UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE CHIMICHE

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE MOLECOLARI

INDIRIZZO: SCIENZE CHIMICHE

XXI CICLO

Production and characterization of SulP anion transporters

Direttore della Scuola: Ch.mo Prof. Maurizio Casarin

Supervisore: Ch.mo Prof. Roberto Battistutta

Dottoranda: Elisa Pasqualetto

2 febbraio 2009

Ai miei genitori

Contents

Summary

The main subject of this thesis is the Sulphate Permease (SulP) protein family that includes more than two hundred members, identified in archaea, bacteria, fungi, plants and animals. Many of these proteins have been functionally characterized: most are anion exchangers or transporters with different substrate specificities and distinct mechanism of action (Saier et *al.*, 1999). In mammals, the SulP family, known as Solute Linked Carrier 26 (SLC26), is composed of eleven members with important roles in normal physiology (Mount and Romero, 2004).

The SulP proteins show a similar structural organization: a hydrophobic central core, which includes ten or twelve membrane helixes, and a less conserved C-terminal cytoplasmic portion that includes a STAS domain (Sulphate Transporter and Anti-Sigma factor antagonist domain). Despite the functional role of the STAS domain is still unclear, it appears to be crucial for the regulation of the transport activity (Ko et *al.*, 2004; Zheng et *al.*, 2005; Shibagaki and Grossman, 2006). Its fundamental role is also underlined by the fact that mutations that alter this domain in the SLC26 family can cause loss of function, resulting in serious genetic diseases, like diastrophic dysplasia or Pendred syndrome (Dawson and Markovich, 2005). No three-dimensional structure of STAS domains or full-length sequences is available for any SulP anion transporter.

One part of the work was focused on the production of different forms of the STAS domain from different species, for the biophysical and structural characterization. Another part of the SulP project was performed at the Johann Wolfgang Goethe University of Frankfurt (Germany) and aimed at the production of some full-length SulP proteins, by a cell-free expression system, an emerging technique for the large-scale production of membrane proteins.

In the last year, I was also involved in the crystallographic study of the Green Fluorescence Protein mutant, GFPmut2, in collaboration with Prof. Stefano Bettati of the University of Parma (Italy). The main aim of this work was the elucidation of the structural basis of the spectroscopic properties of this mutant, in particular with respect to

changes in pH. The GFP chromophore can, in fact, exist either in a protonated or deprotonated state, with distinct spectral properties (Tsien, 1998). In a previous spectroscopic characterization, GFPmut2 (Ser65Ala, Val68Leu, Ser72Ala) was found more sensitive than the wild type GFP to pH changes in the physiological range (Chirico et *al.*, 2002). The structures of GFPmut2 at pH 6 and pH 9 were determined at around 1.6 Å resolution, allowing the correlation between the spectral and structural properties.

Riassunto

L'oggetto principale di questo lavoro di tesi è la famiglia dei trasportatori anionici SulP (*Sulphate Permease*), che comprende più di duecento membri identificati in archea, batteri, funghi, piante e animali. Molte proteine di questa famiglia sono state funzionalmente caratterizzate e agiscono da trasportatori o scambiatori di anioni, e differiscono per l'affinità verso il substrato e il meccanismo di trasporto (Saier et *al.*, 1999). Nei mammiferi la famiglia SulP, conosciuta come *Solute Linked Carrier* 26 (SLC26), è composta di undici membri che svolgono un ruolo fondamentale in molti processi fisiologici nell'uomo (Mount e Romero, 2004).

Tutte le proteine SulP possiedono un'organizzazione strutturale simile: una parte centrale idrofobica, che comprende dieci o dodici eliche di membrana e una porzione C-terminale citoplasmatica meno conservata, che include il dominio STAS (*Sulphate Transporter and Anti-Sigma factor antagonist*). Sebbene non sia ancora chiaro il ruolo funzionale di questo dominio nei trasportatori di anioni, esso sembra essere di cruciale importanza per la regolazione dell'attività di trasporto (Ko et *al.*, 2004; Zheng et *al.*, 2005; Shibagaki e Grossman, 2006). Il suo ruolo fondamentale è rilevato anche dal fatto che mutazioni che alterano questo dominio nei membri della famiglia SLC26 possono comprometterne seriamente la funzionalità, causando malattie genetiche gravi, come la displasia diastrofica o la sindrome di Pendred (Dawson and Markovich, 2005). Non sono ancora note strutture tridimensionali di nessun dominio o intera proteina SulP.

Una parte del lavoro è stata focalizzata sulla produzione di diverse varianti del dominio STAS da specie diverse, finalizzata alla caratterizzazione biofisica e strutturale. Una seconda parte del progetto, svolta presso la *Johann Wolfgang Goethe University* di Francoforte (Germania), ha riguardato la produzione di intere proteine SulP mediante la sintesi *in vitro*, una tecnica molto promettente per la produzione su larga scala di proteine di membrana.

Durante l'ultimo anno, mi sono anche dedicata allo studio cristallografico di un mutante della *Green Fluorescent Protein*, GFPmut2, in collaborazione con il gruppo del

Prof. Stefano Bettati dell'Università di Parma. L'obiettivo principale di questo lavoro è stato definire le basi strutturali delle proprietà spettroscopiche di questo mutante, in particolare al variare del pH. Il cromoforo della GFP può, infatti, esistere sia in forma protonata che deprotonata (Tsien, 1998). Le proprietà spettroscopiche della GFPmut2 (Ser65Ala, Val68Leu, Ser72Ala) sono state in precedenza caratterizzate e, rispetto alla proteina *wild type*, sembra essere più sensibile alle variazioni di pH nell'intervallo fisiologico (Chirico et *al.*, 2002). A tal fine, è stata determinata la struttura della GFPmut2, sia a pH 6 che a pH 9, con una risoluzione di circa 1.6 Å. Il confronto delle due strutture ha consentito la correlazione delle proprietà strutturali con quelle spettroscopiche.

Part A

Production and characterization of SulP anion transporters

1

Introduction

1.1 The Sulphate Permease (SulP) family

The Sulphate Permease (SulP) family is a large and ubiquitous family of membrane proteins with over two hundred sequenced members, identified by sequence homology in archaea, bacteria, fungi, plants and animals. Many of these proteins are functionally characterized: most are anion exchangers (Na^+-independent anion transporters) and transport a wide range of anions, both organic and inorganic, with individual transporters showing different specificities. Many function by SO_4^{2-}/H^+ symport, but SO_4^{2-}/HCO_3^-, or more generally, anion/anion antiport has been reported for several homologues (Saier et *al*., 1999).

In bacteria and plants they are responsible for the uptake of sulphate, a convenient source of sulphur that is a key element in the bacterial as well as in the eukaryotic metabolism (Kertesz, 2001). In mammals, the SulP family, also known as the Solute Linked Carrier 26 (SLC26) family of anion transporters, shows broader anion specificity and more complex functions (Mount and Romero, 2004).

The SLC26 gene family

The SLC26 family is composed of highly versatile anion transporters, with important roles in normal physiology and human pathophysiology. A partial list of physiological processes in which the SLC26 exchangers play critical roles includes outer hair cells (OHCs) electromotility, skeletal development, synthesis of thyroid hormone, transepithelial Na^+-Cl^- transport, bicarbonate excretion by the distal nephron, and bicarbonate secretion by the exocrine pancreas (Mount and Romero, 2004).

SLC26A proteins function as anion exchangers or channels in the luminal or apical membranes of epithelial tissue and are primarily involved in transport of a wide variety of monovalent and divalent anions. Each member has different anion specificity and distinctive tissue distribution; some being expressed in most organs and others with more restricted tissue expression patterns. To date, eleven human SLC26 genes have been identified, ten of which were shown to encode proteins that transport one or more

substrates, including sulphate, chloride, bicarbonate, iodide, oxalate, formate, hydroxyl, mannose and fructose (Table 1; Mount and Romero, 2004). SLC26A5 (prestin) was shown to act as the motor protein of cochlear outer hair cells (Zheng et *al.*, 2000). The SLC26 family thus exhibits an amazing variety of functions, yet the molecular basis of this diversity is poorly understood.

Table 1: The SLC26 gene family.

Gene	Protein name	Reported substrate	Tissue distribution	Disease association
SLC26A1	Sat-1	SO_4^{2-}, oxalate	Liver, kidney	
SLC26A2	DTDST	SO_4^{2-}, Cl^-	Widespread	Chondrodysplasias
SLC26A3	DRA, CLD	SO_4^{2-}, Cl^-, HCO_3^-, OH^-, oxalate, formate	Intestine, sweat gland, pancreas, prostate	Congenital chloride diarrhea
SLC26A4	Pendrin	Cl^-, HCO_3^-, I^-, formate	Inner ear, kidney, thyroid	Pendred syndrome, deafness (DFNB4)
SLC26A5	Prestin	Cl^-?	Inner ear	Deafness?
SLC26A6	CFEX, PAT-1	SO_4^{2-}, Cl^-, HCO_3^-, OH^-, oxalate, formate	Widespread	
SLC26A7	None	SO_4^{2-}, Cl^-, oxalate	Kidney	
SLC26A8	Tat1	SO_4^{2-}, Cl^-, oxalate	Sperm, brain	
SLC26A9	None	SO_4^{2-}, Cl^-, oxalate	Lung	
SLC26A10	None	?	Brain	
SLC26A11	None	SO_4^{2-}	Widespread	

The SLC26 family and genetic diseases

The clinical relevance of the SLC26 gene family was highlighted with the identification of pathogenetic mutations in four of its genes, namely SLC26A2, A3, A4 and A5 (Table 1). Although these four genes share significant sequence homology and encode structurally related proteins, they give rise to distinct clinical phenotypes (Dawson and Markovich, 2005). SLC26A2 is involved in chondrodysplasias that cause skeletal defects, including clubbed feet, cleft palate, and short limbed dwarfism. Mutations in the SLC26A3 gene are linked to congenital chloride-losing diarrhea, a disease in which patients suffer from watery diarrhea containing elevated Cl^- concentrations that can prove

fatal, if left untreated. SLC26A4 is involved in Pendred syndrome, which is the most common form of syndromic deafness, characterized by congenital sensorineural hearing loss and thyroid goiter. SLC26A5 encodes a protein, called prestin that is highly and almost exclusively expressed in the OHCs of the cochlea. The specific expression pattern of prestin in the OHCs suggests that it is a candidate gene for human deafness. Indeed SLC26A5 mutations were identified in individuals with non-syndromic deafness, confirming the physiological role of prestin in human auditory processing (Liu et *al.*, 2003; Toth et *al.*, 2007).

Given the different anion specificity and the distinct tissue distribution of expression for each of these genes, it is not surprising that they are associated with markedly different clinical phenotypes.

Membrane topology of the SulP proteins

With only a few exceptions, the bacterial SulP transporters vary in size from 434 to 573 residues, while the eukaryotic proteins are usually larger, varying from 611 to 893 residues (Saier et *al.*, 1999).

Although the level of amino acid identity between all members of the SulP family is low, around 25%, hydropathy plots of different members of the family from bacteria to humans are clearly similar, suggesting structural and functional similarities. Moreover, blocks of more highly conserved amino acids are present in some transmembrane helices and some of these are functionally important. This implies that there will be common features in the transport mechanisms throughout the family (Loughlin et *al.*, 2002).

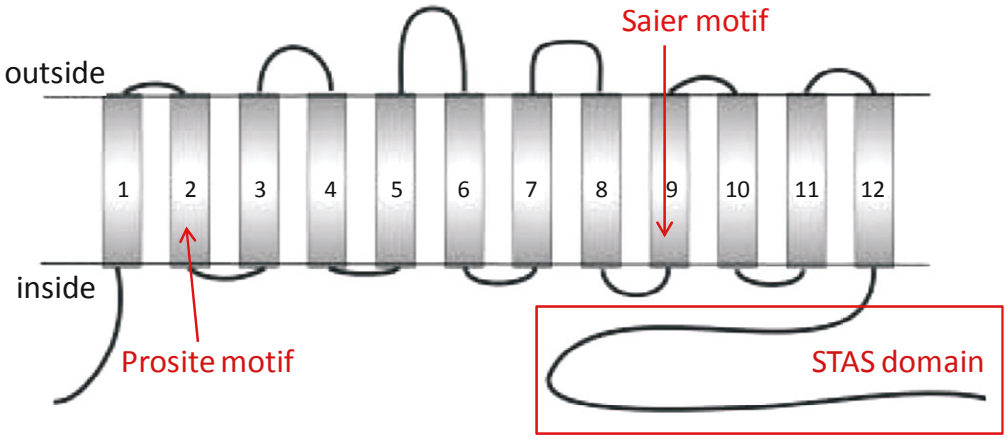

Figure 1: One predicted topology model of the SulP proteins. The position of various conserved motifs and domains is depicted. The number of transmembrane helices can vary from 10 to 14.

The detailed membrane topology of the SulP exchangers has not been determined experimentally, and prediction programs yield highly divergent models (Figure 1). The SulP family proteins are predicted to have 10 to 14 transmembrane spanning α-helices, with intracellular N- and C-termini (Saier et *al.*, 1999; Mount and Romero, 2004).

The transmembrane domain

Much of the homology between SulP exchangers is found within the hydrophobic core of transmembrane domain. The first two putative transmembrane α-helices show a significantly higher level of conservation than that observed for the entire protein. This region includes one of the two "sulphate transporter motifs" that have been used to define the SulP family (Saier et *al.*, 1999). The first *consensus* signature extends across putative helix 2 and comprises 22 amino acids (Prosite, PS01130; Figure 1). Although not all members of the family conform to the exact *consensus* sequence, this region contains several invariant residues that are presumably critical for anion transport. Moreover, an alignment of eukaryotic family members shows that there are also positions in helix 1 with high levels of conservation. In addition to conservation of the residue at each position, the spacing between them, including a short loop between the first two helices, is maintained throughout the eukaryotic members of the family (Leves et *al.*, 2008). Mutagenesis studies on these residues were performed on a plant sulphate transporter SHST1, from the tropical legume *Stylosanthes hamata* (Shelden et *al*., 2001; Loughlin et *al.*, 2002; Leves et *al.*, 2008) and prestin (SLC26A5), a distantly related mammalian member of the SulP family (Rajagopalan et *al.*, 2006). These studies confirm the predicted importance of conserved residues in helices 1 and 2 and suggest that function of the SulP members is dependent on a network of polar and aromatic interactions between these two helices.

The second cluster of invariant residues defined by Saier and colleagues extends across putative helix 9 (Figure 1; Saier et *al.*, 1999). This helix is somewhat atypical in that it contains a great number of polar residues. Two conserved residues in this region (Asn395 and Glu387) were shown to have functional significance in SHST1 (Khurana et *al.*, 2000; Loughlin et *al.*, 2002). These studies suggest that putative helix 9 may be important for stability and/or trafficking of SHST1 to the plasma membrane. Moreover, mutations in the correspondent residues in two members of the SLC26 family result in serious diseases. A severe dysplasia, achondrogenesis type Ib, can be caused by a mutation that affects Asn425 in SLC26A2, equivalent to Asn395 in SHST1. Pendred

syndrome may be the result of the mutation of Glu384 in SLC26A4 (Glu387 in SHST1). These results indicate that conserved residues between distinct members of the family may share essential roles in structure or function.

The STAS domain

The less conserved C-terminal cytoplasmic portion of all SulP proteins extends into the cytoplasm of the cell and includes a so-called STAS domain. The STAS domain (Sulphate Transporter and Anti-Sigma factor antagonist domain) was identified by the sequence analysis of proteins with completely different functions (Aravind and Koonin, 2000). This analysis revealed an unexpected, statistically significant similarity between the carboxy-terminal cytoplasmatic part of SulP transporters (that can vary in length from around 115 to around 250 amino acids) and the bacterial Anti-Sigma factor Antagonists ASA, typified by *Bacillus subtilis* SpoIIAA (117 residues long).

The STAS domain and genetic diseases

The C-terminus is the least conserved region of the protein among different SLC26A family members, therefore it is most likely to be responsible for each protein specific function. Although the STAS domain appears to be of crucial importance for the regulation of transport activity, the functional role of this domain with respect to anion transporters is still poorly understood. Its fundamental role is underlined by the fact that mutations that alter this domain in the SLC26 family can cause loss of function, resulting in serious diseases, like diastrophic dysplasia, Pendred syndrome, and congenital chloride diarrhea (Dawson and Markovich, 2005). The majority of these mutant proteins has improper plasmamembrane targeting and loss of some or full function.

It was shown that mutations in the STAS domain of SLC26A3, which functions as a coupled Cl^-/HCO_3^- exchanger, result in congenital chloride diarrhea, by causing a loss in wild type levels of functional protein at the plasma membrane. This is probably caused by at least two distinct mechanisms: misfolding that prevents the mutant transporters from reaching the native state, and the disruption of important intramolecular interactions critical to form a well folded and functional transporter (Dorwart et *al.*, 2008).

Muallem and colleagues provided clear evidence for a reciprocal regulation between the CFTR chloride channel, implicated in cystic fibrosis, and two members of the SLC26

family (SLC26A3 and SLC26A6). The interaction is mediated by binding of the regulatory (R) domain of CFTR to the STAS domain of SLC26A proteins. The interaction is modulated by PDZ binding scaffold proteins that tether the two transporters into a multimeric complex with other regulatory proteins (Ko et *al.*, 2004). These findings provide new insights into the mechanism of bicarbonate and fluid secretion from epithelial tissues and may lead to better treatments for cystic fibrosis and congenital chloride diarrhea (Gray, 2004).

ASA proteins STAS domain

The bacterial SpoIIAA protein is a key component of the regulation network involved in the induction of sporulation in response to nutrient deficiency. The transcription factor SpoIIAA or Anti-Sigma factor Antagonist (ASA) associates with the complex formed by the sigma factor and the anti-sigma factor SpoIIAB; this association causes the release of the sigma factor from SpoIIAB, triggering sporulation-specific transcription (Diederich et *al.*, 1994; Kroos et *al.*, 1999). SpoIIAB is also a kinase that can phosphorylate and inactivate SpoIIAA (Duncan et *al.*, 1996).

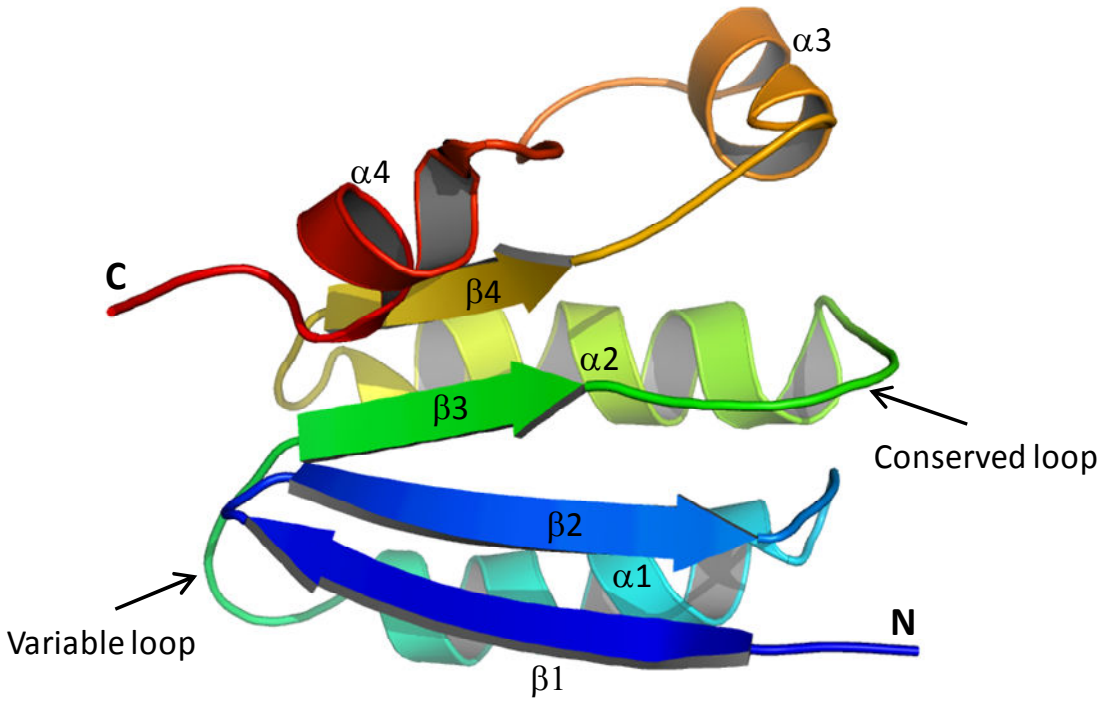

Figure 2: Global fold of the SpoIIAA from *Bacillus subtilis* (Kovacs et *al.*, 1998); α-helices and β-strands are numbered and labelled sequentially. The N- and C-termini are labelled. The position of two relevant loops in anion transporters, the variable and the conserved one, respectively, are indicated by arrows.

The bacterial ASA are structurally well characterized in their 3D structure both by NMR spectroscopy (Kovacs et *al.*, 1998) and X-ray crystallography (Seavers et *al.*, 2001). The SpoIIAA fold consists of four β-strands, forming a β-sheet, surrounded by four α-helices (Figure 2). The β-sheet, in association with hydrophobic surfaces of the α-helices, forms a hydrophobic core that is not readily accessible to the external medium. In contrast, the peripheral exposed surfaces of α-helices and loops are available for interactions with molecules in the environment. The carboxy-terminal region forms a characteristic α-helical handle-like structure.

Anion transporters STAS domain

Unlike the bacterial ASA proteins, the STAS domains present in anion transporters are poorly characterized in terms of both their function and structure; indeed no 3D structure of such domains is known yet.

The STAS domain of anion transporters shows low overall sequence identity with SpoIIAA (about 15-20%). The conservation was traced largely to the four strands that form the scaffold of the STAS domain. In addition, the turn between the two amino-terminal strands and the long loop between strand β3 and helix α2 are strongly conserved in almost all the STAS domains (Figures 2 and 3). This loop and β-pleated sheet were proposed to play a role in nucleotide binding and hydrolysis, by extension from the known biochemistry of the anti-sigma factor antagonists (Aravind and Koonin, 2000). It was shown that SpoIIAA binds GTP and ATP and possesses a weak NTPase activity that is abolished by phosphorylation or by mutation of the phosphorylable serine in the conserved loop (Najafi et *al.*, 1996). The strong conservation of this loop in the STAS domains suggests that it could possess general NTP-binding activity. The conserved loop is probably involved in phosphate binding and the β-sheet scaffold could accommodate the rest of the NTP molecule. The presence of a predicted NTP-binding domain in the cytoplasmic portions of anion transporters indicates that anion transport could be regulated by intracellular concentrations of GTP and/or ATP.

Most of the variability is in the loop between helix α1 and strand β3 (Figures 2 and 3), with inserts of considerable size in some of the anion transporters, of as much as 150 amino acids in the case of SLC26A8 (Aravind and Koonin, 2000). This is immediately apparent from sequence alignment of all the anion transporters STAS domains and their

structural homologs, the SpoIIAA proteins (Figure 4). In the STAS domain of the bacterial transporters the loop is absent, in the plant sulphate transporter SULTR1.2 it comprises around 10 residues, while for the mammalian transporters this loop is invariably longer (for instance around 70 residues for prestin and 150 for SLC26A8). Secondary structure predictions of this region suggest it is largely unstructured (Dorwart et *al.*, 2008). Furthermore, in the transporters, a variable extension at the C-terminal end of the domain is present and the secondary structure predictions of the extreme N- and C-termini do not correlate with that found in the bacterial ASA.

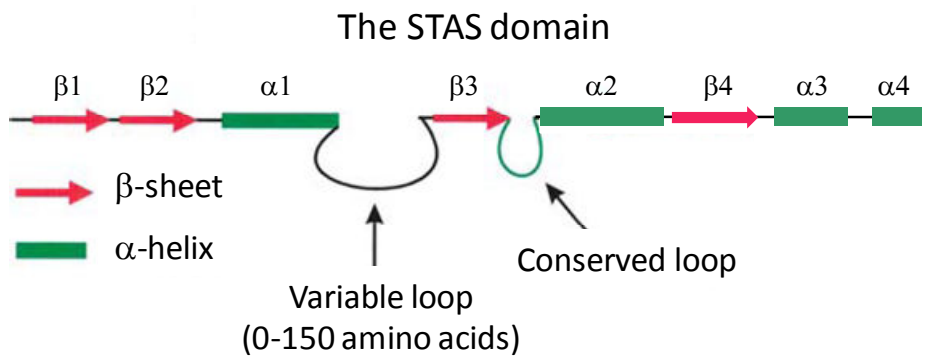

Figure 3: Representation of the secondary structure elements of the anion transporters STAS domain. A highly conserved loop is interspersed between strand β3 and helix α2. The STAS domain also contains a highly variable loop between helix α1 and strand β3. This variable loop is the site of significant insertions in the SLC26A proteins, of as much as 150 amino acids in the case of SLC26A8.

Taking into account these differences in lengths as well as the low amino acid conservation observed, most probably the 3D structure of the anion transporters STAS domains significantly deviates from that of the bacterial ASA, in a way not predictable solely on the basis of the sequence alignment. Presumably, these differences are responsible for the distinct properties of this domain when part of the different transporters.

During the last years, numerous mutagenesis studies were performed on the STAS domains of different SulP transporters, to elucidate their precise function in the transport activity. In the following section, two of these studies, concerning a plant and a bacterial transporter, will be shortly introduced.

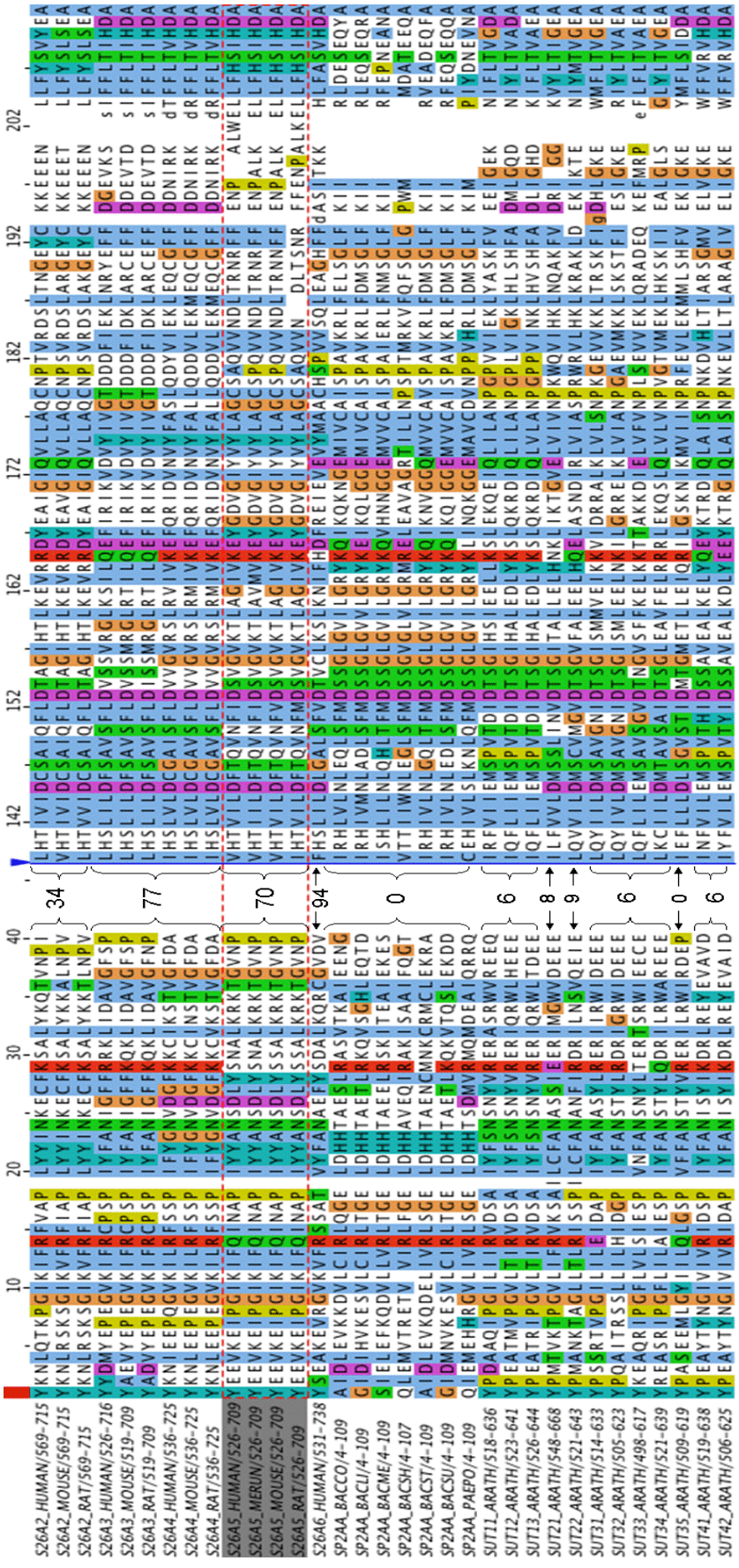

Figure 4: Multiple alignment of a selected set of STAS domains from antisigma factor antagonists and anion transporters. The STAS domains of thirteen human and murin (mouse and rat) SLC26 anion transporters (S26A) were aligned with seven antisigma factor antagonists (SP2AA) from different bacteria and twelve sulphate transporters (SUT) from *Arabidopsis thaliana*. The sequences of the inserts in the variable loop are replaced by the number of amino acid residues (between curly brackets). The alignment was obtained with the program Jalview, using the colour matrix ClustalX.

SULTR1.2 STAS domain

Studies on a sulphate transporter from *Arabidopsis thaliana,* SULTR1.2, examined the effect of deleting or modifying the STAS domain. The results suggest that the STAS domain is essential for facilitating localization of the transporter to the plasma membrane, but it is also critical for the whole sulphate transport activity (Shibagaki and Grossman, 2004).

The STAS domain of the plant sulphate transporter SULTR1.2 was modeled on the basis of the available NMR structure of *B. subtilis* SpoIIAA (1AUZ; Kovacs et *al*., 1998) and the crystal structure of *B. sphaericus* SpoIIAA (1H4Z; Seavers et *al.*, 2001). The structural analysis and modeling suggest that the SULTR1.2 C-terminal STAS domain shares the SpoIIAA fold, although it shows low overall sequence identity with SpoIIAA, around 17% over 130 residues (Figure 5; Rouached et *al.*, 2005). The analysis reveals a compact hydrophobic core at the interface of the α-helices and the β-sheets. This hydrophobic core appears very well conserved between the SULTR1.2 STAS domains and SpoIIAA. The similarity is particularly high in the vicinity of the phosphorylation site, despite the change from a conserved serine (Ser58) in SpoIIAA to the similar amino acid threonine (Thr587) in SULTR1.2.

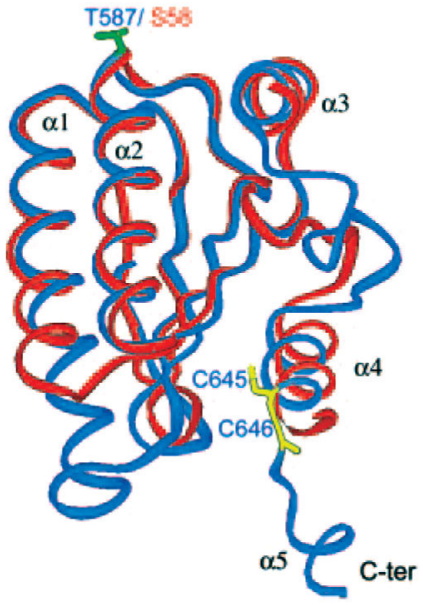

Figure 5: The three-dimensional model of the SULTR1.2 STAS domain. The crystal structure PDB 1H4Z and the deduced model of the SULTR1.2 STAS domain are shown in red and blue ribbons, respectively. The side chain of Thr587 (Ser58 in PDB 1H4Z) is represented in green and the two cysteines (Cys645 and Cys646) of the SULTR1.2 STAS domain are in yellow. α-helices are numbered and labelled sequentially. β-Strands are not labelled for clarity. The C-terminus of the peptide is labelled, whereas its N-terminus is hidden (behind helix α2) by the core of the structure (Rouached et *al.*, 2005).

The major difference between the modeled STAS domain and the SpoIIAA structure lies at the connection between the SULTR1.2 helix α1 and strand β3 (Figure 4). For the STAS domain of SULTR1.2, the variable loop comprises around 10 residues. This variable region lies at the periphery of the domain and is far away from the common phosphorylation region. The modeled STAS domain also differs from the SpoIIAA structure at the very C-terminus. This region is highly variable in length and sequence even between the various sulphate transporters. At the end of the C-terminal helix α4 of its STAS domain, SULTR1.2 possesses a pair of cysteines (Cys645, Cys646) that are not strictly conserved in the paralogs and that are not present in SpoIIAA (Figure 5). The two cysteins seem to play a critical role to maintain the full functionality of SULTR1.2 (Rouached et *al.*, 2005).

An experiment of random mutagenesis in the STAS domain of SULTR1.2 identified domain lesions that altered the transporter biogenesis and/or function (Shibagaki and Grossman, 2006). A number of mutations in the β-sheet that forms the core of the STAS domain prevent plasmamembrane accumulation of SULTR1.2. So the β-sheet seems to serve as a core structure of the STAS domain and lesions within this structure may disrupt proper STAS packing, which could destabilize the entire transporter. In contrast, the N-termini of the first and second α-helices have a number of amino acids critical for the function of the protein; mutations in these regions still allow protein accumulation in the plasmamembrane, but the protein is no longer capable of efficiently transporting sulphate into cells. These results confirm the critical role of the STAS domain for both the activity and biosynthesis/stability of the transporter, and that defined portions of the STAS domain correlate with these specific functions.

Rv1739c STAS domain

The SulP family members have been minimally characterized in bacteria. Anyway, it has been recently shown that induction of Rv1739c expression in *E. coli* increases bacterial uptake of sulphate (Zolotarev et *al*., 2008).

Sulphate uptake was also increased by overexpression of the Rv1739c transmembrane domain, but not of the cytoplasmic C-terminal STAS domain [437-560]. Expression of the isolated C-terminal cytoplasmic domain did not affect sulphate uptake. So, unlike the STAS domain requirement for sulphate transport by *A. thaliana* SULTR1.2

(Shibagaki and Grossman, 2004), the STAS domain was dispensable for the sulphate uptake by Rv1739c.

1.2 The protein prestin

Prestin is the fifth member (A5) of the Solute Linked Carrier 26 (SLC26) family of anion exchangers. It is highly and almost exclusively expressed in the outer hair cells (OHCs) of the organ of Corti in the inner ear of mammals. Although the basic function of SLC26A members is to transport anions (Mount and Romero, 2004), this is not prestin principal role. Unlike the other members of the SLC26 family, mammalian prestin has the unique property of the voltage-dependent conformational changes and it is considered the key player in the OHCs somatic electromotility (Zheng et *al.*, 2000). Since its discovery, it was clear that prestin is fundamentally different from other biological force generators. Its potential nanotechnology applications make it the most interesting subject among SLC26A family members, as shown by the increasing number of publications within recent years. For these reasons, I decided to deal with prestin separately in this section.

OHC electromotility

Cochlear hair cells are non-neuronal epithelial cells that transduce acoustic signals. They are organized in a tonotopic fashion, with those sensitive to high-pitched sounds at the basal end and those sensitive to low pitches at the apical end (Géléoc and Holt, 2003). Perpendicular to this gradient there are four rows of cells: a single row of inner hair cells, and three rows of outer hair cells (Figure 6). The inner hair cells (IHCs) transduce and transmit auditory information to the brain. Outer hair cells (OHCs) provide local mechanical amplification in the form of feedback, thus amplifying the auditory stimuli sensed by the inner hair cells (Dallos, 1992).

In practice, a pure tone stimulus causes the passive basilar membrane of the organ of Corti to resonate at a unique location that depends on frequency (Figure 8a). Active feedback refines or tunes the resonant location and amplifies the membrane motion, thereby enhancing auditory sensitivity to faint sounds by more than 40 decibel (that is 100 fold) (Dallos, 1992). There is a great deal of evidence indicating that OHCs are the principal players providing the feedback that drives cochlear amplification.

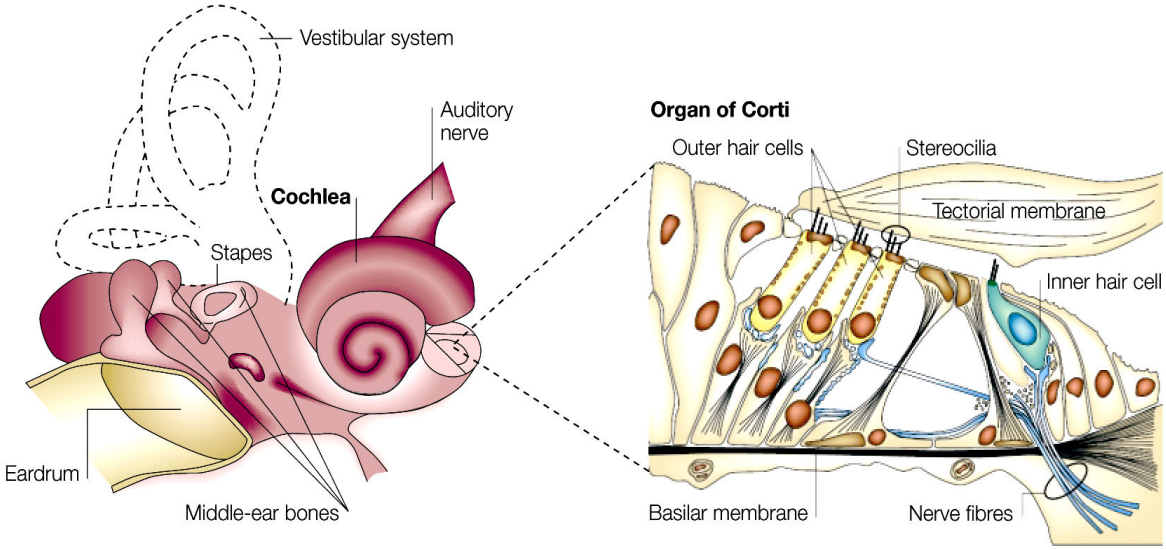

Figure 6: A cross section of the cochlea illustrating the organ of Corti, the sensory epithelium of the inner ear. A single row of inner hair cells and three rows of outer hair cells are located on the basilar membrane. The tectorial membrane overlies the epithelium and normally contacts the stereocilia of the outer hair cells (Dallos and Fakler, 2002).

For the physiological mechanism of amplification two candidate mechanisms have been proposed. One proposal, for which there is evidence in non-mammalian species, is that the apical stereocilia of OHCs act both as the sensors of the motion of the basilar membrane and as a motor source to amplify the motion (Hudspeth et *al.*, 2000). An alternative theory is based on the motility of the mammalian OHCs. OHCs have a distinctive hair (stereocilia) bundle (Figure 7), which is the mechanosensory input organelle of these cells. When mechanically stimulated by incoming sound waves, the ciliary bundle is deflected, and thereby triggers the opening and closing of mechanosensitive ion channels in the stereocilia membrane (Flock et *al.*, 1962; Hudspeth and Corey, 1977). But, unlike all other hair cells, OHCs then translate the resulting changes in membrane potential into macroscopic changes (up to 5%) in the length of their cylindrical cell bodies (Evans and Dallos, 1993). Depolarization triggers cell contraction, whereas hyperpolarization results in cell elongation (Brownell et *al.*, 1985; Kachar et *al.*, 1986). This "electromotility"

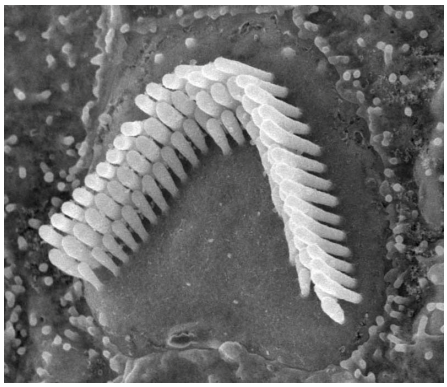

Figure 7: Electron microscopy image of the bundle of the stereocilia on the apical surface of outer hair cell in adult mouse (www.neuroscience.cam.ac.uk).

occurs at acoustic frequencies and generates the mechanical energy that is required for amplifying the sound-induced vibrations in the cochlea (Figure 8b). The local mechanical amplification mechanism that enables the high sensitivity and frequency selectivity of hearing in mammals is known as the "cochlear amplifier" (Ashmore, 1987).

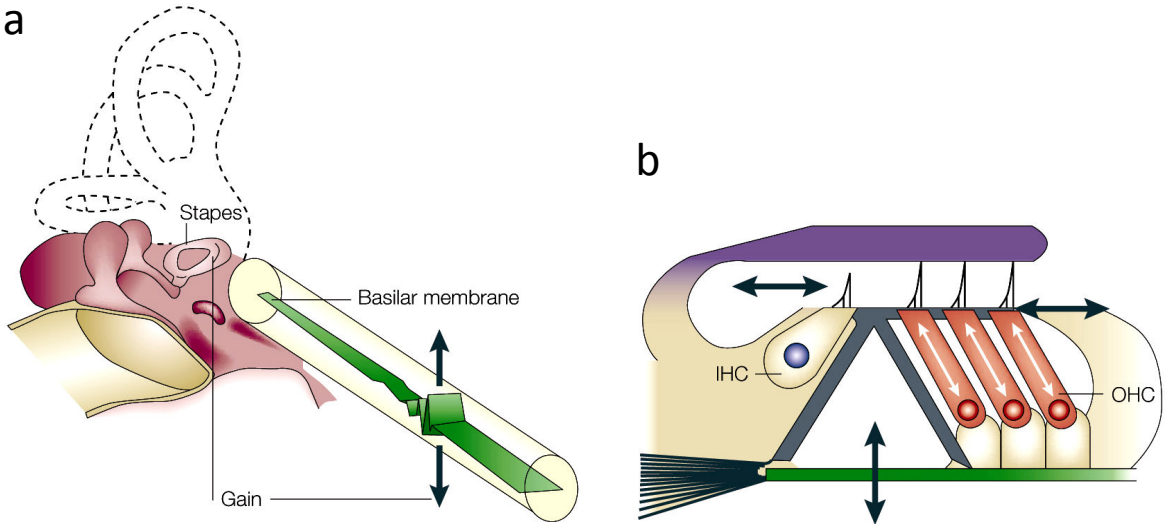

Figure 8: (**a**) Schematic diagram of the middle ear and cochlea, with the cochlea and basilar membrane shown straightened out and a "snap-shot" of a travelling wave at its optimal frequency for the recording location (arrows) superimposed on the basilar membrane. (**b**) Schematic diagram of the organ of Corti, with black arrows indicating the principal directions of motion in response to sound stimulation. OHCs expansion-contraction cycles (white arrows), in appropriate phase with basilar membrane motion, can boost the latter amplitude (Adapted from Dallos and Fakler, 2002).

The main features of OHC electromotility are the following: first, electromotility takes place without hydrolysis of high-energy phosphates such as ATP, but the energy is supplied by the changing membrane potential of the cell (Kachar et *al.*, 1986; Holley and Ashmore, 1988); second, Ca^{2+} ions are not required for the expression of this response, but internal Ca^{2+} levels can modulate it (Ashmore, 1987); third, the electromotile response occurs at microsecond rates (Dallos and Evans, 1995) and works in a cycle-by-cycle mode, up to a frequency of at least 70 kiloHertz (Frank et *al.*, 1999). These microsecond rates are faster than that of any other biological force generator.

The tight coupling between electromotility and transmembrane voltage is reflected by the phenomena of gating charge movements, a charge dislocation across the cell membrane, similar to those observed for voltage-gated ion channels (Armstrong and Bezanilla, 1977). This charge transfer (gating current) gives rise to a voltage-dependent non linear component of membrane capacitance, which shows a bell-shaped dependence on membrane potential (Ashmore, 1990). Since the non linear capacitance (NLC) is easier

to measure than motility, it is widely used as a signature of the electromotile process (Santos-Sacchi, 1991).

As a consequence of all these observations, it is reasonable to assume that the fast mechanical changes in OHCs are powered by a molecular motor that is fundamentally different from other biological force generators, such as the myosin, kinesin or dynein families. The OHC molecular motor performs direct, rapid, reversible electro-mechanical conversion (Zheng et *al.*, 2000).

The discovery of prestin

All these findings led to the hypothesis of an integral membrane protein, termed the motor protein, as the molecular element underlying fast OHC motility (Dallos et *al.* 1991; Kalinec et *al.*, 1992). In response to changes in the transmembrane voltage, the motor protein is thought to undergo a structural rearrangement that changes its area in the plasma membrane (Dallos et *al.*, 1993; Iwasa, 1994). As a result of the concerted action of a large number of motor molecules supposed to be densely packed in the OHCs basolateral membrane, the cell changes its length up to 5%. Indeed, freeze fracture electron microscopy of OHC membranes reveals densely packed 11 nm diameter particles that appear ideally situated to mediate somatic motility (Figure 9; Forge et *al.*,, 1991; Kalinec et *al.*, 1992).

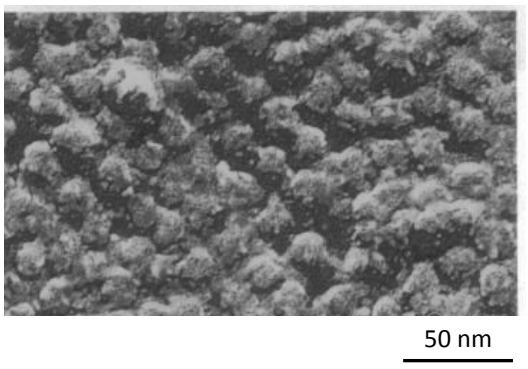

Figure 9: Freeze-etched replicas of the lateral plasma membrane of an OHC partially extracted with Triton X-100 in low salt buffer. The rotary shadowed replicas of the true outer surface of the lateral plasma membrane show large particles forming a regular array. Particle density is about 2500 per μm^2 (Kalinec et *al.*, 1992).

To identify motor protein candidates, Dallos and colleagues used a subtractive cloning strategy to amplify transcripts expressed in OHCs but not in the non-motile IHCs.

A cDNA that is specifically expressed in OHCs was isolated and termed *prestin*, from the musical notation *presto*, meaning fast (Zheng et *al.*, 2000).

Immunolocalization revealed significant staining along the lateral membranes of OHCs, with a developmental time course that paralleled the acquisition of somatic electromotility (Belyantseva et *al.*, 2000). When heterologously expressed in mammalian cell lines, prestin showed all the hallmarks of the OHC motor protein. First, it endowed transfected cells with NLC, which had similar characteristics to that observed in OHCs. Second, prestin expressing cells displayed voltage-dependent shape changes (Zheng et *al.*, 2000). Third, actual force measurements, carried out with an atomic force microscope, showed that prestin generates significant mechanical force, which is independent of frequency up to at least 20 kHz (Ludwig et *al.*, 2001).

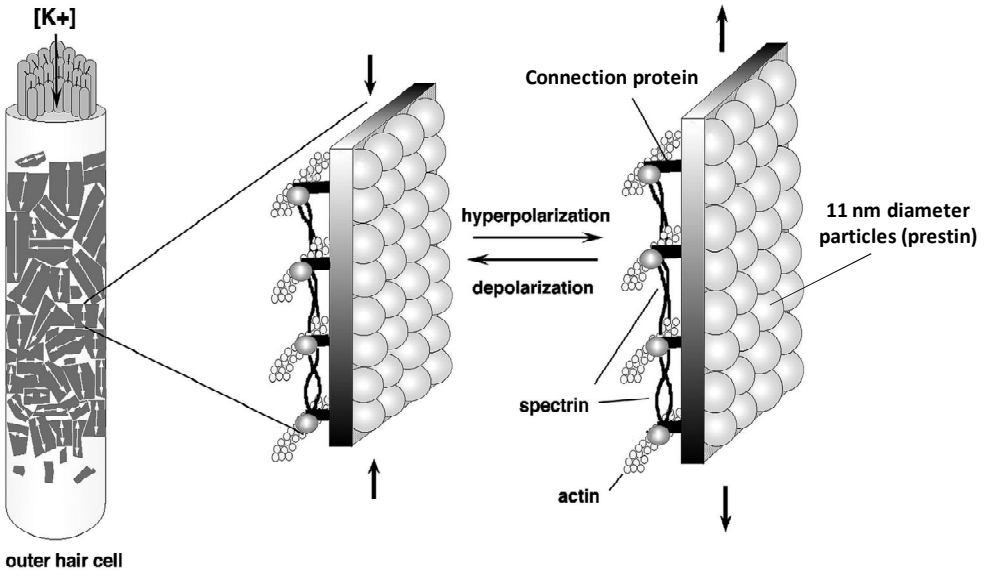

Figure 10: Schematic representation of OHC force generation unit and of the interactions between citoskeletal structures and the 11 nm membrane particles that enclose prestin (Adapted from Frolenkov et *al.,* 1998).

The molecular basis of electromotility can be explained by voltage-dependent conformational rearrangements of prestin, with different conformations occupying different areas in the membrane (Zheng et *al.*, 2000). The specialized cortical cytoskeleton, which lies beneath it, is adapted to maintain the cylindrical shape of the OHC and may ensure that forces produced in the membrane lead to changes of cell length (Figure 10; Frolenkov et *al.*, 1998).

Prestin and deafness

The restricted expression of prestin in OHCs and its proposed function as a mechanical amplifier make it a strong candidate for an association with human deafness. However, the role and the extent of the *prestin* gene defects in human non-syndromic hearing impairment are still poorly understood.

The fundamental role of prestin for normal auditory function was first shown in mice: the deletion of prestin results in the loss of about 40-60 dB in hearing sensitivity (Liberman et *al.*, 2002) and elimination of frequency selectivity (Cheatham et *al.*, 2004).

The human *prestin* gene contains 21 exons and is localized on the long arm of chromosome 7 (7q22.1). A single nucleotide change in the second intron of SLC26A5 was reported to be associated with hearing loss (Liu et *al.*, 2003). This IVS2-2A>G DNA sequence variation occurs in the first coding exon 3 splice acceptor site of the *prestin* gene. It was suggested that this mutation leads to aberrant mRNA splicing and results in non-syndromic moderate-to-profound sensorineural hearing impairment. In addition, a relatively high frequency of heterozygosity for this sequence change was observed in affected subjects, suggesting the possibility of a semi-dominant influence of the mutation. By contrast, further studies demonstrated that the IVS2-2A>G variant may not occur more frequently in hearing impaired patients than in controls, and heterozygosity for this transition may not be sufficient to cause hearing loss (Tang et *al.*, 2005; Teek et *al.*, 2009).

In addition, a heterozygous missense mutation (R150Q) in the sixth coding exon of the *prestin* gene was reported to potentially cause mild to moderate non-syndromic hearing loss (Toth et *al.*, 2007). This is the first genetic and electrophysiological analysis of a human mutation in a coding exon of the *prestin* gene, although the pathogenic role of the R150Q mutation is not unambiguous.

These two changes are, so far, the only ones reported with potential clinical importance. Further studies are needed to clarify the pathogenic role, if any, of these nucleotide substitutions, as well as other SLC26A5 changes, in the etiology of hearing loss.

Reciprocal electromechanical properties of prestin

Prestin, like other transducers, exhibits piezoelectrical properties: it generates mechanical force upon electrical stimulation and may also change its electrical properties

upon mechanical stimulation (Ludwig et *al.*, 2001; Santos-Sacchi et *al.*, 2001). It was estimated that a single prestin molecular assembly produces a force in the OHC axial direction of about 2.4 picoNewtons and a conformational displacement of around 1 nm (Zheng et *al.*, 2000). In turn, the efficiency of conversion from mechanical force to electrical charge was estimated by measuring charge displacement induced by stretching the cell with known force (Dong *et al.*, 2002). The value, around 20 femtoCoulomb per nanoNewton is four orders of magnitude greater than that obtained for the best man-made material. The remarkable properties of prestin make it a candidate for future nanotechnology applications. Prestin ensembles could function as mechanical, voltage-controlled actuators at exceptional speeds.

Prestin topology

Prestin is a transmembrane glycoprotein of 744 residues, with a molecular weight of about 81 kDa (Zheng et *al.*, 2000). It contains about 50% of non-polar residues and it shares the overall structure and specific protein domains of the SLC26 family, such as a highly conserved central core of hydrophobic amino acids, a short N-terminus and a long C-terminus (Zheng et *al.*, 2001). The mostly hydrophilic amino- and carboxy-termini, which flank the hydrophobic core, are located in the cytoplasm, as it was shown by epitope tagging and immunocytochemistry (Ludwig et *al.*, 2001; Zheng et *al.*, 2001).

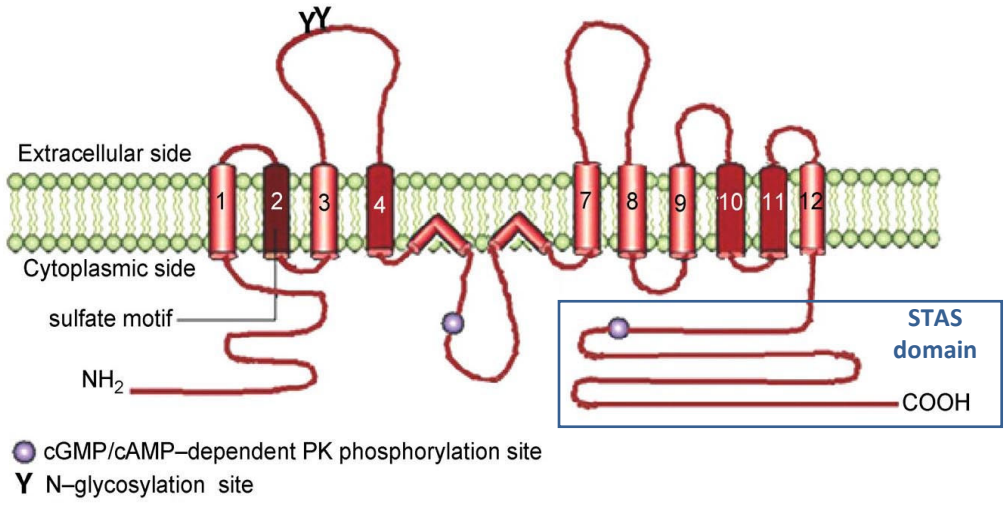

Figure 11: Representation of the membrane topology of prestin, with the 12 membrane helixes, N- and C-terminal cytoplsmic domains. On the basis of the existence of a phosphorylation site at the level of the third loop, helices 5 and 6 are inserted into the membrane, but do not cross it, forming re-entrant loops. The conserved "SulP transporter signature" is present in the second transmembrane domain while a STAS motif is located in the C-terminal region. The two potential N-glycosilation sites Y (Asn163 and Asn166) are labelled on the extracellular surface of the protein (Adapted from Deak et *al.*, 2005).

The number of the membrane helixes is still disputed as topology prediction programs produce ambiguous results: 10 or 12 transmembrane helixes can be hypothesized (Oliver at *al.*, 2001; Zheng at *al.*, 2001; Deak et *al.*, 2005; Navaratnam et *al.*, 2005;). The 12 transmembrane domains model is supported by more experimental evidence and it is, in part, based on placing two potential N-glycosilation sites (Asn163 and Asn166) on the extracellular surface of the protein (Matsuda et *al.*, 2004). In Figure 11, prestin is represented with 12 membrane helixes: on the basis of the existence of a phosphorylation site (cGMP/cAMP-dependent PK phosporylation syte) at the level of the third loop, helices 5 and 6 are inserted into the membrane, but do not cross it, forming re-entrant loops (Deak et *al.*, 2005).

The conserved "SulP transporter signature" is present in the second transmembrane domain, while the C-terminal cytoplasmic region includes the Sulphate Transporter and Anti-Sigma factor antagonist (STAS) domain. Two distinctive charged segments are located in the C-terminal region: a positive-charge cluster is located at residues 557-580; adjacent to this there is a negative-charge cluster at residues 596-613.

Although prestin is most closely related to SLC26A6, the human and mouse orthologs of A6 have only 78% amino acid identity. In contrast, prestin is a highly conserved protein with 92.7% of amino acids being identical among four different mammalian species: human, mouse, rat and gerbil. Such a high degree of conservation is not common among other SLC26A members. Significant changes in prestin primary sequence occurred after the split between mammalian and avian lines, suggesting that prestin evolved in order to fit special mammalian needs (Dallos et *al.*, 2006).

Mechanism of action

Prestin is a new type of biological motor. It is entirely different from the conventional enzymatic-activity-based motor proteins, in that it does not need ATP to function, but it is a direct voltage to force converter. In this case the energy is supplied by the changing membrane potential of the cell and this is probably unique in the animal kingdom (Dallos et *al.*, 2006). The action of prestin is also orders of magnitude faster than that of any other cellular motor protein, as it functions at microsecond rates. In fact, OHC motility works at frequencies up to at least 70 kHz (Frank *et al.*, 1999).

Although prestin possesses all the sequence domains conserved throughout the SLC26 family, it has not yet been shown to function as an anion transporter. Moreover,

neither gating charge movements nor a NLC have been reported for any other member of the SLC26 family, suggesting that prestin may have a unique function within the family.

How the membrane potential change of OHCs results in structural changes in prestin, corresponding to the motor function, is not understood yet. Conceptually, prestin should comprise at least two essential functional domains: the voltage sensor that detects changes in the transmembrane potential of the cell, and the actuator that undergoes a conformational change and thereby facilitates cell contraction or elongation in response to depolarization and hyperpolarization, respectively (Dallos and Fakler, 2002).

Incomplete transporter

It is reasonable to assume that the voltage-sensing charged group is different for prestin than in other SLC26A proteins, which produce no motility. With this in mind, Oliver and colleagues altered each charged, non conserved amino acid in the putative membrane-interacting region of prestin, individually or in groups (Oliver et *al.*, 2001). Surprisingly, no combination of mutations eliminated NLC or altered its gain. These results led to the suggestion that the voltage sensor may not be an intrinsic component of the protein, but an extrinsic ion. Using inside-out and outside-out membrane patches, it was demonstrated that intracellular Cl^- functions as the extrinsic voltage sensor.

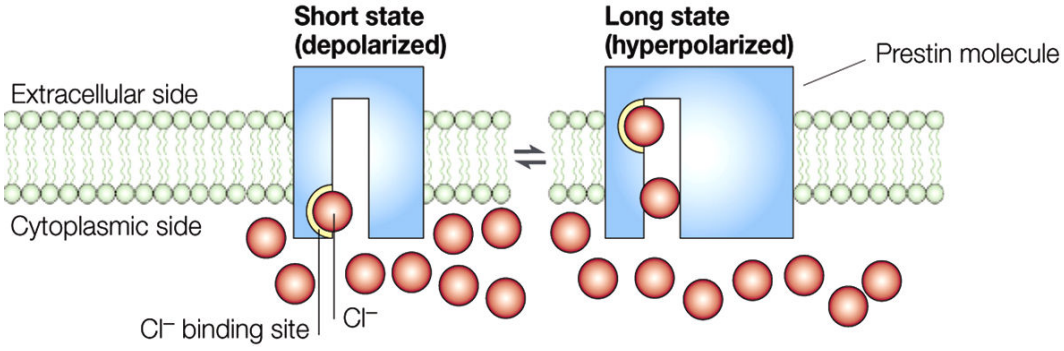

Figure 12: Representation of the incomplete transporter model. When the cell membrane is hyperpolarized, bound Cl^- is translocated across the molecule towards the external face. When bound Cl^- is in the outside position, the molecule assumes its "long" conformation. Conversely, when the bound Cl^- is displaced back towards the cytoplasm, the molecule becomes "short" (Adapted from Dallos and Fakler, 2002).

Detailed analysis showed that the half-activating Cl^- concentration was 6 mM, matching the normal intracellular amount of this anion. They proposed that after binding to a site with millimolar affinity, these anion is translocated across the membrane, without

being released in the extracellular space, by the transmembrane voltage: toward the extracellular surface upon hyperpolarization, toward the cytoplasmic side in response to depolarization. Subsequently, this translocation triggers conformational changes of the protein that finally change its surface area in the plane of the plasmamembrane (Figure 12). The reaction scheme is illustrated in Figure 13a.

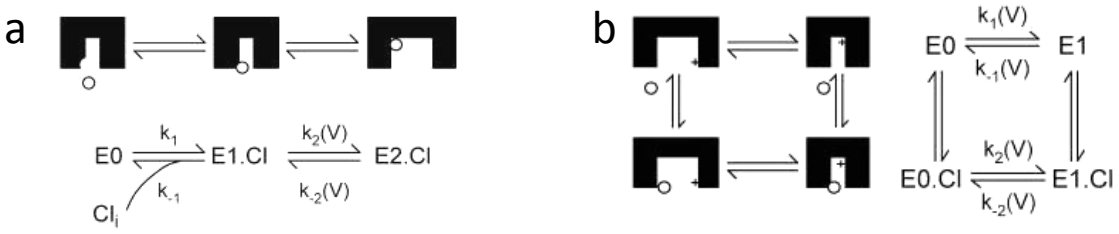

Figure 13: The reaction schemes and the representation of two alternative modes of the incomplete transporter model. (**a**) Prestin changes from a contracted to an expanded state when a Cl^- ion moves from the first binding site at the mouth of the pore to a second site at the top of the pore. (**b**) Prestin changes from an expanded to a contracted state when Cl^- binding enables the allosteric change of an intrinsic positively charged sensor (Muallem and Ashmore, 2006).

Subsequent investigations showed that as intracellular Cl^- concentration decreases, the amount of charge transferred also decreases and voltage sensitivity shifts in the depolarizing direction (Rybalchenko and Santos-Sacchi, 2003; Santos-Sacchi *et al.*, 2006). The direction of shift implies that the net charge moved across the membrane is positive. Thus, two alternatives exist to the idea that Cl^- is the voltage sensor. It is possible that monovalent anions need to attach to a binding site and their combination, with net positivity, is translocated across the membrane. Alternatively, chloride binding could enable an allosteric change, thereby allowing a positive gating charge to be moved (Rybalchenko and Santos-Sacchi, 2003). In this case, all charge movement is provided by the translocation of the intrinsic charged sensor. The reaction scheme of the last proposed mechanism is illustrated in Figure 13b.

Anion antiporter

Recent theoretical work suggests that many experimental data could be better explained if one assume that prestin acts as an electrogenic anion exchanger, exchanging one Cl^- ion for one divalent or two monovalent anions (Muallem and Ashmore, 2006). Unequivocal experimental verification of anion transport is not available yet. According to this model, the charge movement arises as a result of both a Cl^- ion and intrinsic charged

residues moving across the membrane. Thus net positive charge is moved across the membrane as the Cl^- ion is moved towards the extracellular surface. This model is independent of the nature of the Cl^- replacing anion which could be mono- or divalent as long as it guarantees that the reorientation of the intrinsic charged residues is electroneutral. The reaction scheme is illustrated in figure 14.

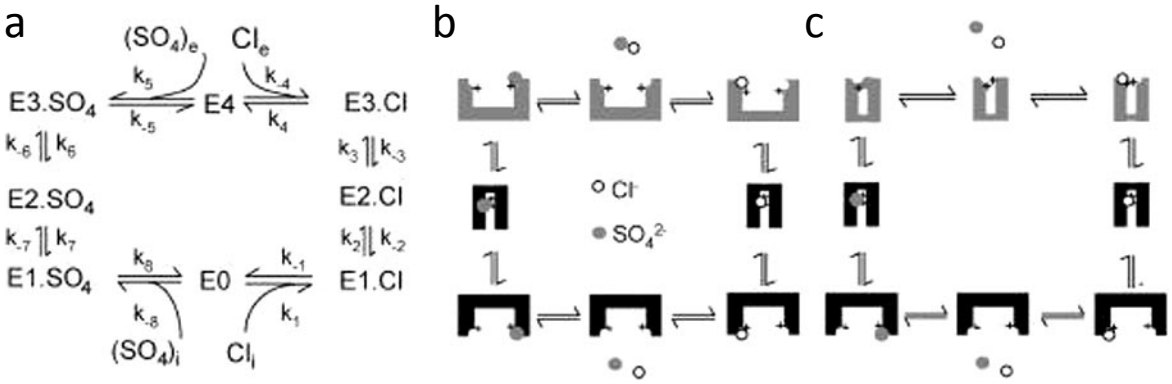

Figure 14: (**a**) The reaction scheme for a Cl^-/SO_4^{2-} exchanger model. Prestin exchanges one Cl^- ion for one SO_4^{2-} ion via an alternating-access mechanism, in which prestin can only change between inward and outward facing states with an anion bound. (**b**, **c**) Two alternative representations of the reaction scheme: both assignments ensure that the critical voltage-dependent transition, E1.Cl⇔E2.Cl, is associated with a conformational change of prestin into a compact state and symmetry is maintained (Muallem and Ashmore, 2006).

Prestin STAS domain

The intracellular C-terminus of prestin is the least well-conserved region compared with other SLC26A proteins and it includes a STAS domain. It has only 25-35% homology with its SLC26A relatives and it is expected to be responsible for the protein specific function. Different experiments showed that changing charged amino acids in the C-terminus to either the opposite charge (R, K > D; E, D > K) or a neutral amino acid (Q) is not able to abolish NLC and does not disrupt plasma membrane (PM) targeting of prestin (Oliver et *al.*, 2001; Bai et *al.*, 2006).

The role of the C-terminus of prestin was investigated in some detail by Dallos and his group with a series of deletion, point and chimeric mutants (Zheng et *al.*, 2005). The function and cellular expression of mutants were examined in a heterologous expression system (TSA-201 and OK cells) by measurement of NLC and confocal immunofluorescence. The subcellular localization of mutant proteins was examined by co-localization experiments of prestin with other subcellular component markers. The

following set of C-terminal truncation mutants was examined in this study: Del516, Del525, Del590, Del630, Del709 and Del719 (Figure 15).

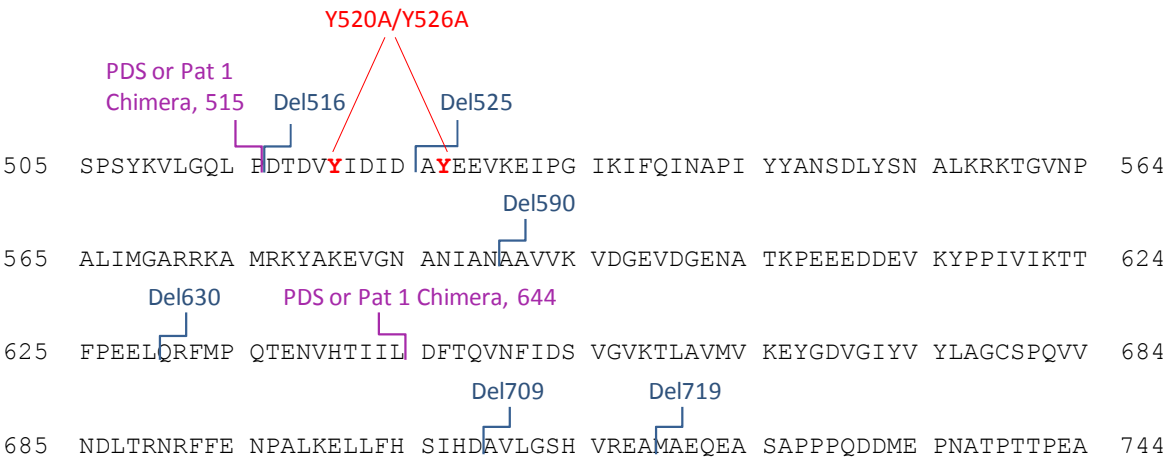

Figure 15: Primary sequence [505-744] of prestin C-terminus from gerbil. The locations of the mutations created and examined in the study are indicated. These include deletion mutants (in blue), chimera junction points (in violet) and double point mutations (in red).

Del719 is the only deletion mutant that retains NLC function and proper PM targeting. The mutants Del516, Del525 and Del590 showed prestin localization consistent with retention of the protein in the endoplasmic reticulum (ER) and in the Golgi apparatus of the cells. Del630, Del709, aside from ER and Golgi retention, displayed widespread cytoplasmic membranous distribution, without apparent PM localization. Thus, deletion of more than 35 C-terminal amino acids results in impaired delivery of prestin to the PM and consequent complete removal of NLC function. This indicates that amino acids between 709 and 719 are required for proper PM targeting and NLC function. A comparison of Del590 and Del630 is also particularly revealing. The subcellular localization results suggest that the region of prestin between amino acids 590 and 630 is necessary for prestin to exit from the ER/Golgi into cytoplasmic vesicles.

In an attempt to restore PM targeting in the truncation mutants, a set of chimeric prestin constructs were created in which the analogous C-terminus portions of PAT1 (SLC26A6) or Pendrin (SLC26A4), the two most closely related proteins to prestin, were exchanged for the prestin C-terminus at sites 515 and 644. The “515” chimera replaced almost the entire C-terminus of prestin with either pendrin or Pat1, while “646” chimera replaced part of the STAS domain (Figure 15). All chimeric proteins lacked NLC, and had altered cellular distribution, with ER and Golgi retention as well as cytoplasmic membranous distribution. Since the chimera mutants (Prestin/Pendrin and Prestin/PAT1)

could not restore prestin PM targeting, the capacity for prestin to insert into the PM of cultured epithelial cells may be dependent on prestin specific C-terminal amino acid residues.

A tyrosine-containing motif (YXXΦ) is one of several well studied motifs that direct the transport of newly synthesized membrane protein from the trans-Golgi network to the basolateral membrane (Keller and Simons, 1997). In this motif Y is tyrosine, X is any amino acid, and Φ is a bulky hydrophobic amino acid. There are seven potential tyrosine containing motifs in the C-terminus of prestin. The aberrant PM targeting seen with the deletion mutants and chimeric proteins may be related to the loss of potential basolateral membrane targeting motifs located in the C-terminus. To approach the question whether these motifs might be involved in membrane targeting of prestin to its PM location, a double point mutant was created: Y520A/Y526A, which abolished two of the seven potential tyrosine-containing motifs (Figure 15). The mutation Y520A/Y526A resulted in lost of NLC function and in intracellular accumulation of prestin. These observations indicate that specific sequences within the C-terminus are essential for the NLC function in addition to its role in membrane targeting.

Altogether these data indicate that the C-terminus of prestin is likely to be intimately involved in anion binding, membrane targeting, and the voltage regulated conformational change of prestin. How the distal amino acids of the C-terminus regulate PM targeting and protein function is not understood yet.

Oligomerization properties

The examination of OHCs membranes by freeze fracture reveals densely packed 11 nm diameter particles (Figure 9; Forge et *al.*, 1991; Kalinec et *al.*, 1992). It has been a consistent assumption that the particles consist of some multimer of the motor protein, inasmuch as the 744 amino acid prestin molecule is too small to produce an 11 nm monomer. How prestin forms oligomers and what part of the molecule is involved in their formation is not completely clear yet, although the involvement of both the N- and the C-terminal domains has been suggested (Navaratnam et *al.*, 2005; Zheng et *al.*, 2005).

The first evidence for prestin multimerization came from fluorescence resonance energy transfer experiments, that showed that homodimerization of prestin depends on an intact N-terminus (Navaratnam et *al.*, 2005; Greeson et *al.*, 2006).

The number of subunits necessary to form a functional motor protein was first addressed by Zheng and colleagues (Zheng et *al.*, 2006). In this study, native and recombinant prestin, obtained from different expression systems, including yeast and mammalian cell lines, was seen resistant to dissociation by lithium dodecyl sulphate (LDS) and behaving as a stable oligomer on LDS-PAGE. Chemical cross-linking and perfluoro-octanoate-electrophoresis (PFO-PAGE) combined with immunoblotting and affinity purification suggest a tetrameric subunit stoichiometry of prestin. Moreover sodium dodecyl sulphate (SDS) dissociates the tetramer into dimers that could be converted to monomers by hydrophobic reducing agents, but not by the hydrophilic ones. These data suggest that prestin monomers are covalently linked to dimers by disulfide bonds located in the hydrophobic membrane core and that these covalently linked dimers associate via hydrophobic interactions to form a tetramer. They proposed that the stable covalent dimer may act as the building block for producing the higher order oligomers that form the 11 nm particles in the OHC basolateral membrane.

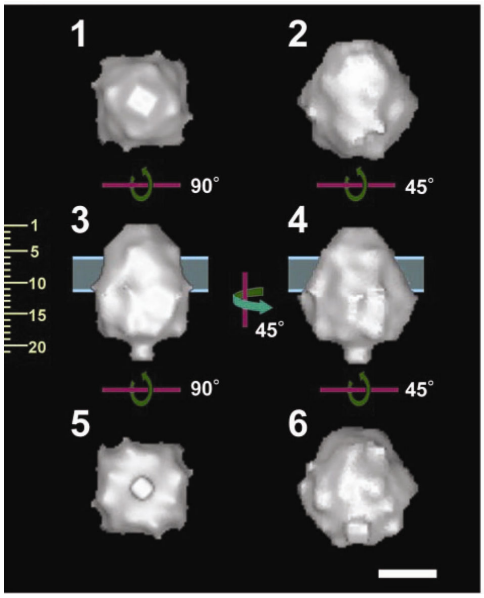

Figure 16: Density map of the 3D reconstruction. The putative position of the transmembrane domain is indicated by two blue lines (~30 Å apart) in side images 3 and 4, which were determined so that the volumes of extracellular, transmembrane, and cytoplasmic domains are almost 15, 33, and 52%. Protein is displayed in bright shades.

By contrast, the experiments of Detro-Dassen, while acknowledging dimers as the functional form, deny that these are formed by covalent bonds (Detro-Dassen et *al.*, 2008). They studied the subunit stoichiometry of rat, zebrafish prestin and of other SulP proteins, SLC26A3 and the bacterial paralog from *Pseudomonas aeruginosa* (PASulP), expressed

in *Xenopus laevis* oocytes or in mammalian cells. According to blue native PAGE and chemical cross-linking experiments, prestin and the other SulP proteins form dimers as predominant oligomeric state. Oligomers dissociate entirely into monomers under non-reducing conditions in the presence of low concentrations of SDS. So they concluded that dimers are held together by non-covalent forces rather than by covalent disulfide bonds.

A preliminary indication of prestin shape was provided by Mio and colleagues who expressed prestin in baculovirus-infected Sf9 cells and purified it (Mio et *al.*, 2008). They observed the negatively stained molecules using electron microscopy, and reconstructed the 3D structure of prestin at 2 nm resolution by single particle analysis. Their result is consistent with prestin being a tetramer, having a large cytoplasmic domain and assuming a "bullet shape", with a fourfold symmetry (Figure 16).

Prestin orthologs

A recent development in the study of prestin is the analysis of its orthologs. Indeed, when assessed by sequence similarity, the closest homolog of mammalian SLC26A5 is that found in the zebrafish hearing organ (Albert et *al.*, 2007). The zebrafish prestin orthologue, zprestin, shares around 50% amino acid identity with mammalian prestin. Like its mammalian orthologue, zprestin is expressed in hair cells of the ear and confers NLC to the membranes of transfected cells, similar to the characteristic electrogenic charge movement that accompanies the prestin-mediated somatic electromotility of mammalian OHCs. Although expression analysis and electrophysiological data show that zprestin properly localizes to the cell membrane upon heterologous expression, it nonetheless fails to generate electromotile responses of the transfected cells. Hence, though displaying a prestin-like voltage sensitivity, zprestin does not seem to be a prestin-like motor, supporting the general view that a prestin-mediated somatic electromotility is a unique feature of mammalian OHCs.

The transport function of zprestin was tested, revealing that it is an electrogenic divalent/chloride anion antiporter, exchanging sulphate or oxalate for chloride in a strictly coupled manner with a 1:1 stoichiometry (Schaechinger and Oliver, 2007). The same result was obtained with the chicken ortholog, which most probably lacks electromotility.

Thus, the prestin orthologs from chicken and zebrafish appear to be an intermediate form, performing voltage-dependent, chloride-sensitive charge dislocations, including full anion transport, but not having acquired motility. The presence of prestin transporter in

non-mammalian hair cells makes sense in evolutionary terms: a protein already present in the hair cells of phylogenetic ancestors to mammals may have adopted a novel, electromotile function during evolution toward the mammalian outer hair cell. Zebrafish and chicken prestin can thus be regarded as the "missing link" that may help us to understand the sequence of events that has led to the emergence of a novel type of motor protein in the course of SLC26 evolution.

2

The project

Aims of this study

Despite the increasing interest in the SLC26 genes and, in general, in the SulP members, a substantial amount of research is still needed to understand the roles of these transporters. In particular, very little is known about the structural organization of these proteins and no three-dimensional structure of domains or full-length sequences are available for any mammalian SLC26 anion transporter or for other members of the SulP family, of any species. The structural characterization is fundamental for the comprehension of the mode of action of a protein and it is an essential step for the understanding of the functional consequences of the mutations responsible for related pathologies. In this context, the long term task is the elaboration of a molecular functional model of the SulP anion transporters, able to explain both the common features of this class of transporters and the peculiar characteristics of the single components.

To this purpose, first important step is the structural characterization of the functionally important C-terminal STAS domain. The work described in this thesis has been focused on the production of different forms of the STAS domain, for the biophysical and structural characterization. The second part of the project is more ambitious and concerns the production of the full-length membrane proteins by a cell-free expression system.

The strategy

To characterize a protein from a structural point of view, it is necessary to produce it in amounts in the order of milligrams. Due to the low abundance of the SulP transporters in natural sources, the unavoidable choice is the production of recombinant material. The strategy adopted includes amplification of the selected genes, starting from the cDNA of the related proteins, cloning into appropriate bacterial expression plasmids, expression in *E. coli* and purification. All the used expression vectors produce a recombinant protein linked to another protein or a short peptide with well-known properties (the so-called "tags"). These fusion tags help to isolate the protein of interest by affinity chromatography. Subsequently, the recombinant protein is excided from the tag by an appropriate proteolytic enzyme and further purified. For the structural and biophysical characterization, several complementary techniques are used, such as circular dichroism

(CD) and fluorescence spectroscopy, dynamic light scattering (DLS), and, if possible, X-ray crystallography, for the characterization at atomic level.

Since the size of the STAS domain is suitable, it is characterized also by solution NMR, in collaboration with the group of Prof. Stefano Mammi from the Department of Chemical Sciences of the University of Padova. The joint crystallographic and NMR efforts may provide a complete structural characterization of the STAS domains, giving complementary information. Crystallography can reveal the high resolution structural details of the STAS domains and their binding properties to tightly bound small molecules and ions. NMR can complete the structural characterization by providing information about the flexibility and dynamics of the domain in solution (most of the variable loop is predicted disordered), and the binding properties to medium or low affinity species.

For the production of the full-length membrane proteins, a variety of intrinsic problems exists with the commonly employed heterologous expression systems. Cell-free expression system represents a recently developed and powerful alternative to *in vivo* expression.

Production and characterization of the STAS domain

The STAS domains of different SulP proteins, from distance-related species, are selected for the structural and biophysical characterization: prestin (SLC26A5) from *Rattus norvegicus* and *Meriones unguiculatus*, pendrin (SLC26A4) from *Homo sapiens* and the bacterial sulphate transporter Rv1739c from *Mycobacterium tuberculosis*.

The attention is focused on mammalian pendrin and prestin, because they share interesting properties: both are expressed in the inner ear and mutations in each of the two genes are linked to deafness or hearing impairment. Pendrin is an extremely interesting target also because the Pendred syndrome is genetically well characterized.

The study of mycobacterial SulP transporters offers potential insight into the sulphur assimilation pathways leading to biosynthesis of sulfolipid pathogenicity determinants. Moreover, the structural characterization of Rv1739c STAS domain may shed light on the evolution mechanisms of this domain and in the different roles played in the various transporters. This approach can be also “technically” useful, since it is well known that even similar polypeptides can have substantially different properties in terms of propensities to crystallize or NMR feasibility.

In order to identify a sequence corresponding to a compact single domain, several types of analyses and predictions are performed on the C-terminal part of each SulP protein, such as multiple sequence analyses, secondary structure predictions, predictions of intrinsic disordered regions and homology modeling outputs. The accurate selection of the N- and C-termini is more critical for domains that are part of a larger protein, as it is the case of the SulP STAS domains, whose boundaries are not clearly defined by sequence alignments. For this reason constructs of different length are selected, for each SulP protein. In parallel, information from literature, in particular functional data on mutations and deletions, are taken into account.

From the sequence alignment in Figure 17 it is immediately evident that the sequence similarity between mammalian transporters and Rv1739c C-terminal region is very low, (less than 20%). Most of the difference is found in the long insertion (70, 80 ammino acids long, between the position 60 and 140 in the alignment) and in the longer C-terminal extension present in the mammalian transporters STAS domain. The evolutionary and functional role of the insertion is unknown; secondary structure predictions suggest it is largely unstructured. According to these considerations, Rv179c STAS domain is more similar to the ASA protein STAS domain (SpoIIAA from *Bacillus subtilis* is reported in the alignment as reference). Taking into account these differences in lengths as well as the low amino acid conservation observed, most probably the 3D structure of SulP STAS is significantly different from that of the ASA proteins. Presumably, the differences are responsible for the distinct properties of this domain when part of the different transporters.

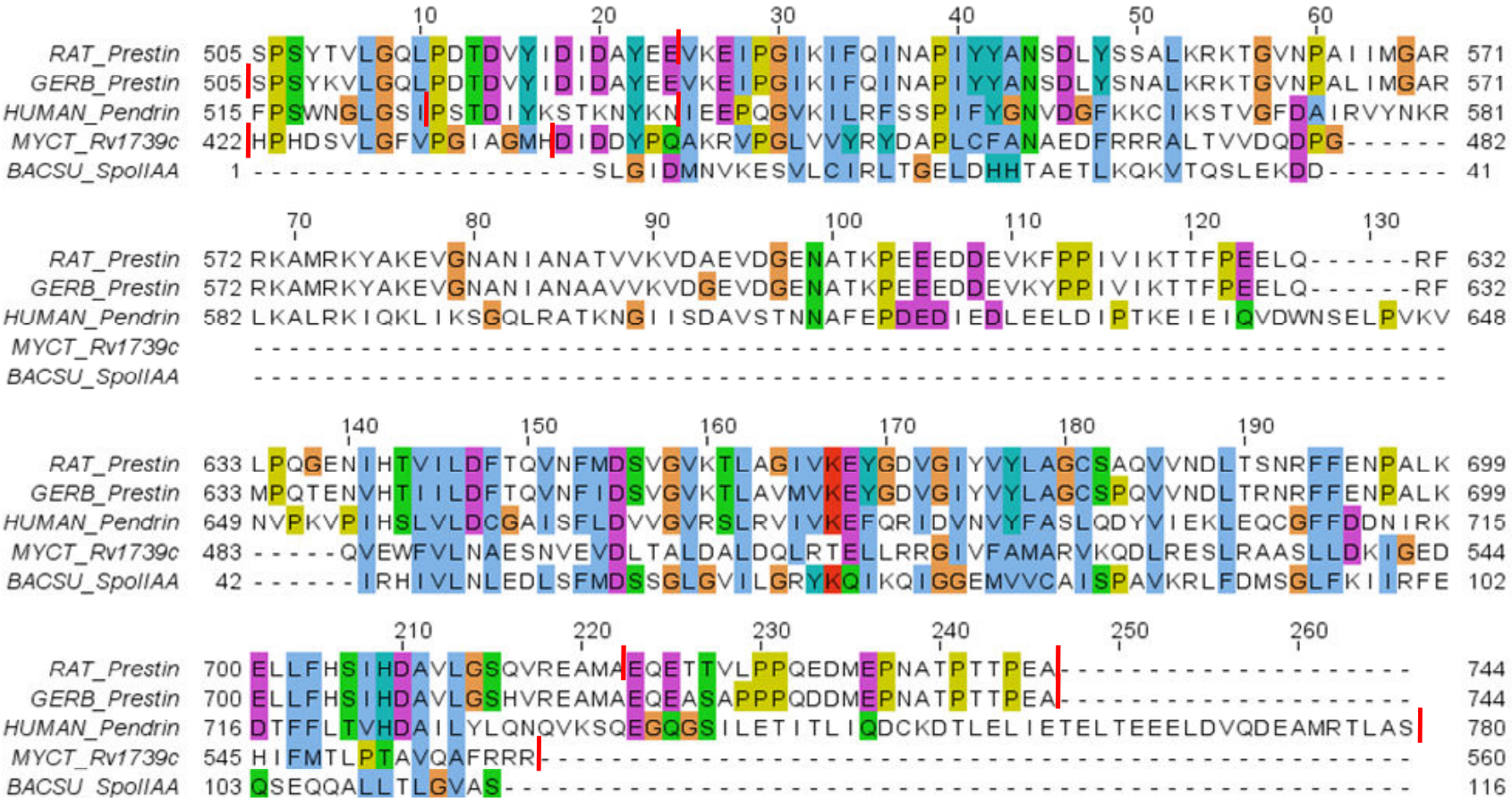

Figure 17: Multiple sequence alignment of the C-terminal portion, comprising the STAS domain, of the four selected anion transporters: Prestin from *Rattus norvegicus* and *Meriones unguiculatus* [505-744]; Pendrin from *Homo sapiens* [515-780]); Rv1739c from *Mycobacterium tuberculosis* [422-560]. As reference, the sequence of the antisigma-factor antagonist from *Bacillus subtilis*, SpoIIAA, is reported. The red lines indicate the various N- and C-terminal boundaries used to screen for the optimal STAS domain boundaries. The alignment was obtained with the program Jalview, using the colour matrix ClustalX.

Production of SulP proteins by cell-free expression system

The structural characterization of membrane proteins is still a challenge mainly because of the difficulty to produce adequate amount (in the order of milligrams) of purified and functional material (Lacapere et *al*., 2007). This is particularly true for eukaryotic membrane proteins, whose high resolution 3D structure is known only for a very few of them.

The production of full-length prestin is first tested for over-expression in classical bacterial systems. The *prestin* gene is cloned into an expression vector (pET151/D-TOPO, Invitrogen) that allows the production of a recombinant fusion protein with a His-tag at the N-terminus. Different *E. coli* strains and different expression conditions are tested, in particular *E. coli* strains C41(DE3) and C43(DE3) (Miroux and Walker, 1996) that were reported successful in the expression of other membrane proteins (Dumon-Seignovert et *al.*, 2004).

As alternative strategy, the protein is produced with cell-free systems based on *E. coli* extracts, an emerging and really promising technique for the large-scale production of membrane proteins. As for the STAS domain, different SulP proteins are selected for cell-free expression: prestin (SLC26A5) from rat, zebrafish and chicken; the transporter Rv1739c from *Mycobacterium tuberculosis*; the plant transporter SULTR1.2 from *Arabidopsis thaliana*.

3
Results and discussion

3.1 Overview

Three variants of prestin C-terminal domain were cloned and produced: one from mongolian gerbil and two from rat. Two constructs were selected either for human pendrin or for the bacterial protein, Rv1739c. (Figure 17). The Table 2 summarizes the cloning and expression results for each of the selected constructs (for a total of 7 constructs).

Table 2: Survey of the STAS domains cloned and expressed in *E. coli*.

STAS domain	Species	Sequence	AA	MW (kDa)	Expression vector	Tag protein	Expression/ Solubility
$PreCD_T$	*M. unguiculatus*	[505-744]	240	26.6	pET-19	$(His)_{10}$	Low/ Good
$PreCD_L$	*R. norvegicus*	[529-744]	216	23.8	pGEX-4T pET-28	GST $(His)_6$	No detectable Good/ High
$PreCD_S$	*R. norvegicus*	[529-720]	192	21.2	pGEX-4T pET-28	GST $(His)_6$	No detectable Good/ Good
$PenCD_L$	*H. sapiens*	[525-780]	256	29.1	pET-151	$(His)_6$	Medium/ Low
$PenCD_S$	*H. sapiens*	[539-780]	242	27.5	pET-151	$(His)_6$	Medium/ Low
$RvCD_L$	*M. tuberculosis*	[422-560]	139	15.7	pET-SUMO	$(His)_6$-SUMO	High/ Good
$RvCD_S$	*M. tuberculosis*	[439-560]	122	14.0	pET-SUMO	$(His)_6$-SUMO	High/ Good

$PreCD_T$: Prestin C-terminal Domain total; $PreCD_L$: Prestin C-terminal Domain long; $PreCD_S$: Prestin C-terminal Domain short; $PenCD_L$: Pendrin C-terminal Domain long; $PenCD_S$: Pendrin C-terminal Domain short; $RvCD_L$: Rv1739c C-terminal Domain long; $RvCD_S$: Rv1739c C-terminal Domain short.

For both the constructs of pendrin, the proteins were produced, starting from the vector pET151/D-TOPO (Invitrogen), with a cleavable His-tag at the N-terminus. Different *E. coli* strains (BL21(DE3), C41/C43(DE3)) were tested for the expression, but the soluble expression yield was always too low for structural studies.

Two constructs of prestin ($PreCD_L$ and $PreCD_S$) and two of the bacterial homologue ($RvCD_L$ and $RvCD_S$) were cloned in different pET vectors and expressed in *E. coli* in a largely soluble form. Interestingly, for $PreCD_L$ and $PreCD_S$ no expression was detected, using the GST (Glutathione S-Transferase) tag protein, introduced by the pGEX plasmids

series. Differently, using the His-tag, given by the plasmid pET-28, both proteins were successfully overexpressed in *E. coli*. $RvCD_L$ and $RvCD_S$ were expressed with the His-SUMO tag to increase the soluble protein expression yield.

The recombinant proteins were purified in milligrams quantities and characterized in solution by classical biophysical methods (gel filtration chromatography, fluorescence and circular dichroism spectroscopy, dynamic light scattering). Although it has not been possible to obtain crystals suitable for X-ray diffraction studies yet, important information on the aggregation propensities, relevant for both crystallography and NMR, were obtained. The results about prestin STAS domain aggregation properties are described in detail in section 3.2. The purification and characterization of the Rv1739c STAS domain are reported in section 3.3.

3.2 Expression, purification and characterization of prestin STAS domain

This section has been adapted from:

Expression, purification and characterization of the C-terminal STAS domain of the SLC26 anion transporter prestin

Elisa Pasqualetto, Anke Seydel, Alberto Pellini, Roberto Battistutta

Protein Expression and Purification 2008, **58**: 249-256.

Experimental procedures

Design of three variants of the C-terminal domain of prestin

Three variants of the C-terminal domain of prestin were designed according to literature functional data on mutations and deletions (Zheng et *al*., 2005) and from predictions based on the primary sequence. The three selected variants, listed in Table 3, are: $PreCD_T$ [505-744] (sequence from mongolian gerbil, *Meriones unguiculatus*); $PreCD_L$ [529-744] and $PreCD_S$ [529-720] (sequences from *Rattus norvegicus*).

All the selected sequences present two distinctive charged regions: a positive-charge cluster is located at residues 557-580; adjacent to this, there is a negative-charge cluster at residues 596-613. With respect to the other two constructs, $PreCD_S$ is devoid of the last 24 ammino acids, 5 of which are prolines.

Table 3: Selected variants of prestin STAS domain.

Construct	Species	Sequence	AA	MW (kDa)
$PreCD_T$	*M. unguiculatus*	[505-744]	240	26.6
$PreCD_L$	*R. norvegicus*	[529-744]	216	23.8
$PreCD_S$	*R. norvegicus*	[529-720]	192	21.2

Cloning of prestin genes into the expression vectors

For $PreCD_L$ and $PreCD_S$, the corresponding DNA sequences were already inserted in pBluescript II SK (+/-) vector and were inserted between the *Nde*I–*Xho*I site of plasmid pET-28b(+) (Novagen). The nucleotide sequence of $PreCD_T$ was amplified by polymerase chain reaction (Vent DNA polymerase, New England Biolabs) from the cDNA of gerbil prestin, using the forward and reverse primers shown in Table 4. They were planned to introduce *Xho*I and *Hind*III restriction sites, respectively, at the 5' and 3' ends. The PCR product was subsequently inserted into the pBluescript II SK (+/-) storage phagemid, previously digested to create blunt ends by the *Eco*RV restriction enzyme (New England Biolabs). Then the *preCD$_T$* sequence was extracted by a *Xho*I/*Hind*III digestion and inserted into the corresponding cloning sites of the expression vector pET-19b (Novagen). All the resulting expression vectors were transformed into the TOP10 *E. coli* strain for the amplification and isolated with commercial kits (Qiagen). The correctness of the sequences was verified by standard sequencing methods

Table 4: Oligonucleotide primers used in this study.

Construct	Primer	Sequence[a]
$PreCD_T$	5′	tcgctcgagagtccgagctacaag
	3′	cgtaagcttatgcctcgggtgt

[a] *Xho*I and *Hind*III restriction sites are underlined.

Proteins expression

For the heterologous expression, the vectors were transformed into the BL21(DE3) *E. coli* strain. All the expression vectors produced a recombinant protein with a cleavable His-tag at its N-terminus. For protein expression, single colonies were grown, for about 16 h, in LB medium (10 g/l tryptone, 5 g/l yeast extract, and 10 g/l NaCl) containing 50 μg/ml kanamicin; this is called overnight culture (ONC). LB medium was inoculated with the ONC (ratio 1:25) and cultivated at 37 °C in a suitable shaker. Protein expression was induced when OD_{600} of the culture reached 0.4, by adding IPTG (IsoPropil-β-D-ThioGalactopyranoside) into a final concentration of 1 mM. After induction, the bacteria were grown at 28 °C for 6 h and then harvested by centrifugation: whole cell lysates were analyzed by SDS-PAGE.

Purification and proteolytic cleavage of fusion proteins

The harvested cells were resuspended in buffer A (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, pH 8.0), and lysed with a French Press at high pressure. The lysate was centrifuged to remove cell debris and applied to a HIS-Select Cartridge column (Sigma-Aldrich). After extensive washing with buffer A containing 20 mM imidazole, the protein was isolated from its $(His)_6$-tag by adding thrombin (Sigma-Aldrich) directly into the column (ON at 4 °C). The eluate was further purified by gel filtration chromatography, using a HiLoad 26/60 preparation grade Superdex 200 column (GE Healthcare), equilibrated with 20 mM Tris-HCl, 300 mM NaCl, 5 mM DTT, pH 7.5. The gel permeation chromatography was performed with a Äkta FPLC chromatographic system (GE Healthcare).

Analytical reverse phase chromatography and mass spectrometry

Analytical reverse phase chromatography was performed using a C4 column (Vydac, 0.46 x 15 cm), and the following solvents: A (5% CH_3CN, 0.1% TFA) and B (90% CH_3CN, 0.08% TFA). The elution of the protein was obtained through a concentration gradient from 25% to 65% of B in 30 min. The reverse phase chromatography was performed with a Äkta purifier chromatographic system (GE Healthcare). Mass spectrometry was performed on an ESI-TOF mass spectrometer.

Circular dichroism (CD) spectroscopy

CD data were collected on a Jasco J-715 spectropolarimeter, using quartz cuvettes of 0.02 cm path-length. Spectra were determined as an average of 10 scans. The protein concentration was 0.6 mg/ml in 20 mM Tris-HCl, 300 mM NaCl, 5 mM DTT, pH 7.5.

Fluorescence spectroscopy

Fluorescence emission measurements (excitation wavelength set at 277 nm) were performed on a Perkin-Elmer model LS-50B spectrofluorimeter, with a 1 cm path-length quartz cuvette. The protein concentration was 30 μg/ml for $PreCD_S$, 20 μg/ml for $PreCD_L$, in 50 mM Na_2HPO_4, 150 mM NaCl, pH 7.5 for native conditions, in 6 M guanidine chloride pH 5.5 for denaturating conditions.

Analytical gel permeation chromatography

Analytical gel permeation was performed using a BioSep-SEC-S2000 column (Phenomenex, 0.46 × 30 cm) equilibrated with 20 mM Tris-HCl pH 7.5, 5 mM DTT and different sodium chloride concentrations (50 e 150 mM).

Dynamic light scattering (DLS)

DLS was performed using a Nicomp model 370 Submicron Particle Sizer instrument. A preliminary set of measurements performed on common protein standards (chymotrypsinogen A, 25 kDa; ovalbumin, 43 kDa; albumin, 67 kDa; aldolase, 158 kDa, catalase, 232 kDa; ferritin, 440 kDa; tyroglobulin, 669 kDa) gave an average coefficient of variation (CV) of 11% on the experimental values of the molecular weight. $PreCD_L$ protein solution was 4.8 mg/ml in 20 mM Tris-HCl at pH 8.5 and 5 mM DTT, with the addition of different additives listed in Table 5. Solutions were filtered with centrifugal filters with a membrane pore of 0.22 μm, and measured on a quartz cuvette at 25 °C. Data presented in the table are the average of at least three independent measurements. The diameter values and the corresponding aggregation states refer to at least the 99% of the particles population present in solution, as determined by a number-weighted distribution analysis.

Crystallization tests

$PreCD_S$ and $PreCD_L$ were concentrated to 6 and 10 mg/ml, respectively, by ultrafiltration. Sparse matrix crystallization tests were performed using the vapour diffusion technique, both with hanging and sitting drop methods. Several standard screens were tested (Hampton Research and Molecular Dimension Ltd.) either manually or with automatic systems. Despite single crystal of reasonable size have not been obtained yet, promising micro-precipitates were observed with reagents such as PEG, especially when pH was around the neutral value.

Results and discussion

The C-terminal cytosolic domain of prestin is usually roughly defined as the last 240 amino acids of the sequence, but the N-terminal boundary is not precisely delineated at least from the structural point of view. In order to identify a sequence corresponding to a

compact single domain, multiple sequence alignments and secondary structure predictions were performed on the C-terminal part of prestin sequence. All the three selected C-terminal variants ($PreCD_T$, $PreCD_L$ and $PreCD_S$, Table 3) comprise the STAS motif common to all SulP anion transporters.

Proteins expression and purification

The three C-terminal variants of prestin were cloned in appropriate vectors for the expression in bacterial systems, with a cleavable N-terminal poly(His)-tag to simplify the purification. The level of expression of the longer form of the cytosolic C-terminal domain, $PreCD_T$, was very low, inadequate for a structural characterization. On the contrary, $PreCD_L$ and $PreCD_S$ were successfully overexpressed mainly in soluble forms (Figure 18). In the optimized protocol the expression was carried out at 28 °C, to enhance the amount of the soluble fraction and the total yield of the crude products was estimated around 15 and 20 mg for 1 L of culture media for $PreCD_S$ and $PreCD_L$, respectively. From the SDS-PAGE runs, the apparent molecular weight of both the variants is higher than expected (24 and 26 kDa for $PreCD_S$ and $PreCD_L$ with the N-terminal $(His)_6$-tag, respectively).

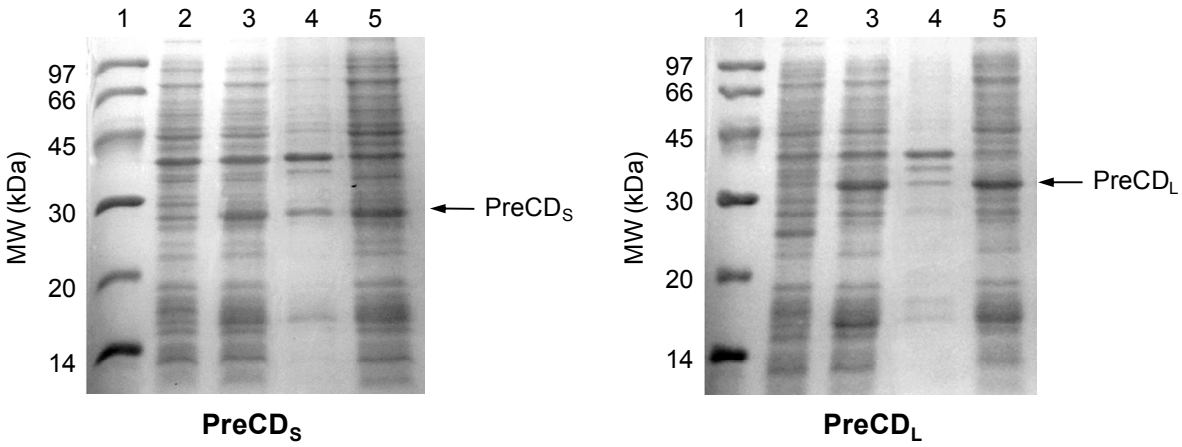

Figure 18: Coomassie-stained SDS-PAGE of protein expression in BL21(DE3) of $PreCD_S$ (left panel) and $PreCD_L$ (right panel). Lane 1: low molecular weight (MW) protein markers. Lane 2: control, not induced bacterial cells. Lane 3: IPTG induced cells. Lane 4: insoluble fraction of bacterial lysate. Lane 5: soluble portion of bacterial lysate.

The soluble fractions of $PreCD_S$ and $PreCD_L$ were purified with a similar final protocol, by a first IMAC affinity step, with the on-column proteolytic cleavage of the $(His)_6$-tag (Figure 19a), followed by a gel permeation step to remove the aggregate fractions and transfer the proteins in the most appropriate buffer for their stabilization and

the following characterization (Figure 19b). The elution profiles of the final gel permeation polishing are shown in Figure 20. Peaks at around 120 ml elution volume (that corresponds to the exclusion volume of the column, MW ≥ 600 kDa) are due to high molecular weight aggregates of the proteins. Fractions corresponding to the peak around 230 ml for $PreCD_S$ and around 220 ml elution volume for $PreCD_L$ were collected and concentrated for the following characterization. The final yield of purified proteins was 3 and 5 mg for 1 L of culture media for $PreCD_S$ and $PreCD_L$, respectively.

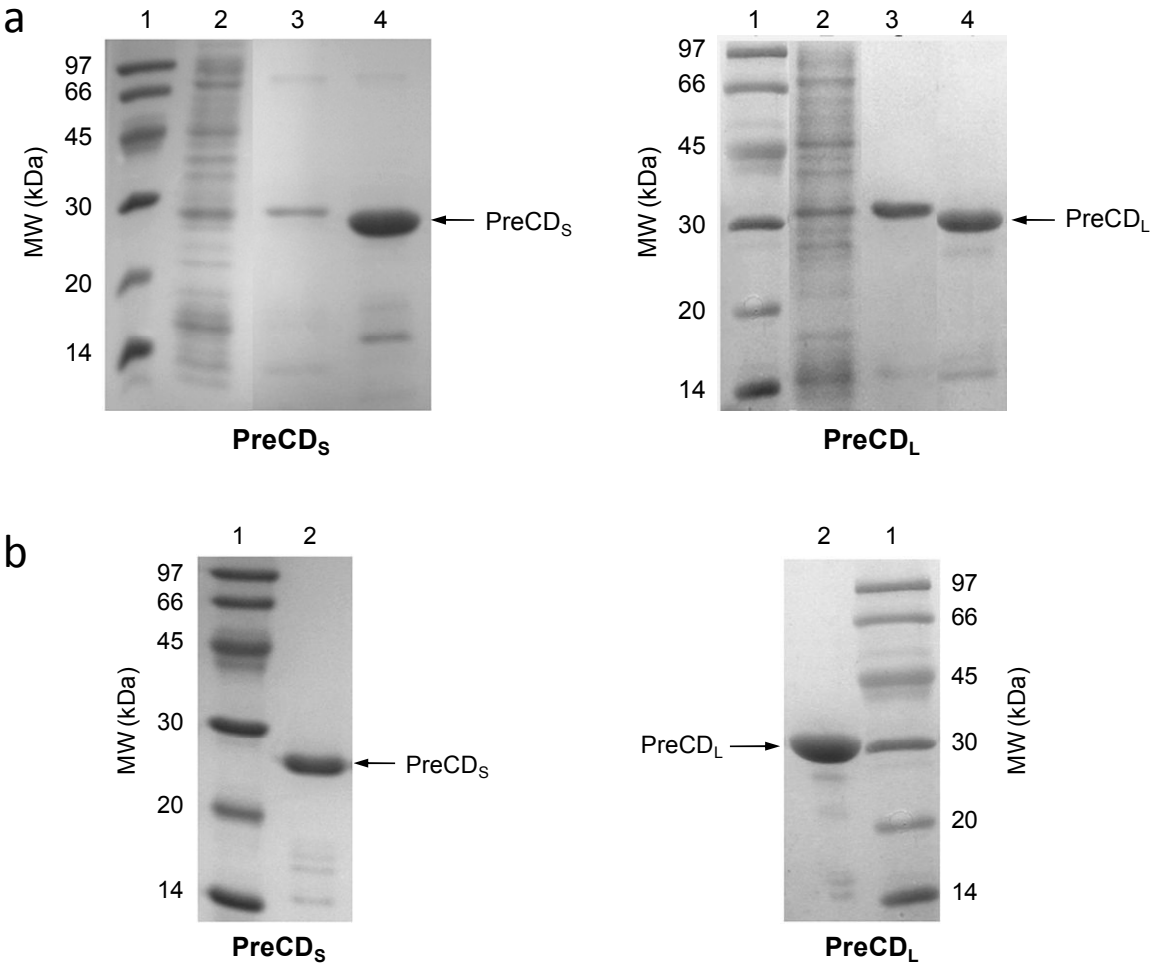

Figure 19: (**a**) Coomassie-stained SDS-PAGE of the purification, by affinity chromatography, of $PreCD_S$ (left panel) and $PreCD_L$ (right panel). Lane 1: low molecular weight (MW) protein markers. Lane 2: sample loaded on the IMAC column. Lane 3: protein after washing and before the proteolytic cleavage. Lane 4: protein after the proteolytic cleavage of the His-tag. (**b**) Coomassie-stained SDS-PAGE after the gel permeation chromatography of $PreCD_S$ (left panel) and $PreCD_L$ (right panel). Lane 1: low molecular weight (MW) protein markers. Lane 2: purified proteins.

Besides by SDS-PAGE, the quality of the final products was checked by reverse phase chromatography that revealed a purity >95% for both samples. ESI-TOF mass analysis confirmed the correct values for the molecular weights of the final cleaved

samples, 21806 Da for $PreCD_S$ (theoretical 21805 Da) and 24418 Da for $PreCD_L$ (theoretical 24413 Da).

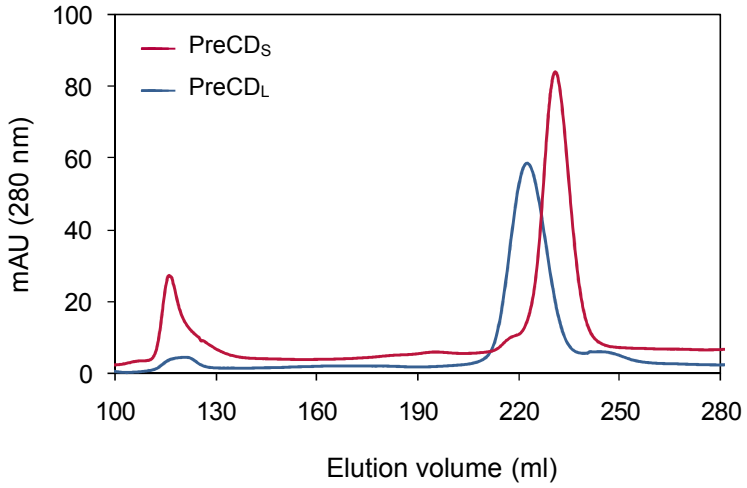

Figure 20: Gel permeation elution profiles of $PreCD_S$ and $PreCD_L$ (red and blue line, respectively) using a HiLoad 26/60 Superdex 200 column equilibrated with 20 mM Tris-HCl, 300 mM NaCl, 5 mM DTT, pH 7.5.

Circular dichroism (CD) and fluorescence spectroscopy

To verify whether the two soluble recombinant variants of prestin C-terminal domain have a defined structure, the purified samples were characterized by circular dichroism and fluorescence spectroscopy. CD measurements (Figure 21a) showed that $PreCD_S$ and $PreCD_L$ have secondary structure (for a total of around 50% based on deconvolution analysis), in an amount compatible with that of typical globular domains. The CD spectra show two negative bands around 207 and 222 nm, indicative of the presence of α-helices. The overall shape of the spectra and the difference in the intensity of the two negative bands indicate the presence also of a certain amount of β-structure. The secondary structure experimentally determined is in accordance with that derived from the 3D structure of bacterial ASA proteins STAS domains. The blue shift of the 207 nm band for $PreCD_L$ is indicative of a somewhat lower relative amount of secondary structure, probably due to the extreme C-terminal tail that is predicted disordered for the presence of several proline residues.

Fluorescence measurements were performed in order to verify the presence of tertiary structure. The fluorescence signal of both $PreCD_S$ and $PreCD_L$ in the native buffer shows a maximum around 305 nm, typical of tyrosine emission. In Figure 21b the $PreCD_L$ spectra are reported. The increase in the fluorescence emission in the presence of the denaturing agent 6 M guanidine chloride indicates that the proteins in the native buffer have tertiary

structure. Taken together, CD and fluorescence data indicate that the recombinant $PreCD_S$ and $PreCD_L$ have a defined 3D structure.

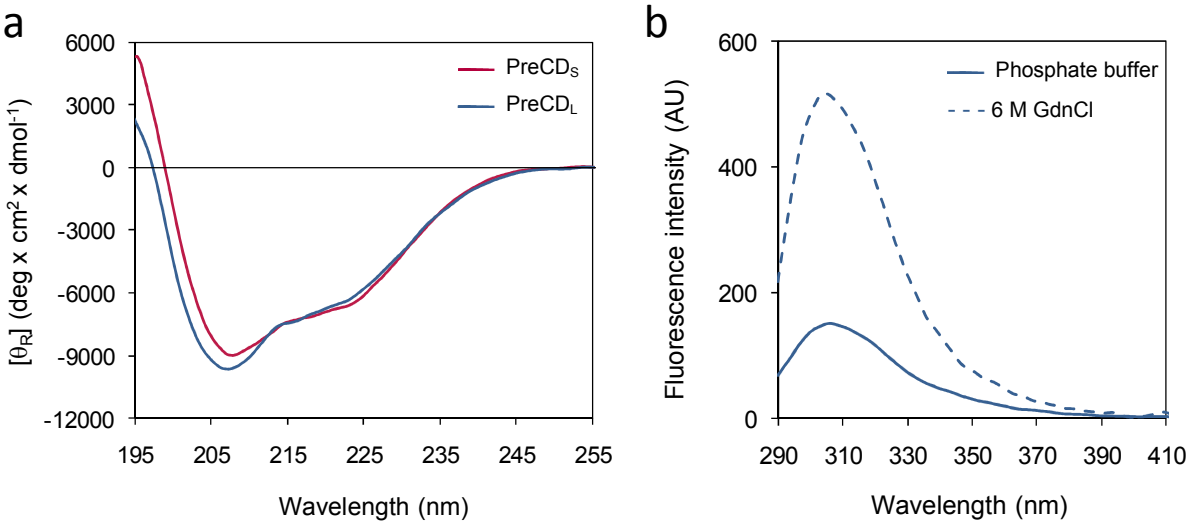

Figure 21: Spectroscopic characterization of the recombinant C-terminal domain of prestin. (**a**) Far-UV circular dichroism spectra of $PreCD_S$ and $PreCD_L$ (red and blue line, respectively). (**b**) Fluorescence spectra of $PreCD_L$ in phosphate buffer and in 6 M guanidinium chloride (continuous and dashed line, respectively).

Oligomerization properties

From gel permeation analysis it was noticed that both cobstructs have a tendency to aggregate, giving rise to oligomers and multimers of high molecular weight (Figure 20, 22). In the elution profile of Figure 22, peaks at around 2.2 ml elution volume (that corresponds to the exclusion volume of the column, MW ≥ 300 kDa) are due to high molecular weight aggregates of $PreCD_L$. These aggregates are sensible to the ionic strengths of the buffer system, as demonstrated by elutions of $PreCD_L$ at different NaCl concentration (Figure 22).

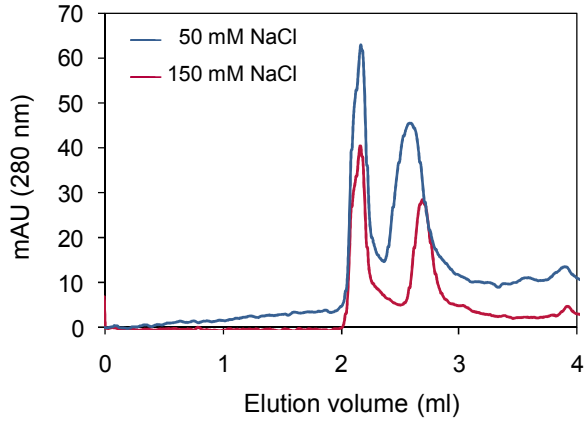

Figure 22: Gel permeation elution profile of $PreCD_L$, using a BioSep-SEC-S2000 (Phenomenex) column equilibrated with 20 mM Tris-HCl pH 7.5, 5 mM DTT and different NaCl concentrations (50 and 150 mM, blu and red line, respectively).

It was reported that full-length prestin aggregates in living cells at the level of the plasma membrane, forming stable dimers or tetramers that are supposed to be essential for the physiological function of the protein (Zheng et *al*., 2006; Detro-Dassen et *al.*, 2008; Mio et *al*., 2008). In this study the aggregation propensity in solution of $PreCD_S$ and $PreCD_L$ was investigated by dynamic light scattering. Preliminary experiments demonstrated that both variants have indeed a great tendency to aggregate, giving rise to oligomers and multimers of high molecular weight that are sensitive to the buffer composition. The $PreCD_L$ variant, which comprises the entire C-terminal portion of prestin, was fully characterize by DLS and the results are reported in Table 5. DLS data show a remarkable difference in the behaviour of $PreCD_L$ and monomeric protein standards in solution (see experimental procedures). Unlike protein standards, $PreCD_L$ was never detected in a single monodisperse state, but it exists in solution in equilibrium between different oligomeric states. The equilibrium can be shifted depending on the presence of a number of chemically different species, such as detergents (octylglucoside and dodecylmaltoside), organic solvents (glycerol), salicylate (which is known to block to some extent the somatic electromotility of prestin (Tunstall et *al*., 1995; Rybalchenko and Santos-Sacchi, 2003; Köppl et *al*., 2004) and sodium salts (NaCl, NaF, NaBr, Na_2SO_4) (Table 5). In the presence of only a buffer system (i.e. in the absence of salts) $PreCD_L$ forms high molecular weight multimers, while in conditions similar to the physiological ones (pH 7.3 and 10 mM NaCl) the aggregation state is 2.6, indicating the presence of an equilibrium between dimers and trimers. The oligomeric state is highly variable; for instance, the two used detergents shift the equilibrium towards different oligomeric forms: octylglucoside (OG) towards the dimer and dodecylmaltoside (DM) towards the tetramer. This is probably due to the fact that at the concentration used (0.1% w/v) DM is present in a micellar form (the CMC of DM is 0.009%) while OG is not (the CMC of OG is 0.7%). In the case of salts, in 100 mM NaCl or NaBr the equilibrium is shifted towards the tetramer while in NaF or Na_2SO_4 towards the dimer. The different behavior of $PreCD_L$ in the presence of different sodium halides at the same concentration indicates that the dissociation effect is primarily due to the anions rather than to the sodium. This is in accordance with the dependence of prestin motor function on the type and concentration of intracellular anions (Oliver et *al*., 2001; Santos-Sacchi et *al*., 2006). The fact that the oligomeric state is influenced by ionic species as well as by apolar substances suggests that the association is mediated by both polar and apolar interactions.

To check whether the aggregation state of $PreCD_L$ is also sensitive to the anion concentration, DLS measurements were performed at different NaCl concentrations, at two different pH values, at the physiological pH of 7.3 and at a pH where $PreCD_L$ is usually slightly less aggregate, pH 8.5 (Table 5). These data indicate that $PreCD_L$ exists in different aggregation states depending on Cl^- concentration. The protein concentration in the DLS measurements is 0.20 mM, a concentration similar to that hypothesised for full-length prestin at the level of the plasma membrane (Köppl et *al*., 2004; He et *al*., 2006). The effect of the chloride, at least at lower concentrations, is possibly due to a specific interaction with the protein. The effect of the variation in the ionic strength seems negligible up to 100 mM salt concentration, considering also the influence of the amount of protein present ($PreCD_L$ is a polyelectrolyte). The increase in the chloride concentration from 10 to 100 mM NaCl shifts the equilibrium from a situation where dimers prevail to one where tetramers are most abundant, at both tested pH.

Table 5: DLS data of $PreCD_L$ in the presence of difference additives.

Additive	Concentration	pH	$PreCD_L$ diameter (nm)	$PreCD_L$ MW (kDa)	Aggregation state
NaCl	10 mM	8.5	5.2	33	1.4
NaCl	10 mM	7.3	7.1	64	2.6
NaCl	100 mM	8.5	8.2	92	3.8
NaCl	100 mM	7.3	7.9	83	3.4
NaCl	300 mM	8.5	9.9	143	5.8
NaF	100 mM	8.5	6.5	58	2.4
NaBr	100 mM	8.5	7.8	84	3.4
Na_2SO_4	100 mM	8.5	6.3	51	2.1
Octylglucoside	0.1%	8.5	5.7	40	1.6
Dodecylmaltoside	0.1%	8.5	7.9	83	3.4
Glycerol	10%	8.5	7.2	68	2.8
Salicylate	1 mM	8.5	5.7	42	1.7

The diameter of the particles present in solution and the corresponding molecular weight (MW) are reported. The aggregation state is calculated as the ratio between the experimental MW and the value for the monomer (24.4 kDa). The diameter values and the corresponding aggregation states refer to at least the 99% of the particles population present in solution, as determined by a number-weighted distribution analysis.

Taken together, these data indicate that the C-terminal domain of prestin has an intrinsic tendency to form oligomers whose nature is highly dependent on the chemical composition of the environment. The prevailing aggregation states are the dimer and the tetramer and a monomeric monodisperse state of the protein has never been detected. Interestingly, the aggregation properties of $PreCD_L$ are dependent on the chloride concentration. Even if it cannot be completely ruled out the possibility that the aggregation properties of the isolated STAS domain can be somewhat different when part of the fully functional prestin, it is noteworthy to note that the properties of $PreCD_L$ *in vitro* show many analogies with those of the full-length protein *in vivo*. In fact, prestin forms oligomers in living cells, possibly dimers and tetramers, which are supposed to be essential for the motor function (Zheng et *al*., 2006; Detro-Dassen et *al.*, 2008; Mio et *al.*, 2008). One plausible interpretation of these data is that the intrinsic tendency to aggregate of prestin C-terminal domain plays a key role in the initial events that lead to the formation of prestin oligomers at the level of the cell membrane. In analogy to what was shown for the full-length protein by Oliver and colleagues (Oliver et *al*., 2001), $PreCD_L$ is sensitive to the presence of different anions. The aggregation properties of $PreCD_L$ at 10 and 100 mM NaCl at pH 7.3, where the equilibrium is shifted between the dimeric and tetrameric form, respectively, are of particular interest in that they may have a physiological relevance. In fact, the OHC intracellular chloride concentration is estimated around 8-20 mM (Ohnishi et *al.*, 1992; Song et *al.*, 2005). Furthermore, a stretch- and voltage-activated conductance in the lateral membrane of the OHC, termed GmetL, has been shown to be permeable to chloride anions (Rybalchenko and Santos-Sacchi, 2003; Song et *al.*, 2005). It has been hypothesized that these channels can induce localized intracellular Cl^- oscillations near prestin and these oscillations may drive motor protein transitions. At 10 mM NaCl $PreCD_L$ is mostly in a dimeric state; the tetrameric form of $PreCD_L$ at 100 mM NaCl could correspond to a state present during a local increment in the chloride concentration near prestin. To note that in the perylinfa surrounding OHCs the concentration of chloride is about 140 mM.

These data confirm the fundamental role of the C-terminal domain in the regulation of the motor function of prestin, in particular in the modulation of the aggregation state of the protein.

3.3 Expression, purification and characterization of Rv1739c STAS domain

Experimental procedures

Design of two variants of the C-terminal domain of Rv1739c

Two variants of the C-terminal domain of Rv1739c from *Mycobacterium tuberculosis* were designed both from literature functional data on deletions (Zolotarev et *al*., 2008) and from predictions based on the primary sequence. The two selected variants, listed in Table 6, are: $RvCD_L$ [422-560] and $RvCD_S$ [439-560].

Table 6: Selected variants of Rv1739c STAS domain.

Construct	Species	Sequence	AA	MW (kDa)
$RvCD_L$	*M. tuberculosis*	[422-560]	139	15.7
$RvCD_S$	*M. tuberculosis*	[439-560]	122	14.0

Cloning of Rv1739c genes into the pET SUMO expression vector

For the production of these constructs, the SUMO system was preferred to the His-tag because it ensures better expression and solubility of the fused construct. The tertiary structure of the SUMO protein is specifically recognized by a protease and this guarantees that the final product starts with the desired amino acid sequence and does not carry the artificial extension coming from the thrombin cleavage site, as the previously utilized pET vectors (Li and Hochstrasser, 1999; Mossessova and Lima, 2000). The nucleotide sequences of the two constructs were generated by PCR (Taq polymerase, Applied Biosystems) using the forward and reverse primers shown in Table 7. The amplified fragments were immediately inserted into the expression vector pET SUMO (Invitrogen), by TA cloning. Taq polymerase has a non template dependent activity that adds a single deoxyadenosine (A) to the 3´ ends of PCR products. The linearized pET SUMO vector has

single 3′ deoxythymidine (T) residues which allow PCR inserts to ligate efficiently into the vector. For optimal efficiencies, the ligation reaction has to be performed immediately after PCR reaction. This is because the single 3′ A-overhangs on the PCR products will be degraded over time, reducing ligation efficiency. The resulting vectors were transformed into the Mach1-T1 *E. coli* strain (Invitrogen) for the amplification. The correctness of the sequences was verified by standard sequencing methods.

Table 7: Oligonucleotide primers used in this study.

Construct	Primer	Sequence
$RvCD_L$	5′	catccgcatgacagcgttctc
	3′	ctattagcgccgacggaacg
$RvCD_S$	5′	gacatcgatgactatccgcag
	3′	ctattagcgccgacggaacg

Proteins expression

For the heterologous expression, the pET SUMO expression vectors were transformed into the BL21(DE3) *E. coli* strain (Invitrogen). The expression vectors produced a recombinant protein with a Small Ubiquitin-like MOdifier (SUMO) and a $(His)_6$-tag at its N-terminus ($(His)_6$-SUMO-tag). SUMO is the *Saccharomyces cerevisiae* Smt3 protein, which is an 11 kDa homolog of the mammalian SUMO-1 protein (Saitoh *et al.*, 1997; Muller *et al.*, 2001).

For protein expression, analogously to what was done for $PreCD_L$ and $PreCD_S$, LB medium was inoculated with the ONC (ratio 1:25) and cultivated at 37 °C, 220 rpm. Protein expression was induced when OD_{600} of the culture reached 0.6, by adding IPTG to a final concentration of 1 mM. After induction the bacteria were grown at 30 °C for 6 h. In the end cells were harvested by centrifugation and whole cell lysates were analyzed by SDS-PAGE.

Purification and proteolytic cleavage of fusion proteins

The harvested cells were resuspended in buffer A (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, pH 8.0), and lysed with a French Press at high pressure. The lysate was centrifuged to remove cell debris and applied to the affinity column (His-Trap, GE Healthcare). After extensive washing with buffer A containing 20 mM imidazole, the fusion protein was eluted with a step gradient procedure, with buffer A

containing first 250 and then 500 mM imidazole. The fractions containing the proteins were pooled, concentrated by ultrafiltration and at the same time the protein buffer was replace to the following: 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7.5). These buffer conditions ensure the optimal SUMO protease activity. The sample was incubated with the SUMO protease (Life sensors) overnight at 16 °C for the proteolytic cleavage of the $(His)_6$-SUMO-tag. The resulting hydrolyzed material was applied to the His-Trap column and immediately recovered to separate the purified protein from the $(His)_6$-SUMO-tag and from the uncleaved fusion protein, which will be retained in the column.

The eluate was further purified using a HiLoad 16/60 preparation grade Superdex 75 column (GE Healthcare), equilibrated with 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7.5. The gel permeation chromatography was performed with a Äkta FPLC chromatographic system (GE Healthcare).

Analytical reverse phase chromatography and mass spectrometry

Analytical reverse phase chromatography was performed using a Jupiter C5 column (0.46 x 30 cm, Phenomenex) and the following solvents: A (100% H_2O, 0.08% TFA) and B (90% CH_3CN, 0.08% TFA). The elution of the protein was obtained through a concentration gradient from 45% to 65% of B in 20 min. The reverse phase chromatography was performed with a Äkta purifier chromatographic system (GE Healthcare). Mass spectrometry was performed on an ESI-TOF mass spectrometer.

Analytical gel permeation chromatography

Analytical gel permeation was performed using a Superdex 200 5/150 column (GE Healthcare) equilibrated with 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7.5.

Circular dichroism (CD) spectroscopy

CD data were collected on a Jasco J-715 spectropolarimeter, using quartz cuvettes of 0.02 cm path-length. Spectra were determined as an average of 10 scans. The protein concentration was 1 mg/ml in 50 mM Na_2HPO_4, 150 mM NaCl, pH 7.5.

Crystallization tests

Sparse matrix crystallization trials were performed, using the vapour diffusion technique, with the sitting drop methods. Using the Oryx8 automatic system (Douglas Instrument), several standard screens were tested (Qiagen and Molecular Dimension Ltd.).

Small needle-like crystals were obtained and used to perform seeding. The crystals grown from the seeds were subjected to diffraction analysis at the X-ray diffraction beam-line of ELETTTRA synchroton in Trieste (Italy), but the images had few and very intense spots tipically generated by the diffraction of salts.

Results and discussion

By analogy to what was done for prestin, multiple sequence alignments and secondary structure predictions were performed on the C-terminal part of Rv1739c sequence. Both the sequences ($RvCD_L$ and $RvCD_S$, Table 6) comprise the STAS motif common to all SulP anion transporters.

Proteins expression and purification

The two C-terminal variants of Rv1739c were cloned and expressed in bacterial systems as a $(His)_6$-SUMO fusion protein. Fusion of a heterologous protein to SUMO can lead to increased expression levels as well as enhanced solubility of the recombinant protein. Both the fusion proteins were, in fact, successfully overexpressed mainly in soluble forms (Figure 23). In the optimized protocol the total yield of the crude products was estimated around 30 mg for 1 L of culture media for both constructs.

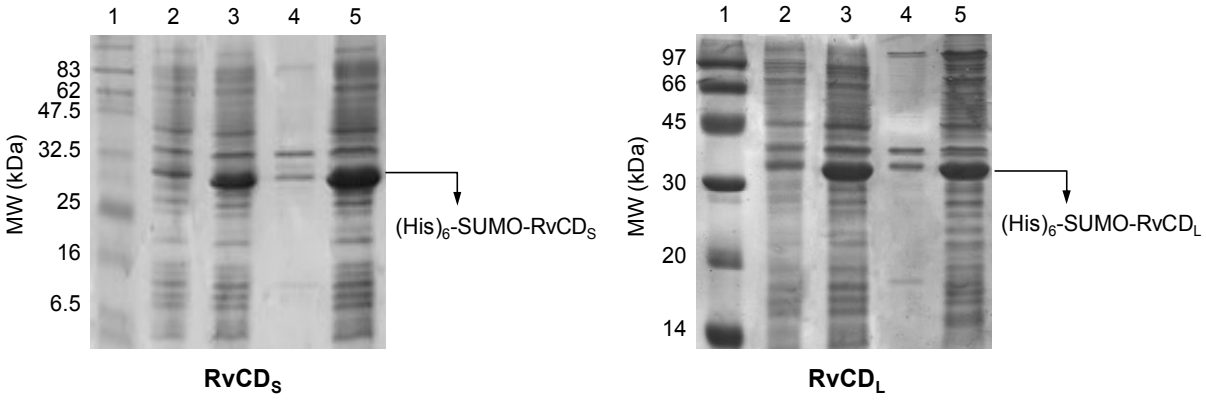

Figure 23: Coomassie-stained SDS-PAGE of protein expression in BL21(DE3) of RvCDS (left panel) and RvCDL (right panel). Lane 1: low molecular weight (MW) protein markers. Lane 2: control, not induced bacterial cells. Lane 3: IPTG induced cells. Lane 4: insoluble fraction of bacterial lysate. Lane 5: soluble portion of bacterial lysate.

The soluble fractions of the $(His)_6$-SUMO fusion proteins were purified by an IMAC affinity step, followed by the proteolytic cleavage of the $(His)_6$-SUMO-tag. Unlikely thrombin enzyme, the SUMO protease recognizes selectively the tertiary structure of the

SUMO protein (Li and Hochstrasser, 1999; Mossessova and Lima, 2000). Therefore, the cleavage results in production of native protein with no extra amino acids added between the cleavage site and the start of the recombinant protein. For $RvCD_S$, the cleavage was successfully: some fusion protein was still present after the proteolytic reaction, but most of $RvCD_S$ was cleaved, still soluble and separated from the $(His)_6$-SUMO-tag (Figure 24). On the contrary, $RvCD_L$ completely precipitated after the proteolitic cleavage of the tag (Figure 25). SUMO fusions may, in fact, increase the expression of recombinant proteins and enhance the solubility of partially insoluble proteins. However, if the protein is completely insoluble it will precipitate, after the proteolytic cleavage of the $(His)_6$-SUMO-tag.

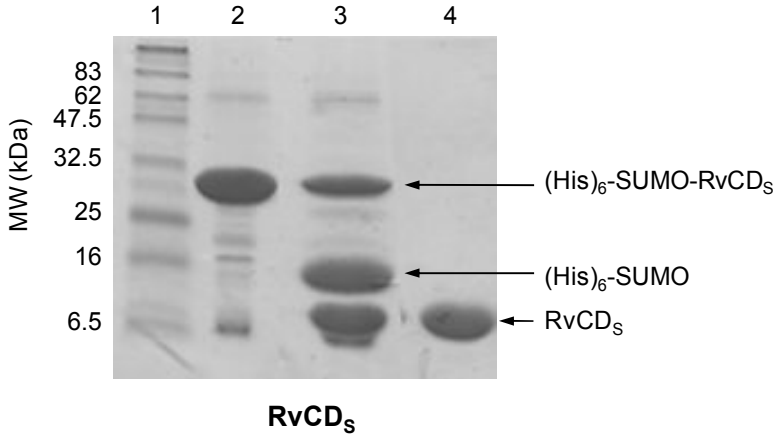

Figure 24: Coomassie-stained SDS-PAGE of the proteolitic cleavage of the $(His)_6$-SUMO tag from $RvCD_S$. Lane 1: low molecular weight (MW) protein markers. Lane 2: fusion protein eluted from the IMAC column. Lane 3: sample after the proteolitic cleavage by the SUMO protease, before the proteins separation. Lane 4: $RvCD_S$ protein after the separation from the $(His)_6$-SUMO-tag and from the uncleaved fusion protein.

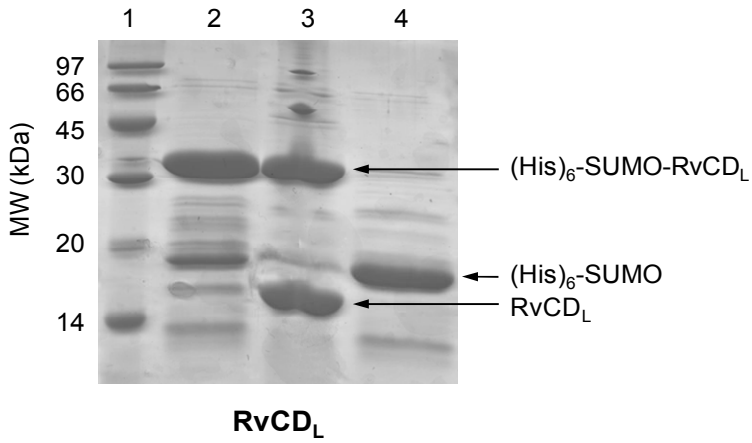

Figure 25: Coomassie-stained SDS-PAGE of the proteolitic cleavage of the $(His)_6$-SUMO tag from $RvCD_L$. Lane 1: low molecular weight (MW) protein markers. Lane 2: fusion protein eluted from the IMAC column. Lane 3 and 4: sample after the proteolitic cleavage by the SUMO protease and separation of the insoluble fraction (Lane 3) from the soluble one (Lane 4). $RvCD_L$ is completely insoluble after the proteolysis of the $(His)_6$-SUMO tag.

$RvCD_S$ was further purified with a gel permeation step (Figure 26b) and the elution profile of the final gel permeation is shown in Figure 26a. Fractions corresponding to the peak around 80 ml were collected and concentrated for the following characterization. The final yield of purified proteins was around 3 mg for 1 L of culture media.

Besides by SDS-PAGE, the quality of the final product was checked by reverse phase chromatography that revealed a purity >95%. ESI-TOF mass analysis confirmed the correct value for the molecular weight of the final cleaved sample, 13976 Da (theoretical 13977 Da).

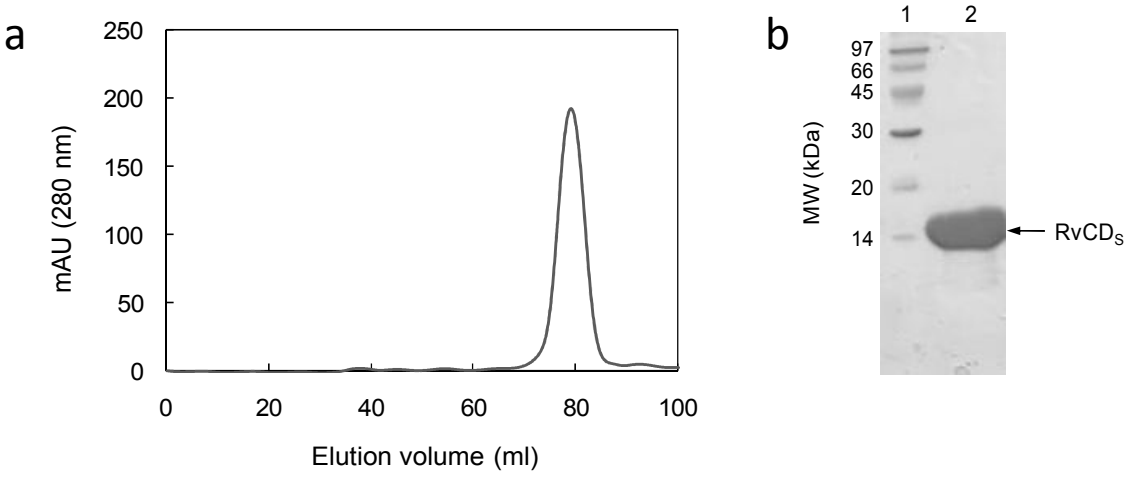

Figure. 26: (**a**) Gel permeation elution profile of $RvCD_S$ using a HiLoad 16/60 Superdex 75 column equilibrated with 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7.5. (**b**) Coomassie-stained SDS-PAGE after the gel permeation chromatography of $RvCD_S$. Lane 1: low molecular weight (MW) protein markers. Lane 2: purified protein.

Analytical gel permeation chromatography

From the analitycal gel permeation (Figure 27), $RvCD_S$ is eluted with a volume corresponding to the monomeric form of the protein in solution.

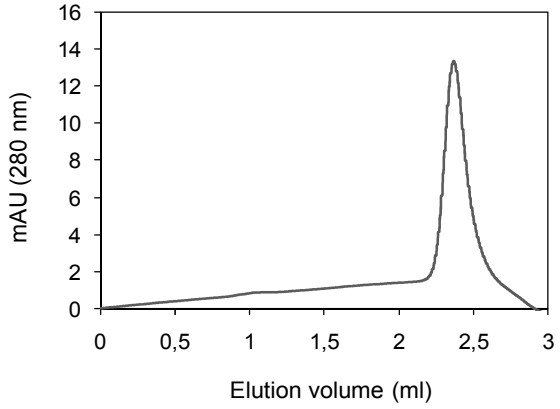

Figure 27: Gel permeation elution profile of $RvCD_S$ using a Superdex 200 5/150 column equilibrated with 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7.5.

Therefore, respect to prestin STAS domain, $RvCD_S$ seems to have a reduced aggregation propensity. This result has to be confirmed by DLS measurements, as it was done for prestin.

Circular dichroism (CD) spectroscopy

To verify whether the recombinant form of Rv1739c C-terminal domain has a defined structure, the purified sample was characterized by circular dichroism spectroscopy in the far-UV region. CD spectrum (Figure 28) showed that $RvCD_S$ has secondary structure, for a total of more than 60% based on deconvolution analysis. The two negative bands around 208 and 222 nm indicate that the secondary structure consists mainly of α-helices. The overall shape of the spectrum also indicates the presence of a small amount of β-structure. From the comparison of the $RvCD_S$ CD spectrum with the spectra of $PreCD_L$ and $PreCD_S$, it is clearly evident that the bacterial STAS domain has a higher percentage of secondary structure in solutions. This is consistent with the fact that the bacterial STAS domain is devoid of the long insertion (unstructured, according to the secondary structure prediction) that in prestin consist of 70 amino acids.

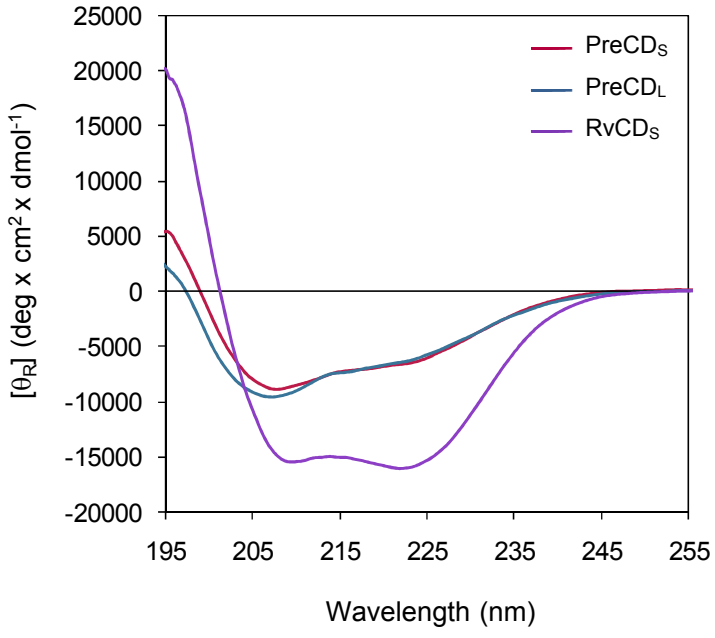

Figure 28: Far-UV circular dichroism spectra of $PreCD_S$, $PreCD_L$ and $RvCD_S$ (red, blue and violet line, respectively).

3.4 Cell-free expression of full-length SulP proteins

Introduction: Cell-free expression of membrane proteins

A variety of intrinsic characteristics make membrane proteins (MPs) to be one of the most problematic targets in expression studies (Junge et *al*., 2008). The hydrophobic nature of MPs often promotes unfolding and aggregation that is frequently followed by proteolysis. Targeting, translocation and stable integration of MPs into native cellular membranes are key problems and often represent the bottlenecks in expression approaches (Wagner et *al*., 2006). Moreover, disintegration and destabilization of host cell membranes as well as toxic effects, due to system overloads or to transport and pore forming activities of the recombinant MPs, could result in growth retardation or even lysis of the host cells. A variety of intrinsic problems therefore exist with the commonly employed heterologous expression host and examples where MPs have been produced in preparative amounts are therefore relatively rare.

Cell-free expression systems have emerged in recent times as promising tools in order to accelerate and to streamline MP expression approaches (Klammt et *al.*, 2006; Schwarz et *al.*, *Methods* 2007; Schwarz et *al*., 2008). The elimination of a living host environment during protein overexpression, in combination with the open accessibility of the reaction, offers a variety of valuable advantages. It is evident that problems with toxic or inhibitory effects of the recombinant MP to the host cell physiology are minimized or even completely eliminated.

Cell-free translation systems are based on the cellular ribosomal protein synthesis system. Generally, the system is composed of a cell extract from *Escherichia coli* or wheat germ. For the preparative scale expression of MPs, mostly CF systems based on *E. coli* extracts have been used so far (Ozawa et *al*., 2004). These extracts are supernatants from a 30000 g centrifugation and contain components such as ribosomes, translation factors and aminoacyl-tRNA synthetases, which are required for production of protein. Efficient protein production requires supplementation of the extract with RNA polymerase, NTPs,

E. coli tRNA mixtures, the 20 amino acids, as well as several enzymes for energy regeneration and their substrates (Schwarz et *al.*, *Nat. Protoc.* 2007).

The highest productivity, i.e. mg amounts of protein in one ml of reaction, can be obtained with the continuous exchange setup of CF expression (CECF) (Spirin et *al.*, 1988; Kigawa and Yokoyama, 1991). The basic reaction principle is that two compartments are separated by a semipermeable membrane that ensures efficient exchange of low molecular compounds. One compartment holds the reaction mixture (RM) with all high molecular weight compounds, like proteins and nucleic acids. The second compartment contains the feeding mixture (FM) with low molecular weight precursors, like nucleotides and amino acids. Intensive agitation during incubation of these CECF reactions ensures the prolonged supply of fresh precursors into the RM concomitant with the continuous removal of inhibitory breakdown products like pyrophosphate. This results in extended reaction times of 12-24 hrs.

A unique characteristic is the open nature of cell-free systems that offers a variety of options to manipulate the reaction conditions in order to protect or to stabilize the synthesized recombinant proteins. In principle, almost any compound that might be beneficial for the stabilization or for the folding of a recombinant protein, like protease inhibitors, cofactors, substrates or any kind of ligands can be added directly into the CF system at any time point of the reaction. Problems with toxicity, instability, or protein folding can therefore be specifically addressed in many cases. In particular, the addiction of detergents into CF reactions offers the completely new approach to synthesize MPs directly into preformed micelles (Berrier at *al.*, 2004; Klammt et *al.*, 2005, 2006; Schwarz et *al*., 2008). This strategy is unique to the CF technique and not possible with any other expression system. It prevents the aggregation of freshly translated MPs and supports their functional folding in the artificial hydrophobic environment of detergent micelles. Alternatively, CF produced MP precipitates formed in the absence of detergent can readily be solubilized in mild detergent without extensive denaturation procedures (Klammt et *al.*, 2004). It is therefore possible to produce functionally folded MPs with both modes of CF expression, while the optimal protocol has to be evaluated for each specific case.

The following work was done in the laboratory of Dr. Frank Bernhard at the Institute of Biophysical Chemistry, Centre of Biomolecular Magnetic Resonance, Johann Wolfgang Goethe University, Frankfurt (Germany).

Experimental procedures

Selection of the SulP proteins for CF expression

The selected SulP proteins for CF expression, listed in Table 8, are: prestin from rat, chicken and zebrafish; SULTR1.2 from *A. thaliana*; Rv1739c from *M. tuberculosis.* For prestin from *R. norvegicus,* three different constructs were selected: one equivalent to the full length protein, and two variants of the transmembrane region, corresponding to the sequences [1-510] ($PreTM_1$) and [71-510] ($PreTM_2$), lacking the N-terminal cytosolic portion. For all the other SulP proteins, only the full length sequences were tested.

Table 8: Selected SulP proteins for CF expression.

Protein	Species	AA	MW (kDa)
Prestin	*Rattus norvegicus*	744	81.3
$PreTM_1$ [71-510]	*Rattus norvegicus*	440	47.2
$PreTM_2$ [1-510]	*Rattus norvegicus*	510	55.4
Prestin	*Gallus gallus*	742	81.1
Prestin	*Danio rerio*	739	81.4
SULTR1.2	*Arabidopsis thaliana*	653	71.7
Rv1739c	*Mycobacterium tuberculosis*	560	59.4

Cloning of the genes into the pET-21cHX expression vector

The coding regions of the selected proteins were amplified by standard polymerase chain reaction (PCR) techniques using Vent DNA polymerase (New England Biolabs) from cDNAs of the related proteins. The purified PCR fragments were cleaved with the restriction enzymes *Xho*I and either *Bam*HI or *Bgl*II and inserted into the corresponding cloning sites of the expression vector pET-21cHx. The plasmid is a derivative of pET-21a(+) (Merck Biosciences) encoding for an extended $(His)_{10}$-tag. Restriction sites were added to the DNA fragments by PCR with suitable oligonucleotide primers (Table 9). The resulting plasmids were isolated with commercial kits (Qiagen) and used as templates for the CF expression of proteins with an N-terminal T7-tag and a C-terminal $(His)_{10}$-tag.

Table 9: Oligonucleotide primers used in this study.

Construct	Primer	Sequence[a]
Prestin (*R. norvegicus*)	5′	aatctcgagtgcttccggtgtggtg
	3′	cttggatccatggatcatgctgaag
PreTM [71-510] (*R. norvegicus*)	5′	cgaggatccttgccagcatataaattca
	3′	aatctcgaggactgtgtagctcgga
PreTM2 [1-510] (*R. norvegicus*)	5′	aatctcgagtgcttccggtgtggtg
	3′	aatctcgaggactgtgtagctcgga
Prestin (*D. rerio*)	5′	caactcgaggtggatgtttgggtggacg
	3′	cgtggatccatggagcacgtaactgttag
Prestin (*G. gallus*)	5′	cgtggatccatggaagatgctcaagaaag
	3′	caactcgaggtggtctaaggcagtctgtg
SULTR1.2 (*A. thaliana*)	5′	caactcgaggacctcgttggagagttttg
	3′	cgtagatctatgtcgtcaagagctcacc
Rv1739c (*M. tuberculosis*)	5′	cgtagatctatgattcccacgatgacatc
	3′	caactcgaggcgccgacggaacgcctg

[a] *Bam*HI, *Bgl*II and *Xho*I restriction sites are underlined.

Western blot analysis

For western blot analysis, the gels were transferred on a 0.45 µm Immobilion-P poly(vinylidene difluoride) membrane (Millipore) in a Hoefer TE22 (GE Healthcare) wet western blot apparatus for 1 h at 400 mA. The membrane was then blocked for 1h in blocking buffer containing PBS (8 mM Na_2HPO_4, 15 mM KH_2PO_4, 0.137 mM NaCl, 3 mM KCl, pH 7.4), 4% (w/v) skim milk powder and 0.05% (w/v) Tween. Horseradish peroxidase-conjugated T7-tag antibody (Merck) was used at a dilution of 1:10000. The anti His-tag antibody (Qiagen) and the anti-mouse antibody (Sigma Aldrich) were used at a dilution of 1:2000 and 1:5000 respectively. All the antibodies were incubated for 1 h with the membrane. After extensive washing with PBS, 0.05% tween, the blots were analyzed by chemiluminescence in a Lumi-imager F1 (Roche Diagnostics).

Preparation of cell-free lysates

Bacterial cell free extracts were prepared from the *E. coli* strain A 19 according to the following protocol (Schwarz et *al*., *Nat. Protoc.* 2007). The cells were grown with good aeration until mid-log phase at 37 °C in rich medium like terrific broth (TB, per litre: 24 g yeast extract, 12 g tryptone, 4 ml 100% glycerol, 100 mM potassium phosphate buffer, pH 7.4), chilled down rapidly and harvested by centrifugation. The time of harvest is

somehow crucial and corresponds in TB medium to an OD_{595} of approximately 3.5. The cell pellet was resuspended and washed three times in ice cold S30-A buffer (10 mM Tris-acetate, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 6 mM β-mercaptoethanol, pH 8.2) and it was finally suspended in S30-B buffer (10 mM Tris-acetate, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 1 mM DTT, 0.1 mM PMSF, pH 8.2) pre-cooled at 4 °C. The cells were disrupted by passing through a pre-cooled French-Press. Cell-debris was removed by centrifugation at 30000 g at 4 °C for 30 min and the upper 2/3rd of the supernatant was transferred into a fresh vial. The centrifugation step and transfer of supernatant was repeated once. For the "run off" step, the lysate was adjusted to a final concentration of 400 mM NaCl and incubated at 42 °C for 45 min in a water bath. Besides the elimination of endogenous mRNA, this treatment causes a considerable precipitate. The turbid solution was filled into a dialysis tube (MWCO 14 kDa) and dialyzed at 4 °C against 60 volumes of S30-C buffer (10 mM Tris-acetate, 14 mM $Mg(OAc)_2$, 0.6 mM KOAc, 0.5 mM DTT, pH 8.2) with gentle stirring. After one further exchange of the dialysis buffer the *E. coli* S30-extract was harvested by centrifugation at 30000 g at 4 °C for 30 min. The clear supernatant was transferred in suitable aliquots and frozen in liquid nitrogen. The final protein concentration in the extract should be between 30-50 mg/ml and could be adjusted by ultrafiltration.

Cell-free expression technique: insoluble expression

The CF expression was performed in the continuous exchange mode. Analytical scale reactions were performed in Plexiglas containers where membranes of regenerated cellulose and a molecular weight cut-off of 12-14 kDa were tightly fixed by a Teflon ring. The RM volume was 70 µl with a RM:FM ratio of 1:14. Preparative scale reactions were carried out in Slide-A-lyser dialysis cassette (Pierce) with a molecular weight cut-off of 10 kDa, in a RM volume of 1-2 ml with a RM:FM ratio of 1:17. The final reaction protocol is given in Table 10. For the optimization screening, the concentration of the ion Mg^{2+} was varied between 13 and 17 mM in order to obtain higher yields. The tRNA mixture and the protease inhibitors concentrations were increased 1.5 and 2 fold the value shown in Table 10, to try to increase the expression yield. The rare codon plasmid was added to a final concentration of 10 µg/ml.

After incubation at 30 °C in a suitable shaker for approximately 20 h, the precipitated proteins were separated from the RM by centrifugation (10 min, 5000 g). The synthesis of

the protein was verified by immunodetection with antibodies directed against the T7-tag and the $(His)_{10}$-tag.

Table 10: Cell-free expression protocol.

Component	Final concentration in Reaction Mix	Final concentration in Feeding Mix
S30 extract	35%	-
S30-C buffer	-	35%
Template DNA	15-20 μg/ml	-
T7-RNA polymerase	6 U/μl	-
RNasin [a]	0.3 U/μl	-
E. coli tRNA-Mix [b]	0.5 mg/ml	-
Pyruvate kinase [b]	40 μg/ml	-
Phosphoenol pyruvate	20 mM	20 mM
Acetyl phosphate	20 mM	20 mM
ATP	1.2 mM	1.2 mM
GTP/CTP/UTP	0.8 mM	0.8 mM
DTT	2 mM	2 mM
Amino acids (A,F,G,H,I,K,L,N,P,Q,R,S,T,V)	0.5 mM	0.55 mM
Amino acids (R,C,W,M,D,E)	1 mM	1 mM
Folinic Acid	0.1 mg/ml	0.1 mg/ml
Complete protease inhibitor [b]	1 tablet/10 ml	1 tablet/10 ml
Hepes-KOH pH 8 EDTA	100 mM	100 mM
KOAc	290 mM	290 mM
$Mg(OAc)_2$	16 mM	16 mM
PEG 8000	2%	2%
NaN_3	0.05%	0.05%

[a] Amersham Biosciences; [b] Roche Diagnostics.

Detergent solubilization of precipitate proteins

After expression in the absence of detergent, the precipitated protein was harvested from the RM by centrifugation. The pellet was suspended in one volume of washing buffer (15 mM Na_2HPO_4, pH 8) and centrifuged for 10 min at 5000 g. The washing step was repeated twice. The pellet was then suspended in one volume of different detergents dissolved in 15 mM Na_2HPO_4, pH 8, at a final concentration of 2%. The mixture was

incubated at 30 °C in a suitable shaker for approximately 90 min to allow the solubilization. Then the suspension was centrifuged at 20000 g, for 20 min, at room temperature to separate the soluble fraction from the precipitate. The protein in both fractions was quantified separately by western blot analysis. The tested detergents were: phosphocholines (DPC, DHPC), long-chain phosphoglycerols (LMPG), alkyl-glucosides (β-OG, DDM), steroids (Chaps, Digitonin), polyoxyethylene-alkyl-ethers (Brij-35, Brij-78) (Table 11). The strong denaturating anionic detergent, sodium dodecyl sulphate (SDS), was included into the analysis as a control for the complete solubilization of the proteins.

Table 11: Detergents tested in this study.

Detergent	Short name	CMC[a] (mM)
n-Dodecyl-β-D-maltoside	DDM	0.12
n-Octyl-β-D-glucopyranoside	b-OG	19
Digitonin	Digitonin	0.73
3-[(3-Cholamidopropyl)dimethylammonio]-1-propansulfonat	Chaps	2.4-8.6
Dodecyl-phosphocholine	DPC	0.9-1.5
1,2-Diheptanoyl-sn-glycero-3-phosphocholine	DHPC	1.4
1-Myristoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)]	LMPG	0.05
Polyoxyethylene-(23)-lauryl-ether ($C_{12/23}$)	Brij-35	0.08
Polyoxyethylene-(20)-cetyl-ether ($C_{16/20}$)	Brij-58	0.075
Polyoxyethylene-(20)-stearyl-ether ($C_{18/20}$)	Brij-78	0.046
Polyoxyethylene-(20)-oleyl-ether ($C_{18\text{-}1/20}$)	Brij-98	0.025

[a] Estimated critical micellar concentrations (CMCs). Averaged values are given if different CMCs for a detergent have been published.

Cell-free expression in the presence of detergents

For soluble expression, long chain Brij derivatives and the steroid detergent, digitonin, were added to the reaction mixtures at the following final concentration: Brij-35 (0.1%), Brij-58 (0.4%), Brij-78 (1%), Brij-98 (1%), digitonin (0.4%, 0.8%). All detergents were supplemented above their specific critical micellar concentrations (CMCs), as specified in Table 11. The production of MPs was analyzed in analytical scale CF reactions after incubation for 20 h, and the protein in the soluble fractions was quantified by western blot analysis.

Results and discussion

Cell-free expression of SulP proteins as precipitate

All the proteins were expressed with an N-terminal T7-tag and with a C-terminal $(His)_{10}$-tag, by using the plasmid pET-21cHx, in order to accelerate their purification and to enable a fast quantification. The synthesis of the protein was verified by immunodetection. An initial CF expression screen in the insoluble form was done for all the selected proteins, following the conditions of Table 10.

With regard to prestin and its truncated derivatives, the CF expression in the absence of detergents showed some expression level (around 200 µg/ml) for the full-length protein (Figure 29), while no detectable expression was seen for the transmembrane region variants ($PreTM_1$ and $PreTM_2$). In contrast, the conventional *in vivo* expression using different *E. coli* strains (BL21(DE3), C41/C43(DE3)) yielded no detectable expression of the full-length prestin by SDS/PAGE and western blot analysis. Concerning all the other SulP proteins, the CF expression screen revealed some detectable expression only for prestin form zebrafish and SULTR1.2 from *A. thaliana.* However, for these proteins only a preliminary screen was performed and the expression conditions have still to be optimized.

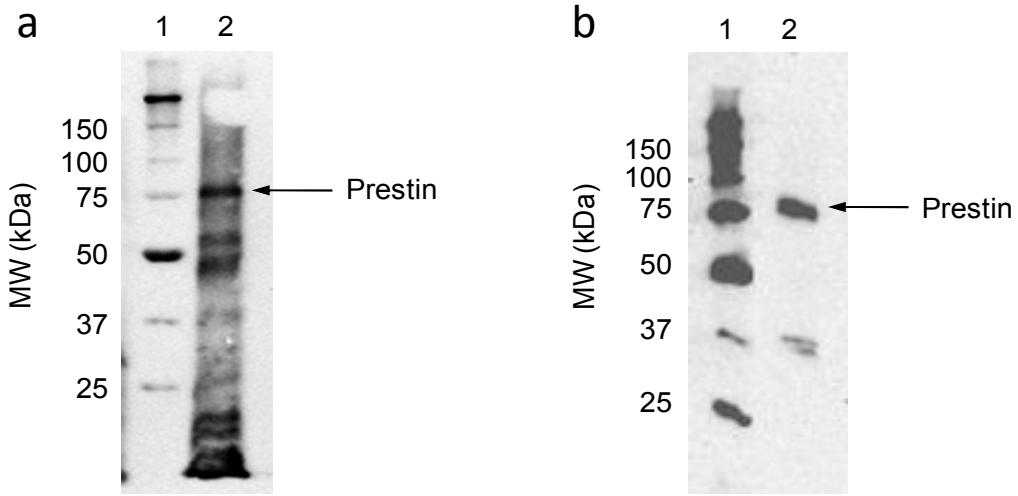

Figure 29: CF production of prestin as precipitate. Western blot analysis with (**a**) the anti T7-tag antibody and (**b**) the anti His-tag antibody. Lane 1: low molecular weight (MW) protein markers. Lane 2: insoluble fraction of the RM.

The anti T7-tag western blot is reported in figure 29a. The band visible around 75 kDa could correspond to the full-length prestin, even if some degradation fragments are

visible around 50 kDa. These fragments could also be the result of premature termination during the CF translation caused, for example, by rare codon usage of the *prestin* gene. The immunodetection with the anti His-tag antibody (Figure 29b) confirms the presence of the full-length prestin close to 75 kDa, but reveals also a C-terminal fragment at approximately 37 kDa. Therefore, the CF expressed prestin has an apparent molecular mass (m) of 75 kDa after SDS/PAGE and western blot analysis (Figure 29), whereas the calculated m should be 81 kDa. Incomplete denaturation of MPs upon SDS/PAGE analysis is often considered to be the reason for variations of the apparent and calculated m values (Klammt et *al*., 2005). Anyway, these data are in accordance with those reported in literature for full-length prestin expressed in yeast and mammalian cell lines (Zheng et *al*., 2006; Mio et *al*., 2008).

With regard to the shorter fragments, the one close to 55 KDa (Figure 29a) could correspond to a truncated prestin derivative ($PreTM_2$), but this disagrees with the lack of CF expression of $PreTM_2$. The shorter fragment, with an apparent molecular mass of 45 kDa (Figure 29a), could be the complementary portion of the 37 kDa C-terminal fragment (Figure 29b). Mass spectrometry measurements will confirm or deny these hypotheses.

Optimization of prestin cell-free expression conditions

For full-length prestin, an extensive screening was performed in order to optimize the reaction conditions and to improve the yield of the protein production.

One of the most critical parameter in the CF expression is the magnesium concentration. Usually, CF reactions have a well defined optimum that correlates with higher protein expression yields. Mg^{2+} and K^+ ions are, in fact, important ions for compensation of negative charges and also for enzyme activities such as pyruvate kinase (Schwarz et *al.*, *Nat. Protoc.* 2007). The Mg^{2+} concentration was screened in the range 13-17 mM, while the K^+ concentration was fixed to 290 mM. The optimal Mg^{2+} concentration, which allowed obtaining the highest yield of protein production, was between 15 and 17 mM (Figure 30). Therefore, all the following tests were performed with 16 mM Mg^{2+} (surrounded by a box in Figure 30).

In order to try to reduce the degradation observed by western blot, the CF reaction conditions were modified varying different parameters, such as the protease inhibitors concentration and the tRNA mixture concentration, and also adding the rare codon plasmid. The addition of broad range protease inhibitors is important for the stabilization of synthesized MPs. Bias in codon usage, especially of larger heterologous genes, might

be considered and could be addressed by increased levels of total tRNA or even by coexpression of rare codon tRNAs (Schwarz et *al.*, *Nat. Protoc.* 2007). However, in this case, no significant improvements in the prestin expression yield or in the degradation were seen. Therefore, the expression yield (around 200 μg/ml) was too low to set up a purification protocol.

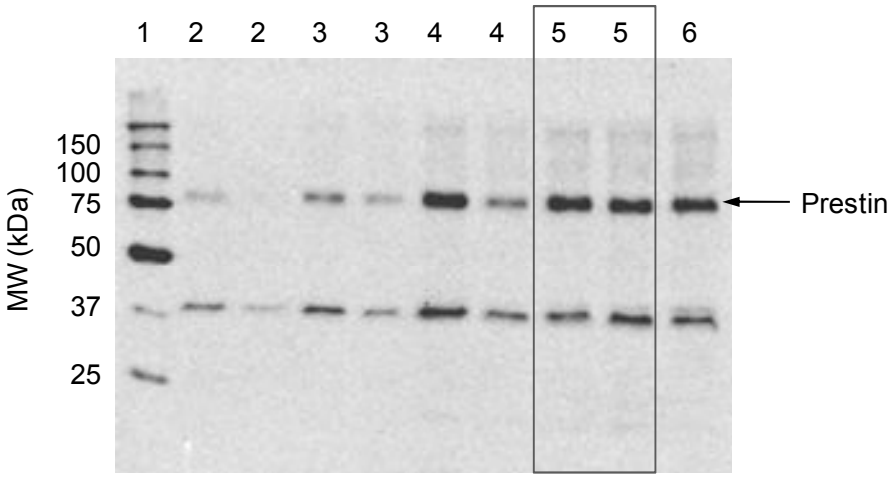

Figure 30: Mg^{2+} concentration screening for CF production of prestin. Western blot analysis, with the anti His-tag antibody, of prestin synthesized in the presence of different Mg^{2+} concentration. Each CF reaction was performed twice. Lane 1: low molecular weight (MW) protein markers. Lanes 2: RM with 13 mM Mg^{2+}; Lanes 3: RM with 14 mM Mg^{2+}; Lanes 4: RM with 15 mM Mg^{2+}; Lanes 5: RM with 16 mM Mg^{2+} (surrounded by a box); Lane 6: RM with 17 mM Mg^{2+}.

Detergent solubilization of precipitate prestin

CF produced precipitates of many MPs are solubilized quickly and effectively in mild detergents during short incubation with gentle shaking for a few minutes (Klammt et *al.*, 2006). This relatively fast solubilization procedure distinguishes CF produced MP precipitates from inclusion bodies that can often be observed after *in vivo* expression of MPs (Kiefer, 2003).

After insoluble expression of prestin, selected detergents were tested for their efficiency in the resolubilization of the CF produced precipitates at a final concentration of 2%. The western blot analysis was used to quantify the protein in the soluble and insoluble fractions, after incubation with detergents. Among all the tested detergents, the quantitative resolubilization of prestin was only possible with the dodecyl-phosphocholine (DPC) and the long-chain phosphoglycerol, LMPG (surrounded by a box in Figure 31). All the other detergents had no, or only minor, resolubilization effects (Figure 31).

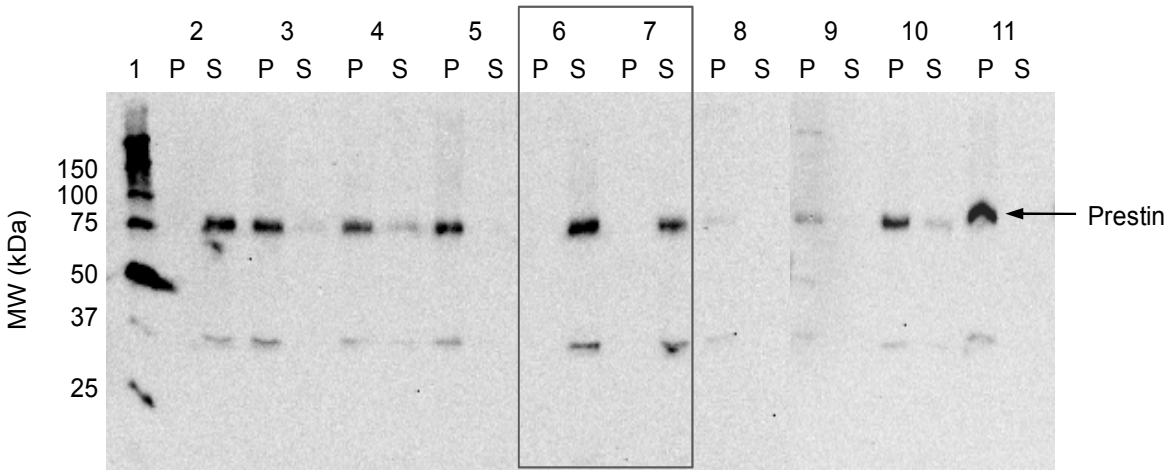

Figure 31: Resolubilization of prestin precipitate with different detergents. Samples of the soluble (pellet, P) and insoluble (supernatant, S) fractions, after incubation with different detergents, were analyzed by western blot with the anti His-tag antibody. Lane 1: low molecular weight (MW) protein markers; Lanes 2: SDS; Lanes 3: Digitonin; Lanes 4: DDM; Lanes 5: DHPC; Lanes 6: DPC; Lanes 7: LMPG; Lanes 8: β-OG; Lanes 9: Chaps; Lanes 10: Brij-35; Lanes 11: Brij-78. The lane 6 (DPC) and 7 (LMPG) are surrounded by a blue box.

Cell-free expression of prestin in the presence of detergents

Brij derivatives and digitonin were used to evaluate prestin expression yield after soluble expression. Long chain Brij derivatives are, in fact, exceptionally tolerated by the CECF system, even at concentrations higher than the CMC, and they exhibit excellent solubilization properties for MPs having diverse topologies and origins. The steroid derivative, digitonin, is another detergent well tolerated by the CECF system, even if the yield obtained is usually lower by comparison with the Brij derivatives (Klammt et *al*., 2005).

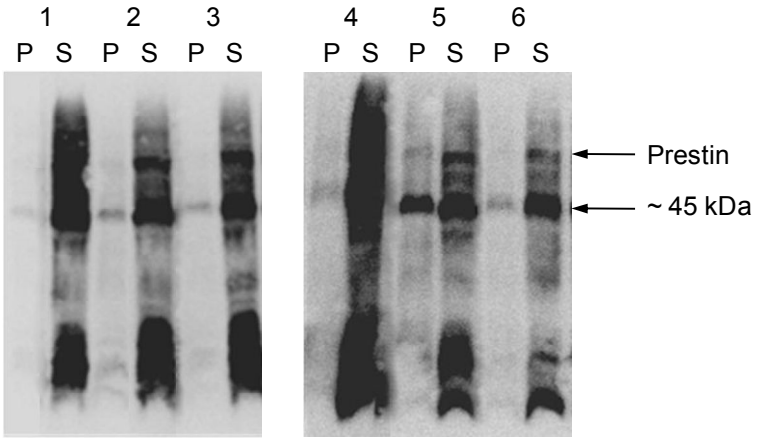

Figure 32: CF expression of prestin in the presence of different detergents. Samples of the soluble (pellet, P) and insoluble (supernatant, S) fractions of the RM were analyzed by western blot with the anti T7-tag antibody. Lane 1: Brij-35; Lanes 2: Brij-58; Lanes 3: Brij-78; Lanes 4: Brij-98; Lanes 5: Digitonin 0.4%; Lanes 6: Digitonin 0.8%.

The soluble expression of prestin increased seriously the degradation. In fact, from the western blot analysis with the anti T7-tag antibody, it is evident that the expression level for the full-length protein is much lower than that of the fragment at around 45 kDa (Figure 32). In addition, the western blot analysis with the anti His-tag antibody revealed no bands corresponding to the antibody recognition. Most likely, the soluble expression caused the loss of the $(His)_{10}$-tag at the C-terminal end of the protein that, being no more "protected" by the precipitated form, becomes more easily subject to degradation

4
Conclusions

With the final aim to get structural information on the SulP family of anion exchangers, different proteins, from distance-related species, were selected, produced and characterized: the two mammal transporters pendrin (SLC26A4) and prestin (SLC26A5) and the bacterial transporter Rv1739c from *Mycobacterium tuberculosis*. Several variants of the STAS domain of each protein were produced in *E. coli,* using the recombinant DNA technology. Among these, two variants of rat prestin and one of the bacterial homologue Rv1739c were expressed in a largely soluble form. For each of these constructs, an optimized two-step purification protocol allowed obtaining a satisfactory yield (3-4 mg of protein per liter of culture) and a high purity level. The proteins were characterized in solution by gel filtration chromatography and classical biophysical methods, such as fluorescence and circular dichroism spectroscopy and dynamic light scattering.

Altogether, results indicate that the C-terminal domain of prestin has an intrinsic tendency to form oligomers, whose nature is highly dependent on the chemical composition of the environment. The aggregation properties at 10 and 100 mM NaCl are of particular interest since they have a physiological relevance. The aggregation properties of prestin STAS domain *in vitro* show many analogies with those of the full-length protein *in vivo,* which forms oligomers in living cells that are supposed to be essential for the motor function. These results confirm the fundamental role of this domain in the regulation of the aggregation state and hence of the motor function of prestin.

The shorter variant of Rv1739c C-terminal domain ($RvCD_S$ [439-560]) was purified in milligrams quantities and characterized in solution by chromatography and circular dichroism spectroscopy. This construct shows a reduced aggregation propensity as pointed out by analytical gel filtration chromatography and recently confirmed by 2D-NMR analysis. Circular dichroism indicates that $RvCD_S$ has an amount of secondary structure higher than that of prestin STAS variants.

The pieces of information we have accumulated about the properties of these constructs are very precious for the structural studies we intend to perform. Some purified constructs have been already submitted to extensive crystallization trials that are currently still under way. In parallel, we have recently started the production of labelled samples for NMR analysis.

The second part of the SulP project aimed at the production of the full-length membrane proteins by cell-free expression system. This part was carried out at the Johann Wolfgang Goethe University of Frankfurt, in collaboration with Dr. Frank Bernhard. The

full-length rat prestin was produced by a cell-free system based on *E. coli* extracts, but serious degradation phenomena hindered to reach the desired yield. Also other SulP members were submitted to cell-free expression trials: zebrafish and chicken prestin, whose properties are interesting in evolutionary terms, the bacterial transporter homologue Rv1739c, and the plant transporter SULTR1.2, whose production yield seems encouraging in preliminary experiments.

References

Albert J.T., Winter H., Schaechinger T.J., Weber T., Wang X., He D.Z.Z., Hendrich O., Geisler H.-S., Zimmermann U., Oelmann K., Knipper M., Gopfert M.C., Oliver D. (2007). Voltage-sensitive prestin orthologue expressed in zebrafish hair cells. *J. Physiol.* **580**: 451-461.

Aravind L., Koonin E.V. (2000). The STAS domain, a link between anion transporters and antisigma-factor antagonist. *Curr. Biol.* **10**: 53-55.

Armstrong, C.M., Bezanilla F. (1977). Inactivation of the sodium channel. II. Gating current experiments. *J. Gen. Physiol*. **70**: 567-590.

Ashmore J.F. (1987). A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier. *J. Physiol.* **388**: 323-347.

Ashmore J.F. (1990). Forward and reverse transduction in the mammalian cochlea. *Neurosci. Res.* **12**: 39-50.

Bai J.P. , Navaratnam D., Samaranayake H., Santos-Sacchi J. (2006). En block C-terminal charge cluster reversals in prestin (SLC26A5): Effects on voltage-dependent electromechanical activity. *Neurosci. Lett.* **404**: 270-275.

Belyantseva I.A., Adler H.J., Curi R., Frolenkov G.I., Kachar B. (2000). Expression and localization of prestin and the sugar transporter GLUT-5 during development of electromotility in cochlear outer hair cells. *J. Neurosci.* **20** RC116: 1-5.

Berrier C., Park K.H., Abes S., Bibonne A., Betton J.M., Ghazi A. (2004). Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent. *Biochemistry* **43**: 12585-12591.

Brownell W.E., Bader C.R., Bertrand D., De Ribaupierre Y. (1985). Evoked mechanical response of isolated cochlear outer hair cells. *Science* **227**: 194-196.

Cheatham M.A., Huynh K.H., Gao J., Zuo J., Dallos P. (2004). Cochlear function in Prestin knockout mice. *J. Physiol.* **560**: 821-830.

Dallos P. (1992). The active coclea. *J. Neurosci.* **2**: 4575-4585.

Dallos P., Evans B.N. (1995). High frequency motility of outer hair cells and the cochlear amplifier. *Science* **267**: 2006-2009.

Dallos P., Evans B.N., Hallworth R. (1991). Nature of the motor element in electrokinetic shape changes of cochlear outer hair cells. *Nature* **350**: 155-157.

Dallos P., Fakler B. (2002). Prestin, a new type of motor protein. *Nat. Rev. Mol. Cell Biol.* **3**: 104-111.

Dallos P., Hallworth R., Evans B.N. (1993). Theory of electrically driven shape changes of cochlear outer hair cells. *J. Neurophysiol*. **70:** 299-323.

Dallos P., Zheng J., Cheatham M.A. (2006). Prestin and the cochlear amplifier. *J. Physiol.* **76**: 37-42.

Dawson P.A., Markovich D. (2005). Pathogenetics of the human SLC26 transporters. *Curr. Med. Chem.* **12**: 385-396.

Deak L., Zheng J., Orem A., Du G.G., Aguinaga S., Matsuda K., Dallos P. (2005). Effects of cyclic nucleotides on the function of prestin. *J. Physiol.* **563**: 483-496.

Detro-Dassen S., Schanzler M., Lauks H., Martin I., zu Berstenhorst S.M., Nothmann D., Torres-Salazar D., Hidalgo P., Schmalzing G., Fahlke C. (2008). Conserved dimeric subunit stoichiometry of SLC26 multifunctional anion exchangers. *J. Biol. Chem*. **283**: 4177-4188.

Diederich B., Wilkinson J.F., Magnin T., Najafi M., Erringston J., Yudkin M.D. (1994). Role of interactions between SpoIIAA and SpoIIAB in regulating cell-specific transcription factor sigma F of *Bacillus subtilis*. *Genes Dev.* **8**: 2653-2663.

Dong X.-x., Ospeck M., Iwasa K.H. (2002). Piezoelectric reciprocal relationship of the membrane motor in the cochlear outer hair cell. . *Biophys. J.* **82:** 1254-1259.

Dorwart M.R., Shcheynikov N., Baker J.M., Forman-Kay J.D., Muallem S., Thomas P.J. (2008). Congenital chloride-losing diarrhea causing mutations in the STAS domain result in misfolding and mistrafficking of SLC26A3. *J. Biol. Chem.* **283**: 8711-8722.

Dumon-Seignovert L., Cariot G., Vuillard L. (2004). The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* **37**: 203-206.

Duncan L., Alper S., Losick R. (1996). SpoIIAA governs the release of the cell-type specific transcription factor sigma F from its anti-sigma factor SpoIIAB. *J. Mol. Biol.* **260**: 147-164.

Evans, B.N., Dallos P. (1993). Stereocilia displacement induced somatic motility of cochlear outer hair cells. *Proc. Natl Acad. Sci. USA* **90**: 8347-8351.

Flock A., Kimura R., Lindquist P.G., Wersäll J. (1962). Morphological basis of directional sensitivity of outer hair cells in the organ of Corti. *J. Acoust. Soc. Am*. **34**: 1351-1355.

Forge A., Davies S., Zajic G (1991). Assessment of ultrastructure in isolated cochlear hair cells using a procedure for rapid freezing before freeze-fracture and deep-etching. *J. Neurocytol*. **20**: 471-484.

Frank G., Hemmert W., Gummer A.W. (1999). Limiting dynamics of high-frequency electromechanical transduction of outer hair cells. *Proc. Natl. Acad. Sci. USA* **96**: 4420-4425.

Frolenkov G. I., Atzori M., Kalinec F., Mammano F., Kachar B. (1998). The membrane based mechanism of cell motility in cochlear outer hair cells. *Mol. Biol. Cell.* **9**: 1961-1968.

Géléoc G.S.G., Holt J.R. (2003). Auditory amplification: outer hair cells pres the issue. *Trends Neurosci.* **26**: 115-117.

Gray M.A. (2004). Bicarbonate secretion: it takes two to tango. *Nat. Cell. Biol.* **6**: 343-350.

Greeson J.N., Organa L.E., Pereirab F.A., Raphael R.M. (2006). Assessment of prestin self-association using fluorescence resonance energy transfer. *Brain Res.* **1091**: 140-150.

He D.Z., Zheng J., Kalinec F., Kakehata S., Santos-Sacchi J. (2006). Tuning in to the amazing outer hair cell: membrane wizardry with a twist and shout. *J. Membr. Biol.* **209**: 119-134.

Holley M.C., Ashmore J.F. (1988). On the mechanism of high-frequency force generator in outer hair cells isolated from guinea pig coclea. *Proc. R. Soc. Lond. B* **232**: 413-429.

Hudspeth A.J., Choe Y., Mehta A.D., Martin P. (2000). Putting ion channels to work: mechanoelectrical transduction, adaptation, and amplification by hair cells. *Proc. Natl. Acad. Sci. USA* **97**:11765-11772.

Hudspeth A.J., Corey D.P. (1977). Sensitivity, polarity and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. *Proc. Natl Acad. Sci. USA* **74**: 2407-2411.

Iwasa K.H. (1994). A membrane motor model for the fast motility of the outer hair cell. *J. Acoust. Soc. Am.* **96:** 2216-2224.

Junge F., Schneider B., Reckel S., Schwarz D., Dötsch V., Bernhard F. (2008). Large-scale production of functional membrane proteins. *Cell. Mol. Life Sci.* **65**: 1729-1755.

Kachar B., Brownell W.E., Altschuler R., Fex J. (1986). Electrokinetic shape changes of cochlear outer hair cells. *Nature* **322**: 365-368.

Kalinec F., Holley M. C., Iwasa K. H., Lim D. J., Kachar B., (1992). A membrane-based force generation mechanism in auditory sensory cells. *Proc. Natl. Acad. Sci USA* **89**: 8671-8675.

Keller P., Simons K. (1997). Post-Golgi biosynthetic trafficking. *J. Cell Sci.* **110**: 3001-3009.

Kertesz M.A. (2001). Bacterial transporters for sulfate and organosulfur compounds. *Res. Microbiol*. **152**: 279-290.

Khurana O.K., Coupland L.A., Shelden M.C., Howitt S.M. (2000). Homologous mutations in two diverse sulphate transporters have similar effects. *FEBS Lett.* **477**: 118-122.

Kiefer H. (2003). *In vitro* folding of alpha-helical membrane proteins. *Biochim. Biophys. Acta* **1610:** 57-62.

Kigawa T., Yokoyama, S. (1991). A continuous cell-free protein synthesis system for coupled transcription-translation. *J. Biochem. (Tokyo)* **110**: 166-168.

Klammt C., Löhr F., Schafer B., Haase W., Dötsch V., Ruterjans H., Glaubitz C., Bernhard F. (2004). High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* **271**: 568-580.

Klammt C., Schwarz D., Fendler K., Haase W., Dötsch V., Bernhard, F. (2005). Evaluation of detergents for the soluble expression of alpha-helical and beta-barreltype integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J.* **272**: 6024-6038.

Klammt C., Schwarz D., Löhr F., Schneider B., Dötsch V., Bernhard F. (2006). Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *FEBS J.* **273**: 4141-4153.

Ko S.B., Zeng W., Dorwart M.R., Luo X., Kim K.H., Millen L., Goto H., Naruse S., Soyombo A., Thomas PJ, Muallem S. (2004). Gating of CFTR by the STAS domain of SLC26 transporters. *Nat. Cell. Biol.* **6**: 343-350.

Köppl C., Forge A., Manley G.A. (2004). Low density of membrane particles in auditory hair cells of lizards and birds suggests an absence of somatic motility. *J. Comp. Neurol.* **479**: 149-155.

Kovacs H., Comfort D., Lord M., Campbell I.D., Yudkin M.D. (1998). Solution structure of SpoIIAA, a phosphorylatable component of the system that regulates transcription factor sigmaF of *Bacillus subtilis. Proc. Natl. Acad. Sci. USA* **95**: 5067-5071.

Kroos L., Zhang B., Ichikawa H., Yu Y.T. (1999). Control of sigma factor activity during *Bacillus subtilis* sporulation. *Mol. Microbiol.* **31**: 1285-1294.

Lacapère J.J., Pebay-Peyroula E., Neumann J.M., Etchebest C. (2007). Determining membrane protein structures: still a challenge! *Trends Biochem. Sci.* **32**: 259-270.

Leves F.P., Tierney M.L., Howitt S.M. (2008). Polar residues in a conserved motif spanning helices 1 and 2 are functionally important in the SulP transporter family. *Int. J. Biochem. Cell. Biol.* **40**: 2596-2605.

Li S.J., Hochstrasser M. (1999). A new protease required for cell-cycle progression in yeast. *Nature* **398**: 246-251.

Liberman M.C., Gao J., He D.Z.Z., Wu X., Jia S., Zuo J. (2002). Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier. *Nature* **419**: 300-304.

Liu X.Z., Ouyang X.M., Xia X.J., Zheng J., Pandya A., Li F., Du L.L., Welch K.O., Petit C., Smith R.J.H., Webb B.T., Yan D., Arnos K.S., Corey D., Dallos P., Nance W.E., Chen Z.Y. (2003). Prestin, a cochlear motor protein, is detective in non-syndromic hearing loss. *Hum. Mol. Genet*. **12**: 1155-1162.

Loughlin P., Shelden M.C., Tierney M.L., Howitt S.M. (2002). Structure and function of a model member of the SulP transporter family. *Cell Biochem. Biophys*. **36**: 183-190.

Ludwig J., Oliver D., Frank G., Klöcker N., Gummer A. W., Fakler B. (2001). Reciprocal electromechanical properties of rat prestin: The motor molecule from rat outer hair cells. *Proc. Natl. Acad. Sci USA* **98**: 4178–4183.

Matsuda K., Zheng J., Du G.G., Klöcker N., Madison L. D., Dallos P. (2004). N-linked glycosylation sites of the motor protein prestin: effects on membrane targeting and electrophysiological function. *J. Neurochem.* **89:** 928-938.

Mio K., Kubo Y, Ogura T., Yamamoto T., Ariska F., Sato C. (2009). The motor protein prestin is a bullet-shaped molecule with inner cavities. *J. Biol. Chem.* **283**: 1137–1145.

Miroux B., Walker J.E. (1996). Over-production of proteins in *Escherichia coli:* mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**: 289-298.

Mossessova E., Lima C.D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory essential for cell growth in yeast. *Mol. Cell Biol.* **20**: 2367-2377.

Mount D.B., Romero M.F. (2004). The SLC26 gene family of multifunctional anion exchangers. *Eur. J. Physiol*. **447**: 710-721.

Muallem D., Ashmore J. (2006). An anion antiporter model of prestin, the outer hair cell motor protein. *Biophys. J.* **90**: 4035-4045.

Muller S., Hoege C., Pyrowolakis G., Jentsch S. (2001). SUMO, ubiquitin's mysterious cousin. *Nature Rev. Mol. Cell Biol*. **2**: 202-210.

Najafi S.M., Harris D.A., Yudkin M.D. (1996). The SpoIIAA protein of Bacillus subtilis has GTP-binding properties. *J. Bacteriol.* **178**: 6632-6634.

Navaratnam D., Bai J.P., Samaranayake H., Santos-Sacchi J. (2005). N-terminal mediated homo-multimerization of prestin, the outer hair cell motor protein. *Biophys. J.* **89**: 3345-3352.

Ohnishi S., Hara M., Inoue M., Yamashita T., Kumazawa T., Minato A., Inagaki C. (1992). Delayed shortening and shrinkage of cochlear outer hair cells. *Am J Physiol.* **263:** 1088-1095.

Oliver D., He D.Z., Klocker N., Ludwig J., Schulte U., Waldegger S., Ruppersberg J.P., Dallos P., and Fakler B. (2001). Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein. *Science* **292**: 2340-2343.

Ozawa K., Headlam M.J., Schaeffer P.M., Henderson B.R., Dixon N.E., Otting G. (2004). Optimization of an Escherichia coli system for cell-free synthesis of selectively N-labelled proteins for rapid analysis by NMR spectroscopy. *Eur. J. Biochem.* **271**: 4084-4093.

Pasqualetto E., Seydel A., Pellini A., Battistutta R. (2008). Expression, purification and characterization of the C-terminal STAS domain of the SLC26 anion transporter prestin. *Protein Expr. Purif.* **58**: 249-256.

Rajagopalan L., Patel N., Madabushi S., Goddard J.A., Anjan V., Lin F., Shope C., Farrell B., Lichtarge O., Davidson A.L., Brownell W.E., Pereira F.A. (2006). Essential helix

interactions in the anion transporter domain of prestin revealed by evolutionary trace analysis. *J. Neurosci.* **26**: 12727-12734.

Rouached H., Berthomieu P., Kassis E.E. Cathala N., Catherinot V., Labesse G., Davidian J.C., Fourcroy P. (2005). Structural and functional analysis of the C-terminal STAS (Sulfate Transporter and Anti-sigma Antagonist) domain of the *Arabidopsis thaliana* Sulfate Transporter SULTR1.2. *J. Bio. Chem.* **280**: 15976-15983.

Rybalchenko V., Santos-Sacchi J. (2003). Cl^- flux through a non-selective, stretchsensitive conductance influences the outer hair cell motor of the guinea-pig. *J. Physiol.* **547**: 873-891.

Saier M.H. Jr., Eng B.H., Fard S., Garg J., Haggerty D.A., Hutchinson W.J., Jack D.L., Lai E.C., Liu H.J., Nusinew D.P., Omar A.M., Pao S.S., Paulsen I.T., Quan J.A., Sliwinski M., Tseng T.T., Wachi S., Young G.B. (1999). Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim. Biophys. Acta* **1422**: 1-56.

Saitoh H., Pu R.T., Dasso M. (1997). SUMO-1: Wrestling with a new ubiquitin-related modifier. *Trends Biochem. Sci*. **22**: 374-376.

Santos-Sacchi J. (1991). Reversible inhibition of voltage-dependent outer hair cell motility and capacitance. *J. Neurosci*. **11**: 3096-3110.

Santos-Sacchi J., Shen W., Zheng J., Dallos P. (2001). Effects of membrane potential and tension on prestin, the outer hair cell lateral membrane motor protein. *J. Physiol*. **531**: 661-666.

Santos-Sacchi J., Song L., Zheng J., Nuttall A.L. (2006). Control of mammalian cochlear amplification by chloride anions. *J. Neurosci.* **26**: 3992-3998.

Schaechinger T.J., Oliver D. (2007). Non mammalian orthologs of prestin (SLC26A5) are electrogenic divalent/chloride anion exchangers. *Proc. Natl. Acad. Sci USA* **104**: 7693-7698.

Schwarz D., Dötsch V., Bernhard F. (2008). Production of membrane proteins using cell-free expression systems. *Proteomics*. **8**: 3933-3946.

Schwarz D., Junge F., Durst F., Frölich N., Schneider B., Reckel S., Sobhanifar S., Dötsch V., Bernhard F. (2007). Preparative scale expression of membrane proteins in *E. coli* based continuous exchange cell-free systems. *Nat. Protoc.* **2**: 2945-2957.

Schwarz D., Klammt C., Koglin A., Löhr F., Schneider B., Dötsch V., Bernhard F. (2007). Preparative scale cell-free expression systems: new tools for the large scale preparation of integral membrane proteins for functional and structural studies. *Methods* **41**: 355-369.

Seavers P.R., Lewis R.J., Brannigan J.A., Verschueren K.H., Murshudov G.N., Wilkinson A.J. (2001). Structure of the Bacillus cell fate determinant SpoIIAA in phosphorylated and unphosphorylated forms. *Structure* **9**: 605-614.

Shelden M.C., Loughlin P., Tierney M.L., Howitt S.M. (2001). Proline residues in two tightly coupled helices of the sulphate transporter, SHST1, are important for sulphate transport. *Biochem. J.* **356**: 589-594.

Shibagaki N., Grossman A.R. (2004). Probing the Function of STAS Domains of the Arabidopsis Sulfate Transporters. *J. Biol. Chem.* **279**: 30791-30799.

Shibagaki N., Grossman A.R. (2006). The Role of the STAS Domain in the Function and Biogenesis of a Sulfate Transporter as Probed by Random Mutagenesis. *J. Biol. Chem.* **281**: 22964-22973.

Song L., Seeger A., Santos-Sacchi J. (2005). On membrane motor activity and chloride flux in the outer hair cell: lessons learned from the environmental toxin tributyltin. *Biophys J.* **88**: 2350-2362.

Spirin A.S., Baranov V.I., Ryabova L.A., Ovodov S.Y., Alakhov Y.B. (1988). A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242**: 1162-1164.

Tang H.Y., Xia A., Oghalai J.S., Pereira F.A., Alford R.L. (2005). High frequency of the IVS2-2A>G DNA sequence variation in SLC26A5, encoding the cochlear motor protein prestin, precludes its involvement in hereditary hearing loss. *BMC Med. Genet.* **6**: 30.

Teek R., Oitmaa E., Kruustük K., Zordania R., Joost K., Raukas E., Tõnisson N., Gardner P., Schrijver I., Kull M., Ounap K. (2009). Splice variant IVS2-2A>G in the SLC26A5

(Prestin) gene in five Estonian families with hearing loss. *Int. J. Pediatr. Otorhinolaryngol.* **73**: 103-107.

Toth T., Deak L., Fazakas F., Zheng J., Muszbek L., Sziklai I. (2007). A new mutation in the human pres gene and its effect on prestin function. *Int. J. Mol. Med.* **20**: 545-550.

Tunstall M.J., Gale J.E., Ashmore J.F. (1995). Action of salicylate on membrane capacitance of outer hair cells from the guinea-pig cochlea. *J. Physiol*. **485**: 739-752.

Wagner S., Bader M.L., Drew D., de Gier J.W. (2006). Rationalizing membrane protein overexpression. *Trends Biotechnol.* **24**: 364-371.

Zheng J., Du G.G., Anderson C.T., Keller J.P., Orem A., Dallos P., Cheatham M.A. (2006). Analysis of the oligomeric structure of the motor protein prestin. *J. Biol. Chem.* **281**: 19916-19924.

Zheng J., Du G.G., Matsuda K., Orem A., Aguinaga S., Levente D., Navarrete E., Madison L.D., Dallos P. (2005). The C-terminus of prestin influences nonlinear capacitance and plasma membrane targeting. *J. Cell Sci.* **118**: 2987-2996.

Zheng J., Long K.B., Shen W., Madison L.D., Dallos P. (2001). Prestin topology: localization of protein epitopes in relation to the plasma membrane. *NeuroReport* **12**: 1929-1935.

Zheng J., Shen W., He D.Z., Long K.B., Madison L.D., Dallos P. (2000). Prestin is the motor protein of cochlear outer hair cells. *Nature* **405**: 149-155.

Zolotarev A.S., Unnikrishnan M., Shmukler B.E., Clark J.S., Vandorpe D.H., Grigorieff N., Rubin E.J., Alper S.L. (2008). Increased sulfate uptake by E. coli overexpressing the SLC26-related SulP protein Rv1739c from Mycobacterium tuberculosis. *Comp. Biochem. Physiol. A* **149**: 255-266.

Part B

Structural studies on the Green Fluorescent Protein mutant, GFPmut2, at different pH

1

Introduction

1.1 The Green Fluorescent Protein (GFP)

GFP was discovered in the early 1960s as an accessory protein of the bioluminescence system of the hydroid jellyfish *Aequorea victoria* based on the protein aequorin (Shimomura et *al.*, 1962). The successful purification of the protein and the subsequent spectral characterization showed that it absorbs blue light and emits green light, thus playing the role of the converter of aequorin blue chemoluminescence into the greenish bioluminescence of the jellyfish (Zimmer, 2002). Thirty years later, the cloning (Prasher et *al.*, 1992) and the successful heterologous expression of the GFP gene (Chalfie et *al.*, 1994) provided the clear demonstration that the green fluorescence is genetically encoded into the primary sequence of the protein and no jellyfish-specific cofactors are needed for the synthesis of the chromophore. Accordingly, GFP can be used as an intrinsic intracellular reporter of target proteins by simple genetic fusion and subsequent gene transfer and expression into cells. Fusion of GFP to a protein rarely affects the proteins activity or mobility and GFP is nontoxic in most cases. Therefore, GFP and its variants have been used for monitoring dynamic processes in cells and organisms, including gene expression, protein localization, and intracellular dynamics. GFP fluorescence is stable, species-independent, and can be monitored noninvasively in living cells by fluorescence microscopy, flow cytometry, or macroscopic imaging techniques (Zimmer, 2002).

Rational and random modifications of the wild-type GFP (wtGFP) primary sequence yielded many mutants with different colors and photophysical properties (Chudakov et *al.*, 2005). The primary advantages of fluorescent protein-based indicators over simple organic dyes are that they can be designed to respond to a much greater variety of biological events and signals, can be targeted to subcellular compartments, can be introduced in a wider variety of tissues and organisms, and seldom cause photodynamic toxicity (Zhang et *al.*, 2002).

Crystal structure and chromofore formation

wtGFP is composed of a single peptide chain of 238 amino acids that has a molecular weight around 27 kDa (Tsien, 1998). As highlighted by X-ray spectroscopy, this sequence

folds in a compact cylindrical three-dimensional structure with a diameter of 24 Å and a height of 42 Å (Figure 1, Ormo et *al.*, 1996). Such a cylindrical structure, often referred to as a β-can or β-barrel, stems from the ordered arrangement of 11-stranded β-sheets. The β-barrel is capped at the extremities by short helices and loops; a single (distorted) helix runs through the centre of the ciylinder (spans the inner portion of the cylinder) and contains the three amino acids forming the chromophore (Figure 1). The chromophore of GFP is buried at the center of the β-barrel and it is highly protected from bulk solvent by the surrounding β-strands. The chromophore is a 4-(4-hydroxyphenyl) methylideneimidazol-5-one group originated from a posttranslational autocatalytic cyclization of the tripeptide segment Ser65-Tyr66-Gly67, with successive dehydrogenation of the α–β bond of the tyrosine in the presence of molecular oxygen (Figure 2; Heim et *al.*, 1994; Cubbit et *al.*, 1995; Reid and Flynn, 1997).

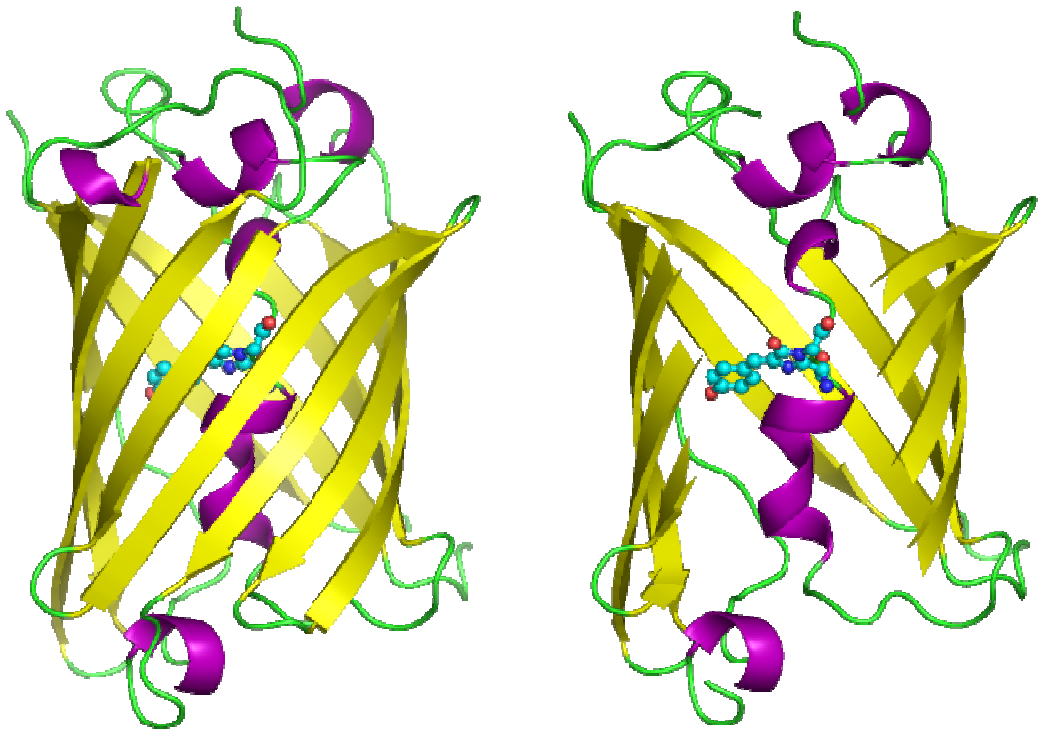

Figure 1: GFP drawn in cartoon style, one fully and one with the side of the beta barrel cut away to reveal the chromophore (highlighted as ball-and-stick). The β-sheets are displayed in yellow, the helixes in purple and the loops in green (Ormo et *al.*, 1996).

The correct folding and configuration of the residues around the chromophore are prerequisites for fluorescence, as the isolated chromophore is not fluorescent in aqueous solution (Niwa et *al.*, 1996). Indeed, the presence of the compact and rigid tertiary structure of GFP is thought to be responsible for the exceptionally high fluorescence and

high quantum yield, around 0.79 (Webber et *al.*, 2001). Owing to the chromophore-enveloping tertiary structure, many of the classical fluorescence quenching agents are almost ineffective in quenching GFP emission, and denaturation occurs only above 76 °C (Ward and Bokman, 1982). Several polar residues and water molecules comprise a hydrogen-bonding network around the chromophore (Brejc et *al.*, 1997). Both the protonation state of the chromophore and its surrounding as well as the hydrogen-bonding network around the chromophore play an important role in the photophysics of GFP (Kummer et *al.*, 2000; Wachter et *al.*, 2000).

Spectral properties as a function of pH

The spectroscopic properties of GFP have been intensively investigated. Wild-type GFP exhibits two absorption maxima at 395 and 475 nm that relax to an emission maximum around 508 nm, emitting a greenish yellow light (Figure 3). These two peaks are attributed to two thermodynamically stable protonation states of the chromophore, in particular of the hydroxyl group of the chromophore phenol. The 395 nm peak is associated with the neutral (protonated) form of the chromophore, whereas the 475 nm peak is associated with the anionic (deprotonated) form (Figures 2 and 3; Heim et *al.*, 1994; Tsien, 1998). The relative intensities of these absorption peaks vary with pH and ionic strength. At physiological pH, GFP has a stronger excitation at 395 nm and a weak amplitude at 475 nm, whereas at high pH (10-11) it loses the 395 nm peak and gains in amplitude at the 475 nm peak (Tsien, 1998). Excitation of either species A (neutral) or species B (anionic) results in similar emission spectra, with a emission maximum around 508 nm. This is presumably due to the fact that the phenolic oxygen of Tyr66 is more acidic in the excited state than in the ground state. After excitation both forms of the chromophore are anionic, which then relax by fluorescence emission (Brejc et *al.*, 1997; Zimmer, 2002).

The equilibrium between these states appears to be governed by a hydrogen bond network that permits proton transfer between the chromophore and neighboring side chains. Indeed, in wtGFP and some mutants, the A and B states differ not only in the protonation of Tyr66, but also in the dissimilar conformation of the residues surrounding the chromophore (Figure 4). This was first revealed by the comparison of the X-ray structures of the wtGFP and the S65T variant (Brejc et *al.*, 1997). In this GFP mutant the absorption peak at 395 nm is suppressed in favor of a peak at 490 nm (Heim et *al.*, 1994;

Delagrave et *al.*, 1995). Analysis of the S65T structure suggests that the change in hydrogen bonding prevents ionization of Glu222 and permits permanent ionization of the chromophore, so that the only excitation peak is in the blue-green region (Ormo et *al.*, 1996; Brejc et *al.*, 1997). Therefore, the S65T structure represents state B in the proposed mechanism.

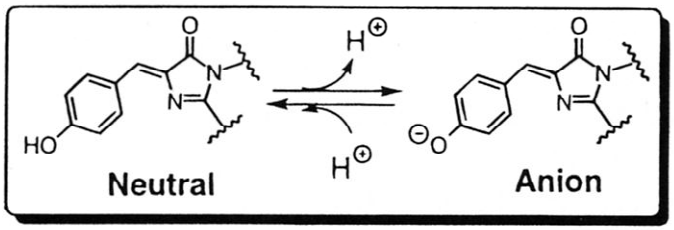

Figura 2: The neutral and anionic forms of the chromophore, traditionally called states A and B, respectively.

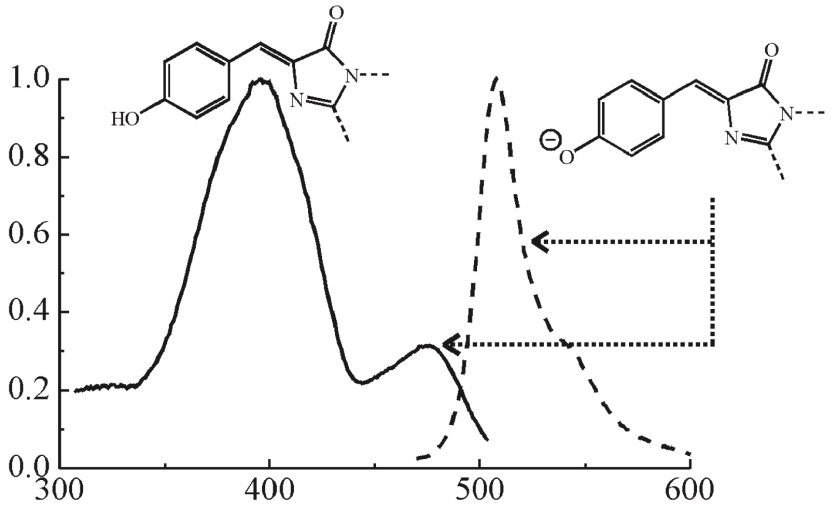

Figure 3: Fluorescence excitation and emission spectra (solid and dashed line, respectively) for wild-type GFP at physiological pH, together with the chromofore structures believed to be responsible for the spectra.

According to the X-ray structures of wtGFP, S65T and other variants, a peculiar hydrogen-bond network connecting the phenol group to residue Glu222 is present in the A state, whereas it is interrupted in the B state (Figure 4; Brejc et *al.*, 1997). The neutral state of the chromophore is maintained by a direct hydrogen bonding network from the deprotonated Glu222 through Ser205 and a water molecule to the hydroxyl group of Tyr66 of the chromophore. In the anionic state Glu222 donates its charge to the fluorophore by proton extraction through a hydrogen bond network involving Ser205 and the bound water. Further stabilization of the ionized state of the fluorophore occurs through a rearrangement of the side chains of Thr203 and His148. The side chain of

Thr203 rotates to solvate the phenolate oxyanion. A similar effect is seen for the His148 side chain, which also stabilizes the phenolic oxygen of Tyr66 though a hydrogen bond.

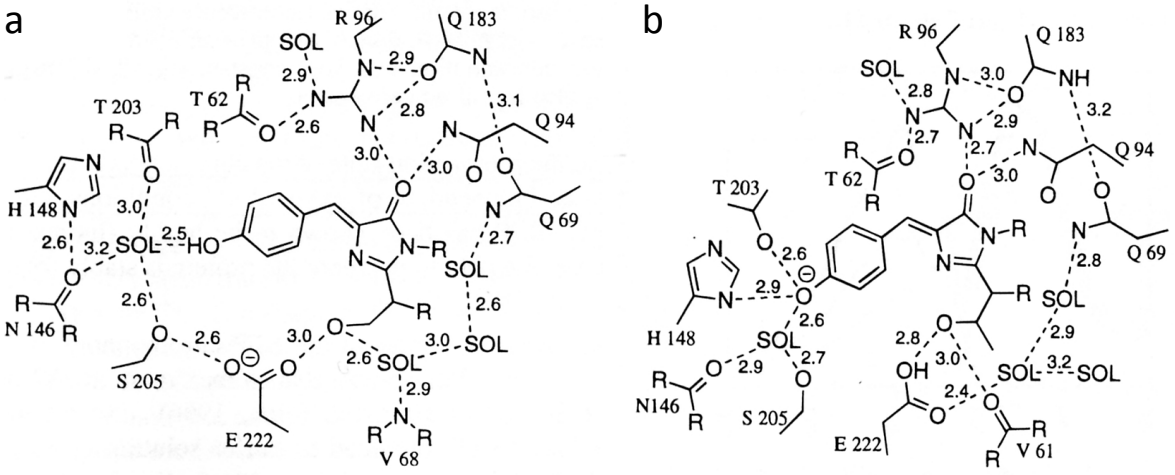

Figure 4: Schematic diagram of the interactions between the chromophore and the surrounding residues and water molecules in (**a**) wtGFP and (**b**) S65T mutant. Hydrogen bonds are shown as dashed lines.

GFP mutants

In wtGFP the equilibrium between the A and B states is affected by a number of factors such as temperature, ionic strength, protein concentration, and, in a peculiar way, pH. Indeed, wtGFP shows pH-dependent absorption/fluorescence spectra at pH around 4.5-5 and 9-10, with impressive absorption/emission stability in the physiological pH range. This behavior is a consequence of molecular optimization of wtGFP by natural selection to play a functional role in *Aequorea victoria* jellyfish. Engineered GFP mutants, however, have no natural role to play and in many cases were found to possess a fast equilibrium between the neutral and the anionic form of the chromophore in the 5-8 pH interval (Bizzarri et *al.*, 2008). The distinct pH-dependences of the fluorescence emission displayed by several GFP variants has been exploited to genetically engineer encodable pH indicators for studies of pH regulation within specific intracellular compartments that cannot be probed using conventional pH-sensitive dyes (Kneen et *al.*, 1998). These pH indicators contributed to shedding light on a number of cell functions for which intracellular pH is an important modulator (Bizzarri et *al.*, 2008).

Several single or multiple mutations of GFP were obtained by random and site-directed mutagenesis to modify the spectral properties and increase the folding efficiency (Tsien, 1998). In particular, as seen before, mutations involving Ser65 are of special

interest since they lead to the selective stabilization of the deprotonated form, shifting the pK_a around neutrality. The most commonly used mutation to favor ionization of the phenol of the chromophore is the replacement of Ser65 by Thr, or S65T (Heim et *al.*, 1994), though several other aliphatic residues such as Gly, Ala, Cys, and Leu have roughly similar effects (Delagrave et *al*., 1995; Cormack et *al.*, 1996). In each of these mutants at physiological pH the 395 nm excitation peak due to the neutral phenol is suppressed, and the 475 nm peak due to the anion is enhanced five- to six-fold in amplitude and shifted to 490 nm. The anionic form of the chromophore has, in fact, enhanced fluorescence emission (Tsien, 1998).

In the S65T mutant, the anionic state of the fluorophore is stabilized by the interruption of the hydrogen-bond network. In this case, Glu222 is forced in the protonated neutral state by a rearrangement of Thr65 that now donates its hydrogen to the backbone carbonyl of Val61, with the consequence that Glu222 can only be an hydrogen–bond donor for Thr65. Furthermore, Glu222 side chain rotates in such a way that it cannot interact with Ser205 anymore, with the breaking of the hydrogen-bond network to the fluorophore. The negative charge, now situated in the hydroxyl group of the fluorophore, is stabilized by Thr203 and His148 whose side chains rotate in such a way to be able to interact directly with the chromophore (Ormo et *al.*, 1996; Brejc et *al.*, 1997).

GFPmut2

Aequorea victoria is found in the cold Pacific Northwest, and fully fluorescent GFP is most efficiently formed at temperatures well below 37 °C. This has limited the uses of GFP and has led to the search for mutants that mature more efficiently at higher temperatures (Tsien, 1998; Zimmer, 2002).

The triple mutant, S65A, V68L, and S72A, called GFPmut2, shows very similar spectroscopic features to those of S65T, but GFPmut2 offers some advantages, like a more efficient folding at 37°C and enhanced fluorescence emission upon excitation of the anionic form of the chromophore (Cormack et *al.* 1996). These properties render GFPmut2 a good candidate for cell biology and biophysical applications.

The spectroscopic properties of GFPmut2 were previously characterized (Chirico et *al.*, 2002): the absorbance spectrum shows a band centered at 485 nm at alkaline pH and the fluorescence emission spectrum, on excitation at 485 nm, shows a band centered at 507 nm (Figure 5). Absorbance and emission are pH dependent: lowering the pH to 4.9

results in 90% decrease of the absorbance intensity. A similar behavior is observed for the fluorescence emission. Below pH 8.4 an absorbance peak at 393 nm appears (band A) and increases substantially on lowering the pH of the solution. Excitation at this wavelength leads to a very weak fluorescence emission with a spectrum similar to the one recorded after 485 nm excitation. The pH dependence of the absorbance and the fluorescence emission of GFPmut2 exhibited an identical pK_a of 6.1±0.1, indicating that only the protonated and deprotonated forms, absorbing at 393 and 485 nm, respectively, are in equilibrium.

Far-UV circular dichroism experiments showed that the observed changes with pH of the spectral properties of GFPmut2 are not correlated with changes in the secondary structure of the protein (Abbruzzetti et *al.*, 2005), as already observed for S65T.

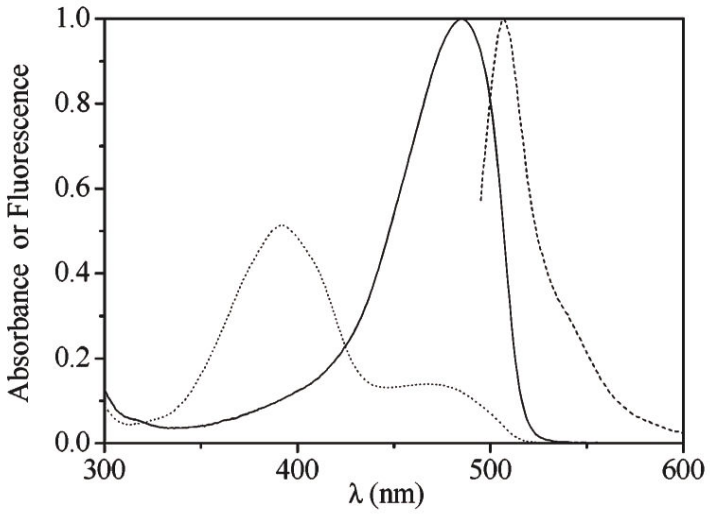

Figure 5: Comparison between the absorbance spectrum (solid line) and the fluorescence emission spectrum (dashed line) of the deprotonated form (pH 8.4) of GFPmut2. The absorbance spectrum of the protonated form (pH 4.9) is reported as dotted line. The excitation wavelength for the emission spectrum was 485 nm (Abbruzzetti et *al.*, 2005).

Abbruzzetti and colleagues performed a detailed kinetic description of the pH-induced transformations at the chromophore site. They used a nanosecond pH-jump technique, coupled with simultaneous transient absorption and fluorescence emission detection, to characterize the dynamics of the acid-induced spectral changes in the GFPmut2 chromophore (Abbruzzetti et *al.*, 2005). The relaxation kinetics can be described by a double exponential decay function with the same lifetimes for the absorbance and fluorescence emission. The response of the GFPmut2 chromophore to a rapid change in pH occurs with a complex kinetics reflecting a diffusion mediated binding

of protons to the phenolate anion and a subsequent process likely due to the rearrangement of the local hydrogen bond network surrounding the chromophore. The protons react with the phenolate with a very low reaction rate, reflecting the protection exerted by the protein fold on the chromophore.

To understand the source of the observed kinetic events, they introduced mutations in amino acid residues that interact with the chromophore (H148G, T203V, and E222Q). When His148 is substituted with Gly, the protonation rate shows a remarkable increase, indicating that the chromophore has a higher direct accessibility to the solvent. On the other hand, the deprotonation rate remains unaffected. The interpretation of this effect is that substitution of the imidazole with a less bulky aliphatic group creates a hole in the β-can, thus allowing an easier entrance way for the protons coming from the solvent. The protonation rate is almost identical for GFPmut2, T203V, E222Q, and S65T GFP, showing that while His148 is a key residue in shielding the chromophore from the bulk solvent, all other mutations do not alter the path of the protons to the chromophore. The deprotonation rate for T203V is an order of magnitude smaller than the corresponding values for the other mutants, showing that the hydrogen bond with the hydroxyl of Thr203 is important in stabilizing the deprotonated form of the chromophore.

They also derived a kinetic model for the protonation reaction of GFP chromofore. According to this, in addition to protecting the chromophore from the solvent, His148 may be involved in the primary binding of protons from the bulk. Protons are then shuttled to the less acidic phenolic oxygen of the chromophore with relatively high efficiency.

Aim of this study

Although a significant amount of research was carried out studying the spectral properties of GFPmut2, many questions regarding the conformational change of the chromophore and its neighboring residues upon changes in pH still remain unresolved. The tridimensional structure of this mutant has not been determined yet and it is fundamental to understand in detail the structural response of GFPmut2 to changes in pH.

The work described in this part of the thesis was carried out in collaboration with Prof. Stefano Bettati's group (*Dipartimento di Biochimica e Biologia Molecolare, Universita` di Parma*) and concerns the crystallographic analyses of GFPmut2 at both acidic (pH 6) and basic pH (pH 9). Comparing the two structures can shed light on the pH induced transformation at the chromop site. This structural information will be correlated

with spectral properties, providing a complete understanding of the response of GFPmut2 to changes in pH.

2

Results and discussion

2.1 Structure of the GFPmut2 at both acidic and basic pH

Experimental procedures

Protein purification and crystallization

The GFPmut2 (S65A, V68L, and S72A) was produced, purified and crystallized by Prof. Bettati's group of the *Università di Parma*. The protein was expressed in *E. coli* with a cleavable $(His)_6$-tag at its N-terminus.

Crystals of GFPmut2 were obtained at 277 K, by using the vapour diffusion technique with hanging drops. Crystals were grown using solutions containing 100 mM MES, pH 6, and 35%-50% (v/v) of 2-methyl-2,4-pentanediol (MPD) as precipitant. For the high pH structure, the mother liquor of the crystals grown at pH 6 was exchanged with 50 mM Tris and 50 mM MES, pH 9, and 75% (v/v) MPD.

Spectroscopic analysis

Absorbance spectra at different pH values of GFPmut2 were performed by Prof. Bettati's group, both in solution and in the crystal. The absorbance was monitored between pH 5 and 9.

Peaks absorbance intensity (at 395 nm and 483 nm for the titration in solution and at 388 nm and 493 nm for the titration on the crystal) was plotted as function of pH and fitted to a titration curve. In solution GFPmut2 has a pK_a around 6.0, consistent with previous report (Chirico et *al*., 2002). In the crystal the pK_a is around 7.0.

Data collection and processing

Diffraction data were measured at the X-ray diffraction beam-line of the ELETTRA synchrotron in Trieste (Italy). The crystals were formed by plates grown one upon the other and it was not possible to separate them and isolate a single one (Figure 6). Nevertheless, at the level of the indexing procedure (Mosflm), it was possible to isolate

the strongest diffraction pattern of a single lattice and datasets at around 1.6 Å resolution could be collected with good final statistics (Table 1).

Crystals were frozen at 100 K under a stream of nitrogen without the need of any cryoprotectant solution, because of the high concentration of MPD in the mother liquor. Wavelengths of 1.1 and 1.2 Å were selected. Diffraction data were measured using MAR-CCD detector (MAR Research) positioned at a distance of 100 and 90 mm from the sample for the structure at pH 6 and 9, respectively. 180 rotations of 1° were performed. Data were processed with the Mosflm software (Leslie, 1991) and reduced and merged with Scala (CCP4 software package, 1994). The statistics are reported in Table 1. Crystals of GFPmut2, at pH 6 and 9, belong to the orthorhombic space group $P2_12_12_1$, with unit cell parameters reported in Table 1. There is a monomer in the asymmetric unit, corresponding to a V_M of 1.99 $Å^3$/Da and a solvent content of 38%.

Structure determination and refinement

The structure of GFPmut2 at pH 6 was solved with the molecular replacement method, using the software Phaser (McCoy et *al*., 2007) from the CCP4 suite, and the GFP mutant S65T/H148D as a search model (PDB code 2 DUI; Shu et *al.*, 2007). The molecular replacement provided a unique solution. Given the isomorphous unit cell, for GFPmut2 at pH 9 an initial rigid body (Refmac from the CCP4 suite), using the GFPmut2 at pH 6 as template, was sufficient to determine the initial phases. The initial models were refined alternating automatic minimization protocols (restrained refinement, software Refmac) with visual inspection of the electron density map and manual adjustment by using the program Coot (Emsley and Cowtan, 2004). During refinement, water molecules were added to the model, both automatically and manually, and those with B factors higher than 50 were excluded. Some molecules of MPD were also introduced and refined.

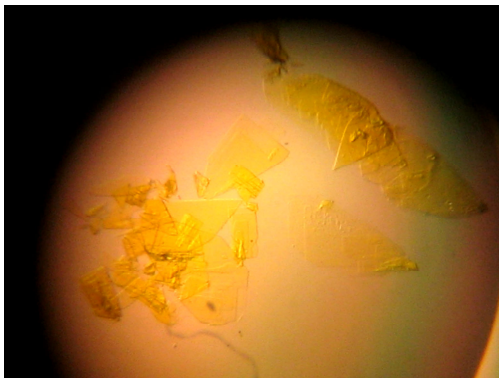

Figure 6: Crystals of GFPmut2.

Table 1: Data collection and refinement statistics of GFPmut2 at both pH 6 and 9.

GFPmut2	**pH 6.0**	**pH 9.0**
Data collection (XRD1, ELETTRA, Trieste)		
Wavelength(Å)	1.1	1.2
Space Group	P $2_12_12_1$	P $2_12_12_1$
Axes a, b, c (Å)	32.5 59.9 112.4	32.5 59.9 112.1
Resolution (Å)	1.64 (1.72)	1.66 (1.75)
Independent reflections	23039 (2269)	21006 (2204)
Multiplicity	4.0 (3.4)	5.7 (5.6)
Completeness (%)	82.9 (82.9)	78.6 (78.6)
R_{sym}	0.079 (0.270)	0.061 (0.241)
R_{meas}	0.090 (0.316)	0.068 (0.265)
$\langle I/\sigma (I)\rangle$	7.3 (2.1)	8.9 (3.0)
B_{Wilson} (Å^2)	11.4	15.1
Refinement		
Protein Residues	233	233
Solvent molecules	287	234
Other molecules	2 MPD	3 MPD
R_{work}/R_{free} [a] (%)	17.2/ 21.1	16.9/ 21.1
B_{mean} (Å^2)	10.84	9.73

Numbers in parentheses refer to the last resolution shell. [a] R_{free} was calculated using 5% of the data.

Result and discussion

Crystal structures

The crystal structure of GFPmut 2 at both pH 6 and pH 9 (above and below the pKa in the crystal) was solved to 1.64 and 1.66 Å resolution, respectively, in order to investigate the structural basis of the spectral response to changes in pH. The overall structure is very similar to that of the wtGFP: 11 β-strands form the classic "β-can" structure and the chromophore, belonging to the central distorted helix, is well buried inside (Figure 7). The quality of the models is good (Figure 8) and the high resolution allows a detailed analysis of the chromophore environment.

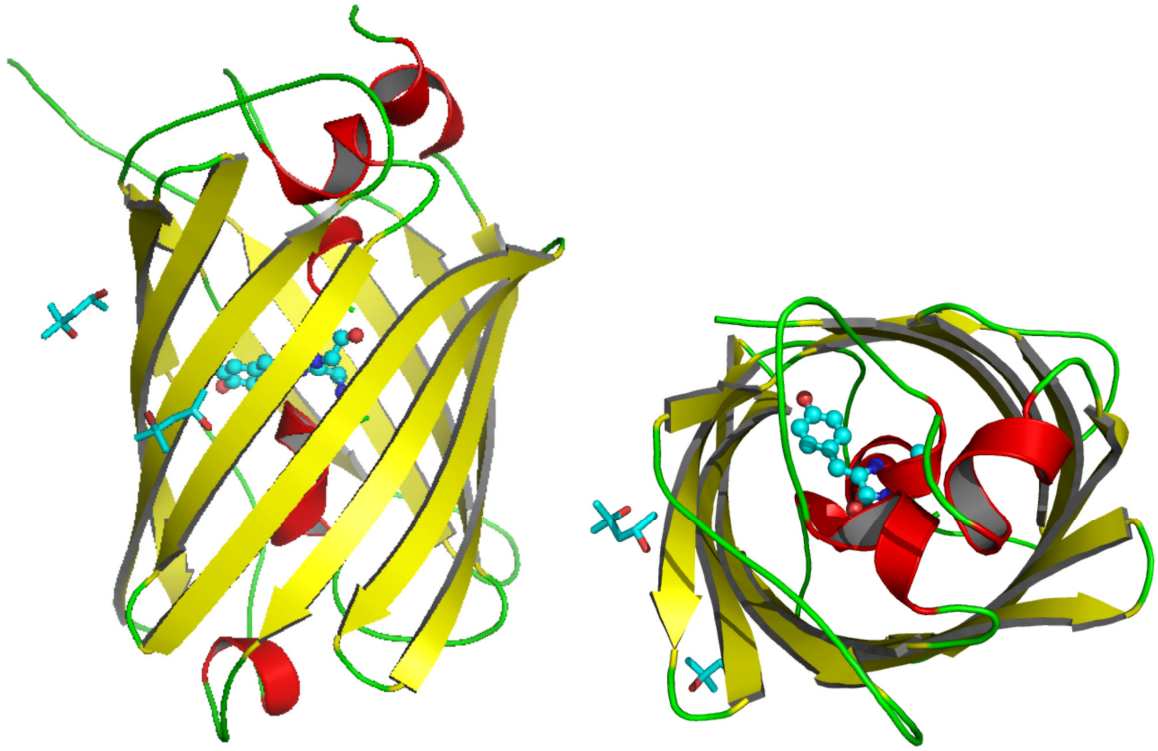

Figure 7: Ribbon view of GFPmut2, seen from the front and from above. The β-sheets are displayed in yellow, the helixes in red and the loops in green. The chromophore is shown as a ball-and-stick model. The figures were produced by PyMol.

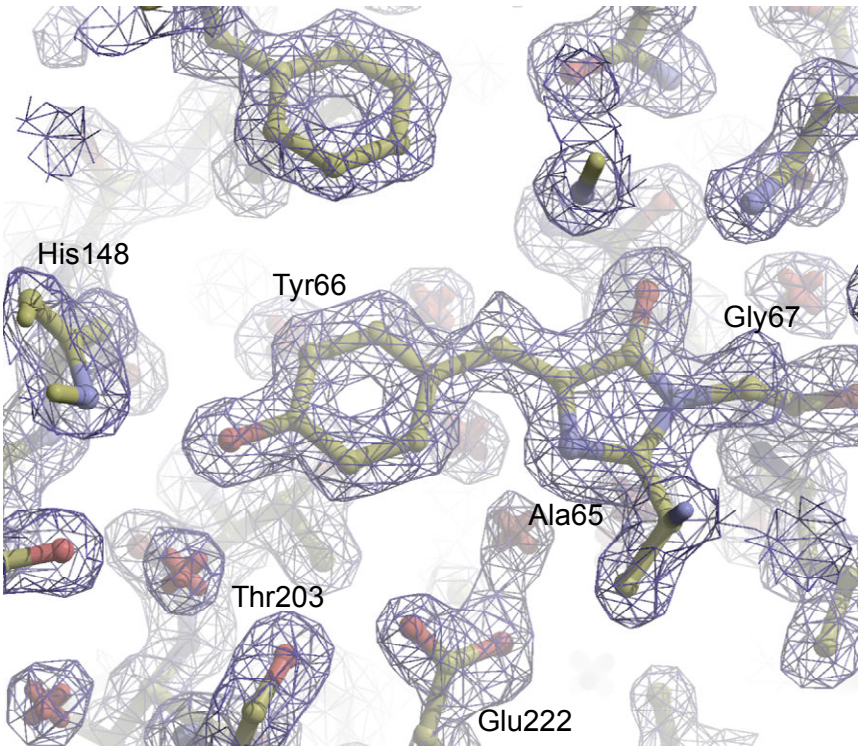

Figure 8: Stereo image of (2Fo-Fc) electron density map, contoured at 1.0 σ, and atomic model of GFPmut2 at pH 6, showing the fluorophore and its environment at around 1.6 Å resolution.

The main differences compared to both the wtGFP (PDB code 1GFL; Yang et *al*., 1996) and the S65T mutant (PDB code: 1EMG; Elsliger et *al*., 1999) are found in the chromophore position and in the surrounding water molecules and can be ascribed to mutations proper to GFPmut2, in particular Ser65Ala and Val68Leu. Due to the bulkier side chain compared to valine, leucine 68 causes a slight shift of the chromophore toward the β-sheet external wall of the protein, in particular toward His148 now at only 3.2 Å in the pH 6.0 structure (with the protonated chromophore). This allows a direct interaction between the two residues (Figure 9), with the formation of a hydrogen bond not possible in other GFP structures with the chromophore in the protonated form, such as the wtGFP (PDB code 1GFL, Figure 11).

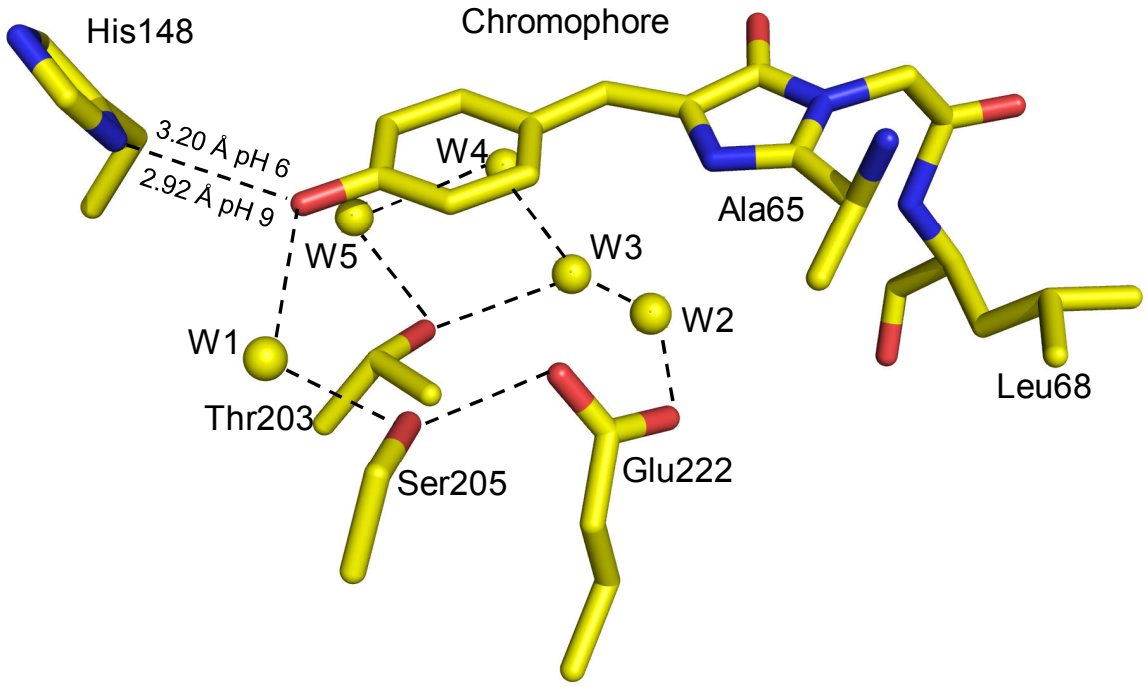

Figure 9: Stereo image of the atomic model of the fluorophore and its environment in the GFPmut2 mutant at pH 6. Ala65, Leu68, His148, Thr203, Ser205, Glu222 and water molecules (W1, W2, W3, W4, W5) are labelled. Dashed lines indicate hydrogen bonds.

Another interesting feature of the GFPmut2 structures is the presence of two additional water molecules (W4 and W5 in Figure 9) in the proximity of the chromophore, in the plane roughly formed by Glu222, Gln69 and Thr203. Usually, here two water molecules (W2 and W3 in Figures 11, 12) are found, connecting Glu222 and Gln69. Thr203 side chain usually interacts directly with only one water molecule (or none in the S65T mutant). In GFPmut2, the position of the two waters (W2 and W3) is different and

two additional water molecules are found (W4 and W5), with the consequences that Glu222 and Thr203 are now directly bridged by W2 and W3 and that Thr203 side chain conformation is stabilized by the interactions with W3 and W5, at 2.6 and 2.7 Å. This is quite relevant because Thr203 side chain usually rotates away from this position in the presence of a deprotonated chromophore (such as in the structure of S65T mutant at pH 8, Figure 12), contributing to its stabilization with a direct H-bond. In GFPmut2 this movement is much more difficult as the Thr hydroxyl group is "fixed" by the two hydrogen bonds with W3 and W5. Another important characteristic of the mut2 structures is that the lack of a hydroxyl function in residue 65 (Ser in wt, Figure 11; Thr in S65T mutant, Figure 12) hampers the possibility to Glu222 to interact with such function. Therefore, Glu222 always interacts with Ser205, taking part in a sort of "circular" H-bond network that now extends from the chromophore hydroxyl to Thr203, through W1, Ser205, Glu222 and two other water molecules,W2 and W3 (Figure 9).

In the structure determined at pH 9 there is only one relevant change, the distance between the hydroxyl chromophore and His148 becomes smaller, from 3.20 to 2.92 Å, indicating a stronger interaction between the two, probably for the deprotonation of the chromophore that donates the proton to the His imidazole ring in an acid-base exchange illustrated in Figure 10. This is well in accordance with the spectroscopic studies of the GFPmut2 properties in solution, which revealed how His148 acts as the primary acceptor for protons coming from the bulk, and their following release to the phenolic oxyanion of the chromophore with relatively high efficiency (Abbruzzetti et *al*., 2005).

At variance with the wt and other GFP mutants studied so far, in mut2 the chromophore deprotonation is not coupled with conformational rearrangements of the surrounding residues, with particular reference to Thr203 and Glu222, now more strongly anchored in their position by the more extended H-bonding network.

The difference between the pKa calculated in solution (around 6) and that in the crystal (around 7) is probably due to the presence of the high MPD concentration in the crystallization medium.

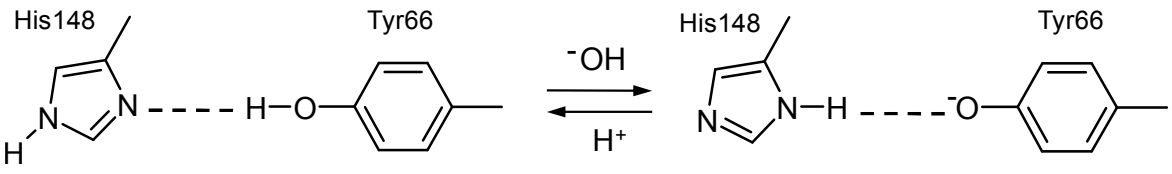

Figure 10: Acid-base equilibrium between His148 and Tyr66.

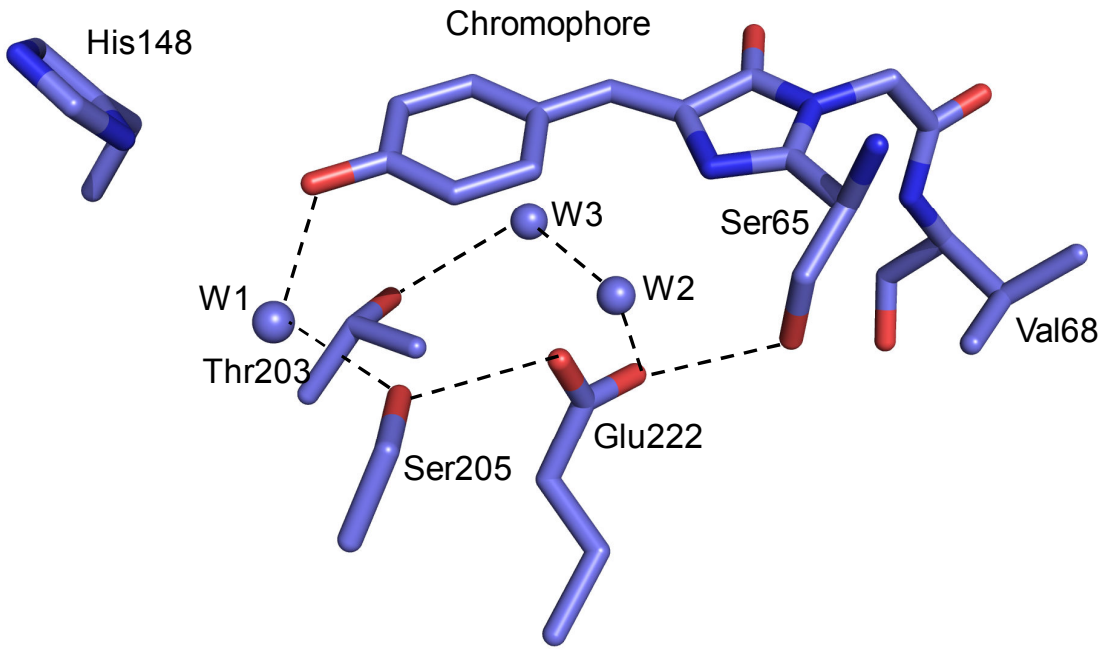

Figure 11: Stereo image of the atomic model of the fluorophore in the protonated form and its environment in the wtGFP (PDB code 1GFL; Yang et *al.*, 1996). Ser65, Val68, His148, Thr203, Ser205, Glu222 and water molecules (W1, W2, W3) are labelled. Dashed lines indicate hydrogen bonds.

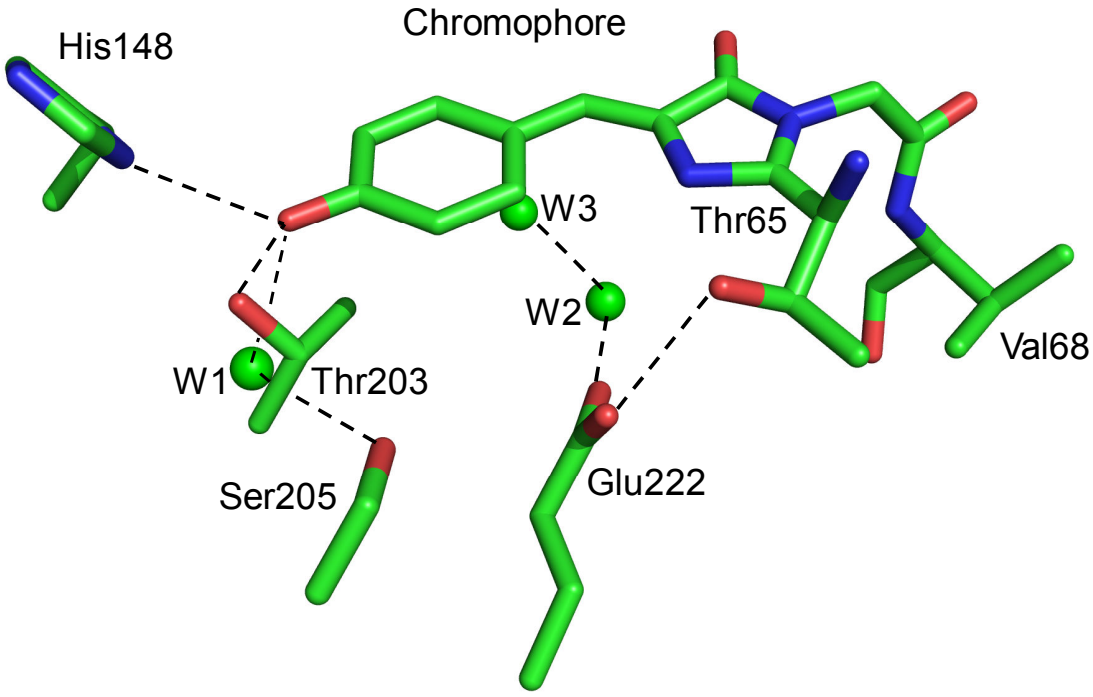

Figure 12: Stereo image of the atomic model of the fluorophore in the deprotonated form and its environment in the S65T mutant at pH 8 (PDB code 1EMG; Elsliger et *al*., 1999). Thr65, Val68, His148, Thr203, Ser205, Glu222 and water molecules (W1, W2, W3) are labelled. Dashed lines indicate hydrogen bonds.

We can conclude that the crystal structures of GFPmut2 at two different pH have indeed revealed the structural bases of the spectral properties of this mutant. The mutations, in particular Ser65Ala and Val68Leu, are responsible for an intrinsically closer interaction between the chromophore and His148. Moreover, the presence of a more extended H-bonding network involving Thr203 and Glu222 is responsible for their diminished conformational freedom.

References

Abbruzzetti S., Grandi E., Viappiani C., Bologna S., Campanini B., Raboni S., Bettati S., Mozzarelli A. (2005). Kinetics of Acid-Induced Spectral Changes in the GFPmut2 Chromophore. *J. Am. Chem. Soc.* **127**: 626-635.

Bizzarri R., Serresi M., Luin S., Beltram F. (2008). Green fluorescent protein based pH indicators for in vivo use: a review. *Anal. Bioanal. Chem.* DOI: 10.1007/s00216-008-2515-9.

Brejc K., Sixma T.K., Kitts P.A., Kain S.R., Tsien R.Y., Ormo M., Remington S.J. (1997). Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **94**: 2306-2311.

Chalfie M., Tu Y., Euskirchen G., Ward W.W., Prasher D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science* **263:** 802-805.

Chirico G., Cannone F., Beretta S., Diaspro A., Campanini B., Bettati S., Ruotolo R., Mozzarelli A. (2002). Dynamics of green fluorescent protein mutant2 in solution, on spin-coated glasses, and encapsulated in wet silica gels. *Protein Sci.* **11**: 1152-1161.

Chudakov D.M., Lukyanov S., Lukyanov K.A. (2005). Fluorescent proteins as a toolkit for *in vivo* imaging. *Trends Biotechnol.* **23**: 605-613.

Collaborative Computational Project N. 4. The CCP4 suite: programs for protein crystallography. (1994). *Acta Crystallogr. D* **50**: 760-763.

Cormack B.P., Valdivia R.H., Falkow S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**: 33-38.

Cubitt A.B., Heim R., Adams S.R., Boyd A.E., Gross L.A., Tsien R.Y. (1995). Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci*. **20**: 448-455.

Delagrave S., Hawtin R.E., Silva C.M., Yang M.M., Youvan D.C. (1995). Red-shifted excitation mutants of the green fluorescent protein. *Biotechnology (N Y).* **13**: 151-154.

Elsliger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J. (1999). Structural and spectral response of green fluorescent protein variants to changes in pH. *Biochemistry* **38:** 5296-5301.

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

Heim R., Prasher D.C., Tsien R.Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **91**: 12501-12504.

Kneen M., Farinas J., Li Y., Verkman A.S. (1998). Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophys. J.* **74**: 1591-1599.

Kummer A.D., Wiehler J., Rehaber H., Kompa C., Steipe B., Michel-Beyerle M.E. (2000). Effects of threonine 203 replacements on excited-state dynamics and fluorescence properties of the green fluorescent protein (GFP). *J. Phys. Chem. B* **104**: 4791-4798

Leslie A.G.W. (1991). Molecular data processing. In: Moras D., Podjarny A.D., Thierry J.P. editors. Crystallographic computing V. Oxford, UK: Oxford University Press; p. 50-61.

McCoy A.J., Grosse-Kunstleve R.W, Adams P.D., Winn M.D., Storoni L.C., Read R.J. (2007). Phaser crystallographic software. *Appl. Cryst.* **40**: 658-674.

Niwa H., Inouye S., Hirano T., Matsuno T., Kojima S., Kubota M., Ohashi M., Tsuji F.I (1996). Chemical nature of the light emitter of the *Aequorea* green fluorescent protein. *Proc Natl Acad Sci USA* **93**: 13617-13622.

Ormo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y., Remington S.J. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**: 1392-1395.

Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G., Cormier M.J. (1992) Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **111**: 229-233.

Reid B.G., Flynn G.C. (1997). Chromophore formation in green fluorescent protein. *Biochemistry* **36**: 6786-6791.

Shimomura O., Johnson F.H., Saiga Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan *Aequorea*. *J. Cell Comp. Physiol.* **59:** 223-240.

Shu X., Kallio K., Shi X., Abbyad P., Kanchanawong P., Childs W., Boxer S.G., and Remington S. J. (2007). Ultrafast Excited-State Dynamics in the Green Fluorescent Protein Variant S65T/H148D 1. Mutagenesis and Structural Studies. *Biochemistry* **46**: 12005-12013.

Tsien R.Y. (1998). The green fluorescent protein. *Annu. Rev. Biochem.* **67**: 509-544.

Wachter R.M., Yarbrough D., Kallio K., Remington S.J. (2000). Crystallographic and energetic analysis of binding of selected anions to the yellow variants of green fluorescent protein. *J. Mol. Biol.* **301**: 157-171.

Ward W.W., Bokman S.H. (1982). Reversible denaturation of *Aequorea* green-fluorescent protein: physical separation and characterization of the renatured protein. *Biochemistry* **21**: 4535-4540.

Webber N.M., Litvinenko K.L., Meech S.R. (2001). Radiationless relaxation in a synthetic analogue of the green fluorescent protein chromophore. *J. Phys. Chem. B* **105**: 8036-8039.

Yang F., Moss L.G., Phillips G.N. (1996). The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14**: 1246-1251.

Zhang J., Campbell R.E., Ting A.Y., Tsien R.Y. (2002). Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol*. **3**: 906-918.

Zimmer M. (2002). Green Fluorescent Protein (GFP): Applications, structure, and related photophysical behavior. *Chem. Rev.* **102**: 759-781.

Abbreviations

ASA	Anti-Sigma factor Antagonist
ATP	Adenosine Tri-Phosphate
Brij-35	Polyoxyethylene-(23)-lauryl-ether ($C_{12/23}$)
Brij-58	Polyoxyethylene-(20)-cetyl-ether ($C_{16/20}$)
Brij-78	Polyoxyethylene-(20)-stearyl-ether ($C_{18/20}$)
Brij-98	Polyoxyethylene-(20)-oleyl-ether ($C_{18\text{-}1/20}$)
CCD	Charge Couple Device
CD	Circular Dichroism
CECF	Continuous Exchange Cell-Free
CF	Cell-Free
CFTR	Cystic Fibrosis Transmembrane conductance Regulator
Chaps	3-[(3-Cholamidopropyl)dimethylammonio]-1-propansulfonat
CMC	Critical Micellar Concentration
DDM	n-Dodecyl-β-D-maltoside
DHPC	1,2-DiHeptanoyl-sn-glycero-3-PhosphoCholine
DLS	Dinamic Light Scattering
DNA	DeoxyriboNucleic Acid
DPC	Dodecyl-PhosphoCholine
DTT	DiThioThreitol
ER	Endoplasmic Reticulum
ESI-TOF	ElectroSpray Ionization Time Of Flight
FM	Feeding Mix
FPLC	Fast Performance Liquid Chromatography
GFP	Green Fluorescent Protein
GFPmut2	Green Fluorescent Protein mutant2
GST	Glutathione S-Transferase
GTP	Guanosine Tri-Phosphate
HPLC	High Perfomance Liquid Chromatography
IHC	Inner Hair Cell
IMAC	Immobilized Metal Ion Affinity Chromatography
IPTG	IsoPropil-β-D-ThioGalactopyranoside

LB	Luria-Bertani liquid medium
LDS	Lithium Dodecyl Sulphate
LMPG	1-Myristoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)]
MES	2-(N-Morpholino)EthaneSulfonic acid
MP	Membrane Protein
MPD	2-Methyl-2,4-PentaneDiol
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off
NLC	Non-Linear Capacitance
NMR	Nuclear Magnetic Resonance
NTP	Nucleoside Tri-Phosphate
OD	Optical Dispersion
OHC	Outer Hair Cell
ONC	Over Night Culture
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEG	PolyEthylene Glycol
$PenCD_L$	Pendrin C-terminal Domain Long
$PenCD_S$	Pendrin C-terminal Domain Short
PFO	PerFluoro-Octanoate
PM	PlamaMembrane
PMSF	PhenylMethane-SulfonylXuoride
$PreCD_L$	Prestin C-terminal Domain Long
$PreCD_S$	Prestin C-terminal Domain Short
$PreCD_T$	Prestin C-terminal Domain total
PreTM	Prestin TransMembrane domain
RM	Reaction Mix
RNA	RiboNucleic Acid
$RvCD_L$	Rv1739c C-terminal Domain Long
$RvCD_S$	Rv1739c C-terminal Domain Short
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	SDS-PolyAcrylamide Gel Electrophoresis
SLC26	Solute Linked Carrier 26
STAS	Sulphate Transporters and AntiSigma factor antagonists
SulP	Sulphate Permease
SUMO	Small Ubiquitin-like Modifier
TB	Terrific Broth
TFA	TriFluoroacetic Acid
TM	TransMembrane
Tris	Tris(hydroxymethyl)aminomethane

tRNA	transporter RiboNucleic Acid
WT	Wild Type
β-OG	n-Octyl-β-D-glucopyranoside

Amino acids

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