THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN NATURAL SCIENCE

Conformational Flexibility in Protein Function

Dynamics of S100A4, Photosynthetic Reaction Centre and the Prenylating Enzymes UbiA and MenA

Annette Duelli

University of Gothenburg

Department of Chemistry and Molecular Biology

Göteborg, Sweden, 2014

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN NATURAL SCIENC

Conformational Flexibility in Protein Function:
Dynamics of S100A4, Photosynthetic Reaction Centre and the Prenylating Enzymes UbiA and MenA

Annette Duelli

Cover: Superposition of S100A4 models from Ensemble Optimization Modeling of SAXS data showing the compact C-terminus (*cyan*) and the extended C-terminus (*green*) upon Ca^{2+}-binding.

ISBN 978-91-628-9141-1
Available online at http://hdl.handle.net/2077/36644

Department of Chemistry and Molecular Biology
Biochemistry and Biophysics
Medicinaregatan 9E
SE-413 90 Göteborg

Printed by Ineko AB
Göteborg, Sweden 2014

Till Staffan och min familj

Abstract

Proteins are the most versatile macromolecules and they are essentially involved in the biological processes throughout all living organisms. The three dimensional structure of proteins and their dynamical properties underlie their biological function and knowledge about protein structure and dynamics contributes to a detailed understanding of biochemical processes. In this work various structural and dynamical methods were applied in the investigation of different protein systems, and in addition the stabilization of two membrane proteins for structural studies was explored.

The first X-ray structure of human S100A4 in complex with a non-muscle myosin IIA (NMIIA) fragment was solved to 1.9Å and contributed to our understanding in the structural mechanism of S100A4 mediated filament assembly which is believed to promote metastasis. The X-ray structure shows that the binding mechanism differs from that of other S100 proteins and that S100A4 adapts its conformation to the chemical properties of the ligand. Further studies on a C-terminal deletion mutant of S100A4 with combined structural high and low resolution methods unveiled a role of the conformational flexible C-terminus in the Ca^{2+}-affinity to S100A4. The results suggest that the reduced metastasis properties that were previously observed in C-terminal deletions mutants of S100A4 might be due to an impaired Ca^{2+}-control.

The nature and the extent of conformational dynamics in photosynthetic reaction centers during the electron transport processes are still not well understood. Differences in the THz absorption spectra of photosynthetic reaction centre from *Rhodobacter sphaeroides* were measured upon light activation and indicate a change in molecular vibrations that occur most probably in LM subunit and are independent of the environment.

Conformational flexible regions influence the formation of protein crystals for structural studies negatively and the structural stabilization of proteins is often applied in protein crystallization. A library of the polyprenyltransfeases UbiA and MenA with rational stabilized, predicted exposed surfaces was produced and eight mutants and the wild type proteins were recombinant overexpressed and purified. Homology models indicate that the mutations of potent mutants are situated in regions that are involved in the formation of crystal contacts. However additional exploration of the buffer environment of UbiA and MenA is required in order to stabilize the proteins for further studies and crystallization.

List of Publications

This thesis is based upon following papers:

Paper I Bence Kiss, **Annette Duelli**, László Radnai, Katalin A. Kékesi, Gergely Katona and László Nyitray. *Crystal structure of the S100A4-nonmuscle myosin IIA tail fragment complex reveals an asymmetric target binding mechanism*. PNAS, 2012, 109(16):p.6048-6053

Paper II **Annette Duelli,** Bence Kiss, Ida Lundholm, Andrea Bodor, Maxim V. Petoukhov, Dmitri I. Svergun, László Nyitray, Gergely Katona. *The C-terminal Random Coil Region Tunes the Ca^{2+}-Binding Affinity of S100A4 through Conformational Activation.* PLOS ONE, 2014, 9(5):p.e97654

PaperIII Ida Lundholm, Weixiao Y. Wahlgren, Federica Piccirilli, Paola Di Pietro, **Annette Duelli**, Oskar Berntsson, Stefano Lupi, Andrea Perucchi and Gergely Katona. *Terahertz absorption of illuminated photosynthetic reaction center solution: a signature of photoactivation?* RSC Advances, 2014, 4:p. 25502-25509

Paper IV **Annette Duelli**, Oskar Berntsson, Emilie Szabo, Weixiao Y. Wahlgren and Gergely Katona. *Exploring the Production and Purification Surface Modified Membrane Protein.* (Manuscript)

Related publications:

Ida V. Lundholm, Helena Rodilla, Weixiao Y. Wahlgren, **Annette Duelli**, Gleb Bourenkov, Josip Vukusic, Ran Friedman, Jan Stake, Thomas Schneider and Gergely Katona. *THz radiation induces non-thermal structural changes in a protein crystal*. Submitted to Science

Other publications:

Maria Ringvall, Elin Rönnberg, Sara Wernersson, **Annette Duelli**, Frida Henningsson, Magnus Åbrink, Gianni Garcia-Feroldi, Ignacio Farjado and Gunnar Pejler. *Serotonin and histamine storage in mast cell secretory granules is dependant on serglycin proteoglycan.* The Journal of allergy and clinical immunology, 2008, 121(4):p.1020-1026

Annette Duelli, Elin Rönnberg, Ida Waern, Maria Ringvall, Svein O. Kolset and Gunnar Pejler. *Mast Cell Differentiation and Activation Is Closely Linked to Expression of Genes Coding for the Serglycin Proteoglycan Core Protein and a Distinct Set of Chondroitin Sulfate and Heparin Sulfotransferases.* The Journal of Immunology, 2009, 183(11):p.7073-83

Astri J. Meen, Inger Øynebråten, Trine M. Reine, **Annette Duelli**, Katja Svennevig, Gunnar Pejler, Trond Jenssen, Svein O. Kolset. *Serglycin is a major proteoglycan in polarized human endothelial cells and is implicated in the secretion of the chemokine GROalpha/CXCL1.* The Journal of Biological Chemistry, 2011, 286(4):p.2636-2647

Contribution report

Paper I I was involved in crystallization set-ups, X-ray data collection, structure solving and preparation of figures.

PaperII I solved the structure, collected SAXS data, analyzed the crystal structure and took part in the analysis of SAXS data, was involved in writing and prepared figures.

Paper III I was involved in data collection at the synchrotron

Paper IV I was involved in the entire project. I isolated and cloned the gene for UbiA, designed the mutations for UbiA and MenA, recombinant overexpression, extraction and purification. I took a major part in interpreting the results, writing and figure preparation.

Abbreviations

ACD	assembly competent domain
BChl	bacteriochlorophyll
Bphe	bacteriopheophytin
CD	circular dichroism
CMC	critical micelle concentration
DDM	n-Dodecyl-β-D-Maltopyranoside
DM	n-Decyl-α-D-Maltopyranoside
DTT	Dithiothreitol
FC-12	fos choline-12
IMAC	immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LDAO	lauryldimethylamine-N-oxide
MD	molecular dynamics
NMIIA	non-muscle myosin IIA
OG	n-Octyl-β-D-Glucopyranoside
OM	n-Octyl-β-D-Maltopyranoside
PDB	Protein Data Base
PDC	protein detergent complex
Q_A/Q_B	quinone A/ quinone B
SAXS	small angle solution X-ray scattering
SEC	size exclusion chromatography
TCEP	Tris(2-carboxyethyl)phosphine
THz	Terahertz
WT	wild type

Contents

1. Introduction

1.1 Protein structural biology

Proteins perform the majority of vital processes in all living organisms and protein dysfunction and missfolding are the direct cause of a number of diseases. Whereas biochemical experimental methods can reveal the type of reactions, the role of certain proteins in complex processes and interaction partners to proteins amongst others, structural information contributes to the understanding in how biological functions are accomplished at the molecular level. Through three dimensional structural protein models the fold of the amino acid chain, the relative orientation of various groups, exposed or buried surfaces or residues, and information about ligand positions and interactions can be identified. This information is highly valuable not only for a deeper understanding but also for the rational design of therapeutic agents that allow us to control disease and pain. The prediction of the three dimensional protein structure by its amino acid sequence is extremely challenging due to the number of geometric possible structures that can be adapted. The combination of several computational methods based on previous knowledge increases the accuracy of predicted models tremendously[1]. However the accuracy and confidence of experimental structure determination can still not be replaced by computational methods. Amongst different experimental high and low resolution methods that coexist for the structural exploration of macromolecules, X-ray crystallography is still the most successful method for the determination of high-resolution protein structures.

1. 2 Protein dynamics

Proteins are dynamic molecules and their biological function depends often on the transition between energetically favorable structural states. Conformational changes in protein molecules include vibrations and side-chain reorientation of amino acids, movements of secondary structure elements and the rearrangement of subunits and play an important role in enzymatic activity, ligand recognition and interaction and protein activation[2]. One of the most dramatic examples of conformational changes might be the folding of intrinsically unstructured proteins to gain biological functionality[3]. Experimentally different conformational states of proteins are often examined by combining high and low resolution and computational methods such as X-ray crystallography, NMR, small – and wide angle X-ray solution scattering (SAXS and WAXS), Infrared spectroscopy, fluorescence labeling methods and molecular dynamics (MD) simulations amongst others. But also X-ray structures obtained from crystals under different conditions or various conformationally trapped states may reveal information about protein dynamics.

In addition to the conformational transition that is typically directly linked to protein activation and function proteins are subjected to constant internal motions that include bond vibrations, side-chain rotamers, hinge bending and loop motions on a timescale of femtoseconds to nanoseconds[4,5], also referred to as “breathing motions”. In X-ray

structures, representing an average of an structure ensemble, atomic motions for instance are partly reflected in the mean square atomic displacement or B-factor[6,7]. However care has to be taken in the interpretation of B-factors, since they display both lattice disorder and thermal motions and are affected by crystal contacts. High resolution NMR structures instead consist of an ensemble of models of a subset of conformational states indicating the internal protein dynamics of the protein molecule. Moreover the methods mentioned above are applied to directly probe internal protein motions. The role of the internal protein dynamics for the biological function of proteins is debated[8] and the question was raised whether these motions are the result of biological constraints or evolutionary optimization[5,9].
Apart from that the importance of protein flexibility, is increasingly recognized not at least in the design of therapeutic agents[10].

1. 3 S100A4

S100A4 is a member of the S100 protein family, the largest subfamily of Ca^{2+}-binding proteins of the EF-hand type. S100 proteins are exclusively expressed in vertebrates where they are distributed over a wide range of different tissues[11,12] to regulate processes such as cytoskeleton dynamics, cell differentiation, cell proliferation, inflammation an cell apoptosis amongst others[11,12,13]. Intracellular, S100 proteins are typically forming homodimers of an X-type for helix bundle, with each subunit containing two Ca^{2+}-binding motives: one high affinity C-terminal canonical EF-hand motif and one N-terminal pseudo EF-hand motif. Upon Ca^{2+}-binding S100 proteins undergo a conformational reorganization that opens up a hydrophobic binding cleft on the protein dimer surface (Figure 1) for the interaction with target proteins both extracellular and intracellular. Through their interaction with target proteins S100 proteins modulate the function of effector proteins and this is believed to play an

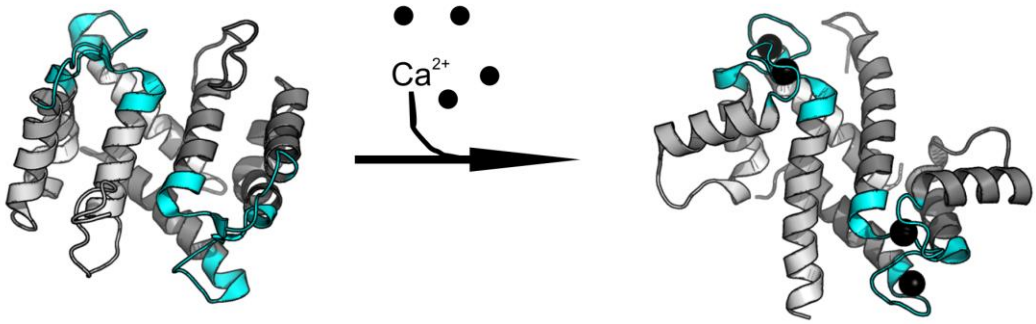

Figure 1. Conformational changes in S100A4 upon Ca^{2+}-binding. The subunits of the dimer are depicted in light and dark grey, Ca^{2+}-in dark in black and the Ca^{2+}-binding motif in cyan

important role in the fine regulation of for instance the enzymatic activity, transcription factors, signal transduction and protein phosphorylation[13]. The protein sequence of the two Ca^{2+}-binding motives is highly conserved amongst members from

the S100 protein family in contrast to the less conserved N- and C-terminal region and the hinge region between helix 2 and 3 that most probably account for the substrate specificity amongst S100 proteins[11]. S100 protein expression levels are often altered in different forms of cancers and they are severly associated with tumorgenesis and tumor progression[12].

S100A4 consisting of 101 amino acids is expressed in various human cell types including fibroblasts[14], monocytes[15], T-lymphocytes[15] and neutrophilic granulocytes[15] where it is found in the nucleus[16], the cytoplasm and the extracellular space[17]. The high resolution structures of S100A4[18] and its activated, Ca^{2+}bound form[19,20,21,22] reveal that conformational changes occur mainly in helix 2 and helix 3 when Ca^{2+}-is bound[20]. Upon Ca^{2+}-activation S100A4 interacts with a variety of proteins such as p53[23], annexin A2 (ANAXA2)[24], F-actin[25] and non-muscle myosin IIA (NMIIA)[26,27,28] to modulate transcription, matrix metalloproteinases and cytoskeleton. S100A4 is strongly associated with metastasis and has been shown to induce a metastatic phenotype in breast cancer models in rats and mice[29,30] and increased cell motility and invasion in Rama37 and epithelial cell lines[29,30] and a human prostate cancer cell line (CaP)[31]. The interaction of S100A4 with non-muscle myosin IIA (NMIIA) is directly correlated with an increased cell motility[32]. NMIIA dimers self-associate to form bipolar filaments that interact with actin filaments in a cross-linking manner to form actomyosin[33] and cell adhesion and the formation of protrusions in migrating cells are dependent on the assembly and disassembly of actomyosin. S100A4 is believed to interfere with cell migration processes through its interaction with NMIIA that has been shown to promote filament dissasembly[27,28,34] and detailed knowledge about the underlying mechanism could contribute to the development of therapeutic agents directed against metastasis properties.

1.4 Photosynthetic reaction centre

The photosynthetic reaction centre is a membrane bound protein complex that converts light energy to chemical energy. Photosynthetic reaction centers are found in green plants and some photosynthetic bacteria and the reaction centre from *Rhodopseudonomas viridis* was the first X-ray structure of a membrane protein that had been determined[35]. The photosynthetic reaction centre of *Rhodobacter sphaeroides* (*R. sphaeroides*) consists of three subunits, heavy (M), light (L) and medium (M) and a set of supporting cofactors: two bacteriochlorophylls P_{870} (special pair), two accessory bacteriechlorophylls, two bacteriopheophytins, two quinones (Q_A and Q_B) and one iron. Figure 2 shows the structure of reaction centre with the three subunits and the cofactors (A) and the cofactors (B) (Q_B is absent). The excitation of photosynthetic reaction centre leads to the transfer of an electron from the special pair (P) to Q_A through bacteriopheophytin[36], resulting in the charge separated state, $P^+BPheQ_A^-$. At this state the electron is further transported to Q_B and in the natural environment P^+ will gain an electron from a reduced cytochrome c2 in order to prevent electron backflow[37]. However if an external electron donator is not present, reaction

centre returns to the ground state through charge recombination and heat is released. With currently 67 reported structures from different organisms in the protein databank (PDB) reaction centre is amongst the best studied membrane proteins, rendering it highly suitable as a model protein. In addition it is convenient to study photosynthetic reaction centre upon activation since the electron transport event can be activated by a laser pulse.

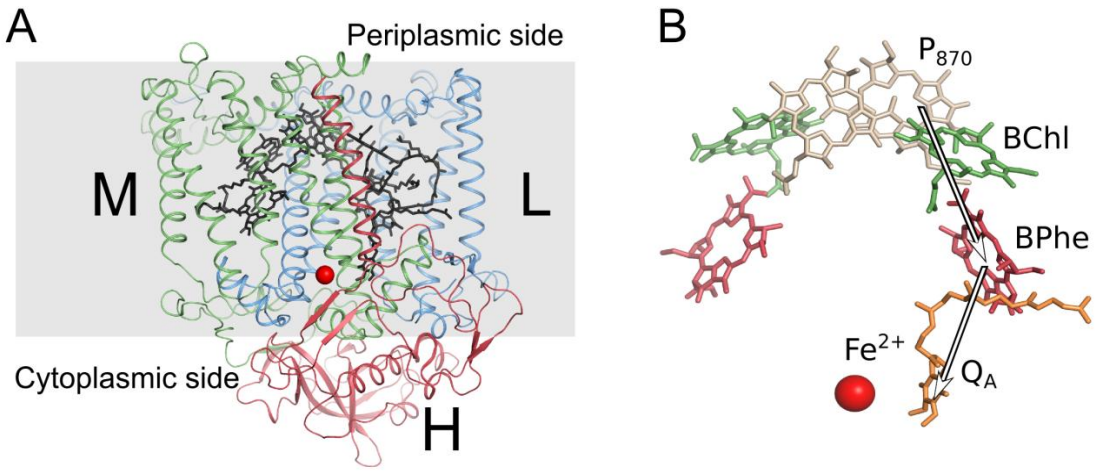

Figure 2. (A) Structure of the photosynthetic reaction center from *R. sphaeroides* (PDB ID 2BNP) showing the three subunits M (*green*), L (blue) and H (*red*) and the cofactors (*black*). The grey region of the picture shows the membrane embedded part of the protein. (B) Cofactors of the reaction center, the special pair (P_{870}), bacteriochlorophyll (BChl), bacteriopheophytin (BPhe) and bound quinine (Q_A). The electron transfer pathway through the cofactors that leads to the charge separated state, $P_{870}^{+}Q_A^{-}$, is shown by the arrows. (Figure from paper III).

1. 5 Membrane protein purification and crystallization: *UbiA* and *MenA*

Membrane proteins account for 30% of the proteome[38]. They perform a variety of essential functions, such as signal transduction, solute transport and catalytic reactions, and their impact in human diseases is not at least reflected in the high number of pharmaceutical agents (over 50%) targeting membrane proteins[39]. Nevertheless structural information about membrane proteins is still difficult to access and high resolution membrane protein structures are highly underrepresented in the protein database (PDB) compared to soluble proteins. The main reason is the difficulty to handle membrane proteins due to the presence of hydrophobic surfaces, but also a higher degree of complexity during their production compared to soluble proteins. Membrane proteins are difficult to produce[40], and the need of detergents to extract membrane proteins can influence the protein activity and stability negatively and does often induce aggregation[41]. Finally hydrophobic surfaces and the requirement of detergents and/or lipids impede the formation of crystal contacts during the crystallization process.

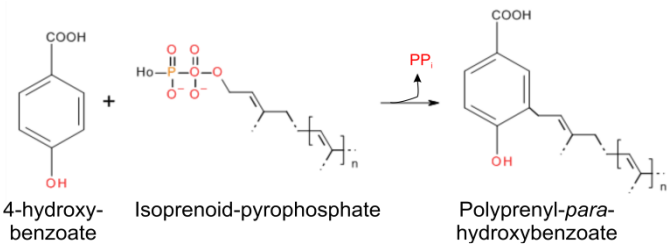

Figure 3. Schematic illustration of the prenylation of 4-hydroxybenzoate

UbiA and Men belong to the UbiA superfamily of membrane embedded prenyltransferases, a group of enzymes that are involved in the biosynthesis of ubiquinone[42,43], menaquinone[44], prenylated hemes and chlorophylls[45,46] and Vitamin E[47]. UbiA fuses an isoprenoid-diphosphate to the *meta* position of 4-hydroxybenzoate, and releases a pyrophosphate[48] as shown in Figure 3. Similarly, MenA catalyzes the prenylation of 1,4-dihydroxy-2-naphtoate in menaquinone biosynthesis[49] and these reactions are considered as key step in the ubiquinone and menaquinone biosynthesis, respectively. Ubiquinone is mainly known as electron transporter between complex I or II and complex III in the respiratory chain and apart from that, ubiquinone acts as an antioxidant[50,51] and is involved in disulfide bond formation in *E. coli* protein folding[52,53]. Menaquinone, in the form of human Vitamin K is involved in the carboxylation of glutamyl residues an important factor in blood coagulation and the dysfunction of the human homolog to MenA, *UBIAD1* is associated bladder tumor progression[54], and Schnyder crystalline cornea dystrophy. To date the availability of structural information about enzymes form the UbiA superfamily is limited to the substrate-bound and apo X-ray structures of two different UbiA horologes from archaea[55,56] and additional structural information could increase our understanding of these enzymes in different organisms.

2. Methodology

2.1 Protein crystallization

Well diffracting protein crystals are a prerequisite for the structure determination of proteins with x-ray crystallography. The formation of protein crystals involves the self-association of protein molecules from a supersaturated protein solution into an ordered array in three dimensions and is highly dependent on a variety of parameters such as temperature, pH, buffer, protein concentration and precipitant amongst others. The conditions that will lead to protein crystals are individual for each protein and cannot be predicted on the protein sequence alone, and especially membrane proteins are difficult to crystallize due to hydrophobic surfaces and the need of detergents. The protein of interest has to be produced, unless obtained from a natural source, in the case of membrane proteins extracted from the membrane, and purified before subjected to crystallization setups (Figure 4). Protein crystallization usually consists of an error –and trial approach combined with accumulated knowledge in an extensive reiterative screening process that hopefully results in crystal leads. Commercial available screens covering a wide range of chemical compositions may facilitate protein crystallization, but if not successful, more rational approaches like the alteration of the protein sequence or the crystallization of domains can be undertaken. In this case it has to be considered, that alterations may affect the biological function of the protein.

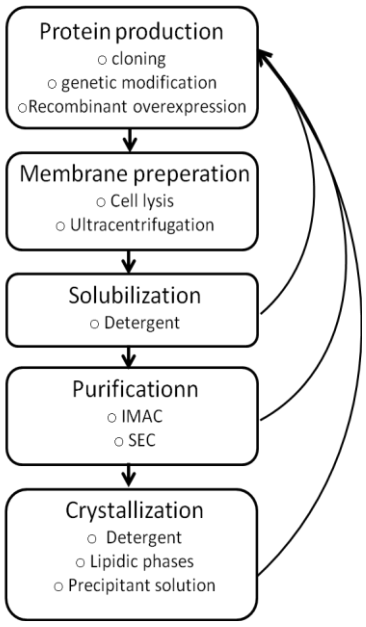

Figure 4. Flowchart for the production of a membrane protein sample for crystallization

2.1.1 Protein production

Growing protein crystals requires huge amounts of protein. As the majority of membrane proteins are not expressed in sufficient amounts naturally, protein production systems such as recombinant protein expression or cell free protein expression are usually utilized to supply the need of protein sample for crystallization. The choice of a suitable host depends on the demands of the target protein regarding translocation and folding machinery, codons, glycosylation and membrane composition amongst others and picking a homologous host to the source of the target protein will ease optimization. Through strong promoters recombinant protein production in *E. coli* is focused on the target protein which can account for up to about 50% of the membrane protein production[57,58]. In contrast to soluble proteins, membrane proteins have to be directed and inserted into the membrane and the recombinant membrane protein production often has to be slowed down in order to prevent the machinery from being overloaded[58]. Lowering the temperature in the case of CorA, an inner membrane magnesium transporter increased the insertion of the

overexpressed protein into the membrane[57] but also reducing the inducing agent (IPTG for promoters under the control of the lac operon) may have positive effects on the protein production. Recombinant protein production through autoinduction[59,60] has been successfully optimized for the production of membrane proteins[61] where the induction is adapted to the cell growth through the enzymatical conversion of lactose to allolactose by the cells themselves.
Besides providing high yields of many proteins, another advantage of the recombinant protein expression is the possibility to introduce genetical modifications in order to design proteins with desired properties.

2.1.2 Rational protein design

Amongst all parameters that influence protein crystal formation, the crystallizability of the protein plays an important role and it has already been recognized early, that variations in the protein sequence in form of homologous proteins influence the formation of protein crystals[62]. Nowadays, with advances in recombinant protein expression, alterations in the protein sequence can be introduced readily and are more common. There are various examples of protein alterations, where sometimes as little as a single point mutation, and truncations leads to the formation of protein crystals where wild type protein crystals were not obtained[63] (and citations within). The truncation of flexible, N- or C-termini or the mutation of cysteine to serine in order to avoid aggregation by disulfide bridge formation are only two out of many approaches to enhance conformational homogeneity and stability in order to facilitate the formation of well-diffracting crystals. Rational surface modifications in order to reduce surface entropy with the aim to increase the propensity of proteins to form crystal contacts have been developed[64,65,66] and the significance of intrinsic protein properties in protein crystallization has been investigated on soluble proteins[67]. In this study it was found, that certain amino acid residues (alanine, glycine and phenylalanine), the mean side chain entropy and well ordered surface epitopes amongst others influence the crystallizability of proteins.
In addition recombinant expression techniques do not only allow for genetical modifications that will influence the protein sequence, but also for modifications that facilitate the purification of proteins like a Histidine-tag. Furthermore Histidine-tags can also be used as markers in the detection of the recombinant protein by WESTERN blotting[68] at various steps throughout protein production and purification.

2.1.3 Detergents

The transfer of membrane proteins from their natural environment, the lipid bilayer, to an aqueous solution is an essential step in producing a protein solution suitable for crystallization. Once membrane proteins are extracted from the biological membrane, their hydrophobic regions must be shielded in order to prevent their hydrophobic collapse in the aqueous buffer solution. Detergents, consisting of a hydrophobic tail- and a hydrophilic head-group are essential tools for both extracting membrane proteins

and substituting lipids around the hydrophobic parts, keeping membrane proteins solubilized. At a certain detergent concentration the critical micelle concentration (CMC) detergents enter into a state, where detergent monomers and micelles exist in equilibrium in solution. It is at concentrations above the CMC detergents exert their effect as solubilizers of membrane components[69]. The extraction of membrane proteins is considered to occur in three steps as depicted in (Figure 5): the uptake of the detergent by the membrane (a), the formation of lipid/detergent/protein micelles and lipid/detergent micelles (b), and finally the formation of detergent micelles, lipid/detergent micelles and protein detergent complexes (PDC) (c)[69,70,71] which are the starting point for further protein purification. The success in obtaining protein crystals is crucial dependant on the choice of the right detergent in the right concentration educing PDCs that posses just the optimal conditions for membrane protein crystallization. Amongst the overwhelming selection of detergents on the market, LDAO, OG, DM and DDM are the ones most frequently found in solved α-helical membrane protein structures[72]. However, being suitable for crystallization does not mean that a detergent is suitable for the solubilization of a given membrane protein which can lead to the application of different detergents in solubilization and crystallization of a single membrane protein.

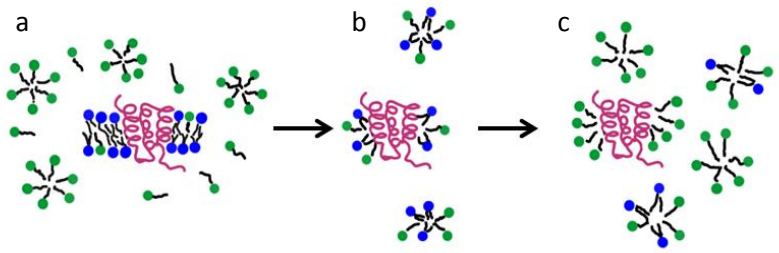

Figure 5. Solubilization of the membrane and extraction of membrane protein. Detergent head groups are represented in green, lipid head groups in blue and the membrane protein in purple. a) detergent is taken up by the membrane; b) lipid/detergent/protein and lipid/detergent micelles; c) protein-detergent-complex, detergent/lipid micelles and detergent micelles.

2.1.4 Chromatographic protein purification and characterization

The purity and homogeneity of the protein solution is generally considered to be an important parameter in protein crystallization, not only because impurities can hamper the growth of monocrystals, but also with respect to reproducibility of the experiments. Following solubilization usually two subsequent chromatographic purification steps for His-tagged proteins are performed: immobilized metal affinity chromatography (IMAC) where the protein is separated from a crude solubilized membrane extract, and size exclusion chromatography (SEC)[73,74,75] in order to remove remaining impurities, aggregates and salts to finally receive a protein solution in a controlled buffer environment. In addition to purification the resulting chromatogram of the SEC reveals properties like aggregation, oligomeric state, the size and stability of PDCs in a given buffer and is therefore a useful tool in screening for suitable buffer components, pH and detergents (so called analytical SEC).

The combination of buffer ingredients like the buffer itself, salts, detergents, possibly additives or cofactors that will yield a protein solution with the desired properties for crystallization set-ups is endless and as with the choice of detergents has to be established individually for each membrane protein. Statistics over buffer components that were most successful in the growth of well-diffracting α-helical membrane protein crystals may aid in finding a starting point for further optimization[72].

2.1.5 Growing protein crystals

The basic principle of protein crystallization is to induce a protein solution to slowly precipitate in an ordered way through the alteration of the solute composition, as described by a phase diagram shown in Figure 6A. The most frequently applied method in protein crystallization is the vapor diffusion method, where a protein solution close to supersaturation is mixed with precipitant solution and placed in a droplet above the precipitant solution as depicted in Figure6B.

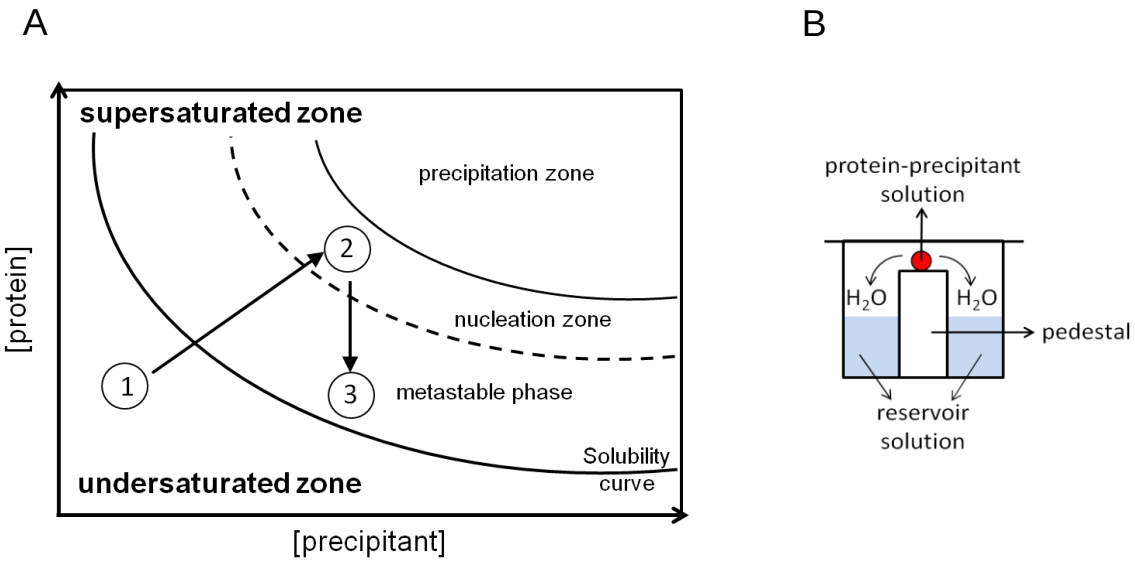

Figure 6. Two dimensional phase diagram (A) and illustration of the vapor diffusion method (B). The phase diagram shows the protein solubility with respect to protein and precipitant concentration. When an undersaturated protein solution (1) is directed towards supersaturation and enters the nucleation zone (2) crystal formation leads to a decrease in protein concentration and the metastable phase can be entered where crystal growth occurs (3). The composition of the protein solution is altered through concentration differences in reservoir solution and protein droplet, that cause water to diffuse to the reservoir solution until an equilibrium is reached.

When tightly sealed, water will diffuse from the droplet, containing a lower precipitant concentration to the reservoir solution, causing precipitant and protein concentration in the droplet to rise until equilibrium between the two solutions is established. Ideally this will create a supersaturated protein solution, as depicted in the phase diagram (1) entering the nucleation zone (2), where the formation of nuclei leads to a reduction of the protein concentration and thus to the entering of the protein solution into the metastable phase (3). The phase diagram is a rather simple illustration, and in practice

nucleation and crystal growth are far more complex putting high demands on precipitant and protein solution. The transformation of the originally undersaturated protein solution to a supersaturated protein solution and the formation of ordered crystals are dependent on the ability of the medium to solubilize the protein and on the attractive interactions between protein molecules themselves. For membrane proteins apart from detergent based crystallization different lipidic environments such as the lipidic cubic phase[76], sponge phase[77] and bicelle method[78] exist.

2.2 X-ray crystallography

Protein molecules are too small to be visualized with a light microscope, since atomic radii and bond lengths are usually in the range of 1-3 Å and the resolution limit of the radiation used is at least half of its wavelength (about 200 nm for visible light). X-rays however, (electromagnetic radiation of wavelengths 0.1-100 Å) fall into the order of magnitude of atomic diameters and bond lengths and can be applied to obtain the resolution required for molecules and macromolecules. X-rays are diffracted by the electrons in atoms, but because unlike visible light, diffracted X-rays cannot be recombined by a lens system to an image, this has to be done by mathematical means from a diffraction pattern.

2.2.1 X-ray diffraction by single, macromolecular crystals

X-rays can basically be diffracted by any atom (or rather electrons in atoms), but it is first through the reinforcement of the diffraction by the repeating units of regularly arranged atoms in crystals and the subsequent interference of scattered rays that distinct diffraction spots, also called reflections are obtained. Through destructive interference waves cancel out each other, and it is only if Bragg's law is fulfilled that constructive interference occurs:

$$2d_{hkl}\sin\theta = n\alpha \qquad \text{(Equation 1)}$$

where n is the order of diffraction, d_{hkl} the interplanar spacing between lattice planes, θ the angle between the incoming beam and the lattice plane and α the wavelength of X-rays (Figure 7).

The

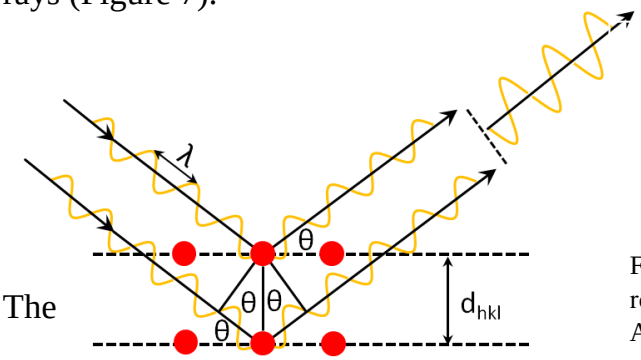

Figure 7. Illustration of two waves of wavelength λ, reflected at the lattice plane that fulfill Bragg's law. Atoms are depicted in read.

Bragg reflections in the resulting diffraction pattern represent each a lattice point related to the Miller indices (hkl) of the crystal lattice planes. The spatial arrangement of the diffraction pattern is reciprocally related to the unit cell dimensions and symmetry rather than to the content of the unit cell. Because of measuring at a single wavelength, a stationary crystal will only produce a small number of Bragg reflections, the crystal is rotated during data collection in order to bring more Bragg reflections into the detector plane. Thus a complete data set for structure determination will be a collection of frames with diffraction patterns of different crystal orientations.

The intensity of each reflection measured in turn is related to the atomic arrangement and the types of atoms in the unit cell and the electron density in the unit cell in turn can be described by the Fourier transform of the structure factor F_{hkl} as followed:

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l \left|F_{hkl}\right| e^{-2\pi i(hx+ky+lz-\alpha^i_{hkl})} \quad \text{(Equation 2)}$$

Where ρ(x,y,z) is the electron density at a position (x,y,z), V the volume of the unit cell, h,k,l the Miller indices, $|F_{hkl}|$ the absolute value of the structure factor and α'_{hkl} the phase angle. The structure factor F_{hkl} itself describes a diffracted ray consisting of amplitude, frequency and phase. The frequencies are experimentally determined through the indices hkl (corresponding to $1/d_{hkl}$) and the structure factor amplitudes through the measured intensities, where $I \sim |F_{hkl}|^2$. The phase information however, is lost during the diffraction experiment, which is referred to as "the phase problem" in X-ray crystallography.

2.2.2 The phase problem

The phase contains information about the exact position of a wave with respect to its origin and each reflection has its own phase. Unless the phase is 0, the electron density giving rise to a measured reflection may not peak at the corresponding crystal lattice planes, but somewhere in-between the crystal lattice planes. With the phase information lacking the exact position of peaks of electron densities remain unknown and the electron density cannot be calculated with equation 2.

Through experimental phasing methods like MAD (multi-wavelength anomalous dispersion), SAD (single wavelength anomalous diffraction), MIR (multiple isomorphous replacement) or SIR (single isomorphous replacement)[79,80] initial phases can be estimated *de novo*. Alternatively the structure factors from a known protein structure can be used as initial estimate of phases, if partial structure homology between the known protein structure (phasing model) and the new protein can be assumed. This method, known as molecular replacement is experimentally less elaborate, since additional data collection on crystal derivates does not have to be performed.

2.2.3 Data collection

Data collection is typically performed at synchrotrons, where electrons circulate at relativistic velocities in a storage ring. When forced into a curved motion, X-rays are emitted, that are directed through a system of collimators, monochromators and focusing mirrors, to produce a parallel, monochromatic X-ray beam. The crystal is aligned to the X-ray beam and rotated about 0.1-2 degrees per frame during data collection. The diffraction pattern of each frame is recorded on a 2D detector (Figure 8) which integrates the intensities of the spots. The angular range of data to be collected, the oscillation range and the alignment of a certain crystal axis can be estimated by the analysis of 3-5 frames recorded prior to data collection. The parameters depend on the unit cell dimension, the internal unit cell symmetry and the mosaicity of the crystal. Software like MOSFLM[81] or HKL-3000[82] amongst others assist in finding a data collection strategy that is suitable for the current crystal.

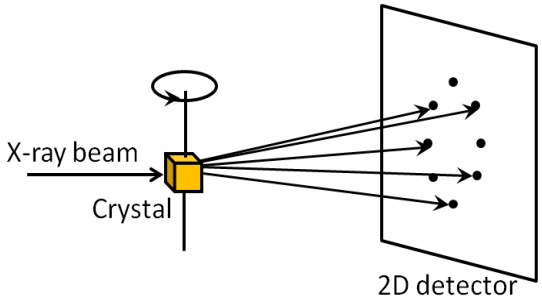

Figure 8. Schematic illustration of a X-ray experiment showing the data collection from a single crystal.

2.2.4 Cryocrystallography

The exposure of protein crystals to X-ray radiation reduces the crystal order through direct ionization of the protein molecules and indirect through chemical modifications by the ionized solute[83] . This radiation damage has a negative impact on the data quality which is observed by the loss of the diffraction intensity, and an increase in the temperature factor amongst others. Through cryocooling and data collection at 100 K, introduced by Hope[84] radiation damage is significantly reduced. In order to prevent the water present in the protein crystal to disrupt the crystal structure when transformed to its crystalline stage, cryoprotectants are added to the protein crystal prior to flash-cooling it in liquid nitrogen. Low molecular weight polyethylene glycols, glycerol or saturated saccharose solution are common examples of cryoprotectants used in protein crystallography.

2.2.5 Data processing

During data processing the reflections are indexed according to the miller indices, unit cell dimensions and space group are determined and the spot intensities are integrated. Subsequent scaling includes the correction of the absolute intensities of the measured reflections for differences in e.g. the X-ray path through the crystal, variability of diffraction powers in different crystals or crystal decay over time, and finally the intensities are converted to structure factors. These processes are performed

computational with software like MOSFLM[81] or XDS[85] and are widely accessible and automated at synchrotrons.

2.2.6 Molecular replacement

The basic principle of molecular replacement is that phases of a structural similar protein (the phasing model) are applied in order to calculate an initial electron density map for the new protein model. During the iterative refinement process the initial phases are converted to the phases of the new protein. For this purpose the phasing model has to be superimposed to the new, unknown target protein in the unit cell. By setting all phases to 0, a map of interatomic vectors, independent of the origin and the orientation of the molecule in the unit cell can be calculated by the Patterson function[86] a variation of equation 2:

$$\rho(u,v,w) = \frac{1}{V}\sum_h \sum_k \sum_l \left|F_{hkl}\right|^2 e^{-2\pi i(hu+kv+lw)} \quad \text{(Equation 3)}$$

The Patterson function can be calculated in the absence of any phasing information to obtain peaks at locations corresponding to interatomic vectors. The correlation of the Patterson maps between phasing model and target protein guides the superposition of the phasing model on the target protein in the unit cell, through a rotation and translation search. The computer program Phaser, in the CCP4 program suite[87] can be used for phasing of macromolecular structures.

2.2.7 Structure refinement

Structure refinement aims the improvement of the phases on one hand, and thus the electron density map and the molecular model on the other hand through a successive iterative refinement process, alternating between the real space and the reciprocal space. Simplified, in the reciprocal space the electron density map is interpreted and a model is build. In the real space the electron density map is improved with the aid of the currently build model through automated refinement. If successful, structure factors obtained from experimentally measured intensities (F_{obs}) and structure factors calculated from the model (F_{calc}) will converge during the refinement process. During the refinement process the refined model is compared for its agreement with the data by means of the crystallographic R-factor:

$$R_{work} = \frac{\sum \left\| F_{obs}(h,k,l) \right| - \left| F_{calc}(h,k,l) \right\|}{\sum \left| F_{obs}(h,k,l) \right|} \quad \text{(Equation 4)}$$

where $|F_{obs}(h,k,l)|$ is derived from the measured intensity of the reflections and $|F_{calc}(h,k,l)|$ the amplitude of the structure factor calculated from the current model. However R_{work} can be artificially reduced through the introduction of too many

parameters, and therefore R_{free} was introduced for statistical cross validation. In principal R_{free} is calculated in the same way as R_{work}, with the exception that a small subset of reflections is used, set aside during the refinement (usually 5%).
R_{work} should not deviate highly from R_{free}.
Computer programs like COOT[88] facilitate the map interpretation (both manual and automated model building) and an example of a computer program that can be used for structure refinement is PHENIX[89].

2.2.8 Data quality and validation

With increasing resolution the diffraction intensity decreases and reflections beyond a certain resolution limit that cannot be distinguished from noise have to be excluded from the data set in order to maintain a good model quality. Traditionally R_{Sym}[90] a measure for the agreement between unique reflections with more than one observation, together with the signal-to-noise ratio ($<I/\sigma(I)>$) are used as criteria for data binning. Due to the dependence of R_{merge} on the redundancy, an alternative form, R_{meas} was suggested as a criterion instead[91] where the multiplicity of reflections is taken into account. However the correlation of these parameters with the final model quality was questioned and a correlation coefficient upon the division of the data in two parts $CC_{1/2}$ was suggested instead[92].
Finally the models root-mean-square (rms) deviations of bond-lengths and bond-angles from an accepted set of values, and the backbone conformational angles Φ and Ψ in accordance with the Ramachandran plot[93] are monitored for physico-chemical reasonable conformations.

2.3 Small angle X-ray scattering

Biological small angle X-ray scattering (SAXS) is a low resolution technique (50-10Å) to investigate macromolecules in solution. The information content that can be extracted from the resulting scattering curve of SAXS experiments includes folding, aggregation, shape, assembly and conformation of macromolecules in solution. Combined with high resolution structures like X-ray crystallography where models often are limited to certain trapped conformational states, the crystal lattice or incomplete complexes, or NMR that faces size limitations, SAXS provides complement biological relevant information. The following sections will focus on proteins, but the principles of biological SAXS can be applied to other biological macromolecules as well.

2.3.1 Scattering by molecules in solution

In contrast to the distinct Bragg reflections obtained in X-ray crystallography, the diffraction of macromolecules in solution is isotropic due to the randomly orientated distribution of particles. The magnitude of the resulting vector q between the incident and the scattered X-ray beam at 2θ (see Figure 9) can be deduced to:

$$q = \frac{4\pi \sin\theta}{\lambda} \quad \text{(Equation 5)}$$

with λ being the wavelength of the incident beam.

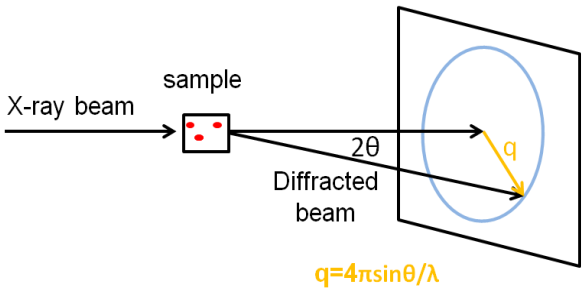

Figure 9. Schematic illustration of X-ray scattering by molecules in solution. The incident X-ray beam is scattered at 2θ of the incident beam in a radial symmetric manner, resulting in the scattering vector q.

X-ray scattering by proteins in solution can be considered as the scattering by assemblies of electrons. The total scattered amplitude is accordingly described by the sum of the scattered waves from all atom pairs in the ensemble:

$$A(q) = \sum_{i=1}^{N} b_i e^{(iq \bullet r_i)} \quad \text{(Equation 6)}$$

Where q is the scattering vector, r_i is the position and b_i the scattering factor of atom i. In order to access the scattering from the protein in solution, the scattering pattern of the solvent is subtracted from the one of the protein solution. For this purpose the scattering length density distribution of the protein solution ρ(r) and the solvent ρ_s are described as the scattering amplitude per volume and the difference $\Delta \rho(r) = \rho(r) - \rho_s$ is the excess scattering length density which is related to the scattering amplitude as follows:

$$A(q) = \int_V \Delta\rho(r) e^{(iqr)} dr \quad \text{(Equation 7)}$$

where V is the particle volume and r interatomic distances. The measured intensity is the product of the amplitude and its complex conjugate, $I(q)=A(q)A(q)^*$ or $\langle I(q)\rangle_\Omega = \langle A(q)A(q)^*\rangle_\Omega$ averaged over all orientations. For monodisperse protein solutions following solvent subtraction $\langle I(q)\rangle_\Omega$ is proportional to the scattering of a single particle averaged over all orientations.

2.3.2 Data collection and reduction

As with for X-ray crystallography, SAXS data is commonly collected at synchrotrons. Foremost the scattering of an empty cell and a standard protein of known concentration (most commonly BSA) are recorded. During the data collection 10 frames of buffer solution and 10 frames of protein solution (~1-10 mg/ml) are alternately collected under continuous flow in order to minimize radiation damage, as the exposure time that varies from 1-10s. The frames are averaged and noise

subtracted (solvent empty cell, background) and finally the scattering of the buffer is subtracted from the scattering of the protein solution. The resulting 2-dimensional data array is circularly averaged and reduced to a one dimensional plot of I(q) vs q. At the SAXS beamline at the ESRF sample uptake, data collection and data reduction are automatically performed in BsxCube.

2.3.3 Instant Sample characterization

Several overall shape parameters for the target protein can be extracted directly from the scattering curve like the radius of gyration (R_g) or the maximum particle diameter (D_{max}). The radius of gyration is the root-mean square of the distances of all electrons from their centre of gravity and thus a measure for the spatial extension of the protein. For an ideal protein the radius of gyration of the protein can be readily extracted from the forward scattering intensity at q=0, I(0) by means of the Guinier approximation[94] :

$$I(q) = I(0)e^{(-\frac{1}{3}R_g^2 q^2)} \quad \text{(Equation 8)}$$

I(0) can be obtained by the intercept of the y-axis of the linear region of the Guinier plot, where lnI(q) is plotted vs q^2. Because of the linear dependence of lnI(q) on q^2 that is not valid at higher q-ranges, the q-range for the estimation of the R_g in biological SAXS should not exceed $1.3/R_g$.
D_{max} can be estimated from the distance distribution function which is a Fourier transformation of the scattering intensity:

$$\rho(r) = \frac{r^2}{2\pi^2}\int_0^\infty q^2 I(q)\frac{\sin qr}{qr}dr \quad \text{(Equation 9)}$$

with r being the interatomic distances. However the computation of ρ(r) is not straight forward, due to the limited range of I(q) available through measurements. Instead an indirect Fourier transformation is applied assuming that ρ(r)=0 for r=0 and r>D_{max} as proposed by Glatter[95], where the ρ(r) function is approximated by a linear combination of a finite number of functions. The ρ(r) function and the R_g are automatically determined by the programs AUTOGNOM and AUTORG[96] implemented in Primus[97] a program in the ATSAS program package.

2.3.4 Theoretical calculated scattering curves from high resolution models

The calculation of theoretical scattering curves of high resolution protein structures and the subsequent comparison to the experimental scattering curve of the corresponding protein in solution is a powerful tool to elucidate divergences between high and low resolution structures. The program CRYSOL[98], part of the ATSAS program package calculates the theoretical scattering curve of high resolution structures with spherical harmonics approximations of the molecular envelope by

taking the scattering of the hydration shell into account. This is of importance, since the solvent scattering differs from the scattering of the hydration shell thus leading to an increase in the protein envelope[99,100]. The agreement between the theoretical calculated and the experimental scattering curve is evaluated through the normalized discrepancy function χ^2:

$$\chi^2 = \frac{1}{N_p - 1} \sum_{i=1}^{N_p} \left[\frac{I(qi)_{\exp} - cI(qi)_{calc}}{\sigma(qi)} \right]^2 \quad \text{(Equation 9)}$$

where N_p is the number of experimental data points, $I(qi)_{exp}$ and $I(qi)_{calc}$ the experimental, respectively calculated I(qi) scattering curve at q=i, c a scaling factor and σ(qi) the experimental error. A χ^2-value of 1 indicates that the theoretical and the calculated scattering curves do not differ.

2.3.5 Modeling with SAXS data

The amount of information that can be extracted from a SAXS experiment is insufficient to directly generate protein models. However in a process of reverse modeling different approaches are applied to reconstitute low resolution shapes with the aid of partial or complete high resolution structures or *ab initio*[101] (and references within). Structural flexible proteins that exist in several conformations in solution can be modeled through the ensemble optimization method (EOM)[102,103] which is implemented in the ATSAS program package. The principle of this method is that a pool of protein structures is generated based upon sequence and/ or structural information. From this pool ensembles of models with different conformations are optimized against the experimental SAXS scattering curve. The theoretical calculated scattering curve of the final model ensemble will have the least discrepancy to the experimental scattering curve, indicated by χ^2, similar to equation 9.

2.3.6 Data quality and validation

Since no universal quality criteria for SAXS raw data exists, it is of great importance to guarantee high sample quality during measurements. Because the excessive scattering length is obtained through solvent subtraction, the solvent must perfectly match the solvent of the protein solution in order to minimize experimental errors. This is achieved through the dialysis of the protein solution against the solvent. Furthermore the sample has to be monodisperse and sample aggregation has to be excluded through for instance filtering the sample prior to measurements. Sample aggregation can be validated with the Guinier plot, where positive or negative deviations from the linear region indicate molecular attraction or repulsion respectively. Furthermore scattering curves have to be investigated for possible radiation damage. Experimentally determined R_g and D_{max} values obtained from the Guinier approximation and the indirect ρ(r) function are most reliable for globular, inflexible proteins in solution.

2.4. Terahertz radiation

Terahertz (THz) radiation ranges from frequencies of 0.1 to 10 THz, thus lying at the interface between electronics (microwaves) and photonics (infrared) (Figure 10), and

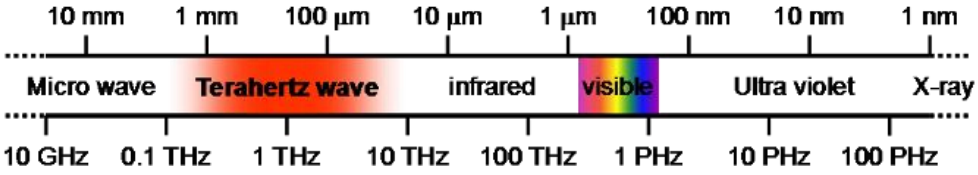

Figure 10. Spectrum of electromagnetic radiation. THz radiation is located in the spectral region within frequencies of 0.1-10 THz [149]

until for about two decades ago it was technically difficult to generate reliable, stable THz radiation at sufficient energies or to detect THz radiation which is generally referred to as the "Terahertz gap". Because THz radiation lies within the frequencies of low frequency internal protein motions[104] it is attractive to explore its application in structural biology.

2.4.1 Terahertz absorption spectroscopy

Terahertz absorption spectroscopy aims to probe vibrational modes of proteins. THz radiation covering a certain wavelength range can be generated at synchrotrons and detected with a bolometer, a thermal detector that senses impinging radiation through temperature dependant electrical resistances. THz radiation is strongly absorbed by water and order to minimize THz absorption by water vapor THz measurements are performed in a water free environment, as for instance a nitrogen purged chamber. Biomolecules absorb less THz radiation compared to water and consequently THz absorption of protein solutions decrease with increasing protein concentration, as shown for BSA and Lysozyme[105,106]. Nevertheless an excess THz absorption can be observed, if the contribution of water to the THz absorption in the buffer solution is considered[105,106]. Light activated proteins such as the photosynthetic reaction centre from *R. sphaeroides* can be activated upon laser illumination facilitating the measurement on activated protein as depicted in Figure 11. By alternating measurements of laser activated (on) and not activated protein (off), two different data sets that differ exclusively in the protein conformation itself can be recorded.

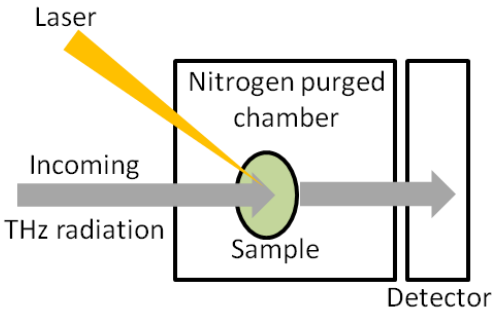

Figure 11. Schematic illustration of the THz absorption data collection. Every second absorption spectrum is recorded under simultaneous laser activation of the protein sample in a nitrogen purged chamber.

THz absorbance of THz radiation underlies Lambert- Beer law[104,106] and the according difference absorbance is accordingly:

$$A = -\log \frac{I_{light}}{I_{dark}} \quad \text{(Equation 10)}$$

Where A is the absorbance, I_{light} is the transmitted THz radiation of the illuminated protein sample and I_{dark} the transmitted THz radiation of the resting protein sample.

3. Results and Discussion

3.1 High resolution structure of S100A4 in complex with a non-muscle myosin IIA fragment (Paper I)

The underlying mechanism for metastasis involves a variety of cellular processes that are still not well understood. Understanding the molecular basis of these processes is a prerequisite for the design of therapeutic agents to control metastasis in affected cancer patients. A number of methods allow us to study molecular events, but it is often the structural information that adds the missing link to complete the whole picture.
S100A4 is strongly associated with metastatic processes. In previous studies it had been shown that S100A4 interacts with the C-terminal region of non-muscle myosin IIA (NMIIA) rods in a Ca^{2+}-dependant manner[107,27] and that NMIIA filament formation was negatively affected *in vitro* by this interaction[27,28,34]. Furthermore the interaction between S100A4 and NMIIA could directly be linked to increased cell motility in S100A4 expressing cells[32]. The structural mechanism underlying filament disassembly however was not known.
Despite several efforts to map the binding sequence and the binding stoichiometry of NMIIA to S100A4 a clear solution was not found. Minimal binding sequences, affinity constants and estimated binding stoichiometries differed amongst studies mainly due to the application of various methods with distinct sensitivities[27,26,108,28,109,110]. Even the Ca^{2+}-activated S100A4 X-ray structures[19,20,21] where the hydrophobic pockets between helix3 and helix4 are in the exposed mode, offer no obvious description of the minimal binding sequence of NMIIA and its binding stoichiometry.

3.1.1 Crystallization of the S100A4 complex

From several crystallization set ups with various S100A4 mutants and the wild type, crystals were only obtained from the quadruple mutant (C3S, C81S, C86S, F45W) in our hands. The cysteine to serine mutations were initially designed in order to prevent disulfide-bridge mediated oligomerization, however when working with S100A4 TCEP was always present in the buffer solution. The binding affinity to MPT was lower in the quadruple mutant, but not in the F45W mutant, and the lower binding affinity can most probably be attributed to the C81S mutation, since C81 has been shown to be critical for S100A4 binding[111]. Interestingly crystal contacts are observed between loop 2 and the MPT of an adjacent molecule in the close vicinity to the C81S mutation. This indicates, that the C81S mutation might provoke a conformational state of the MPT, possibly due to a less tight interaction reflected by the decreased binding affinity, that promote crystal contact formation in this region. Even though the binding affinity of the cysteine to serine mutant was reduced, the binding stoichiometry was found to be equally to the wild type, suggesting that the binding mode is similar. This could later be confirmed by a wild type NMR structure complexed to a non-muscle myosin fragment shortly after our structure was published[22].

3.1.2 Binding modes in the S100 protein family

The crystal structure of the Ca^{2+}-activated S100A4 mutant (F45W, C3S, C81S, C86S) in complex with the NMIIA fragment MPT was solved to 1.9 Å by molecular replacement. The high resolution structure shows that one NMIIA fragment spans one S100A4 dimer as seen in Figure 12, revealing a complete novel binding mode within the S100 protein family. Until this finding structures of related S100 proteins with ligands including α-helical binding motifs were known to bind in a 1:1 stoichiometry

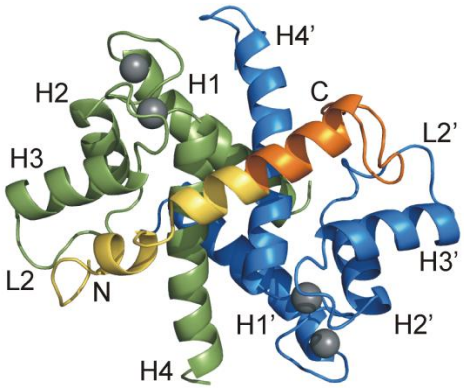

Figure 12. Overview of S100A4 in complex with the 45-residue-long NMIIA tail fragment along the twofold symmetry axis of the dimer. S100A4 subunit A and B are shown in green and blue respectively, while NMIIA peptide in yellow (residues 1893-1913) and orange (residues 1914-1935), and Ca^{2+}-ions are gray. The main secondary structural elements are indicated. (Figure from Paper I)

with one α-helical motif interacting with a single monomeric subunit of the S100 protein[112,113,114,115]. The resulting interaction network involving 36 amino acids (1893-1929) and overall 44 interactions between S100A4 and its ligand (calculated with the protein interaction server[116]) is considerably more extensive compared to the shorter, up to 23 amino acid long interaction sequence in related S100 protein complexes[112,113,114,115]. This may account for the increased binding affinity of S100A4 to one MPT, compared to binding affinities determined in the μM range for other S100 proteins and their ligands[114,117]. Whether this binding mode is unique for the interaction of S100A4 with MPT or whether binding to other targets such as p53 differs is not clear, yet.

3.1.3 Binding sequence and dynamical adaption to the ligand

NMIIA coiled-coils assemble into filaments through the interaction of negatively charged (glutamine 1722-asparagine 1756) with positively charged regions (alanine1868-lysine 1895)[118,119] as depicted in Figure 13, with the assembly competent domain (ACD, in blue in Figure13) being crucial for the formation of filaments[118,119]. Overall observed interactions between S100A4 and residues 1893-1929 of the NMIIA

Figure 13. Schematic illustration of the assembly of NMIIA coiled-coils into filaments. Negatively charged residues (Glu1722-asn1756) (*red*) interact with positvely charged residues (Ala1868-Lys1895) (*blue*) also called the assembly competent domain (ACD).

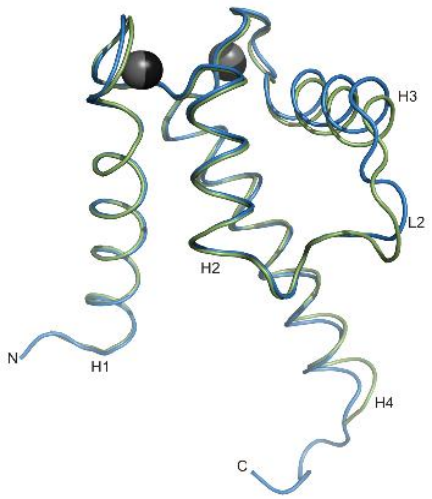

Figure 14. Ribbon representation of the least squares superposition of subunit A (*green*) and subunit B (*blue*) based on the rigid regions of subunit A and B as defined by ESCET[120].(Figure from PaperI).

fragment identify the binding sequence of S100A4 to non-muscle myosin to be part of the ACD and the random coil, resulting in a polar binding sequence with a higher portion of charged residues towards the N-terminus. From the architecture of S100A4 and the polar NMIIA fragment it might be expected that interactions of activated S100A4 are preferably established between hydrophobic regions of the MPT and the hydrophobic cleft, but instead the dimeric S100A4 adapts each subunit to the nature of its ligand, mainly through conformational flexibility of loop 2 and helix 3 as seen by the superposition of subunit A and subunit B after the calculated distance difference matrix (DDM) by ESCET[120] in Figure 14. This in turn allows each subunit to establish distinct interactions to the MPT and as a result hydrophilic interactions are mainly accumulated between one subunit of S100A4 and the N-terminus of the MPT (Figure 15A and left panel), and hydrophobic interactions to the other subunit S100A4 subunit close to the C-terminus of the MPT, as shown in Figure15B and right panel. This is a remarkable example of how identical subunits are able to form distinct characteristic interactions through conformational flexibility and could be the reason for the broad substrate specificity of S100A4. However the overall conformation of the Ca^{2+}-activated peptide-bound S100A4 is very similar to the Ca^{2+}-activated peptide-less S100A4 and the main conformational changes occur upon Ca^{2+}-activation.

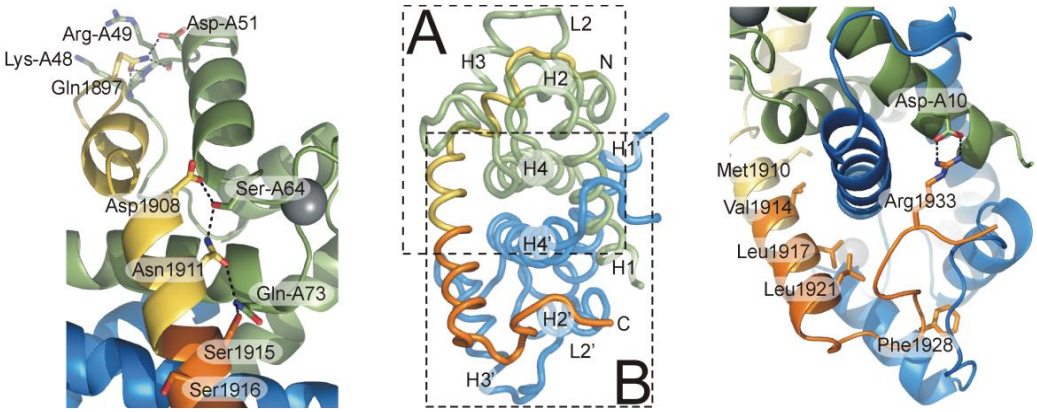

Figure 15. Key interactions within the S100A4-NMIIA MPT complex. (A) Polar interactions at the interface of the N-terminal part of the MPT with subunit A. (B) Interactions of MPT with subunit B are dominated by hydrophobic interactions, where Met 1910, Val1914, Leu1917 and Leu1921 (*a* and *d* positions in the coiled-coil) face inward. Phe1928 is located in a hydrophobic cavity formed by loop 2 (residues 41-51) and helix 3, while Arg 1933 forms ionic interaction with Asp10 of subunit A wrapping around the dimer. The overview in the middle is shown in an orientation perpendicular to the twofold symmetry axis of the S100A4 dimer. (Figure from PaperI).

3.1.4 Structural mechanism for filament disassembly

Thermostability measurements on filament forming NMIIA fragments (Ser1712-Glu1960 and Gln1795-Gln1960) in the absence and presence of S100A4 show that two S100A4 bound to one filament might unwind the coiled-coil to a certain extent. Together with the observation that the ACD and the unstructured region are part of the binding sequence of NMIIA to S100A4 it can be assumed that the disruption of the ACD by S100A4 is mainly responsible for the disassembly of filaments. This could be in combination with bound S100A4 being a spatial hindrance for NMIIA coils to approach each other.

3.1.5 Summary

The structure of the S100A4 mutant (F45W, C3S, C81S, C86S) in complex with a NMIIA fragment was the first reported high resolution structure of a S100A4 protein with a bound peptide derived from a natural interaction partner. Its binding mode is clearly distinct from that shown for other S100 protein family members so far considering the binding stoichiometry and the unusual asymmetric distribution of hydrophobic and hydrophilic between different dimer subunits. The binding region of S100A4 to NMIIA can be assigned to residue 1893-1929, partly coinciding with residues of the ACD of NMIIA that are unwound to a certain extent through the interaction, leading to the disassembly of filaments. The high resolution structure of S100A4 in complex with MPT contributed to our understanding of the binding mechanism to NMIIA and shows the importance of conformational flexibility in the adaption of S100A4 to distinct ligand properties. The conformational flexibility of loop 2 and helix 3 might be the key for the diversity of binding partners to S100A4, but it is not known, yet if the binding mechanism to other ligands resembles the one detected between S100A4 and the NMIIA fragment, or if other ligands bind in a distinct way.

3.2 The role of the C-terminal region of S100A4 (Paper II)

Amongst the S100 protein family loop 2 (situated between helix 2 and helix 3), the N-terminus and especially the C-terminus are sequentially the least conserved regions. A role of the C-terminus of S100A4, that had been shown to become flexible upon Ca^{2+}-binding[121,122] as dynamic surface for target interaction was proposed upon the observation that the helical content at the proximal part of the C-terminus increased[121,122,18]. Zhang *et al*[123] and Ismail *et al*[124] investigated S100A4 C-terminal deletion mutants and found that the C-terminal region was important for the metastatic properties in animal models and migration and invasion in cell lines. By investigation several C-terminal deleted S100A4 mutants it was concluded that it might be the lysine residues 100 and 101 at the tip of the C-terminus that are mainly contributing to the metastatic promoting properties of S100A4. A high resolution structure of a Ca^{2+}-activated S100A4 with the last 8 C-terminal residues removed was reported[125] , but until our, work structural studies have not been performed on a S100A4 deletion mutant with the last 13 C-terminal residues removed in complex to the NMIIA fragment MPT.

3.2.1 Crystallization of the C-terminal deletion S100A4 mutant (Δ13, C3S, C81S, C86S) in complex with a non-muscle myosin IIA fragment

As for the full length construct, no protein crystals of the C-terminal deletion S100A4 mutant in complex with MPT were obtained, but for a triple mutant (C3S, C81S, C86S), which will be denoted Δ13Ser. Δ13Ser crystallized in the space group P1, which requires data collection over an oscillation range of 180°, and in order to obtain a complete data set in our case data from 3 crystals had to be merged. The structure of the Δ13Ser mutant could be solved to 1.4 Å and in addition a further X-ray structure of the S100A4 mutant (F45W, C3S, C81S, C86S) could be solved to 1.4 Å (Figure16), which will be denoted F45WSer. F45WSer crystallized in the same space group as before (Paper I), but under different crystallization conditions.

3.2.2 Comparison of the Δ13Ser and the F45WSer -MPT complexes

Figure 16A and B show overviews of the Δ13Ser and the F45WSer complexes. The overall architecture of the Δ13Ser complex is similar to the one of the F45WSer complex and in addition no interactions can be detected in the C-terminal region of the F45WSer complex to the NMIIA fragment beyond residue 84 as calculated by the protein interaction server[116]. The least square superposition of Subunit A, Subunit B and the MPT as calculated by ESCET[120] reveal that conformational differences between this structures are mainly found in loop 2, helix 3 and the tip of the C-terminus, as the distance difference plot and the superposition of the subunits in Figure16 C-G reveal. Parts of the N-terminus and the C-terminus of the bound MPT are not well defined in the X-ray structure and in addition the conformation of helix 3 does not differ significantly in between the two subunits, as observed in the F45WSer-MPT complex. In addition affinity measurements with the MPT and filament disassembly assays with a dimeric NMIIA rod fragment (1712Q-1960E) did not reveal substantial binding differences or impaired ability to disassemble filaments between

the Δ13 and the wild type. These results suggest that the Δ13Ser-MPT complex represents an alternative conformation that might arise due to the C-terminal truncation and crystal contact differences. For instance helix 3 is involved in the formation of crystal contacts to the central part of the bound MPT in the Δ13Ser complex, and this might influence its conformation.

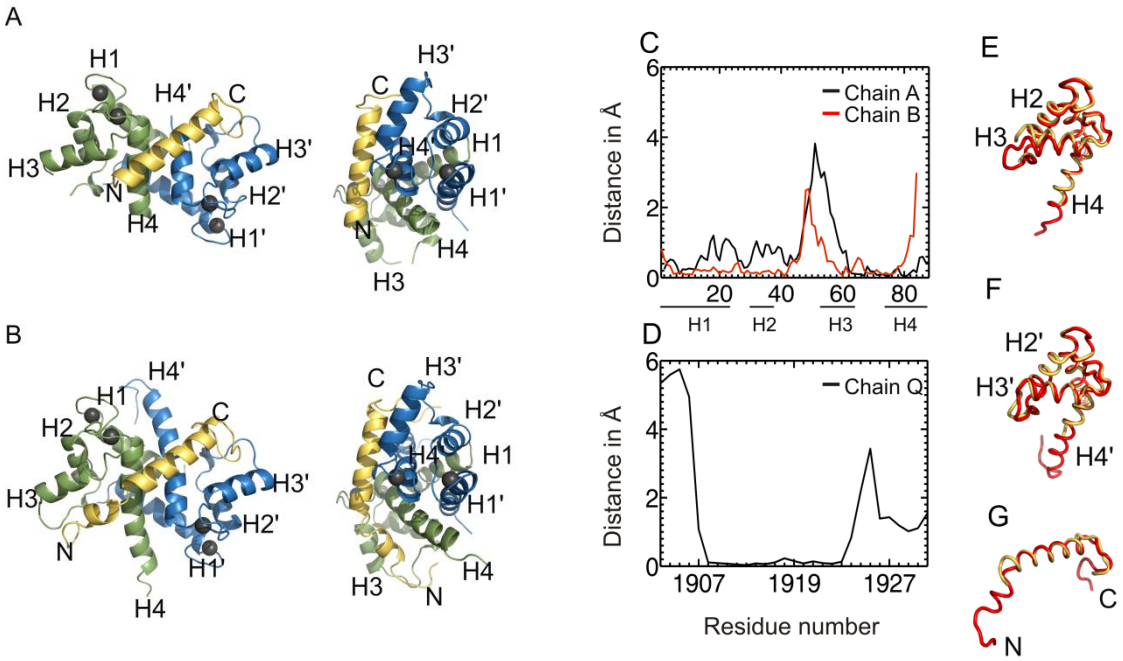

Figure 16. Structure comparison of the D13Ser and the F45WSer in complex with MPT. Three dimensional structure of the Ca^{2+}-activated, MPT-bound Δ13Ser (A) and F45WSer (B) S100A4. Subunit A is shown in *green*, subunit B in *blue* and the bound MPT in *yellow*. Helices (H) and the N-and C-terminus (N, C) of the bound peptide are indicated. Distance differences in chain A and B (C) and the bound MPT (D) between the Δ3Ser and the F45WSer complex and superposition of the subunits A, B and the MPT in (E), (F) and (G), with the subunits of the Δ13Ser shown in *yellow* and the F45Wser shown in *red*. (Figures from Paper II).

3.2.3 Conformational changes in the low resolution solution structure of S100A4 Wild type upon Ca^{2+}-binding

From SAXS measurements of wild type and Δ13 S100A4 in complex with MPT we were not able to assign a role to the C-terminus, since differences in the SAXS curves were rather diffuse. However changes in the SAXS scattering curve in the wild type at 0.15-0.25 reciprocal Ångström could be observed upon Ca^{2+} -binding, that were not present in the Δ13 mutant, as seen in Figure 17A and B. SAXS is a low resolution method and changes in the SAXS region indicate some form of shape remodeling, but changes in the size of wild type S100A4 upon Ca^{2+}-binding were initially difficult to assign. The reason was that on one hand the flexible C-terminus hampered trials to identify the maximum distance within molecules (D_{max}) confidently with the ρ(r) function (see section 2.3.3) and the radius of gyration (R_g) on the other hand differed only slightly. The R_g represents the root-mean-square distance of the centre of mass and may be applied for the size-estimation of globular proteins but it is a poor measure of the average size of elongated proteins. In addition by the instant determination of

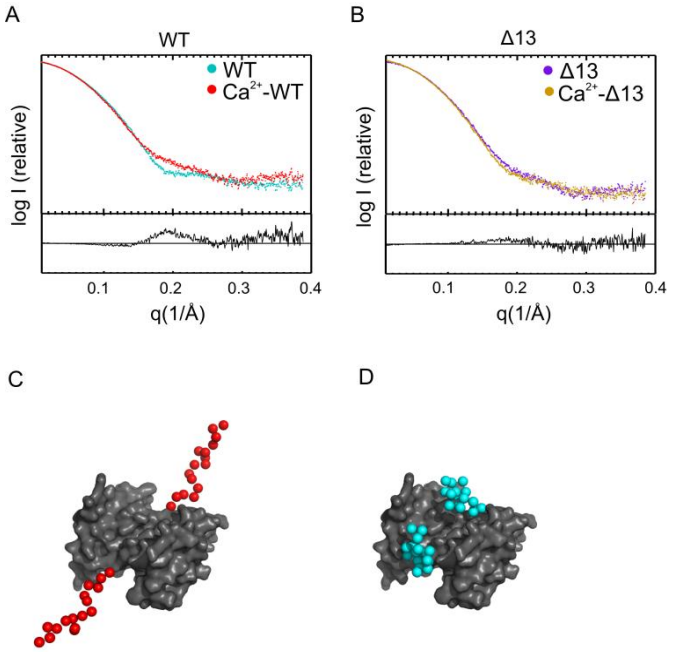

Figure 17. Scattering curves and low resolution models of Ca^{2+}-free and Ca^{2+}-bound S100A4. Differences in the WT (*cyan*) and Ca^{2+}-bound WT (*red*) are mainly observed at 0.15 $Å^{-1}$-0.25$Å^{-1}$(A). The scattering profile of the Δ13 mutant (*purple*) changes less upon Ca^{2+}-binding (*orange*) (B). Typical examples of EOM[101] extended (C) and compact (D) S100A4 models. (Figures from PaperII).

the R_g from SAXS scattering curves the nonlinear data at higher scattering angles is excluded and R_g values might be inaccurate. Therefore the hydrodynamic radius (R_h) was determined by Diffusion NMR and a significant change of the R_h in the wild type upon Ca^{2+}-binding was found (from 25.6±0.4Å to 33.1±1.6Å) but not in the Δ13 mutant. This indicates that the average radius of wild type but not Δ13 S100A4 increases upon Ca^{2+}-binding. Affinity measurements in the presence and absence of Ca^{2+} on MPT binding by wild type and the Δ13 mutant confirm that the conformational change in the core of S100A4 to open up a hydrophobic cleft[19,126] is most probably intact in the Δ13 mutant, as binding to MPT is only observed in the presence of available Ca^{2+}. This means that size changes in the wild type may relate to the C-terminus. A pool of models generated with EOM[102,103] with constant S100A4 core regions and flexible C-termini optimized against the SAXS scattering curves demonstrates clearly that the overall structure of the Ca^{2+}-less wt is rather compact whereas the C-terminus adopts an extremely extended conformation in the Ca^{2+}-bound form as shown in Figure 17C and D. At this stage the cause of the distinct conformations was not obvious.

3.2.3 MD simulations reveal the cause for the distinct C-terminus conformations

From high resolution structures of S100A4[18] and its Ca^{2+}-activated form[20,19,21] different C-terminal conformations could be observed, but the C-terminus is generally not well defined and an explanation for the distinct conformations could not be derived from these structures. In addition, the only high resolution structure of a Ca^{2+}-bound S100A4 with a well defined C-terminus is a NMR structure of a S100A4 complexed to a NMIIA fragment[22]. From SAXS measurements it was observed that the C-terminus adapts an elongated conformation in the presence of Ca^{2+} but the role of this elongated conformation in metastasis promoting properties of S100A4 was still unclear. In order to gain insights in the dynamics of S100A4 molecular dynamics (MD) simulations were performed. MD simulations from energy minimized Ca^{2+}-free and Ca^{2+}-bound S100A4 over 100ns in time steps of 2fs confirmed the observation, that the C-terminus of wild type S100A4 becomes extremely elongated in the presence of Ca^{2+}, whereas in the Ca^{2+}-less form the C-terminus is close to the core region, as depicted in Figure 18A

and B. In addition the MD simulations reveal that the positively charged tip of the C-terminus interacts with the negatively charged Ca^{2+}-binding site (in cyan in Figure 18). From this observation it was concluded, that the interaction of the C-terminus might interfere with the Ca^{2+}-binding, and this is indeed the case as confirmed by ITC measurements that show that the Ca^{2+}-affinity of the high-affinity Ca^{2+}-binding site is 40 times increased in the Δ13 mutant compared to the wild type. Ismail *et al*[124] emphasized the importance of the last basic amino acids of the C-terminus in the role of the metastasis promoting properties of S100A4 and the binding affinity to a NMIIA fragment (the last 149 C-terminal residues) but at this time point the role could not be specified. In MD simulations were the last 3 residues of the C-terminal tip are mutated to alanines, the calculated R_g of S100A4 in the absence of Ca^{2+} (Figure 18C) is more dynamic than in the presence of Ca^{2+} (Figure 18D) and it seems that the C-terminus is interacting with the hydrophobic cleft in the Ca^{2+}-bound S100A4. A rational explanation for the decreased binding affinity of the NMIIA fragment to S100A4 in the absence of the last basic C-terminal residues is that its interaction with the hydrophobic cleft interferes with the binding of the NMIIA fragment in a similar way, as the positively charged residues interfere with the Ca^{2+}-binding in the full length wild type, to reduce the affinity for the ligand.

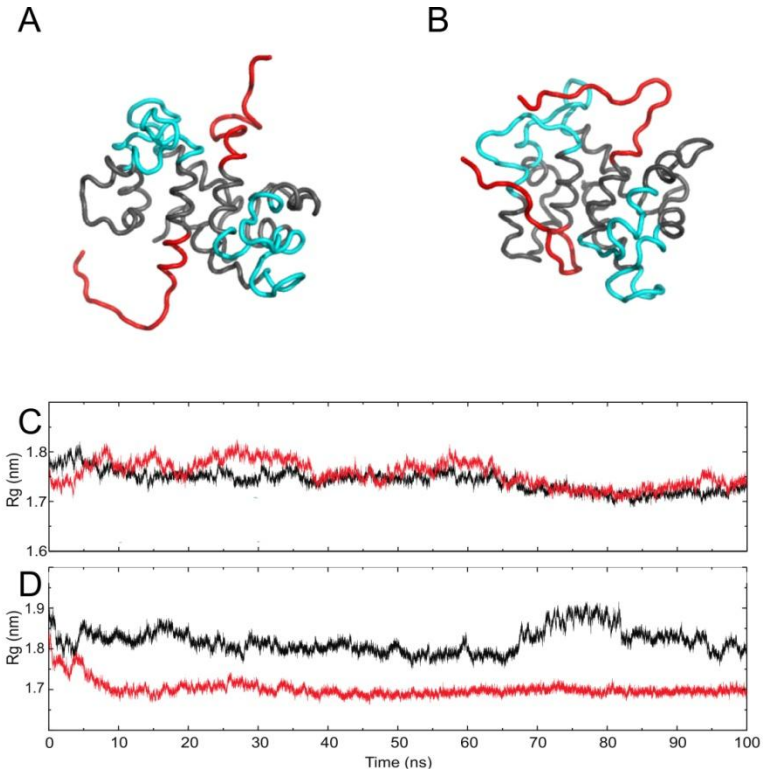

Figure 18. MD simulations of Ca^{2+}-bound and Ca^{2+}-free S100A4. The C-terminus of Ca^{2+}-bound S100A4 becomes elongated (A), whereas Ca^{2+}-free S100A4 is more compact (B). The C-terminal residues are depicted in *red* and the negatively charged EF-hands in *cyan*. The calculated R_g of Ca^{2+}-free WT (*black*) and AAA construct (*red*) (A) and the same constructs in the Ca^{2+}-bound form (B) during the MD simulation. (Figures from Paper II).

3.2.4 Summary

In previous studies it was observed that the C-terminal region had an impact on NMIIA binding and the metastatic promoting properties of S100A4, however no biochemical or structural description of the mechanism had been provided. The C-terminus had been observed to become elongated upon Ca^{2+}-binding amongst others by NMR studies and it was proposed that the C-terminal conformation might interact with the ligand binding. This work shows how the combination of different high – and low resolution and computational methods were gradually guiding us towards the hypothesis, that the C-terminus does not interfere with the ligand, but rather with the Ca^{2+}-binding. And as a result we can propose that the impaired function of C-terminus deleted (Δ13) S100A4 in metastasis might be related to the disruption of its Ca^{2+}-dependant regulation due to increased Ca^{2+}-affinity. In addition it is concluded that the charged residues at the C-terminal tip might prevent the C-terminus to bind to the hydrophobic cleft to cause autoinhibition. Whether the direct interactions between the C-terminus and its natural full length NMIIA ligand or other interaction partners influence metastatic properties or interferes with other regulatory processes as for instance phosphorylation events has to be investigated further.

3.3 THz absorption spectroscopy (Paper III)

Occurring within a time frame of ~200ps, the forward transport of electrons in photosynthetic reaction centre from the light activated special pair (P^*) to form the charge separated state P^+Q_A via P^+BPhe^-[127] is much faster than the back reaction to the ground state P that lies within a time-scale of ~100ms[128]. Based on theoretical calculations it was hypothesized that energetically both the forward and the back reaction are supposed to occur on a faster time-scale, and that the reaction rates in the primary charge separation process might be influenced by conformational changes in photosynthetic reaction centre[36]. Further studies on the temperature dependence of the electron transport also suggest conformational dynamics in the electron transport[129] and in addition conformational changes in subunit H could be observed upon light activation of photosynthetic reaction centre[130]. Finally low frequency modes, of 15 cm^{-1} and 77 cm^{-1} in reaction centre of *Rhodobacter capsulatus* were associated with the charge separated state[131]. However photosynthetic reaction centre dynamics during the electron transport is still not well understood. In this study it was explored whether changes in the vibrational dynamics of photosynthetic reaction centre are induced upon light activation and if so, whether they can be monitored by THz absorption spectroscopy, in the 0-4THz range.

3.3.1 Difference absorption spectroscopy

Because THz radiation is highly absorbed by water, THz absorption spectroscopy measurements of proteins were initially performed on protein films[132,133]. In the first THZ absorption spectroscopy measurements of BSA and lysozyme that were performed in aqueous solution[105,106] a solvent baseline was calculated through the estimation of the water amount in the protein sample. However this method might introduce uncertainties in the results. Photosynthetic reaction centre can be reversibly activated by laser pulses and this has the advantage that the activated and the resting state can be investigated on a single protein sample and subtraction of the two states yield a difference spectrum. In the difference spectrum solvent effects that are not directly related to the protein activation will be subtracted out. This concept has been successfully applied in THz absorption spectroscopy of photoactive yellow protein[134].

3.3.2 Influence of the environment in the activation of photosynthetic reaction centre

THz absorption on photosynthetic reaction centre of *R. sphaeroides* was measured in detergent containing buffer (Figure 19A) and lipidic sponge phase (Figure 19B). The difference absorption spectra for photosynthetic reaction centre measured at different concentrations in detergent containing buffer shows an overall increase in the 20-130cm^{-1} range with an extended peak in the 90-100cm^{-1} range as seen in Figure 19A (two upper lines). A similar difference absorption spectrum is obtained for reaction centre in lipidic sponge phase, as shown in Figure 19B (upper line), and upon scaling and superposition of the absorption spectrum measured in lipidic sponge phase to the one measured in detergent no significant differences can be observed. This might indicate that the vibrations observed could be independent of the environment, and for

instance include the reorganization of bondings that are not solvent slaved[135]. However this has still to be proven, since the solvent dynamics of detergent based solutions or lipidic sponge phase have not been investigated by THz absorption, yet.

3.3.3 Localization of vibrational changes to the LM subunit

In order to investigate whether vibrational changes can be attributed to the cytosolic H- subunit or the intramembrane L and M units (LM) THz difference absorption of

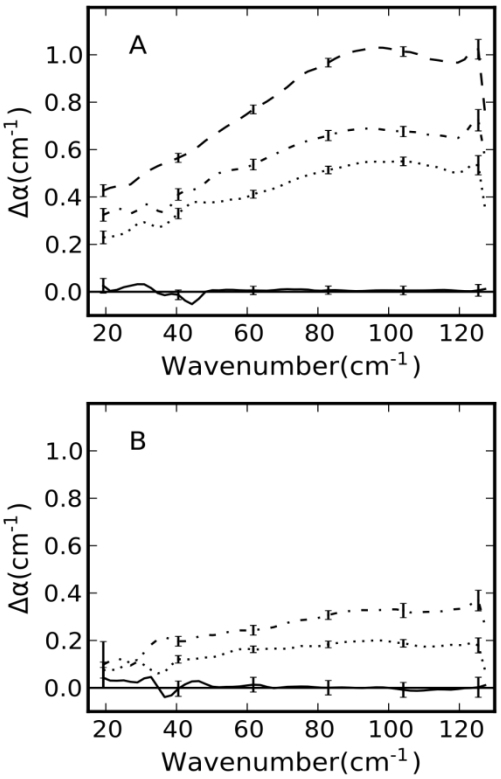

Figure 19. Difference in THz absorption ($\alpha_{illuminated}$- α_{ground}) of RC_{sph} and LM_{sph} in detergent (A) and sponge phase (B) illuminated with laser intensity 7.7 kW m^{-2}. The error bars are representing the standard error ($\sigma/\sqrt{N}$). In panel A the two upper spectra are recorded from two different concentrations of RC_{sph}, 1.1 mM (dashed line), an average of 86 different absorbance spectra and 0.76 mM (dash-dot line), an average of 90 difference absorbance spectra. The spectrum represented as the dotted line shows the difference absorption of the LM_{sph} complex collected from two different samples with a concentration of 0.71 mM, in total 87 difference absorbance spectra. The spectrum represented by a solid line shows the difference absorbance for detergent based buffer, an average of 59 difference absorbance spectra. In pane B the two upper spectra show the difference absorbance for RC_{sph} (0.76 mM) mixed 1:1 with sponge phase, the spectrum represented as a dash dot line is collected from a sample mixed directly prior to measurement (average of 114 spectra) while the spectrum represented as a dotted line is collected from a sample that was incubated for 14 days in RT (an average of 92 spectra). The spectrum represented as a solid line is collected from pure sponge phase, an average of 42 difference absorbance spectra.

LM (here designated RC_{LM}) upon light activation was recorded. The difference absorption spectrum of RC_{LM} in detergent based bufferis shown in Figure 19A. Again, the difference absorption spectrum rises steadily and peaks at 90-100cm^{-1}, in accordance with the difference spectra of the photosynthetic reaction centre measured in detergent buffer and lipidic sponge phase. The equivalent absorbance difference spectra indicate that vibrational changes observed in the THz range measured are located to the LM subunit. The LM unit is highly conserved, and hosts the cofactors. Therefore it is not unlikely that molecular vibrations upon the excitation of photosynthetic reaction centre are concentrated o the LM subunit. It has been previously shown that the charge separation in isolated LM subunits was not reduced compared to the intact reaction centre[136], however binding to ubiquinone was weaker, and the rate of the back reaction slower[136].

Computational normal modes derived from a crystal structure of reaction centre (PDB code 2BNP) with different force constants indicate a similar normal mode distribution as the experimental difference absorption spectra. But it was not possible to assign the measured absorption differences to detailed molecular features.

3.3.4 Temperature effects

The activation of reaction centre with a laser during the data acquisition will rise the sample temperature slightly. Therefore datasets were recorded at different temperatures (25° and 26°). The difference absorption of the temperature profile is distinct from the difference profile of the activated protein as seen in Figure 20, both in

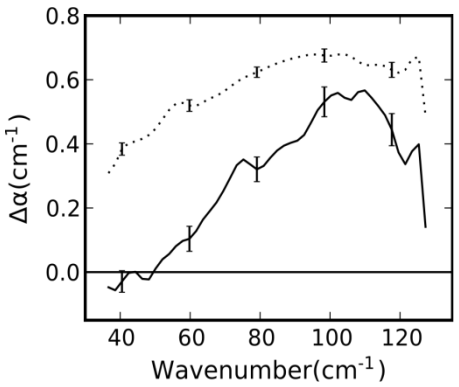

Figure 20. The difference absorption between spectra measured at sample temperatures of 25°C and 26°C (solid line). 10 spectra were used from each temperature when calculating the difference absorbance and the sample used was 0.76 mM RC_{sph} in detergent. This 1°C temperature difference is about ten times higher than the estimated temperature increase upon laser illumination. The difference absorbance spectra for the 0.76 mM RC_{sph} sample illuminated with a 7.7 kW m^{-2} laser is shown as a dotted line for comparison. The error bars are representing the standard error ($\sigma/\sqrt{N}$).

intensity and shape, This indicates that the difference absorption spectrum of reaction centre upon light activation includes non-thermally derived vibrational modes.

3.3.5 Summary

With this experiment it was proven, that non-thermally induced changes in the vibration modes of reaction centre upon light activation can be detected in the 0-4 THz region. The similarity between the reaction centre and the LM difference spectra indicate that vibrational changes might mainly occur in the LM subunit. Furthermore the vibrational changes are most probably not dependant on the solvent environment.
However the featureless shape of the difference spectrum arises from a number of changes in different vibrational modes and it is still difficult to extract key motions.
Interestingly THz absorption spectra of different proteins could be measured and are sensitive to the environment and protein species as several measurements show[132,105,106,133,134,137] and therefore this technique could potentially be explored for the ability to provide more detailed information in combination with computational methods.

3.4. Surface engineering and purification of UbiA and MenA from *E. coli* for crystallization (Manuscript)

When this project was started, no experimental structural information of any member of the UbiA superfamily was available and functional studies had only been performed on extracted membranes containing these proteins[138,44]. The deduction of structural information based on previous knowledge from related proteins was difficult because the functional related soluble aromatic prenylating enzymes differ sequentially and structurally from the intramembrane prenyltransferases of the UbiA superfamily[139]. Furthermore the sequence identity to the closest sequentially related proteins, 5-Epi-aristolochene synthase (*Nicotiana tabacum*), Geranyl-geranyltransferase (*Rattus Norvegicus*) and photosynthetic reaction centre (*Rhodopseudomonas viridis*) is less than 25%[140]. On the basis of two recently solved apo and substrate bound X-ray structures of UbiA homologues from hyperthermophilic organisms[55,56], the position of interacting active site residues could be identified and a reaction mechanism was proposed, where the prenylation of 1,2-dihydroxy-2-naphtoate is accomplished by an ionization-condensation-elimination mechanism, through the interaction of a charged amino acid side chain (here lysine) to the ionized oxygen of the pyrophosphate group of the isoprenoiddiphosphate[56]. However in one organism (*A. pernix*) the active site residues were not directly enzymatically determined, but derived from experimentally determined *E. coli* UbiA active site residues through conserved residues in sequence alignments[55] and in the second organism (*A. fulgidus*) the X-ray structures are lacking the quinone substrate[56]. In addition the natural substrates of these UbiA homologues are not known, and might differ from the natural substrates of UbiA and MenA. Therefore additional structural information will contribute to a better understanding of these enzymes.

3.4.1 Rational approach in the stabilization of entropic surfaces

The importance of the intrinsic protein properties in successful crystallization has long been recognized and various methods like homology screening, directed mutations or the construction of fusion proteins are frequently applied in membrane protein crystallization[141,142]. In this work a library of rational designed *E. coli* UbiA and MenA clones was constructed with the aim to stabilize structurally flexible regions to promote the formation of crystal contacts. In addition UbiA had been shown to oligomerize in a DTT-dependant manner and in an attempt to abolish cysteine-mediated oligomerization by disulfide bridge formation, cysteine residues were mutated to serine, a strategy that has been successfully applied in the crystallization of the S100A4 complex. The theoretical analysis of the mutation sites by means of the corresponding homology models indicate that the stabilized regions are predicted to be situated on exposed loops that are indeed involved in the crystal contact formation in the crystal structures of the two UbiA homologues from hyperthermophilic organisms[55,56], which raises the probability of the mutations to interfere with the formation of crystal contacts.

3.4.2 Conformational flexibility in membrane proteins

In their natural environment membrane proteins are surrounded by often specific lipids that stabilize the protein structure through lateral pressure and hydrophobic head groups[143]. Upon the displacement of the lipid bilayer by detergents in the protein extraction process, membrane proteins gain more conformational freedom because the lateral pressure is released and stabilizing lipids are often removed and this may result in temporary unfolding processes and heterogeneous interactions[144]. In addition poor protein detergent complexes may interact with micelles resulting in increased heterodispersity[144]. Conformationally, membrane proteins can be stabilized by certain detergents, buffers and additives, and the stability in a given condition can be monitored by SEC or through thermostability assays[145]. For this reason in addition to stabilizing exposed surfaces, suitable buffer conditions and additives have to be explored in order to keep the membrane protein itself conformational stable.

3.4.3 Recombinant overexpression and IMAC purification

Although both UbiA and MenA are members of the same protein family and were homologous overexpressed differences in the yield and the behavior during the recombinant overexpression were observed. For UbiA an autoinduction protocol could be established yielding high amounts of protein. In contrast MenA was only successfully overexpressed upon IPTG induction and cells had to be harvested shortly after the induction in order to obtain protein, although both proteins are classified as stable in *E. coli in vivo*[146]. However it is not clear whether the cell growth was inhibited due to the accumulation of missfolded protein, since recombinant expressed protein was found in the membrane. MenA might compete with substrates for ubiquinone biosynthesis and thus influence the viability of the cells by causing ubiquinone deficiency. Similar variations in the recombinant overexpression of various mutants did not arise within one protein species. Differences were also observed in the behavior upon IMAC purification and the detergents needed for the extraction of the proteins. MenA was co-purified with impurities and showed a tendency to aggregate in contrast to UbiA that had a higher degree of purity upon IMAC purification. This indicates that MenA is more instable in the purification buffer than UbiA.

3.4.4 Size exclusion chromatography profiles

The SEC profiles of four UbiA mutants and two MenA mutants were similar to the wild type SEC profiles under equally buffer conditions, as shown for UbiA in Figure 21A-C and MenA in Figure 21D, except for two mutants that were highly polydisperse as seen by the appearance of several protein peeks in Figure 21D. This indicates that the proteins are conformational unstable under the buffer condition tested and aggregation and heterogeneous association occurs. The success in crystallization is negatively influenced by polydispersity[67] and in addition the protein might not preferably exist in a functional conformation. For this reason different detergent and buffer conditions were tested for their ability to improve the quality of the protein solution. The stability of UbiA could be improved in Na-phosphate and FC-12 as the decreased aggregation peak in the SEC profile in Figure 21E indicates. Similar results

were obtained for the wild type MenA in Mes buffer and LDAO as shown in Figure 21F. These results suggest that the mutants can be stabilized in different buffer conditions similar to the wild type. The detergent LDAO is statistically well represented in successfully solved X-ray structures of membrane proteins[147] in contrast to FC-12, that is a harsh detergent that might even have denaturing effects on some proteins. In our hands no enzymatic activity could be detected of UbiA upon the

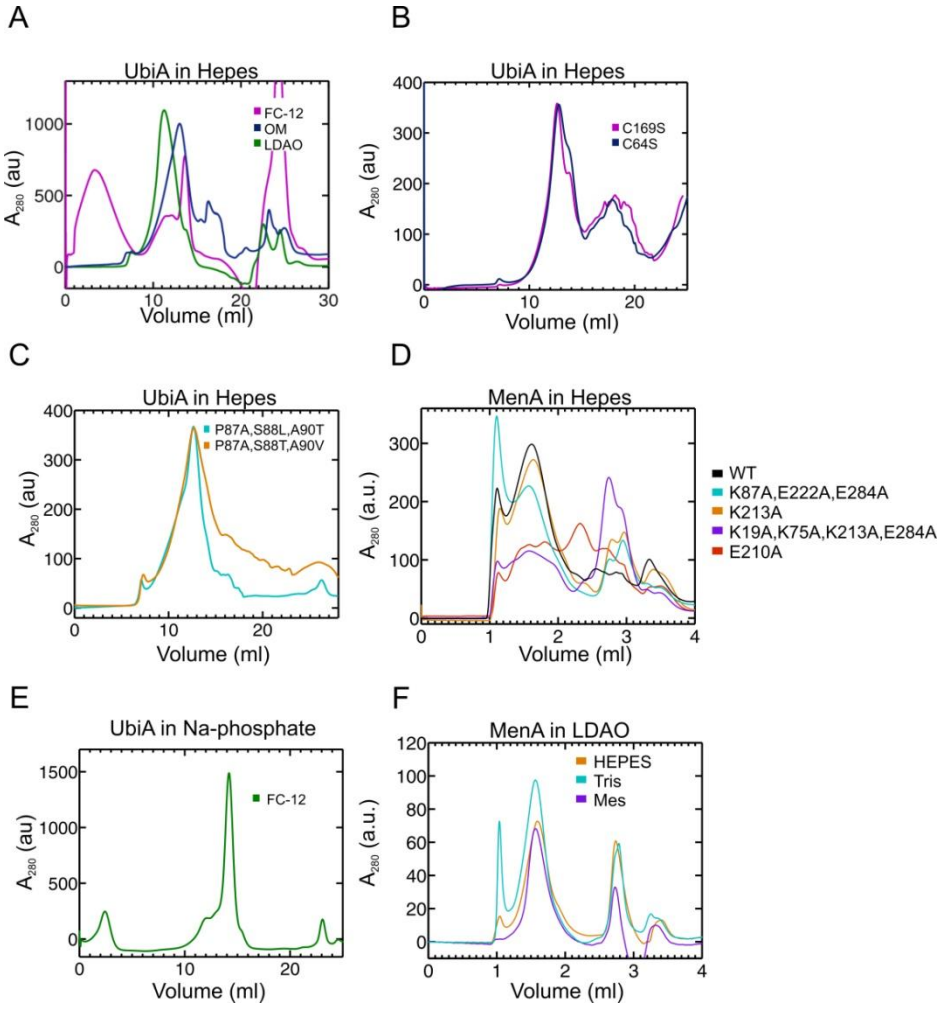

Figure 21. SEC profiles of UbiA and MenA WT and mutants. (A) UbiA in Hepes buffer and the detergentsFC-12 (*magenta*), OM (*blue*) and LDAO (*green*). (B) Mutated UbiA, C169S (*magenta*) and C64S (*blue*) in Hepes. (C) Mutated UbiA, (P87A, S88L, A90T) (*cyan*) and (P87A, S88T, A90V) (*orange*) . (D) MenA WT (*black*) and mutated, (K87A, E222A, E284A) (*cyan*), (K213A) (*orange*), (K75A, E273A, E284A) (*purple*) and (E210A) (*red*). (E) UbiA WT in Na-phosphate buffer and FC-12. (F) MenA WT in LDAO and Hepes (*orange*), Tris (*cyan*) and Mes (*purple*).

purification in Hepes and Tris buffer in FC-12. This could be due to the buffer or detergent conditions, but it cannot be excluded that the secondary structure is disrupted. Therefore the presence of secondary structure should be monitored by for instance circular dichroism (CD). Furthermore the reconstitution into lipids for activity measurements and crystallization methods such as cubic phase[76], lipidic

sponge phase[77], or bicelle[78] crystallization might reduce the influence of the detergent on the protein.

3.4.5 Summary

In this work a library of seven UbiA mutants and nine MenA mutants that were designed for reduced exposed surface entropy properties and cysteine mediated oligomerization was successfully cloned. From this library four mutants each could be recombinant overexpressed and purified by IMAC and SEC. However polydispersity observed in all mutants under the purification condition tested suggests that the proteins and the corresponding mutants have to be stabilized additional by different buffer, detergents and additives. An improvement of the protein stability could already be achieved for the wild type proteins by testing different buffers and detergents, and this indicates that the stability of the mutants could be improved similarly.

4. Conclusions

The conception of proteins being rigid molecules with a predefined structure that recognizes ligands according to the lock and key concept as initially proposed by Fisher has dramatically changed over the last decades and instead the idea that proteins exist in conformational substates in a dynamic energy landscape[5,148] is increasingly gaining attention. Even if structural information is important for the understanding of the protein function, a single three dimensional structure does not provide sufficient information about the protein function and therefore the information of different structures and methods has to be combined in order to gain a full description of a protein.

The high resolution structure of S100A4 in complex with a NMIIA fragment emphasizes the importance of structural information for the understanding of the protein function. Due to a distinct binding mechanism, compared to closely related S100 proteins and the dynamical adaption to its ligand the interaction of S100A4 to NMIIA could not be confidently derived, despite extensive biochemical studies on S100A4. In addition the work on S100A4 also reflects the importance of applying different structures and methods in the understanding of the protein function. Where high resolution structures alone could not deliver an explanation for the role of the C-terminus, the combination with low resolution and computational methods finally guided us towards the role in the Ca^{2+}-binding properties of S100A4. In addition to the role of conformational changes upon Ca^{2+}-binding that serve to activate S100A4 we could identify the importance of the C-terminus dynamics in Ca^{2+}-binding and dynamics of loop 2 and helix 3 in ligand binding, that rather represent conformational isomers of the Ca^{2+}-activated S100A4. The structural and dynamical studies on S100A4 and its interaction with a peptide derived from NMIIA contributed significantly to the basic understanding of the mechanism of S100A4 mediated NMIIA filament disassembly. Nevertheless structural studies of the interaction of S100A4 with NMIIA were performed on a peptide derived from NMIIA, which differs significantly from the full length natural NMIIA, and therefore the interaction with the natural substrate might be more complex. In addition S100A4-mediated filament disassembly might be accompanied with regulatory processes that are not reflected in our structural studies.

The role and nature of conformational changes in the electron transport in photosynthetic reaction centre has been investigated with different structural and biochemical methods, and there is evidence for conformational flexibility. However it is still difficult to probe vibrational modes experimentally and there is a need for additional methods that will facilitate dynamical investigation of proteins and the role and the nature of dynamics in the electron transport in photosynthetic reaction centre are still discussed. The application of THz absorption spectroscopy to proteins is a relative new technique that aims to display vibrational modes that occur on the ps time scale. With the aid of THz absorption spectroscopy it could be shown in this work, that molecular vibrations change upon photoactivation of photosynthetic reaction centre from *R. sphaeroides* and that these changes most probably are independent of the solvent environment. However difficulties in interpreting the often featureless

absorption spectra do still prevent the extraction of key motions or the assignment of vibrations in the protein structure and this suggests that the method has to be developed further to be able to provide more detailed information of vibrations.

As important conformational flexibility is for the protein function, it is a hinder in the production of well diffracting protein crystals for X-ray crystallographic studies, and the purification, especially regarding membrane proteins. Highly entropic surfaces reduce the success in the formation of ordered crystal lattices, and modifications are frequently applied in protein crystallography in order to enhance their conformational stability. In addition to the conformational flexibility of exposed surfaces membrane proteins exhibit often an overall structural enhanced flexibility upon the extraction from their natural environment and challenges in membrane protein purification and stabilization are the main reason for the lack of structural information of membrane proteins. The membrane proteins UbiA and MenA could be overexpressed and purified on IMAC, but both proteins need to be stabilized to obtain protein samples for further investigations.

5. Acknowledgements

During my five years as a PhD-student at "Lundberg lab" I learned a lot and was lucky to meet all of the wonderful people that in one way or another were part of my journey. I want to say "special-thank-you" to following people:

First of all I want to thank you **Gergely** for accepting me as a PhD student in your group. Your great knowledge and skills ranging from biology to physics, including computational skills and especially X-ray crystallography seem to be endless and I am really thankful that you shared part of your knowledge with me during my time as a PhD student.

I also want to thank **Richard** as my co-supervisor for your support during these years and **Martin** for being my examinator and chairing the defense, and for always having an open door for advices when I needed them.

Weixiao, thank you for all you support and advices in difficult situations, you helped me a lot! It was really fun to teach Protein Engineering together with you and I will never forget our first course ;). **Ida**, it was great to have you in the group and I think we had a lot of fun (I don't know why I have to think about the "bushes" in our bathroom now…and the walk to Amundön). **Majo**, you just started too late, but I am happy that we got to know each other and hopefully we will have the opportunity to climb again. **Ying**, **Emilie** and **Oskar**, thank you for working as master students with me on the UbiA/ MenA project. You did a great job all of you! **Raja** and **Cecy** thank you for doing a research project with me.

Bence and **László** thank you for letting me being part of the S100A4 projects

Kristina, thank you for giving me the opportunity to teach the Chalmers students in Protein Engineering during all these years. It was a privileg and I will miss it! **Mikael**, I enjoyed working with you on the Protein Engineering course, and the party. Your Tabouleh is delicious!

Rosie, it is so great to have such a positive person around in the lab, you really improved our working situation by getting the Äkta problem under control! I will miss your laughter and hope that we will have another wine with cheese and olives at the ESRF ;). **Rhawnie**, thank you for all the time we spent together, and all your support. You cheered me up so much with your company, wellness and delicious food! **Elin**, it was great to have you in the lab, I will never forget your Danish skills!

Stephan, I am really happy that you joined the corridor, you made me a (hobby) climber, and without the time on the wall I would have gone crazy. I also appreciate your reintegration trials. **Maria,** artist in your soul**,** you are always up for things, and I really enjoyed our conversations and hanging out with you! **Sebastian P.** thank you for the lunch (and beer) company! **Sebastian W.**, I can still remember you prepping protein and now you have your own big group! **Kiran** and **Sushmita**, thank you for the lunch company it was always nice to talk with you! **Alex**, thank you for all you help with this laser related things! **Petra**, I enjoyed your company (and "Hallsta- and

Surahammar persons are weird). **Madde** it is good to have you on the corridor again! welcome back! **Emil**, always fun to talk and drink beer with you!

David and **Vlad** it was nice to get to know you, thank you for the parties at your place David! **Rob**, in the meanwhile I can understand your English, but I am still disappointed over that you left me to work out with Vlad instead. **Jennie**, it has been nice to get to know you, and to work out with you in the gym! You are a great instructor! And thanks to you and **Mike** for the barbecues at your place and all the fun at conferences! **Cecilia,** you are a true Matlab genious! Thank you for helping me with computer issues! **Peter**, thank you for the interesting conversations, I am sure you will do great in Australia! **Amit**, it was always nice to talk to you, and I really liked this Indian movie! **Rajiv** and **Gisela**, it has been nice to have you around.

Örjan, I would have liked to join the Biochemistry course, and found it really great that you started it! **Lars** and **Bruno**, what would the lab be without you! Lars, you keep everything going, including the Friday fika! Bruno thank you for your help with my computer!

Agneta, **Ann** och **Susanna**, thank you for your help with all this administrative things like reseräkningar and contracts, and that you always had an open door when I needed your help!

The new people **Petra**, **Rebecca**, **Alex**, **Parvin, Raffal**, welcome to the lab and have fun! (I am sure you will!)

"Old" people **Gerhard, Urszula, Linda, Erik, Etienne, Susanna, Anna** thank you for all your advices, and nice conversations during the time you were working here. Susanna, I really appreciated the journal clubs!

All my **Friends**, thank you for always cheering me up!

Meine Familie, **Mama**, **Papa** und **Brigitte**, vielen Dank fuer Eure Unterstützung in aller Form, während der ganzen Jahre, auch wenn ich so weit weg war. Ohne euch hätte ich es nie geschafft!

Staffan, tack för all stöd och tålamod med mig och alla dom näringsrika måtlador som du lagade under dom åren. Jag älskar dig!

6. References

1. Floudas, C. A., Fung, H. K., Mcallister, S. R., Mönnigmann, M. & Rajgaria, R. Advances in protein structure prediction and de novo protein design : A review. *Chem. Eng. Sci.* **61,** 966–988 (2006).
2. Teilum, K., Olsen, J. G. & Kragelund, B. B. Functional aspects of protein flexibility. *Cell. Mol. Life Sci.* **66,** 2231–47 (2009).
3. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293,** 321–31 (1999).
4. McCammon, J. A. Protein dynamics. *Rep. Prog. Phys.* **47,** 1–46 (1984).
5. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–72 (2007).
6. Waller, I. Zur Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen. *Zeitschrift fuer Phys.* **17,** 398–408 (1923).
7. Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. Temperature-dependant X-ray diffraction as a probe of protein structural dynamics. *Nature* **280,** 558– 563 (1979).
8. Kamerlin, S. C. L. & Warshel, A. At the Dawn of the 21 st Century : Is Dynamics the Missing Link for Understanding Enzyme Catalysis? *Proteins* **78,** 1339–1375 (2011).
9. Marsh, J. a & Teichmann, S. a. Parallel dynamics and evolution: Protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. *Bioessays* **36,** 209–18 (2014).
10. Carlson, H. a. Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* **6,** 447–52 (2002).
11. Gross, S. R., Sin, C. G. T., Barraclough, R. & Rudland, P. S. Joining S100 proteins and migration: for better or for worse, in sickness and in health. *Cell. Mol. Life Sci.* **71,** 1551–79 (2014).
12. Chen, H., Xu, C., Jin, Q. & Liu, Z. S100 protein family in human cancer. *Am. J. Cancer Res.* **4,** 89–115 (2014).
13. Donato, R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int. J. Biochem. Cell Biol.* **33,** 637–68 (2001).
14. Strutz, F. *et al.* Identification and characterization of a fibroblast marker: FSP1. *J. Cell Biol.* **130,** 393–405 (1995).
15. Cabezón, T. *et al.* Expression of S100A4 by a variety of cell types present in the tumor microenvironment of human breast cancer. *Int. J. Cancer* **121,** 1433–44 (2007).
16. Flatmark, K. *et al.* Nuclear localization of the metastasis-related protein S100A4 correlates with tumour stage in colorectal cancer. *J. Pathol.* **200,** 589–95 (2003).
17. Helfman, D. M., Kim, E. J., Lukanidin, E. & Grigorian, M. The metastasis associated protein S100A4: role in tumour progression and metastasis. *Br. J. Cancer* **92,** 1955–8 (2005).
18. Vallely, K. M. *et al.* Solution Structure of Human Mts1 (S100A4) As Determined by NMR. *Biochemistry* **41,** 12670–12680 (2002).
19. Pathuri, P., Vogeley, L. & Luecke, H. Crystal structure of metastasis-associated protein S100A4 in the active calcium-bound form. *J. Mol. Biol.* **383,** 62–77 (2008).
20. Malashkevich, V. N. *et al.* Structure of Ca 2 + -Bound S100A4 and Its Interaction with Peptides Derived from. *Biochemistry* **6,** 5111–5126 (2008).
21. Gingras, A. R. *et al.* Crystal structure of the Ca(2+)-form and Ca(2+)-binding kinetics of metastasis-associated protein, S100A4. *FEBS Lett.* **582,** 1651–6 (2008).

22. Elliott, P. R. *et al.* Asymmetric mode of Ca^{2+}-S100A4 interaction with nonmuscle myosin IIA generates nanomolar affinity required for filament remodeling. *Structure* **20,** 654–66 (2012).
23. Grigorian, M. *et al.* Tumor suppressor p53 protein is a new target for the metastasis-associated Mts1/S100A4 protein: functional consequences of their interaction. *J. Biol. Chem.* **276,** 22699–708 (2001).
24. Semov, A. *et al.* Metastasis-associated protein S100A4 induces angiogenesis through interaction with Annexin II and accelerated plasmin formation. *J. Biol. Chem.* **280,** 20833–41 (2005).
25. Watanabe, Y. *et al.* Calvasculin, as a factor affecting the microfilament assemblies in rat fibroblasts transfected by src gene. *FEBS Lett.* **324,** 51–5 (1993).
26. Kriajevska, M. Metastasis-associated Mts1 (S100A4) Protein Modulates Protein Kinase C Phosphorylation of the Heavy Chain of Nonmuscle Myosin. *J. Biol. Chem.* **273,** 9852–9856 (1998).
27. Ford, H. L., Silver, D. L., Kachar, B., Sellers, J. R. & Zain, S. B. Effect of Mts1 on the structure and activity of nonmuscle myosin II. *Biochemistry* **36,** 16321–7 (1997).
28. Li, Z., Spektor, A., Varlamova, O. & Bresnick, A. R. Mts1 Regulates the Assembly of Nonmuscle Myosin-IIA †. *Biochemistry* **42,** 14258–14266 (2003).
29. Lloyd, B. H., Platt-Higgins, a, Rudland, P. S. & Barraclough, R. Human S100A4 (p9Ka) induces the metastatic phenotype upon benign tumour cells. *Oncogene* **17,** 465–73 (1998).
30. Jenkinson, S. R., Barraclough, R., West, C. R. & Rudland, P. S. S100A4 regulates cell motility and invasion in an in vitro model for breast cancer metastasis. *Br. J. Cancer* **90,** 253–62 (2004).
31. Saleem, M. *et al.* S100A4 accelerates tumorigenesis and invasion of human prostate cancer through the transcriptional regulation of matrix metalloproteinase 9. *Proc Natl Acad Sci U S A* **103,** 14825–14830 (2006).
32. Li, Z.-H. & Bresnick, A. R. The S100A4 metastasis factor regulates cellular motility via a direct interaction with myosin-IIA. *Cancer Res.* **66,** 5173–80 (2006).
33. Vicente-Manzanares, M., Ma, X., Adelstein, R. S. & Horwitz, A. R. Non-muscle myosin II takes centre stage in cell adhesion and migration. *Nat. Rev. Mol. Cell Biol.* **10,** 778–90 (2009).
34. Murakami, N., Kotula, L. & Hwang, Y. W. Two distinct mechanisms for regulation of nonmuscle myosin assembly via the heavy chain: phosphorylation for MIIB and mts 1 binding for MIIA. *Biochemistry* **39,** 11441–51 (2000).
35. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of rhodopseudomonas viridis at 3Å resolution. *Nature* **318,** 618–624 (1985).
36. Kirmaier, C. & Holten, D. Primary photochemistry of reaction centers from the photosynthetic purple bacteria. *Photosynth. Res.* **13,** 225–260 (1987).
37. Crofts, A. R. & Wraight, C. A. The Electrochemical Domain of Photosynthesis. *Biochim. Biophys. Acta* **726,** 149–185 (1983).
38. Lundstrom, K. The Ultimate Approach for Rational Drug Design. *Struct. genomics* **34,** 205–212 (2006).
39. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5,** 993–6 (2006).
40. Grisshammer, R. Understanding recombinant expression of membrane proteins. *Curr. Opin. Biotechnol.* **17,** 337–40 (2006).
41. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta* **1666,** 105–17 (2004).

42. Ashby, M. N., Kutsunai, S. Y., Ackerman, S., Tzagoloff, a & Edwards, P. a. COQ2 is a candidate for the structural gene encoding para-hydroxybenzoate:polyprenyltransferase. *J. Biol. Chem.* **267,** 4128–36 (1992).
43. Young, I. G., Leppik, R. a, Hamilton, J. a & Gibson, F. Biochemical and genetic studies on ubiquinone biosynthesis in Escherichia coli K-12:4-hydroxybenzoate octaprenyltransferase. *J. Bacteriol.* **110,** 18–25 (1972).
44. Suvarna, K., Stevenson, D., Meganathan, R. & Hudspeth, M. E. Menaquinone (vitamin K2) biosynthesis: localization and characterization of the menA gene from Escherichia coli. *J. Bacteriol.* **180,** 2782–7 (1998).
45. Saiki, K., Mogi, T., Ogura, K. & Anraku, Y. In Vitro Heme O Synthesis by the cyoE gene Product from Escherichia coli. *J. Biol. Chem.* **268,** 26041–26045 (1993).
46. Oster, U., Bauer, C. E. & Rüdiger, W. Characterization of chlorophyll a and bacteriochlorophyll a synthases by heterologous expression in Escherichia coli. *J. Biol. Chem.* **272,** 9671–6 (1997).
47. Schledz, M., Seidler, a, Beyer, P. & Neuhaus, G. A novel phytyltransferase from Synechocystis sp. PCC 6803 involved in tocopherol biosynthesis. *FEBS Lett.* **499,** 15–20 (2001).
48. Meganathan, R. Ubiquinone biosynthesis in microorganisms. *FEMS Microbiol. Lett.* **203,** 131–9 (2001).
49. Bentley, R. & Meganathan, R. Biosynthesis of vitamin K (menaquinone) in bacteria. *Microbiol. Rev.* **46,** 241–80 (1982).
50. Frei, B., Kim, M. C. & Ames, B. N. Ubiquinol-10 is an effective lipid-soluble antioxidant at physiological concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4879–83 (1990).
51. Bentinger, M., Brismar, K. & Dallner, G. The antioxidant role of coenzyme Q. *Mitochondrion* **7 Suppl,** S41–50 (2007).
52. Bader, M., Muse, W., Ballou, D. P., Gassner, C. & Bardwell, J. C. Oxidative protein folding is driven by the electron transport system. *Cell* **98,** 217–27 (1999).
53. Bader, M. W., Xie, T., Yu, C. a & Bardwell, J. C. Disulfide bonds are generated by quinone reduction. *J. Biol. Chem.* **275,** 26082–8 (2000).
54. Fredericks, W. J. *et al.* The bladder tumor suppressor protein TERE1 (UBIAD1) modulates cell cholesterol: implications for tumor progression. *DNA Cell Biol.* **30,** 851–64 (2011).
55. Cheng, W. & Li, W. Structural insights into ubiquinone biosynthesis in membranes. *Science* **343,** 878–81 (2014).
56. Huang, H. *et al.* Structure of a Membrane-Embedded Prenyltransferase Homologous to UBIAD1. *PLoS Biol.* **12,** e1001911 (2014).
57. Wang, D.-N. *et al.* Practical aspects of overexpressing bacterial secondary membrane transporters for structural studies. *Biochim. Biophys. Acta - Biomembr.* **1610,** 23–36 (2003).
58. Grisshammer, R. & Tate, C. G. Overexpression of integral membrane proteins for structural studies. *Q. Rev. Biophys.* **28,** 315–422 (1995).
59. Kelley, K. C. *et al.* Regulation of sCD4-183 gene expression from phage-T7-based vectors in Escherichia coli. *Gene* **156,** 33–6 (1995).
60. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
61. Gordon, E. *et al.* Effective high-throughput overproduction of membrane proteins in Escherichia coli. *Protein Expr. Purif.* **62,** 1–8 (2008).
62. Campbell, J. W. *et al.* X-ray diffraction studies on enzymes in the glycolytic pathway. *Cold spring Harb. Symp. Quant. Biol.* **36,** 165–170 (1972).

63. Dale, G. E., Oefner, C. & D'Arcy, A. The protein as a variable in protein crystallization. *J. Struct. Biol.* **142,** 88–97 (2003).
64. Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 679–688 (2001).
65. Derewenda, Z. S. Rational protein crystallization by mutational surface engineering. *Structure* **12,** 529–35 (2004).
66. Goldschmidt, L., Cooper, D. R. & Derewenda, Z. S. Toward rational protein crystallization : A Web server for the design of crystallizable protein variants. 1569–1576 (2007). doi:10.1110/ps.072914007.nology
67. Ii, W. N. P. *et al.* Understanding the physical properties controlling protein crystallization based on analysis of large-scale experimental data. **27,** 51–57 (2010).
68. Towbin, H., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. 1979. *Biotechnology* **24,** 145–9 (1992).
69. Kragh-Hansen, U., le Maire, M. & Møller, J. V. The mechanism of detergent solubilization of liposomes and protein-containing membranes. *Biophys. J.* **75,** 2932–46 (1998).
70. Heerklotz, H. & Seelig, J. Titration calorimetry of surfactant-membrane partitioning and membrane solubilization. *Biochim. Biophys. Acta* **1508,** 69–85 (2000).
71. Le Maire, M., Champeil, P. & Moller, J. V. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim. Biophys. Acta* **1508,** 86–111 (2000).
72. Parker, J. L. & Newstead, S. Current trends in α-helical membrane protein crystallization: an update. *Protein Sci.* **21,** 1358–65 (2012).
73. Lathe, G. H. & Ruthven, C. R. L. The seperation of Substances on the Basis of their Molecular weights, Using Columns of starch and Water. *Proc. Biochem. Soc.* (1955). at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1215856&tool=pmcentrez&rendertype=abstract>
74. Lathe, G. H. & Ruthven, C. R. The separation of substances and estimation of their relative molecular sizes by the use of colums of starch in water. *Biochem. J.* **62,** 665–74 (1956).
75. Porath, Jerker Flodin, P. Gel Filtration: A Method For Desalting and Group Seperation. *Nature* **183,** 1657–1659 (1959).
76. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases : A novel concept for the crystallization of membrane proteins. *Biophysics (Oxf).* **93,** 14532–14535 (1996).
77. Wöhri, A. B. *et al.* A lipidic-sponge phase screen for membrane protein crystallization. *Structure* **16,** 1003–9 (2008).
78. Faham, S. & Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **316,** 1–6 (2002).
79. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254,** 51–8 (1991).
80. Dauter, Z. New approaches to high-throughput phasing. *Curr. Opin. Struct. Biol.* **12,** 674–678 (2002).
81. Leslie, A.G.W. Powel, H. R. in *Evol. Methods Macromol. Crystallogr.* **245,** 41–51 (2007).
82. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr. D. Biol. Crystallogr.* **62,** 859–66 (2006).

83. Garman, E. F. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 339–51 (2010).
84. Hope, H. Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Crystallogr. Sect. B Struct. Sci.* **44,** 22–26 (1988).
85. Kabsch, W. Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 125–32 (2010).
86. Patterson, A. L. A fourier series method for the determination of the components of interatomic distances. *Phys. Rev.* **46,** 372–376 (1934).
87. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
88. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 486–501 (2010).
89. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–21 (2010).
90. Arndt, U. W., Crowther, R. A. & Mallett, J. F. W. A computer-linked cathode-ray tube microdensitometer for x-ray crystallography. *J. Phys.* **1,** 510–516 (1968).
91. Diederichs, Kay Karplus, P. A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature* **4,** 269–275 (1997).
92. Karplus, P. A. & Diedrichs, K. Linking crystallographic model and data quality. *Science (80-.).* **336,** 1030–1033 (2012).
93. Ramachandran, G.N. Ramakrishnan, C. Sasisekharan, V. Stereochemistry of polypeptide chain konfigurations. *J. Mol. Biol.* **7,** 95–99 (1963).
94. Guinier, A. La diffraction des rayons X aux tres petits angles: application a l`etude de phenomenes ultramicroscopiques. *Ann. Phys.* 161–237 (1939).
95. Glatter, B. Y. O. A New Method for the Evaluation of Small-Angle Scattering Data. *J. Appl. Cryst.* **10,** 415–421 (1977).
96. Petoukhov, M. V, Konarev, P. V, Kikhney, G. & Dmitri, I. ATSAS 2.1 -towards automated and web supported small-angle scattering data analysis. in *J. Appl. Cryst.* 223–228 (2007).
97. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36,** 1277–1282 (2003).
98. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates D . Svergun , C . Barberato and M . H . J . Koch. *J. Appl. Cryst.* **28,** 768–773 (1995).
99. Hubbard, S. R., Hodgson, K. O. & Doniach, S. Small-angle x-ray scattering investigation of the solution structure of troponin C. *J. Biol. Chem.* **263,** 4151–8 (1988).
100. Fujisawa, T., Uruga, T., Yamaizumi, Z. & Inoko, Y. The Hydration of Ras p21 in Solution during GTP Hydrolysis Based on. **115,** 875–880 (1994).
101. Mertens, H. D. T. & Svergun, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172,** 128–41 (2010).
102. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. Structural characterization of flexible proteins using SAXS. 5656–5664 (2007).
103. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 (2012).
104. Haddad, J. El, Bousquet, B., Canioni, L. & Mounaix, P. Review in terahertz spectral analysis. *Trends anal. chem* **44,** 98–105 (2013).

105. Xu, J., Plaxco, K. W. & Allen, S. J. Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy. *Protein Sci.* **15,** 1175–1181 (2006).
106. Xu, J., Plaxco, K. W. & Allen, S. J. Collective dynamics of lysozyme in water: terahertz absorption spectroscopy and comparison with theory. *J. Phys. Chem. B* **110,** 24255–9 (2006).
107. Kriajevskas, M. V *et al.* Non-muscle Myosin Heavy Chain as a possible Target for Protein Encoded by Metastasis related mts-1 Gene. *J. Biol. Chem.* **269,** 19679–19682 (1994).
108. Chen, H. *et al.* Binding to intracellular targets of the metastasis-inducing protein, S100A4 (p9Ka). *Biochem. Biophys. Res. Commun.* **286,** 1212–7 (2001).
109. Malashkevich, V. N. *et al.* Structure of Ca2+-bound S100A4 and Its Interaction with Peptides Derived from Nonmuscle Myosin-IIA. *Biochemistry* **47,** 5111–5126 (2008).
110. Badyal, S. K. *et al.* Mechanism of the Ca2+-Dependent Interaction between S100A4 and Tail Fragments of Nonmuscle Myosin Heavy Chain IIA. *J. Mol. Biol.* **405,** 1004–1026 (2011).
111. Dulyaninova, N. G. *et al.* Cysteine 81 Is Critical for the Interaction of S100A4 and Myosin-IIA. *Biochemistry* **50,** 7218–7227 (2011).
112. Rustandi, R. R., Baldisseri, D. M. & Weber, D. J. Structure of the negative regulatory domain of p53 bound to S100B (ββ). *Nat. Struct. Biol.* **7,** (2000).
113. Bhattacharya, S., Large, E., Heizmann, C. W., Hemmings, B. & Chazin, W. J. Structure of the Ca2+/S100B/NDR kinase peptide complex: insights into S100 target specificity and activation of the kinase. *Biochemistry* **42,** 14416–26 (2003).
114. Lee, Y.-T. *et al.* Structure of the S100A6 complex with a fragment from the C-terminal domain of Siah-1 interacting protein: A novel mode for S100 protein target recognition. *Biochemistry* **14,** 10921–10932 (2009).
115. Wright, N. T., Varney, K. M., Cannon, B. R., Morgan, M. & Weber, D. J. Ca-S100A1 interacting with TRTK12. *PDB code 2KBM* (2009).
116. Tina, K. G., Bhadra, R. & Srinivasan, N. PIC : Protein Interactions Calculator. *Nucleic Acids Res.* **35,** 473–476 (2007).
117. Gupta, A. a, Chou, R.-H., Li, H., Yang, L.-W. & Yu, C. Structural insights into the interaction of human S100B and basic fibroblast growth factor (FGF2): Effects on FGFR1 receptor signaling. *Biochim. Biophys. Acta* **1834,** 2606–19 (2013).
118. Nakasawa, T. *et al.* Critical regions for assembly of vertebrate nonmuscle myosin II. *Biochemistry* **44,** 174–83 (2005).
119. Ricketson, D., Johnston, C. a & Prehoda, K. E. Multiple tail domain interactions stabilize nonmuscle myosin II bipolar filaments. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 20964–9 (2010).
120. Schneider, T. R. A genetic algorithm for the identification of conformationally invariant regions in protein molecules research papers. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 195–208 (2002).
121. Dutta, K., Cox, C. J., Basavappa, R. & Pascal, S. M. 15N relaxation studies of Apo-Mts1: a dynamic S100 protein. *Biochemistry* **47,** 7637–47 (2008).
122. Dutta, K., Cox, C. J., Huang, H., Basavappa, R. & Pascal, S. M. Calcium coordination studies of the metastatic Mts1 protein. *Biochemistry* **41,** 4239–45 (2002).
123. Zhang, S. *et al.* The C-terminal region of S100A4 is important for its metastasis-inducing properties. *Oncogene* **24,** 4401–11 (2005).
124. Ismail, T. M. *et al.* The basic C-terminal amino acids of calcium-binding protein S100A4 promote metastasis. *Carcinogenesis* **29,** 2259–66 (2008).

125. Ramagopal, U.A., Dulyaninova, N.G., Kumar, P.R., Almo, S.C., Bresnick, A.R., N. Y. S. G. R. C. Structure of truncated (delta8C) S100A4. *PDB code 4HSZ* (2013).
126. Garrett, S. C., Varney, K. M., Weber, D. J. & Bresnick, A. R. S100A4, a mediator of metastasis. *J. Biol. Chem.* **281,** 677–80 (2006).
127. Kirmaier, C., Holten, D. & Parson, W. W. Temperature and detection-wavelength dependence of the picosecond electron-transfer kinetics measured in Rhodopseudomonas sphaeroides reaction centers. Resolution of new spectral and kinetic components in the primary charge-separation process. *Biochim. Biophys. Acta - Bioenerg.* **810,** 33–48 (1985).
128. Clayton, R. K. & Yau, H. F. Photochemical electron transport in photosynthetic reaction centers from Rhodopseudomonas spheroides. I. Kinetics of the oxidation and reduction of P-870 as affected by external factors. *Biophys. J.* **12,** 867–81 (1972).
129. McMahon, B. H., Müller, J. D., Wraight, C. a & Nienhaus, G. U. Electron transfer and protein dynamics in the photosynthetic reaction center. *Biophys. J.* **74,** 2567–87 (1998).
130. Katona, G. *et al.* Conformational regulation of charge recombination reactions in a photosynthetic bacterial reaction center. *Nat. Struct. Mol. Biol.* **12,** 630–631 (2005).
131. Vos, M. H., Rappaport, F., Lambry, J.-C., Breton, J. & Martin, J.-L. Visualization of coherent nuclear motion in a membrane protein by femtosecond spectroscopy. *Nature* **363,** 320– 325 (1993).
132. Whitmire, S. E. *et al.* Protein flexibility and conformational state: a comparison of collective vibrational modes of wild-type and D96N bacteriorhodopsin. *Biophys. J.* **85,** 1269–77 (2003).
133. Balu, R. *et al.* Terahertz spectroscopy of bacteriorhodopsin and rhodopsin: similarities and differences. *Biophys. J.* **94,** 3217–26 (2008).
134. Castro-Camus, E. & Johnston, M. B. Conformational changes of photoactive yellow protein monitored by terahertz spectroscopy. *Chem. Phys. Lett.* **455,** 289–292 (2008).
135. Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Parak, F. G. Slaving: solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16047–51 (2002).
136. Agalidis, I., Nuijs, A. M. & Reiss-Husson, F. Characterization of an LM unit purified by affinity chromatography from Rhodobacter sphaeroides reaction centers and interactions with the H subunit. *Biochim. Biophys. Acta - Bioenerg.* **890,** 242–250 (1987).
137. Ebbinghaus, S. *et al.* Protein sequence- and pH-dependent hydration probed by terahertz spectroscopy. *J. Am. Chem. Soc.* **130,** 2374–5 (2008).
138. Melzer, M. & Heide, L. Characterization of polyprenyldiphosphate: 4-hydroxybenzoate polyprenyltransferase from Escherichia coli. *Biochim. Biophys. Acta* **1212,** 93–102 (1994).
139. Bonitz, T., Alva, V., Saleh, O., Lupas, A. N. & Heide, L. Evolutionary relationships of microbial aromatic prenyltransferases. *PLoS One* **6,** e27336 (2011).
140. Bräuer, L., Brandt, W. & Wessjohann, L. a. Modeling the E. coli 4-hydroxybenzoic acid oligoprenyltransferase (ubiA transferase) and characterization of potential active sites. *J. Mol. Model.* **10,** 317–27 (2004).
141. Alguel, Y. *et al.* New Tools for membrane Protein Research. *Curr. protein Pept. Sci.* **11,** 1–10 (2010).
142. Wiener, M. C. A pedestrian guide to membrane protein crystallization. *Methods* **34,** 364–72 (2004).
143. Lee, A. G. How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta* **1666,** 62–87 (2004).

144. Rosenbusch, J. P. Stability of membrane proteins: relevance for the selection of appropriate methods for high-resolution structure determinations. *J. Struct. Biol.* **136,** 144–57 (2001).
145. Alexandrov, A. I., Mileni, M., Chien, E. Y. T., Hanson, M. a & Stevens, R. C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16,** 351–9 (2008).
146. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112,** 531–52 (1999).
147. Ferrandon, B. & Newstead, S. Rationalizing a -helical membrane protein crystallization. **1,** 466–472 (2008).
148. Kumar, S., Ma, B., Tsai, C.-J., Sinha, N. & Nussinov, R. Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Sci.* **9,** 10–19 (2008).
149. Hayashi, Shin'ichiro Kawase, K. in *Recent Opt. Photonic Technol.* (Kim, K. Y.) Chapter 6 (InTech, 2010).