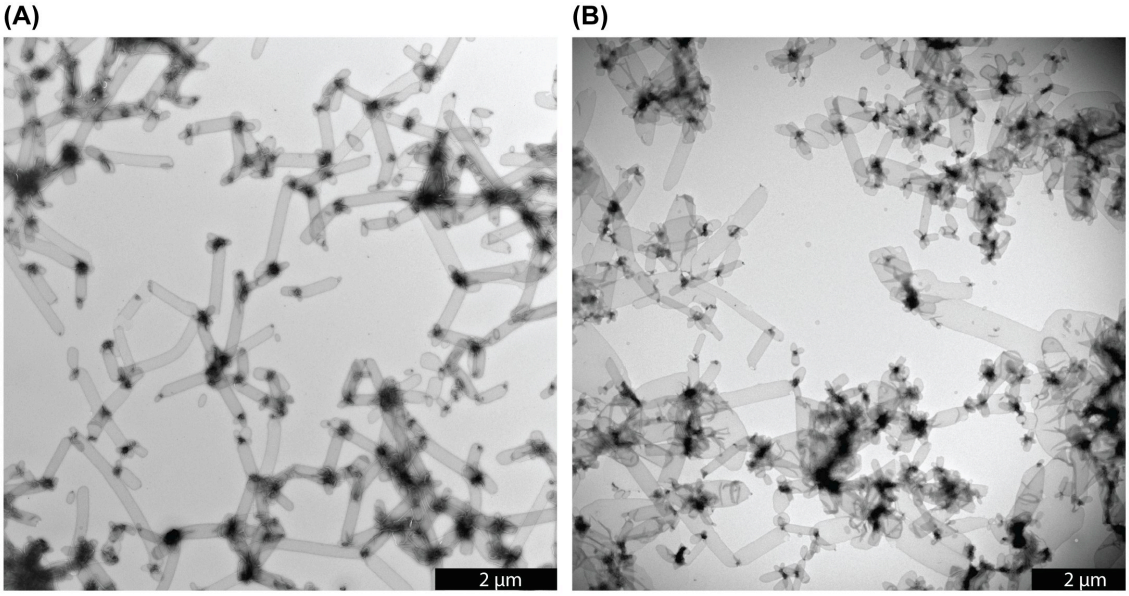


Figure S3.2 Projection maps and lattice lines for 3D map determination. A gallery of 79 projection maps of CitS used for the final 3D dataset is illustrated in (**A**). Three representative lattice lines 2,2 / 1,5 and 6,5 are shown in (**B**) with the corresponding phases (top) and amplitudes (bottom).

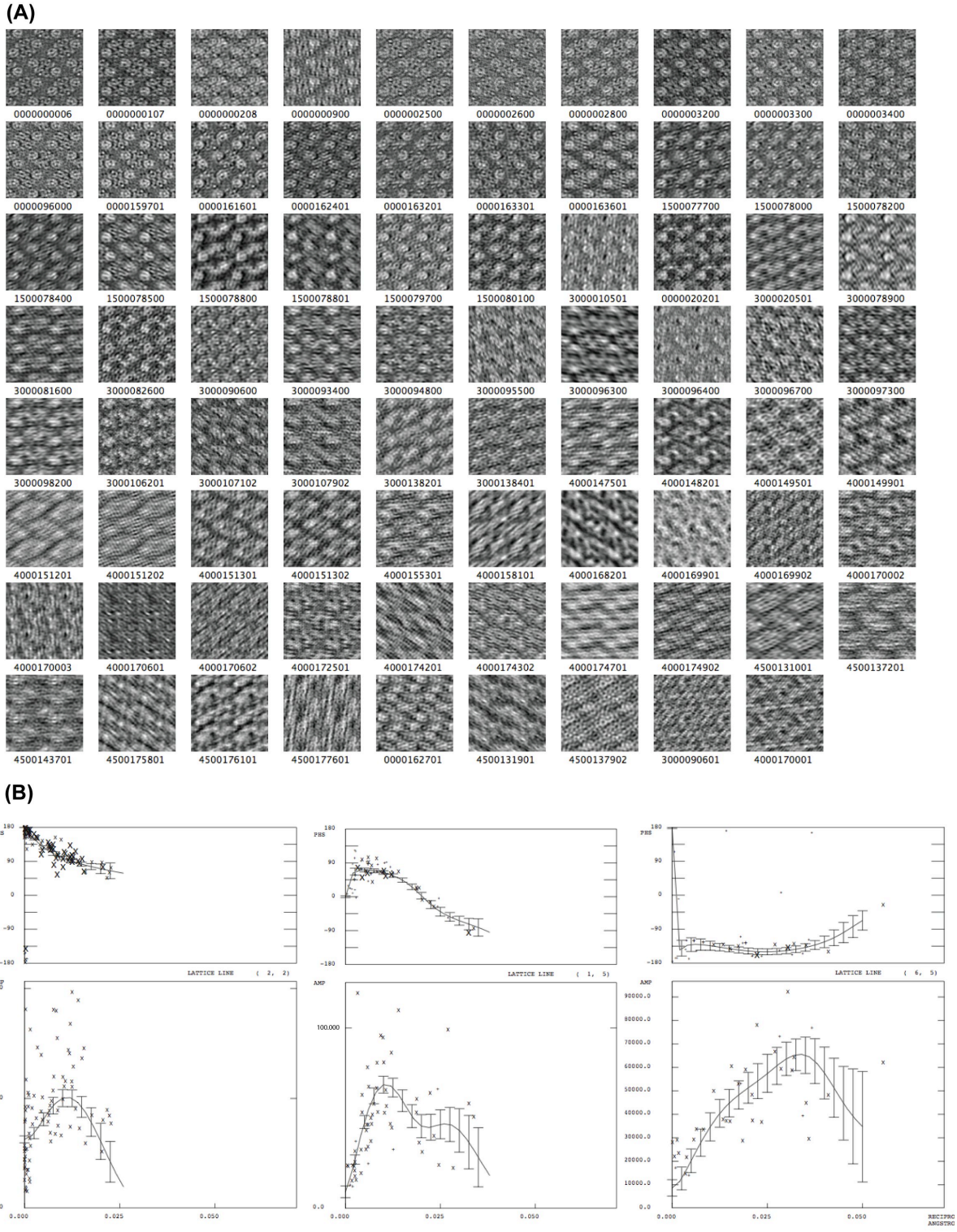


Figure S3.3 Sequence alignment of CitS and *Vc*INDY. CitS and *Vc*INDY show a sequence identity of 13.7 %, which does not allow any homology modeling of CitS.

```
CitS            MTNMSQPPATEKKGVSDLLGFKIFGMPLPLYAFALITLLLSHFYNALPTDIVGGFAIMFI 60
VcIndy          MNRNDSVPLPTNTREWFLHRNSLIVLADVALFLALYHFLPFEHNVVLGISMLAFIAVLWL 60
                *.. .. * . :.    *   .:: :.     :**  :*  ..  .*  .::. :*:::: 

CitS            IGAIFGEIG-KRLPIFNKYIGGAPVMIFLVAAYFVYAGIFTQKEIDAISNVMDKSNFLNL 119
VcIndy          TEALHVTVTAILVPVMAVFFG--IFETQAALNNFANSIIFLFLGGFALAAAMHHQGLDKV 118
                  *:.  :    :*::  ::*   .    .   *. : **      *:: .*.:..: :: 

CitS            FIAVLITGAILSVNRRLLLKSLLGYIPTILMG--IVGASIFGIAIGLVFGIPVDRIMMLY 177
VcIndy          IADKVLAMAQGKMSVAVFMLFGVTALLSMWISNTATAAMMLPLVLGVLSKVDADKQRSTY 178
                :   ::: *  .:.  :::   :  : :: :.   ..* :: :.:*::  : .*:    * 

CitS            VLPIMGGGNGAGAVPLSEIYHS---VTGRSREEYYSTAIAILTIANIFAIVFAAVLDIIG 234
VcIndy          VFVLLGVAYSASIGGIATLVGSPPNAIAAAEVGLSFTDWMKFGLPTAMMMLPMAIAILYF 238
                *: ::* . .*.   :: :  *   . . :.     *    : :.. : ::  *:  : 

CitS            KKHTWLSGEGELVRKASFKVEEDEKTGQITHRETAVGLVLSTTCFLLAYVVAKKILPSIG 294
VcIndy          LLKPTLNGMFELDRAPVNWDKGKVVTLGIFGLTVFLWIFSSPINAALGGFKSFDTLVALG 298
                  :. *.*  ** * .    : .  *  *    . : :. *.    *. . : . * ::* 

CitS            GVAIHYFAWMVLIVAALN----------ASGLCSPEIKAG-------AKRLSDFFSKQLL 337
VcIndy          AILMLSFARVVHWKEIQKTADWGVLLLFGGGLCLSNVLKQTGTSVFLANALSDMVSHMGI 358
                .: :  ** :*      :          ..*** .::          *: ***:.*:  : 

CitS            WVLMVGVGVCYTDLQEIINAITFANVVIAAIIVIGAVLGAAIGGWLMGFFPIESAITAGL 397
VcIndy          FVVILVVATFVVFLTEFASNTASAALLIPVFATVAEAFGMSP-----VLLSVLIAVAASC 413
                :*::: *..  . * *: .  : * ::*..: .:. .:* :       ::.:  *::*. 

CitS            CMANRGGSGDLEVLSACNRMNLISYAQISSRLGGGIVLVIASIVFGMMI 446
VcIndy          AFMLPVATPPNAIVFASGHIKQSEMMRVGLYLNIACIGLLTAIAMLFWQ 462
                .:    .:    :: *..:::  .  ::.  *. . : ::::*.: : 
```

Sequence Identity CitS/*Vc*INDY: 61/446 (13.7%)

Figure S3.4 Merging resolution circle plots for CitS in different substrates. Four resolution circle plots [44], calculated in 2dx, are shown for the merged datasets in (**A**) Na-Acetate, (**B**) Na-Citrate, (**C**) K-Acetate, (**D**) K-Citrate, each limited to 6 Å resolution. Resolutions rings are shown for 36, 24, 18, 12 and 7 Å.

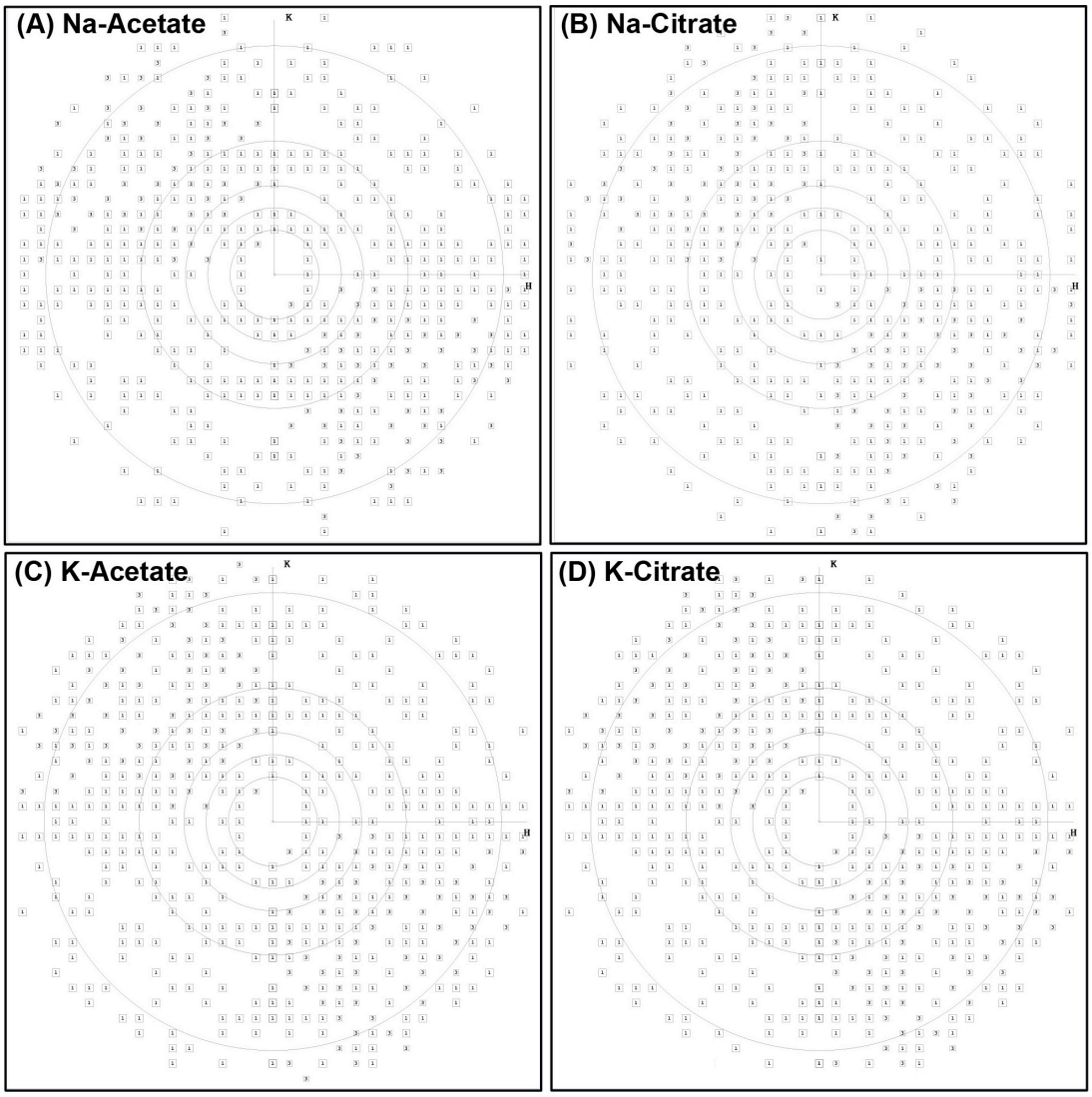
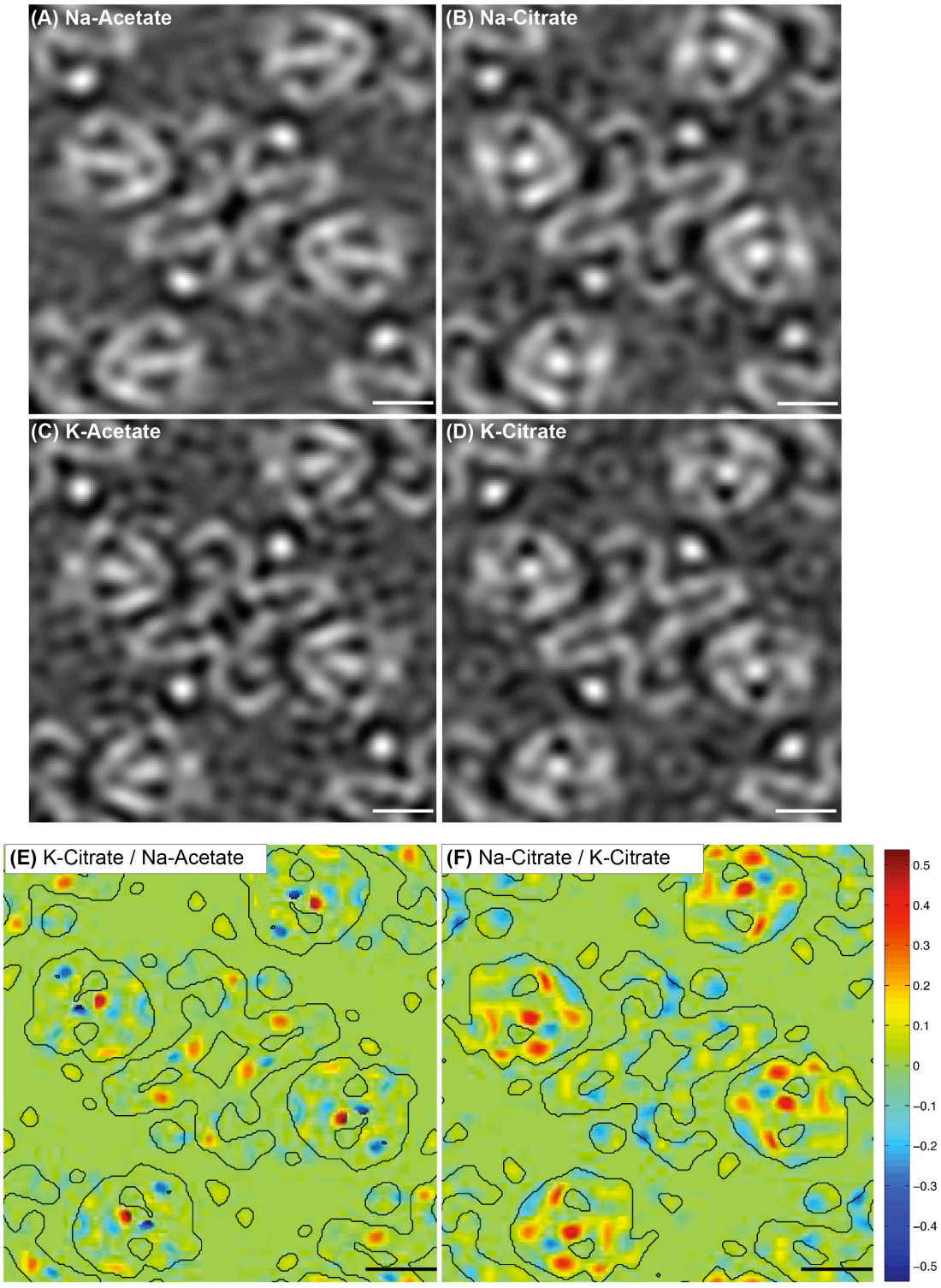


Figure S3.5 CitS projection structures and difference maps in different substrate combinations. Four projection structures of dimeric CitS are shown in (**A**) Na-Acetate, (**B**) Na-Citrate, (**C**) K-Acetate and (**D**) K-Citrate, each at 6 Å resolution. These were used to calculate difference maps for different substrate combinations. Two more difference maps are shown to document the conformational change in (**E**) Na-Acetate/K-Citrate and (**F**) Na-Citrate / K-Citrate. Both maps again reveal helix rearrangements predominantly within the helical cluster; response is minor in the central dimerization interface domain. Scalebar is 2nm.



3.8 References

1. Saier, M. H., Jr. (2000). **A functional-phylogenetic classification system for transmembrane solute transporters**. *Microbiol Mol Biol Rev* **64**, 354-411.
2. Lolkema, J. S. & Slotboom, D. J. (1998). **Estimation of structural similarity of membrane proteins by hydropathy profile alignment**. *Mol Membr Biol* **15**, 33-42.
3. Lolkema, J. S. & Slotboom, D. J. (2003). **Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis**. *J Mol Biol* **327**, 901-909.
4. Ter Horst, R. & Lolkema, J. S. (2012). **Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity**. *Biochim Biophys Acta* **1818**, 72-81.
5. Tsai, C. J. & Ziegler, C. (2010). **Coupling electron cryomicroscopy and X-ray crystallography to understand secondary active transport**. *Curr Opin Struc Biol* **20**, 448-455.
6. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. (2005). **Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters**. *Nature* **437**, 215-223.
7. Forrest, L. R. & Rudnick, G. (2009). **The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters**. *Physiology (Bethesda)* **24**, 377-386.
8. Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y. & Yan, N. (2012). **Crystal structure of a bacterial homologue of glucose transporters GLUT1-4**. *Nature* **490**, 361-366.
9. Forrest, L. R., Krämer, R. & Ziegler, C. (2011). **The structural basis of secondary active transport mechanisms**. *Biochim Biophys Acta* **1807**, 167-188.
10. Abramson, J. & Wright, E. M. (2009). **Structure and function of Na(+)-symporters with inverted repeats**. *Curr Opin Struc Biol* **19**, 425-432.
11. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. (2009). **Unlocking the molecular secrets of sodium-coupled transporters**. *Nature* **459**, 347-355.
12. Jardetzky, O. (1966). **Simple allosteric model for membrane pumps**. *Nature* **211**, 969-970.
13. Law, C. J., Yang, Q., Soudant, C., Maloney, P. C. & Wang, D. N. (2007). **Kinetic evidence is consistent with the rocker-switch mechanism of membrane transport by GlpT**. *Biochemistry* **46**, 12190-12197.
14. Lolkema, J. S., Enequist, H. & van der Rest, M. E. (1994). **Transport of citrate catalyzed by the sodium-dependent citrate carrier of *Klebsiella pneumoniae* is obligatorily coupled to the transport of two sodium ions**. *Eur J Biochem* **220**, 469-475.
15. Pos, K. M. & Dimroth, P. (1996). **Functional properties of the purified Na(+)-dependent citrate carrier of *Klebsiella pneumoniae*: evidence for asymmetric orientation of the carrier protein in proteoliposomes**. *Biochemistry* **35**, 1018-1026.
16. van Geest, M. & Lolkema, J. S. (2000). **Membrane topology of the Na(+)/citrate transporter CitS of *Klebsiella pneumoniae* by insertion mutagenesis**. *Biochim Biophys Acta* **1466**, 328-338.
17. Lolkema, J. S., Sobczak, I. & Slotboom, D. J. (2005). **Secondary transporters of the 2HCT family contain two homologous domains with inverted membrane topology and trans re-entrant loops**. *FEBS J* **272**, 2334-2344.
18. Lolkema, J. S. (2006). **Domain structure and pore loops in the 2-hydroxycarboxylate transporter family**. *J Mol Microbiol Biotechnol* **11**, 318-325.
19. Krupnik, T., Dobrowolski, A. & Lolkema, J. S. (2011). **Cross-linking of dimeric CitS and GltS transport proteins**. *Mol Membr Biol* **28**, 243-253.
20. Kebbel, F., Kurz, M., Grütter, M. G. & Stahlberg, H. (2012). **Projection structure of the secondary citrate/sodium symporter CitS at 6 Å resolution by electron crystallography**. *J Mol Biol* **418**, 117-126.

21. Appel, M., Hizlan, D., Vinothkumar, K. R., Ziegler, C. & Kühlbrandt, W. (2009). **Conformations of NhaA, the Na^+/H^+ exchanger from *Escherichia coli*, in the pH-activated and ion-translocating states**. *J Mol Biol* **388**, 659-672.
22. Goswami, P., Paulino, C., Hizlan, D., Vonck, J., Yildiz, O. & Kühlbrandt, W. (2011). **Structure of the archaeal Na^+/H^+ antiporter NhaP1 and functional role of transmembrane helix 1**. *EMBO J* **30**, 439-449.
23. Mancusso, R., Gregorio, G. G., Liu, Q. & Wang, D. N. (2012). **Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter**. *Nature*.
24. Gipson, B., Zeng, X., Zhang, Z. & Stahlberg, H. (2007). **2dx - User-friendly image processing for 2D crystals**. *J Struct Biol* **157**, 64-72.
25. Gipson, B., Zeng, X. & Stahlberg, H. (2008). **2dx - Automated 3D structure reconstruction from 2D crystal data**. *Microsc Microanal* **14**, 1290-1291.
26. Arheit, M., Castano-Diez, D., Thierry, R., Gipson, B. R., Zeng, X. & Stahlberg, H. (2013). **Image Processing of 2D Crystal Images**. *Methods Mol Biol* **955**, 171-194.
27. Arheit, M., Castano-Diez, D., Thierry, R., Abeyrathne, P., Gipson, B. R. & Stahlberg, H. (2013). **Merging of image data in electron crystallography**. *Methods Mol Biol* **955**, 195-209.
28. Arheit, M., Castano-Diez, D., Thierry, R., Gipson, B. R., Zeng, X. & Stahlberg, H. (2013). **Automation of image processing in electron crystallography**. *Methods Mol Biol* **955**, 313-330.
29. Moscicka, K. B., Krupnik, T., Boekema, E. J. & Lolkema, J. S. (2009). **Projection structure by single-particle electron microscopy of secondary transport proteins GltT, CitS, and GltS**. *Biochemistry* **48**, 6618-6623.
30. Hunte, C., Screpanti, E., Venturi, M., Rimon, A., Padan, E. & Michel, H. (2005). **Structure of a Na^+/H^+ antiporter and insights into mechanism of action and regulation by pH**. *Nature* **435**, 1197-1202.
31. Sobczak, I. & Lolkema, J. S. (2005). **Loop VIII/IX of the Na^+-citrate transporter CitS of *Klebsiella pneumoniae* folds into an amphipathic surface helix**. *Biochemistry* **44**, 5461-5470.
32. Dobrowolski, A., Fusetti, F. & Lolkema, J. S. (2010). **Cross-linking of trans reentrant loops in the Na(+)-citrate transporter CitS of *Klebsiella pneumoniae***. *Biochemistry* **49**, 4509-4515.
33. Sobczak, I. & Lolkema, J. S. (2005). **The 2-hydroxycarboxylate transporter family: physiology, structure, and mechanism**. *Microbiol Mol Biol Rev* **69**, 665-695.
34. Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S. & Tate, C. G. (2003). **Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer**. *EMBO J* **22**, 6175-6181.
35. Perez, C., Khafizov, K., Forrest, L. R., Krämer, R. & Ziegler, C. (2011). **The role of trimerization in the osmoregulated betaine transporter BetP**. *EMBO Rep* **12**, 804-810.
36. Herz, K., Rimon, A., Jeschke, G. & Padan, E. (2009). **Beta-sheet-dependent dimerization is essential for the stability of NhaA Na^+/H^+ antiporter**. *J Biol Chem* **284**, 6337-6347.
37. Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N. G., Hadden, J. M., Sharples, D., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. (2010). **Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1**. *Science* **328**, 470-473.
38. Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K. & Gouaux, E. (2007). **Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter**. *Nature* **445**, 387-393.
39. van der Rest, M. E., Molenaar, D. & Konings, W. N. (1992). **Mechanism of Na(+)-dependent citrate transport in *Klebsiella pneumoniae***. *J Bacteriol* **174**, 4893-4898.
40. Jap, B. K., Zulauf, M., Scheybani, T., Hefti, A., Baumeister, W., Aebi, U. & Engel, A. (1992). **2D crystallization: from art to science**. *Ultramicrosc* **46**, 45-84.

41. (1994). **The CCP4 suite: programs for protein crystallography**. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.
42. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). **UCSF Chimera - a visualization system for exploratory research and analysis**. *J Comput Chem* **25**, 1605-1612.
43. Crowther, R. A., Henderson, R. & Smith, J. M. (1996). **MRC image processing programs**. *J Struct Biol* **116**, 9-16.
44. Cheng, A. & Yeager, M. (2004). **A graphical representation of image quality for three-dimensional structure analysis of two-dimensional crystals**. *Acta Crystallogr A* **60**, 351-354.

Chapter 4 - The G protein-coupled receptor CCR5

The following section has been published in:

Journal of Biomolecular NMR (2013) 55:79-95 (doi:10.1007/s10858-012-9688-4)

Biophysical and structural investigation of bacterially expressed and engineered CCR5, a G protein-coupled receptor

Maciej Wiktor[1], Sébastien Morin[1], Hans-Jürgen Sass[1], Fabian Kebbel[2] and Stephan Grzesiek[1,*]

[1] Focal Area Structural Biology and Biophysics, Biozentrum, University of Basel, Klingelbergstraße 50/70, 4056 Basel, Switzerland
[2] Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland

*Corresponding author

My contribution to this study was the continuous electron microscopic analysis and evaluation of CCR5 preparations, as summarized in figure 4.5.

4.1 Abstract

The chemokine receptor CCR5 belongs to the class of G protein-coupled receptors. Besides its role in leukocyte trafficking, it is also the major HIV-1 coreceptor and hence a target for HIV-1 entry inhibitors. Here, we report *Escherichia coli* expression and a broad range of biophysical studies on *E. coli*-produced CCR5. After systematic screening and optimization, we obtained 10 mg of purified, detergent-solubilized, folded CCR5 from 1L culture in a triply isotope-labeled ($^{2}H/^{15}N/^{13}C$) minimal medium. Thus the material is suitable for NMR spectroscopic studies. The expected α-helical secondary structure content is confirmed by circular dichroism spectroscopy. The solubilized CCR5 is monodisperse and homogeneous as judged by transmission electron microscopy. Interactions of CCR5 with its ligands, RANTES and MIP-1β were assessed by surface plasmon resonance yielding K_D values in the nanomolar range. Using size exclusion chromatography, stable monomeric CCR5 could be isolated. We show that cysteine residues affect both the yield and oligomer distribution of CCR5. HSQC spectra suggest that the transmembrane domains of CCR5 are in equilibrium between several conformations. In addition we present a model of CCR5 based on the crystal structure of CXCR4 as a starting point for protein engineering.

4.2 Introduction

G protein-coupled receptors constitute a large protein superfamily found only in eukaryotes. About 4 % of the protein-coding human genome codes for ~800 GPCRs[1]. Based on phylogenetic analysis human GPCRs cluster into 5 main families: rhodopsin, adhesion, frizzled/taste2, glutamate and secretin, which comprise 701, 24, 24, 15 and 15 members, respectively[2]. The diversity of the GPCR superfamily members is reflected in the variety of their ligand types. Photons, ions, odorants, nucleotides, fatty acids, amino acids, peptides and proteins are only some of the messages that GPCRs can transduce[3]. As GPCRs regulate so many physiological processes such as vision, smell, behavior, mood, immune system, blood pressure, heart rate, digestion or homeostasis, they remain the most commonly drugged protein family[4]. About 40 % of prescribed pharmaceuticals target GPCRs[5]. The structure determination of membrane proteins is notoriously difficult due to the many obstacles impeding membrane protein sample preparation and subsequent structure determination. When this publication was written, the Protein Data Bank[6] contained about 86,000 entries, but only 364 unique membrane protein 3D structures (http://blanco.biomol.uci.edu)[7]. Solved GPCR structures are even sparser. Until now 16 unique GPCR structures have been solved by X-ray crystallography: the first being bovine rhodopsin[8] followed by β_2-adrenergic[9,10], β_1-adrenergic[11], adenosine A_{2A}[12,13], dopamine D3[14], CXCR4[15] and several others. To obtain high-resolution structural data the replacement of the intracellular (IC) loop three with T4 lysozyme[9], thermostabilization[11] or stabilization by anti- or nanobodies[9] proved to be successful strategies. Additionally, all crystallized GPCRs were bound to an agonist[13], an inverse agonist[9,10,16] or most often to an antagonist[11,12,14,15,17-21]. Although not GPCRs, prokaryotic sensory rhodopsin II[22] and proteorhodopsin[23] are examples of 7-TM domain proteins solved by solution NMR spectroscopy. Very recently the structure of *E. coli*-expressed and refolded CXCR1 has been determined in phospholipid bilayers using solid state NMR spectroscopy[24].

CCR5 (CC chemokine receptor 5) belongs to the γ-group of the rhodopsin family of GPCRs. It is found in the plasma membrane of Th1 lymphocytes, macrophages, NK cells and immature dentritic cells and is involved in various infectious and inflammatory diseases as well as cancer[25]. Since humans carrying the Δ32 allele of the CCR5 gene, a 32-base pair deletion resulting in a premature stop codon in the extracellular (EC) loop 2 and a nonfunctional receptor, are healthy, the exact role of CCR5 is not completely understood. The main interest in CCR5 is, however, a consequence of its involvement in AIDS. R5-tropic HIV-1 infection necessitates the sequential interaction of viral envelope glycoprotein gp120 with CD4 and CCR5[26]. Two copies of the CCR5-Δ32 allele confer nearly complete resistance to HIV-1 infection[27,28]. Δ32 occurs at 5–14 % frequency in European Caucasians but not in African, Native American, and East Asian populations[29]. This is hypothesized to be a result of pandemics that took place in Europe in medieval ages[30]. Successful strategies to block HIV-1 entry have been developed based on small-molecule inhibitors of CCR5[31] as well as derivatives of its natural chemokine ligand RANTES[32-35].

High-resolution structural data would greatly improve the understanding of CCR5 function and the nature of its interaction with the chemokine ligands RANTES, MIP-1α, and MIP-1β, as well as substantially enhance possibilities for anti-HIV-1 drug discovery. So far it has been very challenging to obtain sufficient amounts of this protein suitable for structural studies. Large-scale CCR5 expression at the yield of 1 mg/L was reported in insect cells[36] where

screening for mutants is time-consuming and isotope labeling is very costly and has not been achieved for deuterium. Alternatively, 1–3 mg/L of CCR3 but only 0.1–0.3 mg/L of CCR5 was obtained from *E. coli* after fusing the N-terminus of the chemokine receptor to the C-terminus of thioredoxin[37]. However, the described expression system relied on the usage of rich TB medium and ligand binding of the expressed receptors was not shown. Nevertheless, there is a growing number of various GPCRs functionally expressed in *E. coli* [38-44], including the chemokine receptor CXCR1, which was expressed as a GST-CXCR1 fusion construct in $^{15}N/^{13}C$-labeled form at 5 mg/L and after reconstitution to proteoliposomes could bind IL-8 and activate G protein[24].

Petrovskaya et al. have compared direct expression of 17 diverse GPCRs in *E. coli* to hybrid expression with the N-terminal fusion partners OmpF or Mistic[45]. Interestingly, almost all GPCRs expressed in the presence of a fusion partner at 5 mg/L yield, but for most the expression was severely reduced in its absence. Thanks to a better access to isotope labeling bacterial or yeast expression systems are preferred for NMR, however, a significant progress has been recently made in isotope labeling in mammalian cells, which, unlike bacteria or yeast, provide possibilities to obtain human posttranslational modifications[46].

Here we report a CCR5 production platform that yields up to 10 mg of purified protein per 1 L of bacterial culture. CCR5 is solubilized from *E. coli* without the requirement of refolding. As the expression conditions were optimized in minimal medium, triple isotope ($^{2}H/^{13}C/^{15}N$) labeling does not compromise the yield. In order to boost the expression, we fused the N-terminus of CCR5 to well expressing small proteins or signal sequences. A C-terminal 10His-tag and rigorous washing conditions yield over 90 % purity after a single IMAC purification step. The fusion partner can be readily and quantitatively cleaved off by thrombin and separated on a size exclusion column, where CCR5 monomers and dimers migrate as separate symmetric peaks. Both monomers and dimers are monodisperse and homogeneous as judged from electron micrographs. The expected α-helical secondary structure content is confirmed by circular dichroism (CD) spectroscopy. When solubilized in a DDM/CHAPS/CHS/DOPC mixture CCR5 interacts with RANTES, MIP-1β and 2D7 with nanomolar affinities. Recorded ^{1}H-^{15}N HSQC spectra suggest that the TM domains of CCR5 are in equilibrium between several conformations. We also show that the number of cysteine residues has a severe impact on both protein yield and oligomeric state. Following Hernanz-Falcon et al.[47], two point mutations I52V and V150A were introduced to reduce the tendency of dimer formation, but no such reduction was observed. Our system establishes a high-yield platform for biophysical and structural studies on CCR5.

4.3 Materials & Methods

4.3.1 Generation of expression constructs

Plasmids pET28F10 and pMT10H10 containing the CCR2b sequence fused to OmpF and Mistic were a generous gift from Prof. A. Arseniev (Russian Academy of Sciences, Moscow, Russia). Plasmid pCA528 was kindly provided by Prof. A. Spang (Biozentrum, Basel, Switzerland). pET vectors were obtained from Novagen. The *E. coli*-optimized CCR5 DNA sequence in the pQE-T7 vector was generated by GeneArt. The CCR5 gene was cloned using standard molecular biology

techniques. Plasmid DNA was amplified with the QIAprep Spin Miniprep Kit (Qiagen). Point mutations were carried out using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). DNA sequences of the cloned constructs can be found in the supporting information Text S4.1 - S4.10.

4.3.2 Protein expression

Freshly transformed Rosetta 2 (DE3) Competent Cells (Novagen) were transferred to 1–2 L of M9 medium after overnight growth on LB agar plates. The cultures were shaken in 5 L baffled flasks at 100 rpm at 37 °C until OD_{600} = 2.6–2.8. The cultures were cooled down on ice with occasional shaking until the temperature dropped to 20–25 °C. CCR5 expression was induced with 1 mM IPTG and the cultures were shaken at 100 rpm at 20 °C. After harvesting, cells were pelleted and stored at -70 °C. For expression in D_2O transformed cells were grown on LB agar plates prepared in 50 % D_2O. 1–2 L cultures were preceded by 100 mL precultures grown until OD_{600} = 1. All compounds used in the preparation of M9 medium in D_2O (including trace elements, vitamins, antibiotics) were prepared in 99.8 % D_2O. Uniform ^{15}N- and ^{13}C-labeling was carried out using $^{15}NH_4Cl$ (98 % ^{15}N, 1 g/L), and [$^{1}H/^{13}C_6$]-D-glucose (99 % ^{13}C, 4 g/L) as the sole nitrogen and carbon sources, respectively. Hence, the labeling efficiency is expected as 98 % for ^{15}N and 99 % for ^{13}C. Judging from strong peaks in the HN(CO)CA spectrum, which showed no signs of $^{1}J_{CH}$ splitting in the absence of ^{1}H decoupling during $^{13}C^{\alpha}$ evolution, the deuteration efficiciency is estimated as 80 %. This is in agreement with the data of Otten et al.[48], and is consistent with bacterial metabolism[49]. Details of the M9 medium composition can be found in Text S4.11. Expression of WT CCR5 in insect cells was performed as described previously[36].

4.3.3 Membrane fraction preparation

Frozen *E. coli* cell pellet (1 g) was suspended in 6–8 mL of buffer A (20 mM HEPES pH 7, 150 mM NaCl, 10 % (v/v) glycerol) supplemented with 0.5 mM PMSF, 5 mM benzamidine and EDTA-free complete protease inhibitor cocktail (Roche). Cells were broken using a French press at 31,600 psi. Cell debris was removed by centrifugation at 6,600g for 15 min. The supernatant was centrifuged at 126,000g for 15 min, and the resulting pellet (from now on called membrane fraction) collected. After suspending in buffer A, a 20 % (w/v) solution of the membrane fraction was stored at -70 °C. The preparation of the insect cell membrane fraction was carried out as described previously[36].

4.3.4 Detergent screening

Frozen 20 % (w/v) solutions of the membrane fraction were thawed, diluted twice and supplemented with detergent to the final concentration of 2 %. Solubilization was carried out at RT for 2 h with 1,000 rpm shaking. Unsolubilized material was removed by centrifugation at 100,000g for 30 min. The clarified supernatant (2 μL) was loaded onto a Protran BA85 nitrocellulose membrane (Whatman) and dried at RT. Dot blots were blocked, labeled with anti-His-tag antibody, developed and quantified in the same way as western blots described below. Detergents were obtained from Anatrace with the exception of 1,2-diheptanoyl-sn-glycero-3-phosphocholine (DHPC, Avanti Polar Lipids).

4.3.5 Protein purification

A frozen 20 % (w/v) solution of membrane fraction was thawed and supplemented to a final concentration of 0.5 M NaCl, 20 mM KCl, 10 mM $MgCl_2$ and 2.5 % FosCholine-12 (FC-12). Protein solubilization was carried out at 4–8 °C for 1–2 h. Unsolubilized material was removed by centrifugation at 126,000g for 30 min. The clarified supernatant was supplemented with 35 mM imidazole and bound to Ni–NTA beads (Qiagen) for 2 h. The resin was washed with 100 column volumes of buffer B (20 m M HEPES pH 7, 1 M NaCl, 60 mM imidazole, 10 % (v/v) glycerol, 0.1 % FC-12). The protein was then eluted with buffer C (20 mM HEPES pH 7, 150 mM NaCl, 0.4 M imidazole, 0.15 % FC-12). Protein-rich fractions were pooled and dialyzed against buffer D (20 mM Tris pH 8, 150 mM NaCl, 0.5 mM EDTA, 0.1 % FC-12). To cleave the fusion partner, 2 U of thrombin per 1 mg of purified protein was sufficient to complete the cleavage over 16 h at RT. The protein was concentrated using a 30 kDa molecular weight cut off (MWCO) concentrator and injected onto Superdex 200 10/300 GL (analytical run) or Superdex 200 26/60 HiLoad (preparative run) columns equilibrated in buffer E (20 Na_2HPO_4 pH 7.4, 180 mM NaCl, 0.1 % FC-12).

4.3.6 Gel electrophoresis and western blotting

Protein samples for SDS-PAGE were mixed with 5x SDS loading buffer (312.5 mM Tris–HCl pH 6.8, 50 % (v/v) glycerol, 25 % β-mercaptoethanol, 10 % SDS, 0.0125 % bromophenol blue), incubated at 30 °C for 15 min and centrifuged at 17,000g for 5 min prior to loading on a 4–20 % gradient precast gel (Pierce). The electrophoresis was performed at 100 V constant voltage. Gels were stained using 0.25 % solution of Coomassie Brilliant blue R-250 (AppliChem) in 25 % isopropanol and 10 % acetic acid and destained in 10 % acetic acid. For western blotting onto PVDF membrane (Bio-Rad), a Criterion Blotter (Bio-Rad) was used. The transfer was performed at 0.5 A constant current for 1 h in the transfer buffer (48 mM Tris–HCl pH 9.2, 39 mM glycine, 0.375 % SDS, 20 % methanol). The membrane was blocked with 3 % BSA in TBST buffer (10 mM Tris–HCl pH 8, 150 mM NaCl, 0.5 % Tween-20). Subsequently, the membrane was incubated with mouse monoclonal HIS-1 anti-polyhistidine-peroxidase antibody (Sigma-Aldrich) at 1:6,000 dilution for 1 h. After washing 4 x 2 min with TBST buffer, the blot was developed using chemiluminescent HRP substrate (Roche). The signal was recorded using a BioMax XAR Film (Kodak) or using a LAS-4000 luminescent image analyzer (Fujifilm). The signal intensities were quantified using ImageJ 1.43r [50].

4.3.7 Transmission electron microscopy

For transmission electron microscopy (TEM) analysis 5 μL of 10 μg/mL protein solution was adsorbed on carbon-coated copper 200 mesh grids rendered hydrophilic by glow discharge in air during 20 s. The grids were washed in five drops of double distilled water and negatively stained with two drops of 2 % uranyl acetate. Electron micrographs were recorded on a Philips CM10 instrument equipped with a LaB_6 filament operating at an accelerating voltage of 80 kV. Images were recorded at nominal defocus values of 0.5 μm on a Veleta CCD camera at a nominal magnification of 130,000, corresponding to a pixel size of 3.7 Å at the sample level.

4.3.8 CD spectroscopy

Circular dichroism spectra were recorded on 3–13 μM monomeric CCR5 fractions. Measurements were performed on a Chirascan CD spectrometer (Applied Photophysics) at 20 °C in 1 mm quartz Suprasil cuvettes (Hellma). Typically, spectra in a wavelength range of 195–260 nm spectra were recorded in triplicates and averaged. After baseline (buffer) subtraction, the mean residue molar ellipticity Θ_{MRM} was calculated from the following equation $\Theta_{MRM} = \Theta / (C * n * l)$, where Θ is the ellipticity (deg), C is the concentration (mol/L), n is the number of residues and l is the optical path length (cm). The relative α-helical content α_r was calculated as follows $\alpha_r = (-\Theta_{MRM,222\ nm} + 3{,}000)/ 39{,}000$ [51], where Θ_{MRM} is given in units of deg * cm^2 * $dmol^{-1}$.

4.3.9 Surface plasmon resonance

Surface Plasmon resonance (SPR) interaction assays were performed using a T100 Biacore instrument (GE Healthcare) at 20 °C. The setup consisted of a CM5 chip on which an antibody against the His-tag (Qiagen) was immobilized, using amine coupling chemistry. The antibody (4,000–10,000 RU) could capture ~2,000–5,000 RU of recombinant His-tagged CCR5, solubilized from membranes using a detergent mixture of 1 % DDM, 1 % CHAPS, 0.2 % CHS, and 1 mM DOPC at pH 7. Thioredoxin removal was performed on the chip using 5 U of thrombin injected in 300 μL over 60 min (5 μL/min). Experiments were performed in buffer F (20 mM HEPES pH 7.0, 150 mM NaCl, 0.1 % DDM, 0.1 % CHAPS, 0.02 % CHS, 50 nM DOPC, 0.1 mg/mL BSA) with a flow rate of 50 μL/min. Signals were processed with the Biacore T100 Evaluation Software using double referencing with both a reference channel and blank injections.

4.3.10 NMR

Several samples of FC-12-solubilized m11CCR5 (monomeric fraction) produced in isotope labeled M9 medium were concentrated in a 30 kDa MWCO Ultracel-30 K Amicon Ultra Centrifugal Filter (Millipore) to 100–200 μM (~2–3 % FC-12) and supplemented with 5 % D_2O. All spectra were recorded in Shigemi tubes on a Bruker DRX800 spectrometer equipped with a triple resonance Z-gradient TCI cryoprobe. ^{1}H-^{15}N TROSY[52] spectra were acquired under various buffer and temperature conditions (see text) as data matrices of 63*(^{15}N, t1) x 512*($^{1}H^{N}$, t2) data points (where n* refers to the number of complex points) with acquisition times of 25 ms (^{15}N) and 40 ms ($^{1}H^{N}$). Standard three-dimensional triple resonance TROSY spectra for backbone assignment [53] were recorded on a sample of 200 μM uniformly $^{2}H/^{13}C/^{15}N$-labeled CCR5 in ~3 % FC-12 Foscholine at 20 °C. Experimental times were HNCO: 5.5 days, HNCA: 2.7 days, HN(CO)CA: 2.7 days, HN(CA)CO: 6.4 days, and HNCACB: 7.3 days. All spectra were processed using NMRPipe[54].

4.3.11 CCR5 model building

The core of CCR5 (residues 19–298) was built using the SWISS-MODEL server (http://swissmodel.expasy.org/workspace) with the crystal structure of CXCR4 (3ODU[15], 32 % sequence identity) as a template. At the C-terminus of CCR5, helix H8 modeled based on the rhodopsin structure 3C9L[55] was added using VMD 1.9[56]. In addition, the N-terminus of CCR5

(residues 1–18) and another part of the C-terminus including palmitoylated cysteines (residues 312–331) were added as an extended amino acid chain. Residues 332–352 were not included to reduce computational time. Finally, sulfate groups were added to Tyr10 and Tyr14 as well as palmitoyl groups to Cys321, Cys323 and Cys324. After each manipulation step the structure was energy-minimized and relaxed by a short molecular dynamic simulation (MD) run using NAMD 2.7[57]. For these MD runs the protein was embedded in a lipid bilayer of 137 POPC molecules, hydrated with 10,774 TIP3 water molecules and neutralized by adding Na^+ and Cl^- ions. The final structure was embedded in a bilayer of 188 POPC molecules, hydrated with 20,781 TIP3 water molecules, relaxed with several short (≤1 ns) equilibration steps and finally equilibrated with a 10 ns MD run.

4.4 Results

4.4.1 Protein expression

Even though many approaches are described in the literature, there is no universally applicable strategy to obtain a high yield GPCR expression system. The selection of expression vector, bacterial strain, culturing conditions etc. remains largely empirical. To increase the chance of achieving high yield, we tested the expression of CCR5 cloned into several different T7-inducible vectors containing various N- and C-terminal fusion partners/tags. The summary of tested constructs can be found in table 4.1.

Table 4.1. Summary of GPCR constructs tested for expression

Vector	N-term. Tag	Fusion partner	Cleavage site	GPCR[b]	cDNA	C-term. Tag	Expression
pET28F10	–	OmpF (1–34/362)[a]	–	CCR2b (1)	*H. sapiens*	6His	+++
pMT10H10	–	Mistic (1–110/110)	Thrombin	CCR2b (3)	*H. sapiens*	10His	+++
pET-22b	–	pelB (1–22/374)	pelB	CCR5 (8)	*H. sapiens*	8His	+
pGEV2	–	GB1 (1–56/56)	Thrombin	CCR5 (7)	*H. sapiens*	8His	+++
pQE-T7	6His	–	TAGZyme	CCR5 (6)	*E. coli*	–	+
pET28F10	–	OmpF (1–34/362)	–	m7CCR5^{306} (2)	*E. coli*	6His	+++
pMT10H10	–	Mistic (1–110/110)	–	m7CCR5^{306} (4)	*E. coli*	10His	+++
pET-41a	–	GST (1-218/218)	–	m7CCR5^{306} (10)	*E. coli*	10His	+++
pCA528	6His	SUMO (1–98/101)	Ulp1	m7CCR5^{306} (8)	*E. coli*	10His	+++
pET-32b	–	TrxA (1–109/109)	Thrombin	m7CCR5^{306} (5)	*E. coli*	10His	+++

[a]Residues 1–34 from 362 total. [b]The most frequently used constructs are listed. A comprehensive list of constructs with their DNA sequences is given in Supplementary Text S4.1-S4.10. () refer to the numbers in Text S4.1-S4.10.

As we intended to use the expression system also for isotope labeling, expression was carried out in M9 minimal medium supplemented with Hutner's trace elements[58]. To neutralize the codon bias in some of the constructs we used Rosetta 2 (DE3) cells carrying the pRARE plasmid encoding for rare tRNAs.

GPCR overexpression was assayed by western blot for each of the cloned constructs. The expression in pET-22b and pQE-T7 vectors, which provide no or only a very small fusion partner, was clearly lower than in the others. This suggests that CCR5 expression yield benefits from the N-terminal fusion partner. However, the type of the fusion partner seems of much less importance than expected (Table 4.1). Therefore, shortly after the preliminary screening, the work was restricted to the TrxA-CCR5 fusion construct, which was selected because of its high yield, purity and convenience of separation, and since it can be directly compared to the analogous expression system for chemokine receptors developed by Ren et al.[37]. For every tested fusion construct, the yield was significantly higher at 20 °C than at 37 °C (Figure 4.1a/b). A further decrease of the temperature to 12 °C or a decrease of IPTG concentration from 1 mM to 0.1 mM resulted in a lower yield (data not shown). The highest yield was achieved at 20 °C at 24–48 h after induction (Figure 4.1a–d).

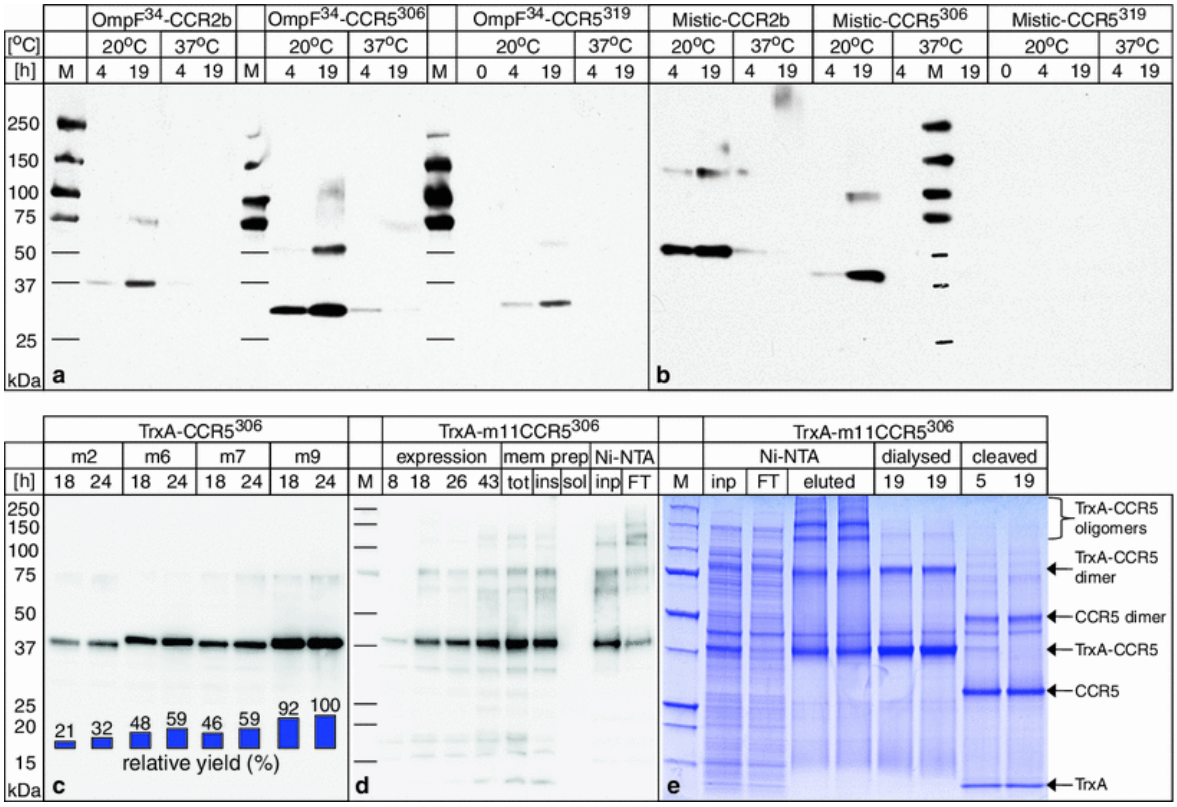


Figure 4.1 Summary of the expression and purification of various CCR5 constructs in *E. coli* monitored by western blot and SDS-PAGE. Comparison of the expression of longer (1–319) and shorter (1–306) versions of OmpF^{34}-m2CCR5 (**a**) and Mistic-m2CCR5 (**b**) constructs at 20 °C and 37 °C. CCR2b constructs are used as a positive control. (**c**) Comparison of the expression of various Cys mutants of TrxA-CCR5^{306}. (**d**) Expression, membrane preparation and binding to Ni–NTA of TrxA-m11CCR5^{306}. Broken *E. coli* cells expressing CCR5 were centrifuged to remove cell debris. The remaining suspension (tot) was subsequently separated into insoluble membrane (ins) and soluble cytoplasmic (sol) fractions. CCR5 was found in the membrane fraction (ins) but not in the cytoplasmic fraction (sol). Solubilized membranes (inp) were loaded on Ni–NTA. (**e**) Purification of m11TrxA-CCR5^{306}. After elution from Ni–NTA oligomerized m11TrxA-CCR5^{306} was dialyzed and digested with thrombin.

For further optimization of the protein construct, it was important to anticipate the sequence-specific position of the secondary structure elements. Initially, the constructs were based on the two-dimensional topology predicted by Oppermann[59]. However, after the crystal structure of CXCR4[15] became available, we generated a homology model based on the latter structure and the C-terminal helix H8 of rhodopsin[55] using state-of-the-art molecular dynamics energy minimization in explicit solvent of CCR5 embedded in a lipid bilayer. The result of the simulation is shown as a full structural model in figure 4.2 and the subsequently derived secondary structure topology in figure 4.3a.

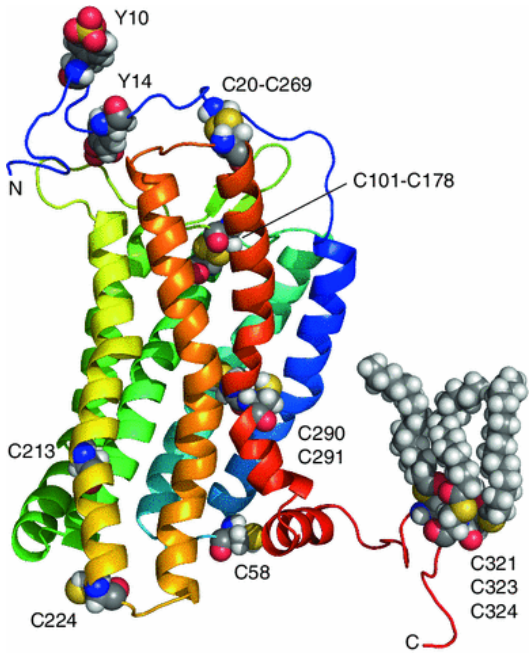
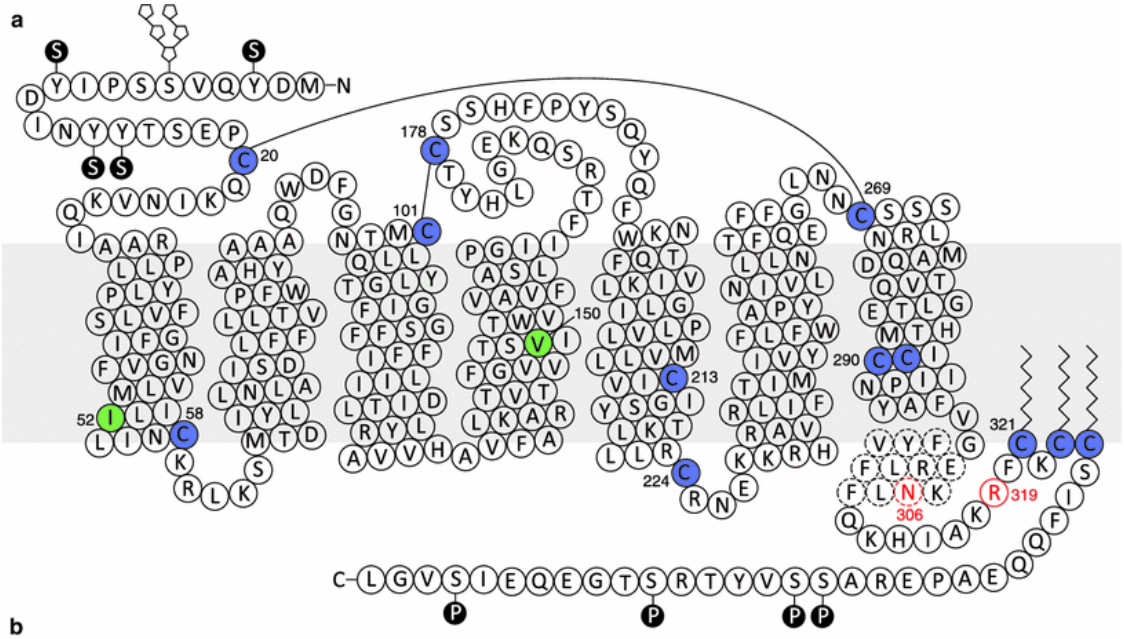


Figure 4.2 Modeled 3D structure of CCR5 (residues 1-331) based on the CXCR4 structure. Sulfation of Tyr10 and Tyr14 as well as palmitylated Cys321, Cys323 and Cys324 are depicted as spheres.

Anticipating problems with the formation of intermolecular disulphide bridges we have systematically tested the role of all 12 cysteines by the truncation of the cysteine-containing C-terminus (after N306 or R319) and site-directed mutagenesis of the remaining 9 cysteines in other regions. In these regions, solvent-exposed cysteines were mutated to serines, whereas cysteines in the TM domains were replaced by alanines. The locations of the respective residues are highlighted in figures 4.2/4.3a, and the naming convention of the various mutants is listed in figure 4.3b.

The expression of these cloned constructs was monitored by western blotting against the C-terminal His-tag. The signal from the shorter (1–306) OmpF^{34}-m2CCR5^{306} and Mistic-m2CCR5^{306} constructs was stronger than from the longer (1–319) OmpF^{34}-m2CCR5^{319} and Mistic-m2CCR5^{319} constructs (Figure 4.1a/b). From this observation, we concluded that the shorter constructs were either expressing better or were more resistant to C-terminal degradation. Therefore, further work was limited to the shorter (1–306) CCR5 constructs (CCR5^{306}). Within the latter, a negative correlation exists between the expression yield and the number of cysteine residues (Figure 4.1c). Thus, TrxA-m2CCR5^{306} construct containing 7 cysteines (Figure 4.3b) expressed worse than TrxA-m7CCR5^{306} (4 Cys) or TrxA-m6CCR5^{306} (3 Cys), and much worse than TrxA-m9CCR5^{306} (0 Cys).



Mutant	Extracellular Cys	Transmembrane Cys	Intracellular Cys	Other Mutations
m2CCR5^{306}	----------	----------	C58S C224S	----------
m6CCR5^{306}	C20S C101S C178S C269S	----------	C58S C224S	----------
m7CCR5^{306}	----------	C213A C290A C291A	C58S C224S	I52V V150A
m9CCR5^{306}	C20S C101S C178S C269S	C213A C290A C291A	C58S C224S	----------
m11CCR5^{306}	C20S C101S C178S C269S	C213A C290A C291A	C58S C224S	I52V V150A

Figure 4.3. CCR5 topology and engineered mutations. (a) Membrane topology prediction of the human CCR5 according to the CXCR4 homology model (Figure 4.2). The grey rectangle approximates the position of the membrane. EC (IC) space is at the top (bottom). The potential posttranslational modifications include sulfation of Y3, Y10, Y14 and Y15, phosphorylation of S336, S337, S342 and S349 (both marked as black circles), palmitoylation of C321, C323 and C324 as well as glycosylation of S6. The positions of mutated residues are highlighted (C in blue, other in green). C-terminal truncations are marked with red circles and potential helix H8 with dashed lines. Disulphide bridges form between C20 and C269 and between C101 and C178. **(b)** Table summarizing the introduced point mutations of the listed CCR5 mutants.

4.4.2 Detergent screening

A good detergent for membrane protein studies should be able to solubilize the protein, keep it stable and functional in solution as well as allow structural studies. In order to explore the possible detergent space, we performed a systematic screen by solubilizing *E. coli* membrane fractions in various detergents at 2 % (w/v) concentration. After removal of the unsolubilized material, the clarified solutions were dried on a nitrocellulose membrane and analyzed by dot blot using an anti-His antibody. The chemiluminescent signal was quantified densitometrically and normalized to the maximum value (Figure 4.4a). The results indicate that OmpF^{34}-m7CCR5^{306} was efficiently solubilized by anionic (sodium dodecanoyl sarcosine and SDS) and zwitterionic detergents (FosCholines and dimethyl glycines) with aliphatic chains. The cationic trimethylammonium chlorides and the zwitterionic Anzergents were intermediate to moderate in their solubilization efficiency. Nonionic detergents (maltosides and Anapoes) turned out to solubilize OmpF^{34}-m7CCR5^{306} extremely poorly with the single exception of tetradecylmaltoside, which solubilized about a third as much as FosCholines.

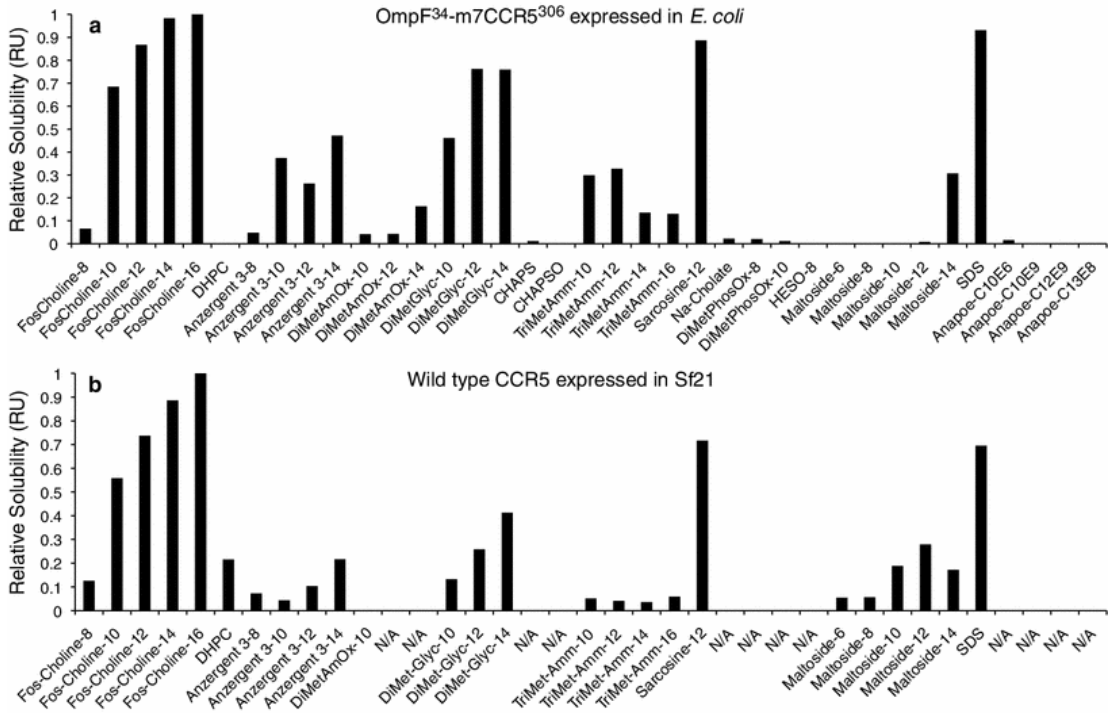


Figure 4.4 Detergent screening for solubilization of OmpF^{34}-m7CCR5^{306} expressed in *E. coli* (a) and wild-type CCR5 expressed in *Sf*21 cells (b). Values were normalized against FC-16. DHPC, DiMetPhOx-10, n-decyl-N,N-dimethylamine-N-oxide; TriMetAmm-10, N-dodecyltrimethylammonium chloride; Sarcosine-12, sodium dodecanoyl sarcosine; DiMetPhOx-8, dimethyloctylphosphine oxide; HESO-8, N-octyl-2-hydroxyethyl sulfoxide; Maltoside-6, n-hexyl-β-d-maltopyranoside.

These results on *E. coli* OmpF^{34}-m7CCR5^{306} are similar to a solubility screen carried out on wild-type CCR5 expressed in *Sf*21 cells (Figure 4.4b). Analogous to *E. coli* CCR5, the insect cell protein was efficiently solubilized by sodium dodecanoyl sarcosine, SDS and FosCholines. Dimethyl glycines, Anzergents and trimethylammonium chlorides solubilized relatively worse and maltosides somewhat better, but still not very efficiently. Due to its relatively mild character and lipid-like headgroup we picked FC-12 as the main working detergent. Even though FosCholines with longer hydrocarbon tails performed better, they are much less suitable for NMR due to their high aggregation number and lower solubility.

4.4.3 Protein purification and identity confirmation

Considering a broad scope of applications we sought to establish a simple, robust and efficient purification scheme. Fractionation by centrifugation of the disrupted *E. coli* cells showed that the expressed TrxA-m11CCR5^{306} was only present in the membrane fraction and the heavier cell debris fraction, but not in the soluble, cytosplasmic fraction (Figure 4.1d). The isolated membrane fraction was readily solubilizable by a number of detergents (see detergent screening section). Similarly, the receptor could also be solubilized from the cell debris. However, for most applications only the preparation from the lighter fraction was used.

The solubilized TrxA-m11CCR5^{306} was purified in FC-12 using Ni–NTA chromatography resulting in up to 10 mg of ~90 % pure (as estimated from SDS-PAGE) receptor per 1 L of *E. coli* culture (Figure 4.1d/e). Interestingly, purification by Ni–NTA triggered TrxA-m11CCR5^{306}

oligomerization on SDS-PAGE, which was reversible by dialysis (Figure 4.1e). The fusion partner was cleavable with thrombin (Figure 4.1e). Other proteases were also tested (data not shown) including TEV and 3C protease with no (TEV) or partial success (3C).

Trials to solubilize the receptor in maltosides failed (data not shown). Some TrxA-m11CCR5^{306} could be purified in tetradecylmaltoside but precipitated within few hours after elution from the Ni–NTA column. TrxA-m11CCR5^{306} solubilized in FC-12 followed by a detergent exchange to dodecylmaltoside on Ni–NTA also resulted in nearly complete protein precipitation. The purified, cleaved m11CCR5^{306} migrated on SDS-PAGE as a mixture of partially stable dimers at apparent MW of ~50 kDa and monomers at ~30 kDa (Figure 4.1e). Both MW values are smaller than expected. This phenomenon is common for membrane proteins and can be caused by incomplete unfolding by SDS and/or by a larger relative amount of SDS bound as compared to the soluble protein standard. Besides monomers and dimers also higher order oligomers were often observed (Figure 4.1e), especially after protein concentration. Discrete and sharp bands of CCR5 monomer and oligomers on the SDS-PAGE suggest that the primary structure of the protein is maintained (Figures 4.1e and S4.1). The identity and integrity of the C-terminus of the expressed constructs were confirmed by anti-His antibody western blotting (Figure 4.1a–d). To further confirm the protein identity, trypsinized TrxA-m7CCR5^{306} and Mistic-m7CCR5^{306} were analyzed by mass spectrometry. We were able to identify large stretches of fusion partners and the N-terminal fragment of the receptor in both monomer and oligomer (Figure S4.1). Peptides from TM domains were not detectable, which suggests that the CCR5 core was resistant to proteolysis.

4.4.4 Characterization of CCR5 size distribution, stability and homogeneity

It is commonly observed that GPCRs form homo- and heterodimers as well as higher oligomeric structures. For both *E. coli* (Figure 4.1) and insect cell expressed CCR5[36], besides monomers also oligomers are detected on SDS gels. The biological relevance of GPCR oligomerization is not clear. Since this heterogeneity also presents a problem for structural studies, the question of oligomerization was further investigated under non-reducing conditions using size exclusion chromatography. After Ni–NTA purification and digestion by thrombin, cleaved TrxA-m11CCR5^{306} was concentrated and injected onto a Superdex 200 column. The receptor migrated as a mixture of monomers, dimers and higher order oligomers (Figure 4.5a). This observation is consistent with the results of the SDS-PAGE (Figure 4.1e). Good protein separation was achieved on a 60 cm long size exclusion column. According to a column calibration with standard soluble proteins, the monomer and dimer peaks migrated similarly to particles of about 95 ± 3 (SD) kDa and 184 ± 9 kDa MW, respectively (N = 7). This suggests that the monomeric (dimeric) protein micelle contains ~165 (~313) FC-12 molecules. The ratio of monomer and dimer micelles depended on the stringency of Ni–NTA washing conditions, since higher imidazole concentrations depleted the monomeric fraction (data not shown). Apparently, this is due to the weaker binding of monomers to Ni–NTA. Relative to the monomers and dimers, the fraction of higher order oligomers was much smaller.

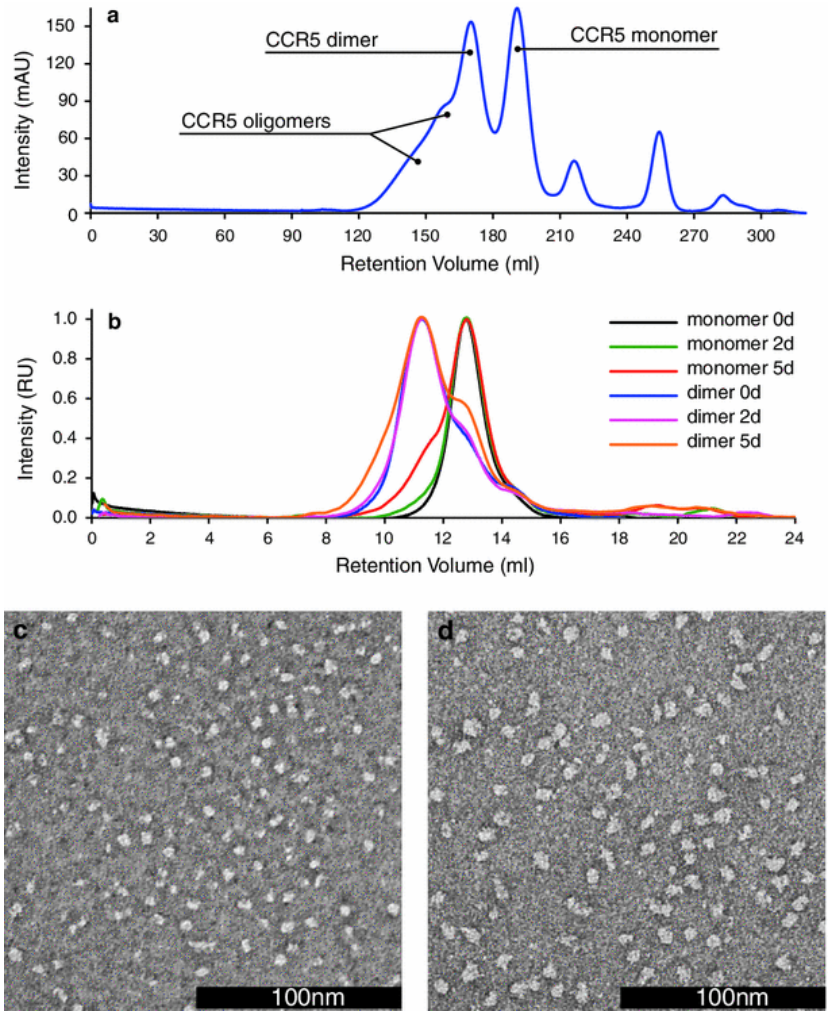


Figure 4.5 Monomers and dimers of m7CCR5^{306} and m11CCR5^{306}. (**a**) Size exclusion chromatography of m11CCR5^{306} on a Superdex 200 HiLoad 26/60 column. The 60 cm long column enables isolation of monomers and dimers. (**b**) Stability test of m7CCR5^{306} monomers and dimers. To prevent Cys oxidation 1 mM TCEP was included. Purified monomers and dimers were concentrated separately to ~40 µM and re-run on a Superdex 200 10/300 GL column. For easier comparison all six chromatograms were scaled to one. Negative stain pictures of m7CCR5^{306} monomers (**c**) and dimers (**d**).

In order to assay the influence of disulphide formation on the quality of the preparation, several different cysteine containing CCR5^{306} mutants were compared to the cysteine-free mutant under non-reducing conditions by size exclusion chromatography (Figure 4.6). The number of cysteines clearly correlates with enhanced oligomerization. The m2CCR5^{306} mutant (7 Cys) formed the most oligomers, whereas m6CCR5^{306} (3 Cys) and m7CCR5^{306} (4 Cys) mutants were less oligomerized. Interestingly, the effect of EC Cys mutations (m6CCR5^{306}) seems similar to the effect of TM Cys mutations (m7CCR5^{306}), which suggests that both EC and TM Cys may mediate disulphide bond formation. The higher oligomer formation of the cysteine-containing mutants could be suppressed by the addition of a reducing agent (Figure S4.2). Mutation of all Cys residues (m9CCR5^{306} and m11CCR5^{306}) resulted in a significant reduction of oligomerization, essentially rendering most of the protein monomeric or dimeric. Hence, it is likely that the remaining dimers and the residual higher oligomers are stabilized by non-disulphide interactions, presumably between the TM domains.

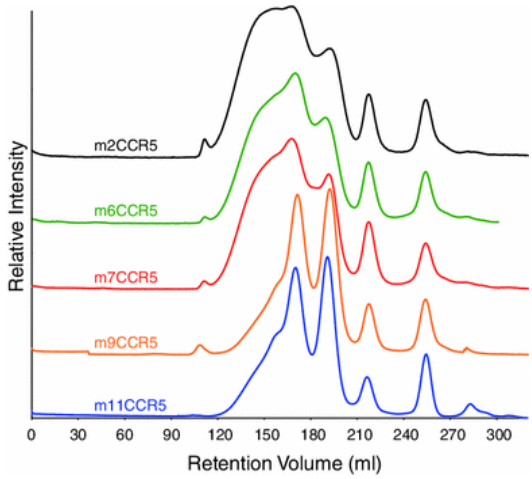


Figure 4.6 Size exclusion chromatography of various CCR5^{306} mutants demonstrates the impact of Cys residues on the oligomeric state of the purified receptor. No observable difference between m9CCR5^{306} and m11CCR5^{306} suggests that I52V and V150A mutations are not involved in CCR5 dimerization.

As non-dimerizing CCR5 would be of advantage for structural studies, following the findings by Hernanz-Falcon et al. that point mutations I52V and V150A strongly reduce dimer formation in HEK-293 cells[47], we tested these mutations in the m9CCR5^{306} mutant, which does not contain cysteines that could lead to intermolecular disulphide bridges. In contrast to the *in vivo* findings[47], these mutations did not reduce the dimerization propensity of the receptor (Figure 4.6). To assess the stability of m7CCR5^{306} monomer and dimer preparations under reducing conditions, both fractions were concentrated to ~40 µM and incubated for 5 days at RT. After 2 days of incubation, almost no change in the size distribution was detected, whereas after 5 days only a small fraction of monomers interconverted to dimers and some of dimers fell apart to monomers or formed higher order oligomers (Figure 4.5b). We tested a maximum monomer m7CCR5^{306} concentration of 137 µM, which also did not show any significant oligomerization after 4 days of incubation. Thus on the time scale of several days, both monomer and dimer preparations are very stable. The homogeneity of the monomeric and dimeric m7CCR5^{306} was confirmed by negative stain TEM. Monomeric (Figure 4.5c) and dimeric (Figure 4.5d) particles had average diameters of ~6.6 and ~8.3 nm, respectively.

4.4.5 Characterization of CCR5 secondary structure

The secondary structure content of several m7CCR5^{306} monomer preparations was assessed by CD. For all studied constructs we observed double minima at about 208 and 222 nm characteristic for α-helical proteins (Figure 4.7). The helical content derived from the mean residue molar ellipticity $\Theta_{MRM,222\ nm}$ for Mistic-m7CCR5^{306} (46 %) was slightly larger than for OmpF^{34}-m7CCR5^{306} (43 %) and TrxA-m7CCR5^{306} (42 %). This can be explained by the fact that Mistic is a purely helical bundle and increases the Θ_{MRM} of the whole fusion construct. This is not the case for the other fusion constructs, where the fusion partners contribute much less to Θ_{MRM} due to their mixed α/β (TrxA-m7CCR5^{306}) or likely β secondary structure (OmpF^{34}-m7CCR5^{306}). The 42 % α-helical content of TrxA-m7CCR5^{306} is similar to the value of ~40 % obtained by Ren et al. for the thioredoxin-CCR3 fusion construct[37]. For the m7CCR5^{306} monomer, that is after removal of the fusion partner from TrxA-m7CCR5^{306}, the CD signal was the strongest and indicated an α-helical content of 52 %. This is in a good agreement with the ~50 % helical content of a typical GPCR[39].

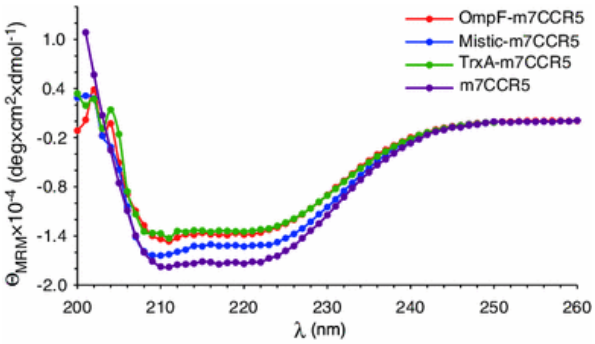


Figure 4.7 CD spectra of OmpF[34]-m7CCR5[306], Mistic-m7CCR5[306] and TrxA-m7CCR5[306] fusion constructs and of m7CCR5[306] (cleaved receptor without fusion partner) normalized to the protein concentration. The data was recorded on 3–13 µM samples in 20 Na_2HPO_4 pH 7.4, 180 mM NaCl, 0.1 % FC-12 at 20 °C on the monomeric receptor fractions isolated on SEC. Each spectrum shows characteristic features of α-helical secondary structure.

To assess the thermal stability of the CCR5 preparation, the CD spectrum of TrxA-m7CCR5[306] was followed over the range from 5 to 95 °C in 5 °C increments (Figure S4.3a). With increasing temperature the spectrum lost amplitude and its characteristic double minima. Decreasing the temperature from 95 °C back to 5 °C did not restore the initial shape and intensity, which indicates that denaturation was irreversible. The plot of the ellipticity at 222 nm against temperature (Figure S4.3b) shows a very broad thermal transition between 20 and 80 °C. Low thermal stability is commonly observed for GPCRs. In the present case, this problem may be aggravated by a non-optimal membrane-mimicking detergent system, which lacks important lipids and the rigidity of the two-dimensional membrane, as well as the absence of stabilizing ligands.

4.4.6 Functional studies on CCR5

Due to the numerous differences in the expression machinery and the cellular environment, the production of functional GPCRs in heterologous systems is very challenging. To prove the proper folding and the functionality of our CCR5 preparation, we tested binding of several ligands to the receptor using SPR. High sensitivity, automation and high-throughput makes this method widely used in the GPCR field for screening ligands[60], solubilization[61] and crystallization[62] conditions. For the SPR experiments, the receptor was solubilized in a DDM/CHAPS/CHS/DOPC mixture since a similar detergent/lipid composition was demonstrated to give best ligand binding activity for CCR5 and CXCR4[61] as opposed to FC-12 where little binding could be detected.

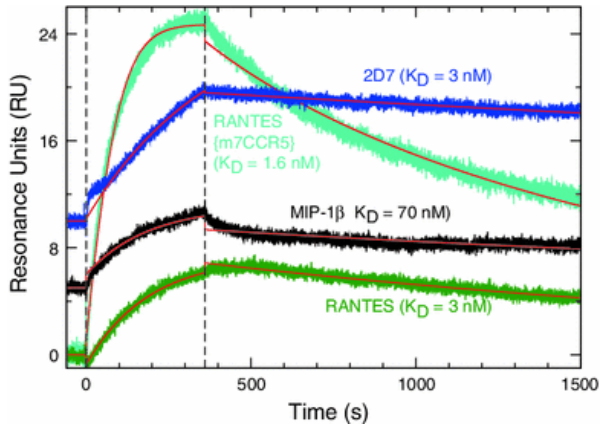


Figure 4.8 SPR binding assay of m7CCR5[306] (cyan) and m11CCR5[306] (other colors) solubilized in a DDM/CHAPS/CHS/DOPC mixture. The graph contains four overlaid independent runs, normalized to the amount of immobilized receptor and plotted to the same scale. Each run is composed of three phases separated by the dashed lines: equilibration, binding and dissociation. Data were fitted to a simple 1:1 binding model including a correction term for mass transport (red).

The protein was immobilized on the sensorchip via an anti-His-tag antibody. Subsequently, TrxA was cut off from the N-terminus of the receptor by an injection of thrombin. As monitored by the decrease of the SPR signal, the cleavage efficiency was estimated to be typically about 70 % (Figure S4.4). Binding was assayed for the CCR5 chemokine ligands RANTES and MIP-1β as well as for the conformation-dependent antibody 2D7 (Figure 4.8), which recognizes several residues from the second EC loop[63].

Each ligand showed fast binding and slow dissociation reactions. K_D values obtained from fitted k_{on} and k_{off} rates were all in the nanomolar range. m7CCR5^{306} bound RANTES with a K_D of 1.6 nM. m11CCR5^{306}, in which cysteines involved in disulphide bridge formation are absent, showed a two-fold decreased affinity (K_D = 3.1 nM) and a three-fold decrease of the response amplitude. MIP-1β and 2D7 bound with 71 nM and 2.8 nM affinity, respectively. For comparison, an identical experiment was performed with the wild-type CCR5 expressed in *Sf*21 cells. The obtained K_D values for RANTES, MIP-1β, 2D7 binding are 2.6, 200, and 0.1 nM, respectively. A summary of the performed experiments can be found in table S4.1. Considering the differences in the protein constructs, the K_D values for *E. coli* and insect cell expressed CCR5 are in reasonable agreement. However, when refractive index amplitudes for ligand binding are normalized to the refractive index amplitudes of bound CCR5 (Table S4.1), it is evident that the amount of bound RANTES and MIP-1β is about 2–3 fold and of 2D7 about 15 fold reduced for *E. coli* m11CCR5^{306}. We attribute this reduction to the lack of closed disulphide bridges at the extracellular side and the missing tyrosine sulfation in *E. coli*, which is important for chemokine binding[64].

4.4.7 NMR studies of CCR5

As opposed to crystal structures, which provide frozen snapshots of GPCR structures, NMR in principle can give simultaneous access to protein structure, dynamics and interactions. Thus, it emerges as a promising method to rationalize GPCRs' function. However, due to the numerous challenges in the sample preparation, the success of NMR studies on GPCRs has been very limited so far. To make our system suitable for NMR, the expression optimization was carried out directly in minimal medium. In this way isotope labeling does not compromise the final yield, which for detergent-solubilized, cleaved, monomeric m11CCR5^{306} was 2 mg per 1L of cell culture in triply isotope-labeled ($^{2}H/^{15}N/^{13}C$) minimal medium. For NMR measurements, samples were prepared from monomeric CCR5 fractions of the m11CCR5^{306} mutant. To estimate the quality of the preparation, ^{1}H-^{15}N correlation spectra were recorded (Figures 4.9 and S4.5). To optimize spectral quality, a variation of salt (0–180 mM NaCl), pH (4.2–7.4) and temperature (5–35 °C) was carried out. Optimal conditions were found at 20 °C, 0 mM NaCl and pH 4.2. Under these conditions, the spectra did not change over a period of few months. An increase in temperature to 35 °C gave only marginal improvement (data not shown). However, it had a destabilizing effect on the protein and caused a decrease of the NMR signal over time. The spectrum of m11CCR5^{306} under optimal conditions (Figure 4.9a) has a narrow dispersion, characteristic for an α-helical protein. It contains on the order of 60–80 intense and narrow resonances that presumably correspond to flexible backbone amides in the N- and C-terminal tails and the interhelical loops. Furthermore, a background of many more broad resonances is observed that most likely correspond to protein core residues. The line broadening in this region may be related to intermediate conformational exchange and/or to the large size of the

protein/detergent micelle. An attempt was made to assign at least some of the better-resolved backbone resonances by three-dimensional triple resonance TROSY HNCO, HN(CA)CO, HNCA, HN(CO)CA, and HNCACB experiments[53].

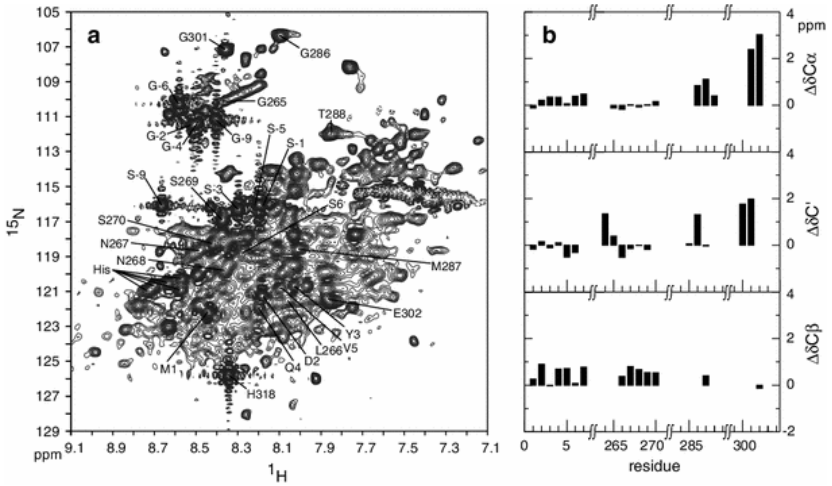


Figure 4.9 (**a**) ^{1}H-^{15}N TROSY spectrum of 112 µM monomeric ^{2}H/^{15}N-labeled m11CCR5^{306} (5 mM sodium acetate pH 4.2, 5 % D_2O, ~3 % FC-12) recorded at 20 °C on an 800 MHz spectrometer equipped with a cryoprobe with a total experimental time of 18 h. Assigned resonances are labeled. (**b**) Secondary $^{13}C^{\alpha}$, $^{13}C'$ and $^{13}C^{\beta}$ shifts for residues in the CCR5 sequence, for which backbone assignments could be established.

Due to the low signal to noise ratio, unambiguous assignments could only be achieved for 21 residues within the CCR5 amino acid sequence. These are located at the N-terminus (M1-S7), in the loop between helix 6 and 7 (F264-S270), in helix 7 (L285-T288) as well as in the putative helix 8 (V300-E302). The secondary $^{13}C^{\alpha}$, $^{13}C'$ and $^{13}C^{\beta}$ shifts for these residues are indicated in figure 4.9b. It is obvious that most residues have close to random coil shifts consistent with higher flexibility and concomitant higher resonance intensity. However, residues L285-T288 show moderately positive (~1–2 ppm) and residues V300- E302 larger positive (~2–3 ppm) $^{13}C^{\alpha}$ and $^{13}C'$ secondary shifts, which are consistent with a helical structure. Since besides the flexible N-terminus only residues in the region of helix 7 had a high enough signal to noise ratio for assignment, one may speculate that the region of helix 7 displays increased flexibility or more generally a different time scale of motional averaging. However, due to the highly limited assignment, this statement should be considered as very preliminary.

4.5 Discussion

Due to its involvement in HIV-1 infection, CCR5 is a major target for structural biology and the pharmaceutical industry. Despite that expression and purification schemes have been described for numerous GPCRs, there is a lack of an efficient isotope labeling platform for CCR5. 1 mg/L expression of CCR5 was reported in insect cells[36] where screening for mutants is time-consuming and isotope labeling very costly. On the other hand, so far no high-yield expression in isotope-labeled form has been reported for CCR5 in *E. coli* where these limitations are not present[37]. Our goal was to develop methods that allow structural and biophysical characterization in particular by NMR for CCR5 and potentially other GPCRs. Here, we have achieved large overexpression of CCR5 by fusing small stable protein domains or signal sequences to its N-terminus.

As the induction of CCR5 expression essentially arrested *E. coli* growth, increasing cell density proved to be a successful strategy to maximize the yield. The highest receptor overexpression was observed 24–48 h post induction at OD_{600}≈3. The induction at earlier or later phase of growth resulted in lower yields. Temperature had a dramatic effect on the expression level with the optimum ~20 °C. Variation of the CCR5 sequence also influenced the final yield. Thus, the expression of the longer CCR5 constructs (1–319) seemed much lower than the expression of the shorter ones (1–306). The number of cysteines in the CCR5 sequence correlated negatively with the expression level. When all 9 Cys residues were mutated (TrxA-m9CCR5^{306} and TrxA-m11CCR5^{306}), the yield was highest, while it was lowest for TrxA-m2CCR5^{306} (2 IC Cys mutated), i.e. ~1/3 of TrxA-m11CCR5^{306}.

A detergent screen revealed that charged detergents, especially anionic and zwitterionic were very efficient in OmpF^{34}-m7CCR5^{306} solubilization. Nonionic detergents, with the exception of tetradecylmaltoside, which solubilized about ~1/3 of available receptor, worked very poorly. A very similar solubility pattern was observed for the wild-type CCR5 receptor expressed in insect cells, i.e. there is good solubility in charged detergents and low solubility in nonionic detergents. In addition, exchange trials from FC-12 to maltoside consistently failed for material from both expression systems. These observations suggest that the poor CCR5 solubility in maltoside detergents, which are widely used in GPCR research, is not unique to the receptor expressed in *E. coli* and therefore rather a consequence of the receptor's low stability than a problem specific to the bacterial expression. In this respect it should be noted that the homology to other chemokine receptors like CXCR4 and CXCR1, which have more favorable solubilization properties, is not very high, i.e. about 30 %. In particular, larger differences exist at the CCR5 C-terminus, which harbors 3 cysteine palmitoylation sites not present in CXCR1 and CXCR4.

The detergent screening results are in agreement with previous screens proposing FosCholines as promising candidates for CCR5 solubilization[37]. Unfortunately, a good surfactant for solubilization is not always also well suited for other purposes. For some applications, like the SPR functional assay, other detergents or detergent/lipid mixtures provide better receptor activity[60]. Thus the search for an optimal detergent system or efficient detergent exchange protocols is still ongoing in our laboratory.

Protein oligomerization can severely decrease homogeneity of a sample and in this way compromise the quality of a sample for structural studies. In the case of CCR5 expressed in *E. coli*, the Cys residues, besides affecting the yield, also mediate oligomerization. Using size exclusion chromatography we have shown that the number of cysteines in CCR5^{306} constructs correlates with the amount of oligomerized protein (Figure 4.6). The fact that Cys-mediated oligomerization was also observed in the case of m7CCR5^{306}, for which all but the EC Cys were mutated, may suggest that in our system, at least to some extent, EC disulphide bridges are not properly formed. On the other hand, the oligomerization of m6CCR5^{306}, for which all but the TM cysteines C213, C290, C291 were mutated, implies that also TM Cys residues are reactive. This observation is consistent with the homology model (Figure 4.2), where C213 and C291 are located on the surface of the CCR5 core and accessible for intermolecular disulphide formation. When not jeopardized by intermolecular disulphide bridge formation, CCR5 forms a mixture of monomers, dimers and higher order oligomers. Due to their high stability, dimers and oligomers are also visible on SDS-PAGE. Both monomeric and dimeric species can be separated, concentrated and studied separately. The interconversion between monomers and dimers occurs after few days and goes both possible directions. As judged by TEM, both fractions are

homogenous and monodisperse with a clear difference in size. Based on the retention volume, the size of monomers and dimers was estimated to be 95 ± 3 (SD) kDa and 184 ± 9 kDa, respectively (N=7).

Based on computer modeling followed by the crosslinking of CCR5-transfected cells it was proposed that two point mutations together I52V and V150A yield a nonsignaling, non-dimerizing mutant of CCR5[47]. Such a non-dimerizing CCR5 mutant would be highly desirable for NMR studies. Unfortunately, the I52V and V150A mutants (m7CCR5^{306} or m11CCR5^{306}) did not exhibit significantly smaller propensity for dimerization in comparison to the non-mutated forms of CCR5 (Fig. 4.6). This is in line with results of co-immuno-precipitation and BRET experiments[65] that contradict the impairment of CCR5 dimerization for these mutants. The involvement of these two residues in dimerization is further challenged by the recently published CXCR4 structure, which shows dimer interactions at unrelated surfaces by helices V and VI (CXCR4 bound to IT1t) or by the intracellular ends of helices III and IV (CXCR4 bound to CVX15)[20].

Due to its robustness, polyhistidine-tag chromatography is widely used as a first purification step. Using a 10His-tag we achieved strong binding and could apply more rigorous washing conditions without compromising the final yield. This resulted in ~10 mg of purified TrxA-m11CCR5^{306} from 1 L of *E. coli* culture. This is a considerable improvement over the previously described system, where ~0.3 mg of Trx-hCCR5 per L was reported[37]. Importantly, this yield is not compromised when isotope labeling including D_2O is applied, which makes our system fully suitable for NMR studies. Out of 10 mg of TrxA-m11CCR5^{306} oligomeric mixture it is possible to isolate 2 mg of cleaved monomeric m11CCR5^{306}.

The quality of our preparations was assessed by CD, where all m7CCR5^{306} constructs showed the characteristic features of an α-helical secondary structure. NMR supports this observation as the HSQC spectrum of m11CCR5^{306} exhibits, typical for α-helical proteins, rather narrow peak dispersion (~2 ppm). Based on the circular dichroism data, we estimate that α-helices constitute ~52 % of the sequence of m7CCR5^{306} monomer which suggests that the receptor produced with our method has a correct secondary structure. The CD data indicate that the thermal stability of CCR5 is not very high. Some secondary structure is already lost at 5 °C but, as the amplitude of these changes is relatively small, it is difficult to judge their consequence on CCR5 structure and activity. Low thermal stability can be explained by several factors, most importantly suboptimal detergent system, lack of important lipids, absence of a ligand, nano-/antibody or a small molecule drug, that would stabilize CCR5.

The quality of the protein preparation was further validated by an SPR interaction assay. We observed nanomolar binding of RANTES to m7CCR5^{306} (K_D = 1.6 nM) and m11CCR5^{306} (3.1 nM) and of MIP-1β to m11CCR5^{306} (70 nM). These affinities are comparable to affinities of insect cell expressed CCR5 and within one order of magnitude to values observed in cellular binding assays (RANTES 0.38 nM, MIP-1β 7.2 nM)[64]. The observed affinities may be affected by the lack of posttranslational modifications (tyrosine sulfation) in *E. coli*, which increase the affinity of CCR5 for chemokines [64]. Furthermore, compared to m7CCR5^{306} RANTES affinity is weaker for m11CCR5^{306}, which lacks the extracellular cysteines. This is consistent with the reported importance of the extracellular disulphide bridges for chemokine binding[66].

High-affinity (2.8 nM) binding of m11CCR5^{306} was also observed for the 2D7 antibody, which is commonly used as a native conformation probe. Nevertheless, this affinity is considerably lower than for the insect cell CCR5 (0.1 nM). In addition, the refractive signal

amplitudes (Table S4.1) also suggest that the amount of bound ligand relative to m11CCR5^{306} is about 2–3 times smaller for chemokines and about 15 times smaller for 2D7. Again the lack of tyrosine sulfation and extracellular disulphide bridges may be the reason for this finding. Thus, further efforts are necessary to obtain higher activity, e.g. by proper refolding of disulphide bridges under controlled conditions.

Due to the substantial challenges in the preparation of isotope labeled samples, NMR spectra of GPCRs are very sparse in the literature. Therefore, only few HSQC spectra of ^{15}N labeled GPCRs have been reported, including the vasopressin V2 receptor[67], bovine rhodopsin[68], and the chemokine CXCR1 receptor[69].

Here, we present a spectrum of uniformly ^{15}N-labeled m11CCR5^{306}. Our initial HSQC spectrum had low dispersion and very broad lines besides for a number of apparently mobile terminal or loop residues. Similar observations have been made for other GPCRs[67-69]. However, the quality of the m11CCR5^{306} spectrum could be improved by a decrease of the pH from 7.4 to 4.2 and the removal of salt, which reduced hydrogen exchange and increased the sensitivity of the measurement. Unfortunately, even with these improvements the quality of the spectra is still not sufficient for structural analysis and needs further improvement but presents a starting point in the NMR investigation of CCR5. Obviously, the key bottleneck is the severe line broadening, which may be the result of conformational heterogeneity of the TM domains and/or chemical exchange on an intermediate time scale in the microsecond to millisecond range. Therefore, CCR5 stabilization by locking it in a single conformation, may be an important step towards the improvement of the NMR spectra.

Due to major difficulties in sample preparation for structural studies, protein engineering is very common in the GPCR field. This process alters the protein sequence and may modify its native properties, but so far has been indispensible for gaining insights into the structure and function of this important class of proteins. Since a vast majority of GPCR structures were solved by X-ray crystallography, not surprisingly, alterations comprised stabilization (rigidification, fixation in selected conformations), removal of unstructured regions, introduction of a soluble domain into a loop, etc. NMR spectroscopy on the other hand requires isotope labeling. However, so far efficient isotope labeling has not been possible for insect cells or natural tissues from which all except CXCR1 solved GPCRs were derived. Thus, simple prokaryotic organisms, like *E. coli* are often the system of choice for an NMR spectroscopist, as they allow cost-effective isotope labeling in addition to fast access to protein engineering. Due to the size limitations of NMR, the preparation of stable, monomeric and non-aggregating GPCRs is vital. Taking advantage of *E. coli,* we developed an efficient and robust CCR5 expression platform, which may find applications in biophysical, functional as well as structural characterizations of CCR5. We also believe that many of our observations have more general character and may be useful and applicable for other GPCRs.

4.6 Acknowledgements

We thank Prof. A. Arseniev for providing the plasmids pET28F10 and pMT10H10 and Prof. A. Spang for the plasmid pCA528, Dr. Paul Jenö and Suzanne Moes for mass spectrometry analysis as well as Dr. Marcel Blommers, Dr. Lukasz Skora and Prof. Sebastian Hiller for stimulating discussions. This work was supported by the EU FP7 Combined Highly Active Anti-Retroviral Microbicides (CHAARM), SNF Grant 31-109,712 and SystemsX.ch (C-CINA).

4.7 Supplemental material

Text S4.1: DNA sequence of pET28F10: OmpF^{34}-CCR2b

atggtgaagcgcaatattctggcagtgatcgtccctgctctgttagtagcaggtactgcaaacgctgcagaaatctataacaaagatggcaacaaagtagatcatatg
ctgtccacatctcgttctcggtttatcagaaataccaacgagagcggtgaagaagtcaccacctttttgattatgattacggtgctccctgtcataaatttgacgtgaag
caaattggggcccaactcctgcctccgctctactcgctggtgttcatctttggttttgtgggcaacatgctggtcgtcctcatcttaataaactgcaaaaagctgaagtgc
ttgactgacatttacctgctcaacctggccatctctgatctgctttttcttattactctcccattgtgggctcactctgctgcaaatgagtgggtctttgggaatgcaatgtgc
aaattattcacagggctgtatcacatcggttattttggcggaatcttcttcatcatcctcctgacaatcgatagatacctggctattgtccatgctgtgtttgctttaaaagcc
aggacggtcacctttggggtggtgacaagtgtgatcacctggttggtggctgtgtttgcttctgtcccaggaatcatctttactaaatgccagaaagaagattctgtttat
gtctgtggcccttattttccacgaggatggaataatttccacacaataatgaggaacattttggggctggtcctgccgctgctcatcatggtcatctgctactcgggaatc
ctgaaaaccctgcttcggtgtcgaaacgagaagaagaggcatagggcagtgagagtcatcttcaccatcatgattgtttactttctcttctggactccctataatattgtc
attctcctgaacaccttccaggaattcttcggcctgagtaactgtgaaagcaccagtcaactggaccaagccacgcaggtgacagagactcttgggatgactcactg
ctgcatcaatcccatcatctatgccttcgttggggagaagttcagaaggtatctctcggtgttcttccgaaagcacatcaccaagcgcttctgcaaacaatgtccagtttt
ctacagggagacagtggatggagtgacttcaacaaacacgccttccactggggagcaggatgtctcggctgctttactcgagcaccaccaccaccaccactag

Text S4.2: DNA sequence of pET28F10: OmpF^{34}-m7CCR5^{306}

Atggtgaagcgcaatattctggcagtgatcgtccctgctctgttagtagcaggtactgcaaacgctgcagaaatctataacaaagatggcaacaaagtagatcatat
ggattatcaggttagcagcccgatttatgatattaattattataccagcgaaccgtgccagaaaattaatgtgaaacaaattgcagcacgtctgctgcctccgctgtata
cctggtttttattttcggctttgtgggcaatatgctggttgttctgattctgattaatagcaaacgcctgaaaagcatgaccgatatttatctgctgaatctggcaattagcga
cctgtttttttctgctgaccgttccgttttgggcacattatgcagcagcacagtgggattttggtaataccatgtgtcagctgctgaccggtctgtatttattggctttttagc
ggcatttttttattattctgctgaccattgatcgttatctggcagttgttcatgcagttttgcactgaaagcacgcaccgttacctttggtgttgttaccagcgctattacctg
ggttgttgccgtttttgcaagcctgcctggcattatttttacccgtagccagaaagaaggtctgcattatacctgtagcagccattttccgtatagccagtatcagttttgga
aaaattttcagaccctgaaaattgttattctgggtctggttctgccgctgctggttatggttattgcctatagcggcattctgaaaaccctgctgcgtagtcgcaatgaaaa
aaaacgtcatcgtgccgttcgtctgatttttaccattatgattgtgtattttctgttttgggcaccgtataatatcgttctgctgctgaataccttttcaggaatttttttggcctgaa
taattgcagcagcagcaatcgtctggatcaggcaatgcaagttaccgaaaccctgggtatgacacatgctgccattaatccgattatttatgcgtttgtgggcgaaaaa
tttcgcaatctcgagcaccaccaccaccaccactag

Text S4.3: DNA sequence of pMT10H10: Mistic-CCR2b

atgggcttttgtacatttttttgaaaaacatcaccggaagtgggacatactgttagaaaaaagcacgggtgtgatggaagctatgaaagtgacgagtgaggaaaagga
acagctgagcacagcaatcgaccgaatgaatgaaggactggacgcgtttatccagctgtataatgaatcggaaattgatgaaccgcttattcagcttgatgatgatac
agccgagttaatgaagcaggcccgagatatgtacggccaggaaaagctaaatgagaaattaaatacaattattaaacagattttatccatctcagtatctgaagaagg
agaaaaagaaggttctggttctggtctggttccgcgtggatctcatatgctgtccacatctcgttctcggtttatcagaaataccaacgagagcggtgaagaagtcacc
accttttttgattatgattacggtgctccctgtcataaatttgacgtgaagcaaattggggcccaactcctgcctccgctctactcgctggtgttcatctttggttttgtgggc
aacatgctggtcgtcctcatcttaataaactgcaaaaagctgaagtgcttgactgacatttacctgctcaacctggccatctctgatctgctttttcttattactctcccattg
tgggctcactctgctgcaaatgagtgggtctttgggaatgcaatgtgcaaattattcacagggctgtatcacatcggttattttggcggaatcttcttcatcatcctcctga
caatcgatagatacctggctattgtccatgctgtgtttgctttaaaagccaggacggtcacctttggggtggtgacaagtgtgatcacctggttggtggctgtgtttgctt
ctgtcccaggaatcatctttactaaatgccagaaagaagattctgtttatgtctgtggcccttattttccacgaggatggaataatttccacacaataatgaggaacatttt
ggggctggtcctgccgctgctcatcatggtcatctgctactcgggaatcctgaaaaccctgcttcggtgtcgaaacgagaagaagaggcatagggcagtgagagtc
atcttcaccatcatgattgtttactttctcttctggactccctataatattgtcattctcctgaacaccttccaggaattcttcggcctgagtaactgtgaaagcaccagtcaa
ctggaccaagccacgcaggtgacagagactcttgggatgactcactgctgcatcaatcccatcatctatgccttcgttggggagaagttcagaaggtatctctcggtg
ttcttccgaaagcacatcaccaagcgcttctgcaaacaatgtccagttttctacagggagacagtggatggagtgacttcaacaaacacgccttccactggggagca
ggaagtctcggctggtttactcgagcaccatcaccaccatcaccaccatcaccactaa

Text S4.4: DNA sequence of pMT10H10: Mistic-m7CCR5^{306}

atgggcttttgtacatttttttgaaaaacatcaccggaagtgggacatactgttagaaaaaagcacgggtgtgatggaagctatgaaagtgacgagtgaggaaaagga
acagctgagcacagcaatcgaccgaatgaatgaaggactggacgcgtttatccagctgtataatgaatcggaaattgatgaaccgcttattcagcttgatgatgatac
agccgagttaatgaagcaggcccgagatatgtacggccaggaaaagctaaatgagaaattaaatacaattattaaacagattttatccatctcagtatctgaagaagg
agaaaaagaaggttctggttctggtctggttccgcgtggatctcatatggattatcaggttagcagcccgatttatgatattaattattataccagcgaaccgtgccagaa
aattaatgtgaaacaaattgcagcacgtctgctgcctccgctgtatagcctggttttattttcggctttgtgggcaatatgctggttgttctgattctgattaatagcaaacg
cctgaaaagcatgaccgatatttatctgctgaatctggcaattagcgacctgtttttttctgctgaccgttccgttttgggcacattatgcagcagcacagtgggattttggt
aataccatgtgtcagctgctgaccggtctgtatttattggctttttagcggcatttttttattattctgctgaccattgatcgttatctggcagttgttcatgcagttttgcac
tgaaagcacgcaccgttacctttggtgttgttaccagcgctattacctgggttgttgccgtttttgcaagcctgcctggcattatttttacccgtagccagaaagaaggtct
gcattatacctgtagcagccattttccgtatagccagtatcagttttggaaaaattttcagaccctgaaaattgttattctgggtctggttctgccgctgctggttatggttatt
gcctatagcggcattctgaaaaccctgctgcgtagtcgcaatgaaaaaaaacgtcatcgtgccgttcgtctgatttttaccattatgattgtgtattttctgttttgggcacc
gtataatatcgttctgctgctgaataccttttcaggaatttttttggcctgaataattgcagcagcagcaatcgtctggatcaggcaatgcaagttaccgaaaccctgggtat
gacacatgctgccattaatccgattatttatgcgtttgtgggcgaaaaatttcgcaatctcgagcaccatcaccaccatcaccaccatcaccactaa

Text S4.5: DNA sequence of pET-32b: TrxA-m7CCR5[306]

atgagcgataaaattattcacctgactgacgacagttttgacacggatgtactcaaagcggacggggcgatcctcgtcgatttctgggcagagtggtgcggtccgtg
caaaatgatcgccccgattctggatgaaatcgctgacgaatatcagggcaaactgaccgttgcaaaactgaacatcgatcaaaaccctggcactgcgccgaaatat
ggcatccgtggtatcccgactctgctgctgttcaaaaacggtgaagtggcggcaaccaaagtgggtgcactgtctaaaggtcagttgaaagagttcctcgacgctaa
cctggccggttctggttctggccatctggtacctcgaggctcggggcccggctcaggttctggctcaatggattatcaggttagcagcccgatttatgatattaattatt
ataccagcgaaccgtgccagaaaattaatgtgaaacaaattgcagcacgtctgctgcctccgctgtatagcctggttttttattttcggctttgtgggcaatatgctggttg
ttctgattctgattaatagcaaacgcctgaaaagcatgaccgatatttatctgctgaatctggcaattagcgacctgtttttttctgctgaccgttccgttttgggcacattatg
cagcagcacagtgggattttggtaataccatgtgtcagctgctgaccggtctgtattttattggctttttttagcggcatttttttttattattctgctgaccattgatcgttatctg
gcagttgttcatgcagtttttgcactgaaagcacgcaccgttacctttggtgttgttaccagcgctattacctgggttgttgccgtttttgcaagcctgcctggcattatttttt
acccgtagccagaaagaaggtctgcattatacctgtagcagccatttttccgtatagccagtatcagttttggaaaaattttcagaccctgaaaattgttattctgggtctg
gttctgccgctgctggttatggttattgcctatagcggcattctgaaaaccctgctgcgtagtcgcaatgaaaaaaaacgtcatcgtgccgttcgtctgatttttaccatta
tgattgtgtattttctgttttgggcaccgtataatatcgttctgctgctgaataccttttcaggaatttttttggcctgaataattgcagcagcagcaatcgtctggatcaggcaa
tgcaagttaccgaaaccctgggtatgacacatgctgccattaatccgattatttatgcgtttgtgggcgaaaaatttcgcaatctcgagcaccaccatcatcaccaccac
caccaccactga

Text S4.6: DNA sequence of pQE-T7: CCR5

Atgaaacaccatcaccatcaccatatgaaacaggattatcaggttagcagcccgatttatgatattaattattataccagcgaaccgtgccagaaaattaatgtgaaac
aaattgcagcacgtctgctgcctccgctgtatagcctggttttttattttcggctttgtgggcaatatgctggttattctgattctgattaattgcaaacgcctgaaaagcatg
accgatatttatctgctgaatctggcaattagcgacctgtttttttctgctgaccgttccgttttgggcacattatgcagcagcacagtgggattttggtaataccatgtgtca
gctgctgaccggtctgtattttattggctttttttagcggcatttttttttattattctgctgaccattgatcgttatctggcagttgttcatgcagtttttgcactgaaagcacgcac
cgttacctttggtgttgttaccagcgttattacctgggttgttgccgtttttgcaagcctgcctggcattatttttacccgtagccagaaagaaggtctgcattatacctgtag
cagccattttccgtatagccagtatcagttttggaaaaattttcagaccctgaaaattgttattctgggtctggttctgccgctgctggttatggttatttgctatagcggcat
tctgaaaaccctgctgcgttgtcgcaatgaaaaaaaacgtcatcgtgccgttcgtctgatttttaccattatgattgtgtattttctgttttgggcaccgtataatatcgttctg
ctgctgaataccttttcaggaatttttttggcctgaataattgcagcagcagcaatcgtctggatcaggcaatgcaagttaccgaaaccctgggtatgacacattgttgcat
taatccgattatttatgcgtttgtgggcgaaaaatttcgcaattatctgctggtgtttttttcagaaacatattgccaaacgcttttgcaaatgttgcagcatttttcagcagga
agcaccggaacgcgcaagcagcgtttatacccgtagcaccggtgaacaggaaattagcgttggtctgtag

Text S4.7: DNA sequence of pGEV2: GB1-CCR5

atgcagtacaagcttgctctgaacggtaaaaccctgaaaggtgaaaccaccaccgaagctgttgacgctgctaccgcggaaaaagttttcaaacagtacgctaacg
acaacggtgttgacggtgaatggacctacgacgacgctaccaaaaccttcacggtaaccgaactggttccgcgtggatccaccatggattatcaagtgtcaagtcca
atctatgacatcaattattatacatcggagccctgccaaaaaatcaatgtgaagcaaatcgcagcccgcctcctgcctccgctctactcactggtgttcatctttggttttg
tgggcaacatgctggtcatcctcatcctgataaactgcaaaaggctgaagagcatgactgacatctacctgctcaacctggccatctctgacctgtttttccttcttactg
tccccttctgggctcactatgctgccgcccagtgggactttggaaatacaatgtgtcaactcttgacagggctctattttataggcttcttctctggaatcttcttcatcatc
ctcctgacaatcgataggtacctggctgtcgtccatgctgtgtttgctttaaaagccaggacggtcacctttggggtggtgacaagtgtgatcacttgggtggtggctgt
gtttgcgtctctcccaggaatcatctttaccagatctcaaaaagaaggtcttcattacacctgcagctctcattttccatacagtcagtatcaattctggaagaatttccaga
cattaaagatagtcatcttggggctggtcctgccgctgcttgtcatggtcatctgctactcgggaatcctaaaaactctgcttcggtgtcgaaatgagaagaagaggca
cagggctgtgaggcttatcttcaccatcatgattgtttatttttctcttctgggctccctacaacattgtccttctcctgaacaccttccaggaattctttggcctgaataattgc
agtagctctaacaggttggaccaagctatgcaggtgacagagactcttgggatgacgcactgctgcatcaaccccatcatctatgcctttgtcggggagaagttcag
aaactacctcttagtcttcttccaaaagcacattgccaaacgcttctgcaaatgctgttctatttttccagcaagaggctcccgagcgagcaagctcagtttacacccgat
ccactggggagcaggaaatatctgtgggcttgcaccatcaccatcaccatcaccattga

Text S4.8: DNA sequence of pET-22b: CCR5

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgcccagccggcgatggccatggatatcggaattaattcggatccaatggattatcaagtg
tcaagtccaatctatgacatcaattattatacatcggagccctgccaaaaaatcaatgtgaagcaaatcgcagcccgcctcctgcctccgctctactcactggtgttcat
ctttggttttgtgggcaacatgctggtcatcctcatcctgataaactgcaaaaggctgaagagcatgactgacatctacctgctcaacctggccatctctgacctgttttt
ccttcttactgtcccccttctgggctcactatgctgccgcccagtgggactttggaaatacaatgtgtcaactcttgacagggctctattttataggcttcttctctggaatctt
cttcatcatcctcctgacaatcgataggtacctggctgtcgtccatgctgtgtttgctttaaaagccaggacggtcacctttggggtggtgacaagtgtgatcacttgggt
ggtggctgtgtttgcgtctctcccaggaatcatctttaccagatctcaaaaagaaggtcttcattacacctgcagctctcattttccatacagtcagtatcaattctggaag
aatttccagacattaaagatagtcatcttggggctggtcctgccgctgcttgtcatggtcatctgctactcgggaatcctaaaaactctgcttcggtgtcgaaatgagaa
gaagaggcacagggctgtgaggcttatcttcaccatcatgattgtttatttttctcttctgggctccctacaacattgtccttctcctgaacaccttccaggaattctttggcc
tgaataattgcagtagctctaacaggttggaccaagctatgcaggtgacagagactcttgggatgacgcactgctgcatcaaccccatcatctatgcctttgtcgggg
agaagttcagaaactacctcttagtcttcttccaaaagcacattgccaaacgcttctgcaaatgctgttctatttttccagcaagaggctcccgagcgagcaagctcagtt
tacacccgatccactggggagcaggaaatatctgtgggcttgcaccatcaccatcaccatcaccattga

Text S4.9: DNA sequence of pCA528: SUMO-m7CCR5[306]

atgggtcatcaccatcatcaccatggttcggactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactcacatcaatttaaaggt
gtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaaaagacagggtaaggaaatggactccttaag
attcttgtacgacggtattagaattcaagctgatcagacccctgaagatttggacatggaggataacgatattattgaggctcacagagaacagattggtgggagacc
ggatccgatggattatcaggttagcagcccgatttatgatattaattattataccagcgaaccgtgccagaaaattaatgtgaaacaaattgcagcacgtctgctgcctc
cgctgtatagcctggttttattttcggctttgtgggcaatatgctggttgttctgattctgattaatagcaaacgcctgaaaagcatgaccgatatttatctgctgaatctgg
caattagcgacctgtttttctgctgaccgttccgttttgggcacattatgcagcagcacagtgggattttggtaataccatgtgtcagctgctgaccggtctgtattttatt
ggctttttagcggcatttttttattattctgctgaccattgatcgttatctggcagttgttcatgcagttttgcactgaaagcacgcaccgttacctttggtgttgttaccag
cgctattacctgggttgttgccgtttttgcaagcctgcctggcattatttttacccgtagccagaaagaaggtctgcattatacctgtagcagccattttccgtatagccag
tatcagttttggaaaaattttcagaccctgaaaattgttattctgggtctggttctgccgctgctggttatggttattgcctatagcggcattctgaaaaccctgctgcgtagt
cgcaatgaaaaaaaacgtcatcgtgccgttcgtctgatttttaccattatgattgtgtattttctgttttgggcaccgtataatatcgttctgctgctgaataccttcaggaat
tttttggcctgaataattgcagcagcagcaatcgtctggatcaggcaatgcaagttaccgaaaccctgggtatgacacatgctgccattaatccgattatttatgcgtttg
tgggcgaaaaatttcgcaatctcgagcaccaccatcatcaccaccaccaccaccactga

Text S4.10: DNA sequence of pET-41a: GST-m7CCR5[306]

atgtcccctatactaggttattggaaaattaagggccttgtgcaacccactcgacttcttttggaatatcttgaagaaaaatatgaagagcatttgtatgagcgcgatgaa
ggtgataaatggcgaaacaaaaagtttgaattgggtttggagtttcccaatcttccttattatattgatggtgatgttaaattaacacagtctatggccatcatacgttatata
gctgacaagcacaacatgttgggtggttgtccaaaagagcgtgcagagatttcaatgcttgaaggagcggttttggatattagatacggtgtttcgagaattgcatata
gtaaagactttgaaactctcaaagttgattttcttagcaagctacctgaaatgctgaaaatgttcgaagatcgtttatgtcataaaacatatttaaatggtgatcatgtaacc
catcctgacttcatgttgtatgacgctcttgatgttgttttatacatggacccaatgtgcctggatgcgttcccaaaattagtttgttttaaaaaacgtattgaagctatccca
caaattgataagtacttgaaatccagcaagtatatagcatggcctttgcagggctggcaagccacgtttggtggtggcgaccatcctccaaaatcggatggttcaact
agtatggattatcaggttagcagcccgatttatgatattaattattataccagcgaaccgtgccagaaaattaatgtgaaacaaattgcagcacgtctgctgcctccgct
gtatagcctggttttattttcggctttgtgggcaatatgctggttgttctgattctgattaatagcaaacgcctgaaaagcatgaccgatatttatctgctgaatctggcaat
tagcgacctgtttttctgctgaccgttccgttttgggcacattatgcagcagcacagtgggattttggtaataccatgtgtcagctgctgaccggtctgtattttattggctt
ttttagcggcatttttttattattctgctgaccattgatcgttatctggcagttgttcatgcagtttttgcactgaaagcacgcaccgttacctttggtgttgttaccagcgctat
tacctgggttgttgccgtttttgcaagcctgcctggcattatttttacccgtagccagaaagaaggtctgcattatacctgtagcagccattttccgtatagccagtatcag
ttttggaaaaattttcagaccctgaaaattgttattctgggtctggttctgccgctgctggttatggttattgcctatagcggcattctgaaaaccctgctgcgtagtcgcaa
tgaaaaaaaacgtcatcgtgccgttcgtctgatttttaccattatgattgtgtattttctgttttgggcaccgtataatatcgttctgctgctgaataccttcaggaatttttgg
cctgaataattgcagcagcagcaatcgtctggatcaggcaatgcaagttaccgaaaccctgggtatgacacatgctgccattaatccgattatttatgcgtttgtgggc
gaaaaatttcgcaatctcgagcaccaccatcatcaccaccaccaccaccactaa

Text S4.11: The exact composition of m9 medium used in this study

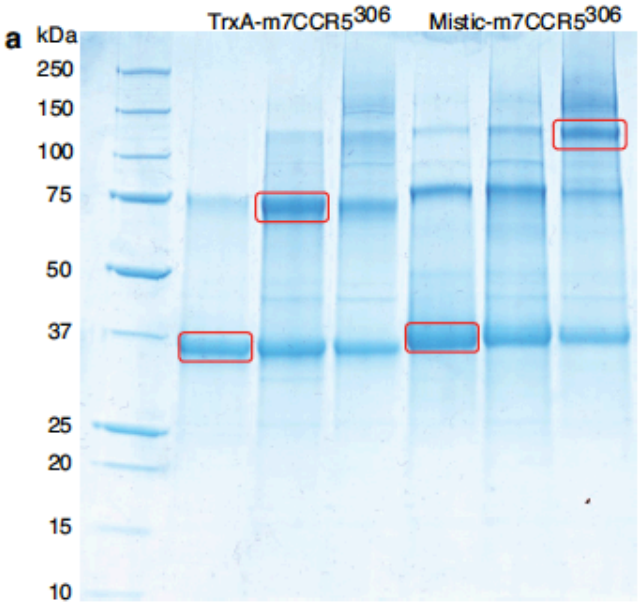
To make 1L of M9 medium mix:

- 851 mL autoclaved ddH_2O,
- 0.1 mL 0.22 µm-filtered 1 M $CaCl_2$,
- 2mL 0.22 µm-filtered 1 M $MgSO_4$,
- 10 mL 0.22 µm-filtered Hutner's trace elements,
- 3 mL 0.22 µm-filtered 10 mg/mL thiamine hydrochloride,
- 4 mL 0.22 µm-filtered 0.25 mg/mL biotin,
- 20 mL 0.22 µm-filtered 20% glucose,
- 10 mL 0.22 µm-filtered 10% NH_4Cl pH 7.4,
- 100 mL 0.22 µm-filtered 10 x M9 salts (67.8 g Na_2HPO_4, 30 g KH_2PO_4, 5 g NaCl, pH 7.4).

To prepare 200 mL of 100 x Hutner's trace elements use the procedure below:

1. Dissolve 1 g $FeSO_4$ and 10 g EDTA in 80 mL of ddH_2O. Adjust pH. A golden yellow solution results above around pH 5.5 and this is sufficient to proceed.
2. Dissolve the listed salts in 80 mL of ddH_2O:
   - 4.4 g $ZnSO_4 \cdot 7H_2O$,
   - 2.2 g H_3BO_3,
   - 1 g $MnCl_2 \cdot 4H_2O$,
   - 0.32 g $CoCl_2 \cdot 6H_2O$,
   - 0.235 g $CuCl_2 \cdot 2\ H_2O$,
   - 0.22 g $(NH_4)6MO_7O_{24} \cdot 4H_2O$.
3. Combine solutions 1 and 2 and adjust pH to 6.9 using KOH and bring volume to 200 mL. Solution is bright green.
4. Filter through 0.22 µm and store at 4°C. Solution turns purple.

Figure S4.1 Identity confirmation of TrxA-m7CCR5^{306} and Mistic-m7CCR5^{306} by mass spectrometry. (a) Monomer and dimer (TrxA-m7CCR5^{306}) or monomer and trimer (Mistic-m7CCR5^{306}) CCR5 bands (red rounded rectangles) were excised from SDS-PAGE gel stained with Colloidal Blue Stain Kit (Novex) and digested with trypsin prior to mass spectrometry analysis. **(b)** Identified peptide fragments of the analyzed fusion constructs were marked in bold. Individual components of the fusion constructs (fusion partner, linker, cleavage site, CCR5 sequence, His-tag) were marked with colors.



b

TrxA-m7CCR5^{306} monomer

MSDKII HLTDDSFDTDVLKADGAILVDFWAEWCGPCK**MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPK**YGIRG**IPTLLLFKNGEVAA TKVGALSKGQLK**EFLDANLAGSGSGSGHMDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVVLILINSKRLKSMTDI YLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSAITWVVAVFAS LPGIIFTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVIAYSGILKTLLRSRNEKKRHRAVRLIFTIMIVYFLFWAPYNIVL LLNTFQEFFGLNNCSSSNRLDQAMQVTETLGMTHAAINPIIYAFVGEKFRNLE**HHHHHH**

TrxA-m7CCR5^{306} dimer

MSDKII HLTDDSFDTDVLKADGAILVDFWAEWCGPCK**MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPK**YGIRG**IPTLLLFKNGEVA ATKVGALSKGQLK**EFLDANLAGSGSGSGHMDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVVLILINSKRLKSMT DIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSAITWVVAVF ASLPGIIFTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVIAYSGILKTLLRSRNEKKRHRAVRLIFTIMIVYFLFWAPYNI VLLLNTFQEFFGLNNCSSSNRLDQAMQVTETLGMTHAAINPIIYAFVGEKFRNLE**HHHHHH**

Mistic-m7CCR5^{306} monomer

MGFCTFFEKHHR**KWDILLEKSTGVMEAMKVTSEEKEQLSTAIDR**MNEGLDAFIQLYNESEIDEPLIQLDDDTAELMKQAR**DMYGQEKL NEKLNTIIKQILSISVSEEGEKE**GSGSG**LVPRGSH**MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVVLILINS KRLKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSAIT WVVAVFASLPGIIFTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVIAYSGILKTLLRSRNEKKRHRAVRLIFTIMIVYFL FWAPYNIVLLLNTFQEFFGLNNCSSSNRLDQAMQVTETLGMTHAAINPIIYAFVGEKFRNLE**HHHHHHHHHH**

Mistic-m7CCR5^{306} trimer

MGFCTFFEKHHR**KWDILLEKSTGVMEAMKVTSEEKEQLSTAIDR**MNEGLDAFIQLYNESEIDEPLIQLDDDTAELMKQAR**DMYGQEKL NEKLNTIIKQILSISVSEEGEKE**GSGSG**LVPRGSH**MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVVLILINS KRLKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSAIT WVVAVFASLPGIIFTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVIAYSGILKTLLRSRNEKKRHRAVRLIFTIMIVYFL FWAPYNIVLLLNTFQEFFGLNNCSSSNRLDQAMQVTETLGMTHAAINPIIYAFVGEKFRNLE**HHHHHHHHHH**

Figure S4.2 Effect of β-mercaptoethanol (BME) on the oligomeric state of TrxA-m7CCR5^{306}

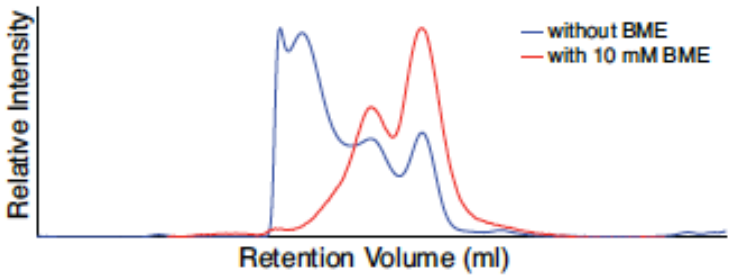


Figure S4.3 Heat denaturation of TrxA-m7CCR5^{306} measured using CD. (a) 200-250 nm CD spectra of TrxA-m7CCR5^{306} at series of temperatures in the range of 5-95°C. Heat irreversibly changes the shape and the amplitude of the spectrum. **(b)** Plot of mean residue molar ellipticity at 222 nm versus temperature shows a broad transition.

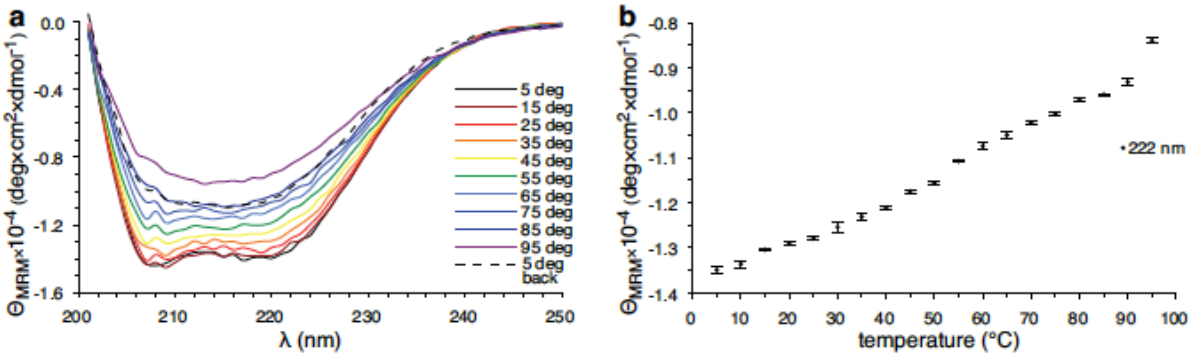


Figure S4.4 Thioredoxin removal from the N-terminus of m11CCR5^{306} immobilized on the sensor chip (~3400 RU) as monitored by the decay of the SPR signal. The arrows indicate the beginning and the end of the thrombin injection. When the refractive signal of the deposited CCR5 micelles is taken as a reference, the decrease of ~260 RU corresponds to about 70 % cleavage efficiency assuming a molecular weight of 100 kDa for the m11CCR5^{306} micelle and of 12 kDa for thioredoxin.

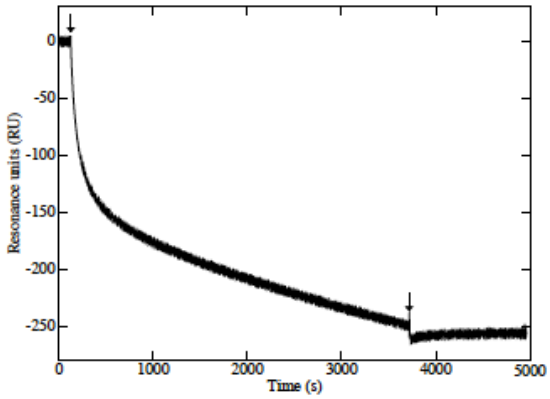


Figure S4.5 HSQC spectra of 112 µM monomeric $^{2}H/^{15}N$-labeled m11CCR5^{306} at various buffer conditions (listed in the insets). At 180 mM NaCl a pH change from 7.4 **(a)** to 5.8 **(b)** and 4.2 **(c)** leads to the appearance of additional resonances. An increase in sensitivity is achieved by the removal of 180 mM NaCl without a significant change in spectral dispersion **(d)**. The spectra were recorded at 20 °C on a 800 MHz spectrometer equipped with a cryoprobe. Typical experimental times were ~6 hours.

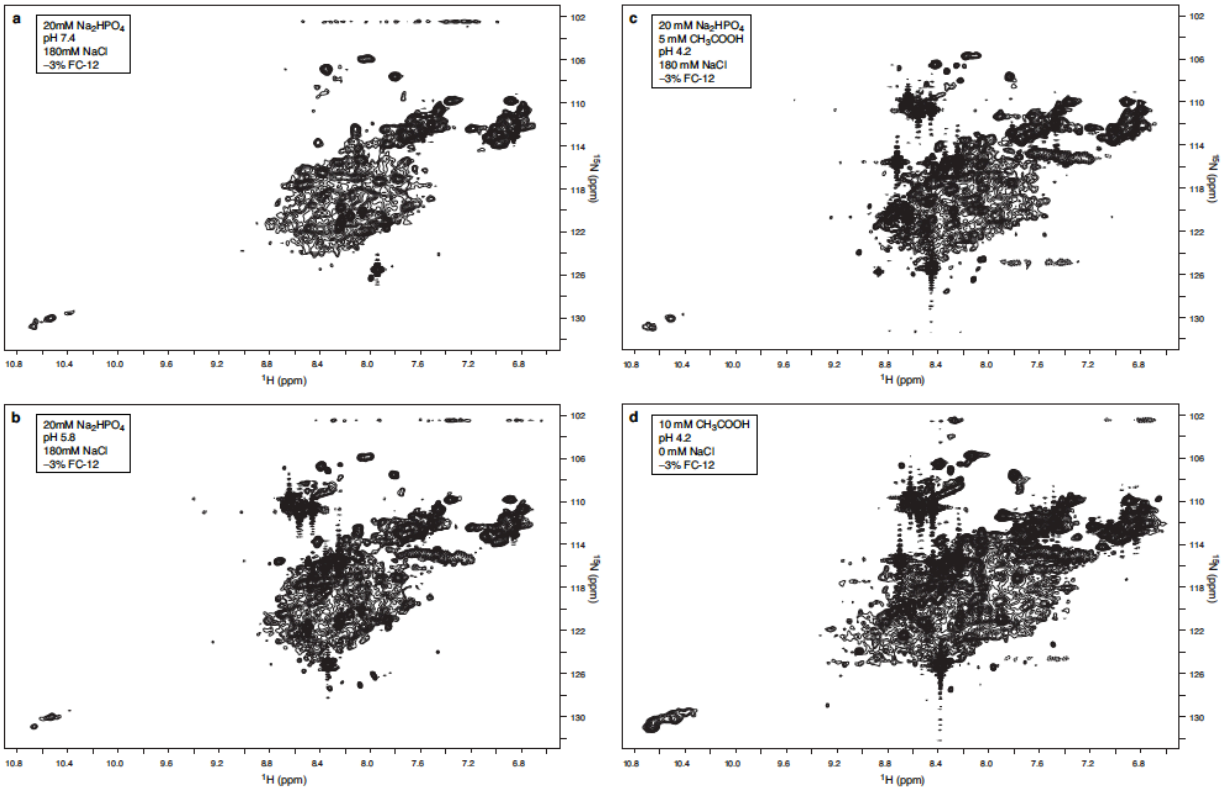


Table S4.1 Comparison of the binding parameters of RANTES, MIP-1β and 2D7 to m11CCR5^{306} expressed in *E. coli* (blue) and to the wild type CCR5 expressed in *Sf*21 (green) obtained by SPR.

Ligand	**MW** [kDa]	**Conc.** [nM]	$\mathbf{k_{on}}$ [$M^{-1}s^{-1}$]	$\mathbf{k_{off}}$ [s^{-1}]	$\mathbf{K_D}$ [nM]	$\mathbf{R_{max}}$ [RU]	$\mathbf{R_{max}}$/ **R** (CCR5)
2D7	150	5	45000	0.00013	2.8	17.57	0.00861
RANTES	8	36	140000	0.00042	3.1	8.73	0.00428
MIP-1β	8	1000	5000	0.00035	71	5.48	0.00269
2D7	150	5	500000	0.00005	0.1	640	0.128
RANTES	8	36	100000	0.00030	2.6	60	0.012
MIP-1β	8	1000	1200	0.00024	200	29	0.0058

4.8 References

1. Takeda, S., Kadowaki, S., Haga, T., Takaesu, H. & Mitaku, S. (2002). **Identification of G protein-coupled receptor genes from the human genome sequence**. *FEBS Lett* **520**, 97-101.
2. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schioth, H. B. (2003). **The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints**. *Mol Pharmacol* **63**, 1256-1272.
3. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. (2006). **How many drug targets are there?** *Nat Rev Drug Discov* **5**, 993-996.
4. Bockaert, J. & Pin, J. P. (1999). **Molecular tinkering of G protein-coupled receptors: an evolutionary success**. *EMBO J* **18**, 1723-1729.
5. Filmore, D. (2004). **It's a GPCR world**. *Mod Drug Discov* **7**, 24-28.
6. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). **The Protein Data Bank: a computer-based archival file for macromolecular structures**. *J Mol Biol* **112**, 535-542.
7. White, S. H. (2004). **The progress of membrane protein structure determination**. *Protein Sci* **13**, 1948-1949.
8. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). **Crystal structure of rhodopsin: A G protein-coupled receptor**. *Science* **289**, 739-745.
9. Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I. & Kobilka, B. K. (2007). **Crystal structure of the human beta2 adrenergic G-protein-coupled receptor**. *Nature* **450**, 383-387.
10. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). **High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor**. *Science* **318**, 1258-1265.
11. Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G. & Schertler, G. F. (2008). **Structure of a beta1-adrenergic G-protein-coupled receptor**. *Nature* **454**, 486-491.
12. Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P. & Stevens, R. C. (2008). **The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist**. *Science* **322**, 1211-1217.
13. Lebon, G., Warne, T., Edwards, P. C., Bennett, K., Langmead, C. J., Leslie, A. G. & Tate, C. G. (2011). **Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation**. *Nature* **474**, 521-525.
14. Chien, E. Y., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V. & Stevens, R. C. (2010). **Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist**. *Science* **330**, 1091-1095.
15. Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V. & Stevens, R. C. (2010). **Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists**. *Science* **330**, 1066-1071.
16. Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., Green, H. F., Liu, T., Chae, P. S., Dror, R. O., Shaw, D. E., Weis, W. I., Wess, J. & Kobilka, B. K. (2012). **Structure and dynamics of the M3 muscarinic acetylcholine receptor**. *Nature* **482**, 552-556.
17. Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., Haga, T. & Kobayashi, T. (2012). **Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist**. *Nature* **482**, 547-551.

18. Hanson, M. A., Roth, C. B., Jo, E., Griffith, M. T., Scott, F. L., Reinhart, G., Desale, H., Clemons, B., Cahalan, S. M., Schuerer, S. C., Sanna, M. G., et al. (2012). **Crystal structure of a lipid G protein-coupled receptor**. *Science* **335**, 851-855.
19. Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K. & Granier, S. (2012). **Crystal structure of the micro-opioid receptor bound to a morphinan antagonist**. *Nature* **485**, 321-326.
20. Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C. & Iwata, S. (2011). **Structure of the human histamine H1 receptor complex with doxepin**. *Nature* **475**, 65-70.
21. Granier, S., Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Weis, W. I. & Kobilka, B. K. (2012). **Structure of the delta-opioid receptor bound to naltrindole**. *Nature* **485**, 400-404.
22. Gautier, A., Kirkpatrick, J. P. & Nietlispach, D. (2008). **Solution-state NMR spectroscopy of a seven-helix transmembrane protein receptor: backbone assignment, secondary structure, and dynamics**. *Angew Chem Int Ed Engl* **47**, 7297-7300.
23. Reckel, S., Gottstein, D., Stehle, J., Lohr, F., Verhoefen, M. K., Takeda, M., Silvers, R., Kainosho, M., Glaubitz, C., Wachtveitl, J., Bernhard, F., Schwalbe, H., Guntert, P. & Dotsch, V. (2011). **Solution NMR structure of proteorhodopsin**. *Angew Chem Int Ed Engl* **50**, 11942-11946.
24. Park, S. H., Das, B. B., Casagrande, F., Tian, Y., Nothnagel, H. J., Chu, M., Kiefer, H., Maier, K., De Angelis, A. A., Marassi, F. M. & Opella, S. J. (2012). **Structure of the chemokine receptor CXCR1 in phospholipid bilayers**. *Nature* **491**, 779-783.
25. Balistreri, C. R., Caruso, C., Grimaldi, M. P., Listi, F., Vasto, S., Orlando, V., Campagna, A. M., Lio, D. & Candore, G. (2007). **CCR5 receptor: biologic and genetic implications in age-related diseases**. *Ann N Y Acad Sci* **1100**, 162-172.
26. Choe, H., Martin, K. A., Farzan, M., Sodroski, J., Gerard, N. P. & Gerard, C. (1998). **Structural interactions between chemokine receptors, gp120 Env and CD4**. *Semin Immunol* **10**, 249-257.
27. Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A. & Landau, N. R. (1996). **Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection**. *Cell* **86**, 367-377.
28. Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, C. M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., et al. (1996). **Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene**. *Nature* **382**, 722-725.
29. Stephens, J. C., Reich, D. E., Goldstein, D. B., Shin, H. D., Smith, M. W., Carrington, M., Winkler, C., Huttley, G. A., Allikmets, R., Schriml, L., et al. (1998). **Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes**. *Am J Hum Genet* **62**, 1507-1515.
30. Duncan, S. R., Scott, S. & Duncan, C. J. (2005). **Reappraisal of the historical selective pressures for the CCR5-Delta32 mutation**. *J Med Genet* **42**, 205-208.
31. Kondru, R., Zhang, J., Ji, C., Mirzadegan, T., Rotstein, D., Sankuratri, S. & Dioszegi, M. (2008). **Molecular interactions of CCR5 with major classes of small-molecule anti-HIV CCR5 antagonists**. *Mol Pharmacol* **73**, 789-800.
32. Gaertner, H., Cerini, F., Escola, J. M., Kuenzi, G., Melotti, A., Offord, R., Rossitto-Borlat, I., Nedellec, R., Salkowitz, J., Gorochov, G., Mosier, D. & Hartley, O. (2008). **Highly potent, fully recombinant anti-HIV chemokines: reengineering a low-cost microbicide**. *Proc Natl Acad Sci U S A* **105**, 17706-17711.
33. Lederman, M. M., Veazey, R. S., Offord, R., Mosier, D. E., Dufour, J., Mefford, M., Piatak, M., Jr., Lifson, J. D., Salkowitz, J. R., Rodriguez, B., Blauvelt, A. & Hartley, O. (2004). **Prevention of vaginal SHIV transmission in rhesus macaques through inhibition of CCR5**. *Science* **306**, 485-487.

34. Lusso, P., Vangelista, L., Cimbro, R., Secchi, M., Sironi, F., Longhi, R., Faiella, M., Maglio, O. & Pavone, V. (2011). **Molecular engineering of RANTES peptide mimetics with potent anti-HIV-1 activity**. *FASEB J* **25**, 1230-1243.
35. Nardese, V., Longhi, R., Polo, S., Sironi, F., Arcelloni, C., Paroni, R., DeSantis, C., Sarmientos, P., Rizzi, M., Bolognesi, M., Pavone, V. & Lusso, P. (2001). **Structural determinants of CCR5 recognition and HIV-1 blockade in RANTES**. *Nat Struct Biol* **8**, 611-615.
36. Nisius, L., Rogowski, M., Vangelista, L. & Grzesiek, S. (2008). **Large-scale expression and purification of the major HIV-1 coreceptor CCR5 and characterization of its interaction with RANTES**. *Protein Expr Purif* **61**, 155-162.
37. Ren, H., Yu, D., Ge, B., Cook, B., Xu, Z. & Zhang, S. (2009). **High-level production, solubilization and purification of synthetic human GPCR chemokine receptors CCR5, CCR3, CXCR4 and CX3CR1**. *PLoS One* **4**, e4509.
38. Attrill, H., Harding, P. J., Smith, E., Ross, S. & Watts, A. (2009). **Improved yield of a ligand-binding GPCR expressed in *E. coli* for structural studies**. *Protein Expr Purif* **64**, 32-38.
39. Baneres, J. L., Martin, A., Hullot, P., Girard, J. P., Rossi, J. C. & Parello, J. (2003). **Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B4 binding to recombinant BLT1**. *J Mol Biol* **329**, 801-814.
40. Dodevski, I. & Plückthun, A. (2011). **Evolution of three human GPCRs for higher expression and stability**. *J Mol Biol* **408**, 599-615.
41. Furukawa, H. & Haga, T. (2000). **Expression of functional M2 muscarinic acetylcholine receptor in *Escherichia coli***. *J Biochem* **127**, 151-161.
42. Krepkiy, D., Wong, K., Gawrisch, K. & Yeliseev, A. (2006). **Bacterial expression of functional, biotinylated peripheral cannabinoid receptor CB2**. *Protein Expr Purif* **49**, 60-70.
43. Shibata, Y., White, J. F., Serrano-Vega, M. J., Magnani, F., Aloia, A. L., Grisshammer, R. & Tate, C. G. (2009). **Thermostabilization of the neurotensin receptor NTS1**. *J Mol Biol* **390**, 262-277.
44. Weiss, H. M. & Grisshammer, R. (2002). **Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli***. *Eur J Biochem* **269**, 82-92.
45. Petrovskaya, L. E., Shulga, A. A., Bocharova, O. V., Ermolyuk, Y. S., Kryukova, E. A., Chupin, V. V., Blommers, M. J., Arseniev, A. S. & Kirpichnikov, M. P. (2010). **Expression of G-protein coupled receptors in *Escherichia coli* for structural studies**. *Biochemistry (Mosc)* **75**, 881-891.
46. Sastry, M., Xu, L., Georgiev, I. S., Bewley, C. A., Nabel, G. J. & Kwong, P. D. (2011). **Mammalian production of an isotopically enriched outer domain of the HIV-1 gp120 glycoprotein for NMR spectroscopy**. *J Biomol NMR* **50**, 197-207.
47. Hernanz-Falcon, P., Rodriguez-Frade, J. M., Serrano, A., Juan, D., del Sol, A., Soriano, S. F., Roncal, F., Gomez, L., Valencia, A., Martinez, A. C. & Mellado, M. (2004). **Identification of amino acid residues crucial for chemokine receptor dimerization**. *Nat Immunol* **5**, 216-223.
48. Otten, R., Chu, B., Krewulak, K. D., Vogel, H. J. & Mulder, F. A. (2010). **Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins**. *J Am Chem Soc* **132**, 2952-2960.
49. Gottschalk, G. (1986). **Bacterial Metabolism**. 2 edit, Springer, New York.
50. MD Abramoff, P. M., SJ Ram. (2004). **Image processing with imageJ**. *Biophotonics Int* **11**, 36-42.
51. Morrow, J. A., Segall, M. L., Lund-Katz, S., Phillips, M. C., Knapp, M., Rupp, B. & Weisgraber, K. H. (2000). **Differences in stability among the human apolipoprotein E isoforms determined by the amino-terminal domain**. *Biochemistry* **39**, 11657-11666.

52. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. (1997). **Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution**. *Proc Natl Acad Sci U S A* **94**, 12366-12371.
53. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. (1999). **TROSY-type triple-resonance experiments for sequential NMR assignment of large proteins**. *J Am Chem Soc* **121**, 844-848.
54. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). **NMRPipe: a multidimensional spectral processing system based on UNIX pipes**. *J Biomol NMR* **6**, 277-293.
55. Stenkamp, R. E. (2008). **Alternative models for two crystal structures of bovine rhodopsin**. *Acta Crystallogr D Biol Crystallogr* **D64**, 902-904.
56. Humphrey, W., Dalke, A. & Schulten, K. (1996). **VMD: visual molecular dynamics**. *J Mol Graph* **14**, 33-38, 27-38.
57. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. & Schulten, K. (2005). **Scalable molecular dynamics with NAMD**. *J Comput Chem* **26**, 1781-1802.
58. Hutner, S. (1950). **Some approaches to the study of the role of metals in the metabolism of microorganisms**. *Proc Am Philos Soc* **94**, 152-170.
59. Oppermann, M. (2004). **Chemokine receptor CCR5: insights into structure, function, and regulation**. *Cell Signal* **16**, 1201-1210.
60. Navratilova, I., Besnard, J. & Hopkins, A. L. (2011). **Screening for GPCR Ligands Using Surface Plasmon Resonance**. *ACS Med Chem Lett* **2**, 549-554.
61. Navratilova, I., Sodroski, J. & Myszka, D. G. (2005). **Solubilization, stabilization, and purification of chemokine receptors using biosensor technology**. *Anal Biochem* **339**, 271-281.
62. Navratilova, I., Pancera, M., Wyatt, R. T. & Myszka, D. G. (2006). **A biosensor-based approach toward purification and crystallization of G protein-coupled receptors**. *Anal Biochem* **353**, 278-283.
63. Khurana, S., Kennedy, M., King, L. R. & Golding, H. (2005). **Identification of a linear peptide recognized by monoclonal antibody 2D7 capable of generating CCR5-specific antibodies with human immunodeficiency virus-neutralizing activity**. *J Virol* **79**, 6791-6800.
64. Bannert, N., Craig, S., Farzan, M., Sogah, D., Santo, N. V., Choe, H. & Sodroski, J. (2001). **Sialylated O-glycans and sulfated tyrosines in the NH2-terminal domain of CC chemokine receptor 5 contribute to high affinity binding of chemokines**. *J Exp Med* **194**, 1661-1673.
65. Lemay, J., Marullo, S., Jockers, R., Alizon, M. & Brelot, A. (2005). **On the dimerization of CCR5**. *Nat Immunol* **6**, 535; author reply 535-536.
66. Blanpain, C., Lee, B., Vakili, J., Doranz, B. J., Govaerts, C., Migeotte, I., Sharron, M., Dupriez, V., Vassart, G., Doms, R. W. & Parmentier, M. (1999). **Extracellular cysteines of CCR5 are required for chemokine binding, but dispensable for HIV-1 coreceptor activity**. *J Biol Chem* **274**, 18902-18908.
67. Tian, C., Breyer, R. M., Kim, H. J., Karra, M. D., Friedman, D. B., Karpay, A. & Sanders, C. R. (2005). **Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor**. *J Am Chem Soc* **127**, 8010-8011.
68. Werner, K., Richter, C., Klein-Seetharaman, J. & Schwalbe, H. (2008). **Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy**. *J Biomol NMR* **40**, 49-53.
69. Park, S. H., Casagrande, F., Das, B. B., Albrecht, L., Chu, M. & Opella, S. J. (2011). **Local and global dynamics of the G protein-coupled receptor CXCR1**. *Biochemistry* **50**, 2371-2380.

Chapter 5 – General discussion & outlook

5.1 The secondary citrate/Na^+ symporter CitS

The contribution of secondary active transporters in numerous essential physiological processes makes these biomolecular machines highly interesting targets for structural studies. Any kind of structural information is expected to significantly extend the understanding of their versatile functionality, which in turn may support the prospective design of pharmaceuticals against various diseases[1,2].

The major part of this PhD thesis was dedicated to structural studies on the secondary citrate/Na^+ symporter CitS from the bacterium *Klebsiella pneumoniae*. As the best-characterized 2-hydroxycarboxylate transporter within structural class ST[3] of the MemGen classification system, CitS was predicted to represent a novel structural fold for secondary transporters[3,4]. As a complementary technique to ongoing x-ray crystallographic studies at the University of Zürich, we chose electron crystallography of membrane-embedded two-dimensional crystals.

The presented studies on CitS well emphasized both, the strengths and bottlenecks of electron crystallography of membrane proteins. Within a relative short period of six months, highly ordered 2D crystals were obtained. However, crystallization was only achieved within a very narrow range of conditions (especially pH, lipid chemistry and lipid-protein ratio). This required the screening of hundreds of different crystallization conditions using a total of 70 mg of purified CitS. Fortunately, due to very high expression levels of CitS in *E. coli* and a well established purification protocol, this huge amount of pure protein was available[5]. The most tedious part of the electron crystallographic workflow on CitS was found to be the sample preparation for cryo-EM analysis. Surprisingly, the most common protocols of sugar embedding (trehalose, glucose and tannic acid at various concentrations[6]) only led to unusable results. As a novel method, CitS 2D crystals were plunge frozen in liquid ethane in the absence of any cryo-protectants. A crucial washing step in low salt buffer prior to freezing additionally emphasized the unusual high sensitivity of CitS crystals for its physico-chemical environment. Subsequent imaging by low-dose imaging techniques provided a convenient and straightforward process. However, the microscopes used during this thesis were not yet equipped with state-of-the-art direct electron detectors, but with CMOS cameras exhibiting less efficient optical transfer functions. Therefore, data collection was performed on photographic film, a time and resource-consuming step that is expected to be significantly improved by future hardware developments. During the acquisition of images from tilted specimen in the microscope, beam induced sample drift emerged as main bottleneck[7]. This well-known phenomenon required the recording of several hundreds of images, especially at tilt angles >30 degree, to end up with 79 images in the final dataset. Recent and ongoing developments in electron crystallographic image processing, especially in the *2dx* software package enabled a very fast and reliable extraction of structural information from acquired cryo-electron micrographs.

The electron crystallography pipeline then finally provided substantial new insights into the structure and function of CitS that were inaccessible by all other techniques applied prior to this thesis, including extensive biochemical experimentation[8], single-molecule fluorescence

spectroscopy[9], single particle EM[10] and x-ray crystallography (personal communication). In a first step, we determined the projection structure of the symporter embedded in a lipid bilayer, which enabled us to develop detailed models concerning the monomer-monomer interface and domain organization of dimeric CitS[11]. The models presented in this first step were then complemented by the very first 3D data of CitS and 2-HCTs in general. Our study then well demonstrated the need and impact of 3D information, since our initial 2D models of CitS could be significantly improved and refined. Our investigations on substrate induced conformational changes powerfully demonstrated the accessibility of membrane proteins in 2D crystals, compared to 3D crystals.

One goal of this study was to use the obtained 3D map of CitS as search model for molecular replacement to solve the atomic structure using the existing x-ray diffraction dataset. This complex and tedious procedure was still in progress, when this thesis was finished. A further optimization of the CitS crystal size (> 1μm) would facilitate its structural analysis by electron diffraction. The expected improvement in both, the resolution and completeness of the data would further alleviate the structure determination to atomic resolution.

The 3D structure determination of membrane proteins by electron crystallography mainly suffers from (1) charge induced drift of tilted samples during image acquisition[7], (2) unevenly adsorbed crystals and (3) tilt-limited data completeness referred to as 'missing cone'[12]. Fortunately, several hard- and software developments are in progress that should overcome these limitations in prospective studies. In this context, the obtained 3D dataset of CitS has served as model record for an innovative software solution that treats single unit cells or even single molecules within a 2D crystal as single particles. Each of these is then corrected for its own tilt geometry variation within the crystal. This is expected to significantly improve the resolution and completeness of a 3D dataset. Recently, a further project was initiated where CitS 2D crystals are imaged with a novel dose-fractionation procedure in the electron microscope. Thereby, the usual electron dose of one image (~10 $e^-/Å^2$) is split into 5-10 subsequent images, each with an extremely low dose of 1-2 $e^-/Å^2$. Computational alignment of these then would effectively eliminate drift induced data loss[7]. During the analysis of substrate induced conformational changes of CitS, the dataset was additionally used to develop the novel procedure of calculating difference maps as described in chapter 3. All three described software developments already showed promising results and are expected to be refined and finalized in the near future.

The lack of hydrophilic interfaces in membrane proteins often impedes the formation of intermolecular contacts in 3D crystals, which is one major hurdle in obtaining well diffracting 3D crystals. This can be significantly improved by co-crystallization with specific soluble binding proteins such as antibodies, antibody fragments and designed ankyrin repeat proteins (DARPINs)[13,14]. For CitS, a highly specific DARPIN was developed at the University of Zürich [5,15]. The knowledge of the molecular and structural details of this protein-protein interaction is expected to considerably support ongoing x-ray crystallographic studies on CitS. Therefore, a sub-project has already been initiated that aims to identify the binding interface of CitS and its DARPIN, again by electron crystallography. Based on promising initial results, we expect to identify the interface in the near future.

5.2 The G protein-coupled receptor CCR5

Despite their importance in a huge variety of physiological processes and pharmaceutical developments, structural studies on GPCRs still remain tedious. This mainly results from relatively low expression levels and the highly flexible character of this class of integral membrane proteins[16]. Therefore, the study presented in chapter 4 introduced a novel high-titer expression platform for CCR5 derived from *E. coli*.

As part of this thesis, negative stain transmission electron microscopy was frequently used to assess the overall homogeneity of CCR5 protein preparations. This included the screening of several protein constructs and detergent systems and well supported the identification of the most promising preparation conditions, as presented here. The direct microscopic visualization of single protein particles again proofed to be well complementary to standard techniques such as SDS-PAGE and size-exclusion chromatography.

Additional experiments not shown in this thesis include reconstitution trials of CCR5 into lipid bilayers, including liposomes, 2D crystals and lipidic nanodiscs[17]. This is expected to propel its structure determination using different crystallographic and/or NMR techniques[18]. However, so far no reconstitution could be achieved and further optimization is still required. Thereby, EM could again serve as ideal technique, e.g. by the visualization and identification of specifically gold labeled CCR5 in lipid bilayers *via* its histidine affinity tag [19].

5.3 References

1. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. (2009). **Unlocking the molecular secrets of sodium-coupled transporters**. *Nature* **459**, 347-355.
2. Forrest, L. R., Krämer, R. & Ziegler, C. (2011). **The structural basis of secondary active transport mechanisms**. *Biochim Biophys Acta* **1807**, 167-188.
3. Ter Horst, R. & Lolkema, J. S. (2012). **Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity**. *Biochim Biophys Acta* **1818**, 72-81.
4. Lolkema, J. S., Sobczak, I. & Slotboom, D. J. (2005). **Secondary transporters of the 2HCT family contain two homologous domains with inverted membrane topology and trans re-entrant loops**. *FEBS J* **272**, 2334-2344.
5. Huber, T., Steiner, D., Röthlisberger, D. & Plückthun, A. (2007). **In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na(+)-citrate symporter CitS as an example**. *J Struct Biol* **159**, 206-221.
6. Chiu, P. L., Kelly, D. F. & Walz, T. (2011). **The use of trehalose in the preparation of specimens for molecular electron microscopy**. *Micron* **42**, 762-772.
7. Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., Henderson, R. & Grigorieff, N. (2012). **Beam-induced motion of vitrified specimen on holey carbon film**. *J Struct Biol* **177**, 630-637.
8. Krupnik, T., Dobrowolski, A. & Lolkema, J. S. (2011). **Cross-linking of dimeric CitS and GltS transport proteins**. *Mol Membr Biol* **28**, 243-253.
9. Kästner, C. N., Prummer, M., Sick, B., Renn, A., Wild, U. P. & Dimroth, P. (2003). **The citrate carrier CitS probed by single-molecule fluorescence spectroscopy**. *Biophys J* **84**, 1651-1659.
10. Moscicka, K. B., Krupnik, T., Boekema, E. J. & Lolkema, J. S. (2009). **Projection structure by single-particle electron microscopy of secondary transport proteins GltT, CitS, and GltS**. *Biochemistry* **48**, 6618-6623.
11. Kebbel, F., Kurz, M., Grütter, M. G. & Stahlberg, H. (2012). **Projection structure of the secondary citrate/sodium symporter CitS at 6 Å resolution by electron crystallography**. *J Mol Biol* **418**, 117-126.
12. Gipson, B. R., Masiel, D. J., Browning, N. D., Spence, J., Mitsuoka, K. & Stahlberg, H. (2011). **Automatic recovery of missing amplitudes and phases in tilt-limited electron crystallography of two-dimensional crystals**. *Phys Rev E Stat Nonlin Soft Matter Phys* **84**, 011916.
13. Sennhauser, G. & Grütter, M. G. (2008). **Chaperone-assisted crystallography with DARPins**. *Structure* **16**, 1443-1453.
14. Monroe, N., Sennhauser, G., Seeger, M. A., Briand, C. & Grütter, M. G. (2011). **Designed ankyrin repeat protein binders for the crystallization of AcrB: plasticity of the dominant interface**. *J Struct Biol* **174**, 269-281.
15. Röthlisberger, D., Pos, K. M. & Plückthun, A. (2004). **An antibody library for stabilizing and crystallizing membrane proteins - selecting binders to the citrate carrier CitS**. *FEBS Lett* **564**, 340-348.
16. Tate, C. G. (2012). **A crystal clear solution for determining G-protein-coupled receptor structures**. *Trends Biochem Sci* **37**, 343-352.
17. Bayburt, T. H. & Sligar, S. G. (2010). **Membrane protein assembly into Nanodiscs**. *FEBS Lett* **584**, 1721-1727.
18. Park, S. H., Berkamp, S., Cook, G. A., Chan, M. K., Viadiu, H. & Opella, S. J. (2011). **Nanodiscs versus macrodiscs for NMR of membrane proteins**. *Biochemistry* **50**, 8983-8985.
19. Hainfeld, J. F. & Powell, R. D. (2000). **New frontiers in gold labeling**. *J Histochem Cytochem* **48**, 471-480.

List of Acronyms

2D / 3D	two-/three-dimensional
2-HCT	2-hydroxycarboxylate transporter
ABC	ATP binding casette
ADP/ATP	adenosine-di/-triphosphate
AH	amphipathic helix
BAD	biotin affinity domain
BRET	bioluminescence resonance energy transfer
BSA	bovine serum albumine
CCR2/5	CC chemokine receptor 2/5
CD	circular dichroism
(Cryo-)EM	(cryo-)electron microscopy
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHS	cholesterol hemi-succinate
CMC	critical micellar concentration
CT	C-terminal
CTF	contrast transfer function
CXCR1/4	CXC chemokine receptor 1/4
ΔG	free Gibbs energy
ΔV	electric potential
DARPIN	designed ankyrin repeat protein
DASS	divalent anion sodium symporter
DM	*n*-Decyl-β-D-maltoside
DDM	*n*-Dodecyl-β-D-maltoside
DOPC	1,2-dioleoyl-*sn*-glycero-3-phosphocholine
DHPC	1,2-diheptanoyl-*sn*-glycero-3-phosphocholine
E. coli	*Escherichia coli*
EC	extracellular
EDTA	Ethylendiamintetraacetat
EM	electron microscopy
FC-12	Fos-Choline-12
FFT	fast Fourier Transform
GABA	gamma-amino butyric acid
GPCR	G protein-coupled receptor
GST	Glutathion-S-transferase
HEK	human embryonic kidney
HEPES	2-(4-(2-hydroxyethyl)- 1-piperazinyl)-ethansulfonic acid
HIV	human immuno-deficiency virus
HRP	horseradish peroxidase
HSQC	heteronuclear single quantum coherence
IC	intracellular
IMAC	immobilized metal ion chromatography
INDY	I'm not dead yet

IPTG	isopropyl-thio-galactoside
kDa	kilo Dalton
K_D	dissociation constant
K_M	Michaelis-Menten constant
LDR	lipid to detergent ratio
LPR	lipid to protein ratio
MCF	mitochondrial carrier family
MD	molecular dynamic
MFS	major facilitator superfamily
MIP	macrophage inflammatory protein
(MW)CO	(molecular weight) cutoff
nm	nanometer
NMR	nuclear magnetic resonance
NT	N-terminal
NTA	nitrilotriacetic acid
OD	optical density
Omp	Outer membrane protein
PMSF	Phenylmethylsulfonylfluorid
POPC	1-palmitoyl-2-oleyl-*sn*-glycero-3-phosphocholine
ppm	parts per million
PVDF	Polyvinylidenfluorid
RANTES	regulated on activation, normal T-cell expressed and secreted
RF	radio frequency
RMSD	root-mean-square deviation
RND	resistance nodulation cell devision
rpm	rotations per minute
RU	relative units
s/sec	second
SBCGP	single binding center gated pore
SDS-PAGE	sodium-dodecylsulfate-polyacrylamide gel electrophoresis
Sf	*Spodoptera frugiperda*
SMR	small multidrug-resistance
SPR	surface plasmon resonance
[ST]	structural class of the MemGen classification system
TBST	Tris buffered saline with tween
TC	transporter classification
TCEP	Tris(2-chlorethyl)phosphat
TEM	transmission electron microscopy
TMS	transmembrane segment
v/v	volume by volume
WT	wild-type
w/v	weight by volume
w/w	weight by weight
XRD	x-ray diffraction

List of Figures

Acknowledgements

First of all, my biggest 'Thank you' goes to my mentor and thesis supervisor **Henning Stahlberg**. He supported me throughout all the different stages of this PhD project by providing great scientific advice and vision, freedom, understanding and funding. He additionally taught me so much about membrane proteins, electron microscopy, computers, a good work & life balance and much more. Thank you so much!

I also would like to thank my co-supervisors and co-referees **Markus Grütter** and **Sebastian Hiller**. During numerous discussions both could give me great scientific input and feedback for all aspects of the different projects.

Thank you so much, **Mareike Kurz**. It was so nice to work with you on the CitS project. You were not only an inexhaustible source for CitS - I could also enjoy being part of an exemplary scientific collaboration: efficient, reliable, honest and personal.

Without the help from **Marcel Arheit** I would still be in the stage of processing my electron micrographs. Thank you so much for transferring all your knowledge about image processing and for all of your efforts to further improve this.

This work would not have been possible without the help of **Mohamed Chami** and **Kenny Goldie**. Mohamed introduced me to the magic of sample preparation for Cryo-EM of 2D crystals. Kenny continuously supported me with great ideas and help for using all the different electron microscopes. Many thanks to both of you!

Bill Anderson, you must have been an electron microscope in your earlier life. This is the only explanation for your incredible knowledge about these machines. Thank you so much for keeping the microscopes running all the time.

I could additionally enjoy numerous very inspiring scientific discussions (and coffee breaks) with **Philippe Ringler, Sebastian Scherer, Shirley Müller, Alexandra Graff-Meyer** and **Thomas Braun**. Thank you all very much.

During my work on CCR5 and Nanodiscs I could interact with numerous excellent scientists, whom I would like to thank: **Stephan Grzesiek**, **Sebastién Morin**, **Maciej Wiktor**, **Lydia Nisius** and **Ricardo Adaixo**.

The initial part of my PhD studies at the Biozentrum was performed in the lab of Dagmar Klostermeier. **Lenz** and **Manu**, thanks for your great friendship.

Finally, I would like to thank my family and, especially, my wife **Stefanie**. Your continuous support and your understanding made this thesis possible, more than everything else.