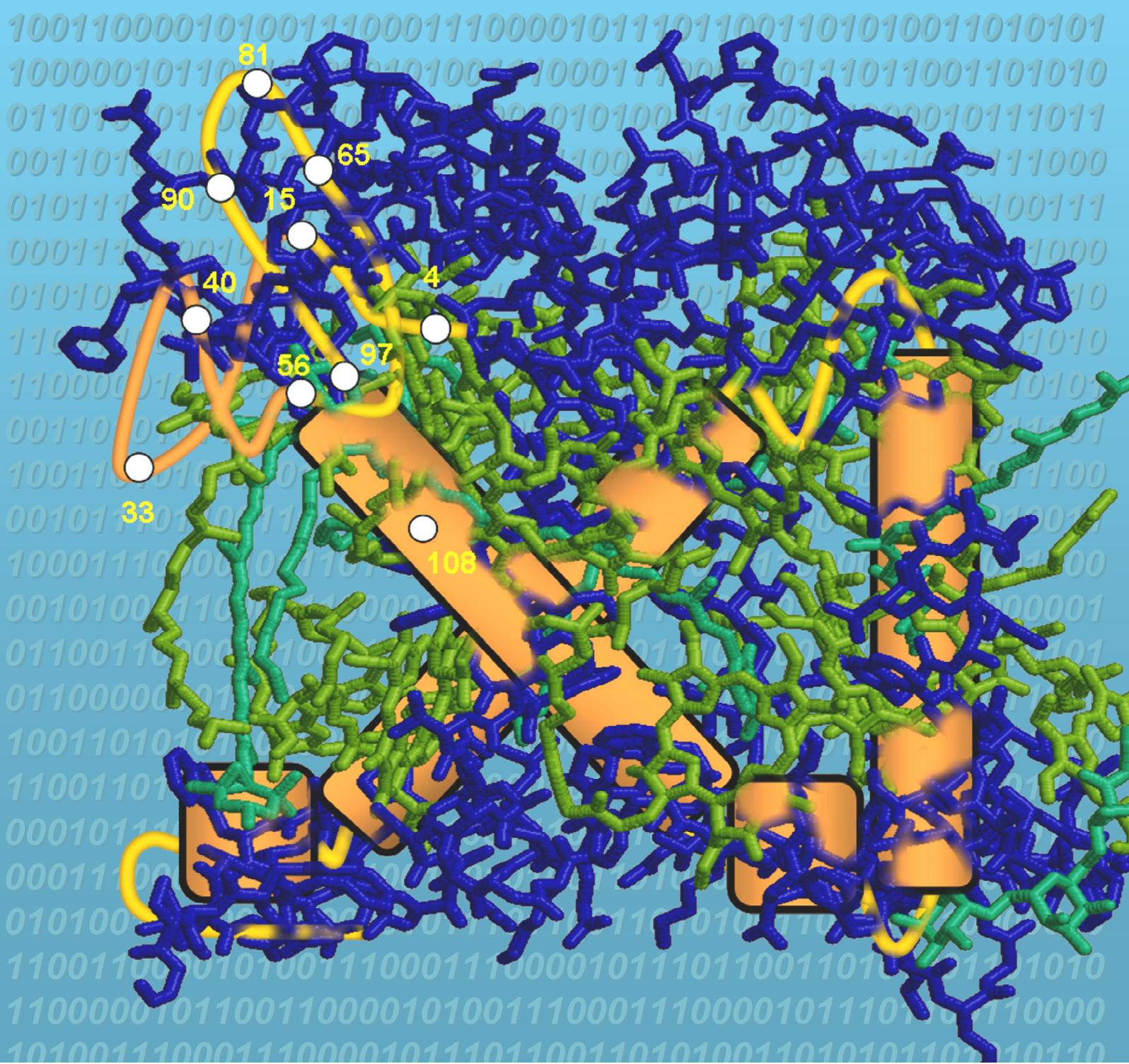


Modeling membrane protein structure using site-directed ESR spectroscopy

Aleh Kavalenka



Modeling membrane protein structure using site-directed ESR spectroscopy

Aleh Kavalenka

Promotor:

Prof. Dr. H. van Amerongen,
Hoogleraar in de Biofysica
Wageningen Universiteit, Nederland

Co-promotoren:

Dr. M. A. Hemminga,
Universitair Hoofddocent, Laboratorium voor Biofysica
Wageningen Universiteit, Nederland

Dr. J. Štrancar,
Laboratory of Biophysics
Jožef Stefan Institute, Slovenia

Promotiecommissie:

Prof. Dr. F. A. M. Leermakers (Wageningen Universiteit)
Prof. Dr. J. M. Vlak (Wageningen Universiteit)
Dr. M. I. Huber (Universiteit Leiden)
Prof. Dr. R. Boelens (Universiteit Utrecht)

Dit onderzoek is uitgevoerd binnen de onderzoeksschool EPS

Modeling membrane protein structure using site-directed ESR spectroscopy

Aleh Kavalenka

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof.dr. M.J. Kropff,
in het openbaar te verdedigen
op dinsdag 30 september 2009
des namiddags te half twee in de Aula

Kavalenka A.

Modeling membrane protein structure using site-directed ESR spectroscopy

Thesis, Wageningen University, 2009.

ISBN: 978-90-8585-424-1

*“Всякое знание, которое человек открывает путем науки, есть дар Божий людям и
возвещает присутствие Божие в сем мире“
Elder Tadej*

Contents

	Abbreviations	
Chapter 1	General introduction	1
Chapter 2	Speeding-up a genetic algorithm for ESR-based spin label characterization of biosystem complexity	21
Chapter 3	Site-directed spin labeling study of the light-harvesting complex CP29	37
Chapter 4	Analysis of side chain rotational restrictions of membrane-embedded proteins by spin label ESR spectroscopy	57
Chapter 5	Optimization of membrane protein structure based on SDSL-ESR constraints and conformational space modeling	67
Chapter 6	Summarizing discussion	95
	Summary	115
	Acknowledgments	116
	Publications	118
	Curriculum vitae	119
	Education Statement of the Graduate School Experimental Plant Sciences	120

Abbreviations

14:1 PC	1,2-dimyristoleoyl-sn-glycero-3-phosphocholine
22:1 PC	1,2-dierucoyl-sn-glycero-3-phosphocholine
Chl	chlorophyll
CP29	chlorophyll-a/b-binding protein 29
DM	n-Dodecyl β -D-maltoside
DOPC/DOPG	dioleoylphosphatidylcholine/ dioleoylphosphatidyl-glycerol
ESR (EPR)	electron spin (paramagnetic) resonance
FRET	Förster (or fluorescence) resonance energy transfer
GA	genetic algorithm
GHOST	condensation algorithm that filters and groups multiple solutions found during optimization of the simulated spectra
HEO	hybrid evolutionary optimization
LDS	lithium dodecyl sulfate
Lhc	light-harvesting complex
LHCII	light-harvesting chlorophyll-a/b-binding protein of photosystem II
MD	molecular dynamics
MTS-SL	(1-Oxyl-2,2,5,5-tetrame-thylpyrroline-3-methyl) methanethiosulfonate spin label
NMR	nuclear magnetic resonance
PC	phosphatidylcholine
PS	photosystem
RGB	red green blue color model
SDSL	site-directed spin-labeling
SGA	simple genetic algorithm
SL	spin label

Chapter 1

General introduction

Introduction

Proteins are the key molecules in cells of living organisms, including human beings. They participate for instance in signaling pathways, intracellular and extracellular transport, mechanical work, and can act as sensors. Biochemistry, and particularly the knowledge about protein structure and function, provides important insights and practical applications in medicine, agriculture, nutrition, and industry¹. Our special interest in this thesis concerns the study of structural characteristics of membrane proteins and intrinsically disordered proteins, which have been characterized far less well than water-soluble proteins.

We describe a new modeling approach for membrane protein structure characterization based on a) structural constraints experimentally determined by site-directed labeling; b) conformational space modeling; c) protein structure optimization. The first chapter is a general introduction to the structural organization of proteins, in particular membrane proteins, and to the techniques, which have been used to study these protein systems. Later on in the thesis the focus will be on site-directed spin labeling (SDSL) electron spin resonance (ESR) spectroscopy, an alternative and complementary technique with respect to well-established and powerful techniques like X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy.

First, in section 1.1 some background information is presented about proteins and biological membranes. Section 1.2 describes different experimental techniques used for protein structure determination. In section 1.3 the protein structure modeling is overviewed. The next two sections describe the two membrane proteins: bacteriophage M13 coat protein (section 1.4), and light-harvesting protein complex CP29 (section 1.5), which are studied in detail in this thesis. Finally, section 1.6 introduces the overview of the thesis.

1.1 Protein: life's workhorse

Interestingly, the whole wealth and diversity of proteins and protein functions (i.e., regulation, signaling, transport, catalysis, etc.) primarily stems from a combination of 20 basic amino acids, sequentially linked one to another. Due to particular combinations of amino acids, a specific environment, and interactions with many other molecules, the primary protein sequence leads to the formation of a 3D structure. The structural organization and the corresponding dynamical properties together define the functionality of the proteins.

1.1.1 Protein structure and dynamics

Already the primary sequence of the protein mainly defines its 3D structure. The organization of the amino acid chain into a particular pattern (examples of helices, sheets, and turns are presented in Figure 1C) is referred to as the secondary structure. The

secondary structure folds into the three-dimensional one, which is called the tertiary structure. Finally, if several folded proteins are organized into a functional protein complex, one speaks of the quaternary structure (see Figure 1A).

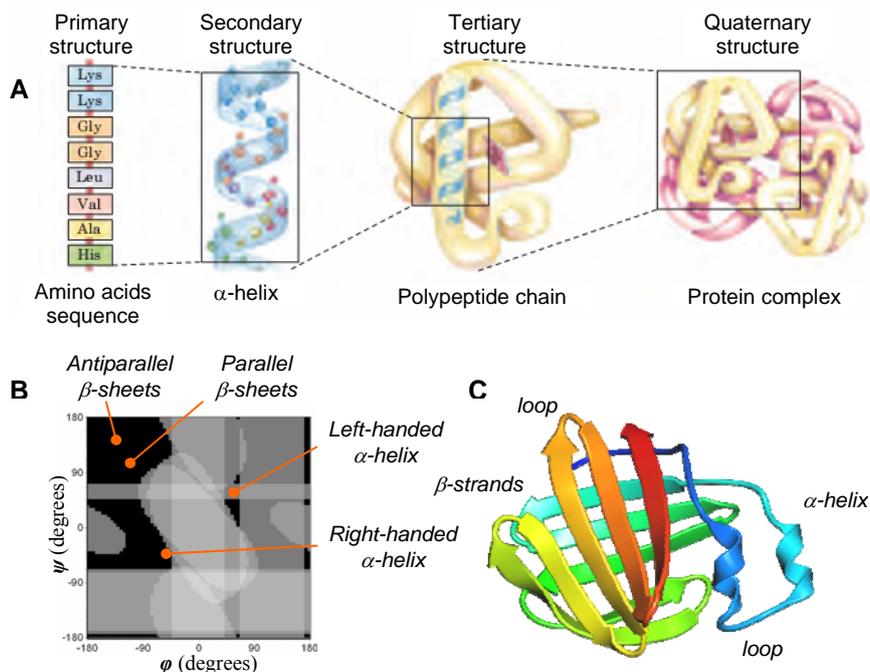


Figure 1. Protein structure. **A.** Levels of structure in proteins¹. The primary structure consists of a sequence of amino acids linked together by peptide bonds. The resulting polypeptide can be coiled into units of secondary structure, such as an α -helix. The helix is a part of the tertiary structure of the folded polypeptide, which itself can be one of the subunits that make up the quaternary structure of a multisubunit protein, in this case hemoglobin. **B.** Parameterization of the secondary structure of a protein: the distribution of ϕ and ψ backbone dihedral angles is given in the schematic Ramachandran plot² and defines the secondary structure. Black areas indicate the allowed combinations of ϕ and ψ dihedral angles. Values of ϕ and ψ corresponding to the secondary structure motifs, α -helix and β -sheet, are marked with orange lines. **C.** An example protein structure, an acid-binding protein, rich in β -sheets, also having two short α -helices and several turns (loops)³.

In proteins, the amino acids are covalently linked in linear peptide chains, via peptide bonds, in which the carboxyl group of one amino acid is joined with the amino group of another amino acid. The peptide unit is rigid and planar; however, the bonds at the end of the peptide unit are free to rotate. This allows polypeptide chains to form a wide range of three-dimensional protein structures⁴. The bond angles arising from rotations at the C_{α} atom are identified as ϕ and ψ rotations. Allowed values for these angles are graphically represented on a Ramachandran plot^{2,5}, which also identifies regions of different motifs of the secondary structure, e.g. α -helix and β -sheets (see Figure 1B). The backbone of the protein (the regular repeating main chain) formed out of amino acid bases

is more rigid than the protein side chain, which is made of amino acid residues. Due to single bonds free rotation within the side chain, each amino acid is described by a set of possible side chain rotamers^{6,7}.

Interestingly, forces that control secondary, tertiary and quaternary structure are much weaker than the ones that constitute a covalent bond. Secondary structure motifs, α -helix and β -sheet, are entirely governed by hydrogen bonding. A notable exception concerns transmembrane proteins, in which the secondary structure is strongly stabilized by hydrophobic interactions due to membrane lipids. Hydrogen bonding, hydrophobic forces, electrostatic forces and van der Waals forces, in combination with steric hindrance effects, together govern protein folding and stabilize tertiary and quaternary structures.

Although static structures are known for many proteins, the functions of proteins are often governed by their dynamics⁸. Protein dynamics is characterized by the timescale, the amplitude and the direction of the fluctuations⁹.

1.1.2 Biological membrane and membrane proteins

Biological membranes form boundaries between different compartments. Membranes are composed of lipids and protein molecules that form a thin hydrophobic barrier. Carbohydrates are also present as part of glycoproteins and glycolipids. Transport proteins in the plasma membrane allow the passage of certain ions and molecules; receptor proteins transmit signals into the cell; and membrane enzymes participate in some reaction pathways. Because the individual lipids and proteins of the plasma membrane are not covalently linked, the entire structure is remarkably flexible, allowing changes in the shape and size of the cell¹.

The complex structure of a biological membrane can be described with a fluid mosaic model of Singer and Nicholson^{10,11}. The acyl chains in the interior of the membrane form a fluid, hydrophobic region. Integral proteins float in this sea of lipids (see Figure 2A). Both proteins and lipids are free to move laterally (the speed depends strongly on size of protein, packing and temperature) in the plane of the bilayer, but movement from one face of the bilayer to the other is restricted.

Subsequent refinements of the fluid mosaic model were suggested, based on the results of experimental and theoretical studies focusing on some specific aspects of the membrane structure¹². For example, Israelachvili¹³ proposed that proteins and lipids need to 'adjust' to each other (a concept of 'hydrophobic matching'¹⁴⁻¹⁶, see Figure 2D). It was suggested that small proteins consisting of one or a few α -helical chains are likely to accommodate to the bilayer thickness by helix tilting, whereas larger proteins or less flexible proteins induce changes in bilayer thickness. A more advanced picture of the fluid mosaic model which is still under discussion contains patches of lipids, the composition of which differs from the average for the bilayer¹¹.

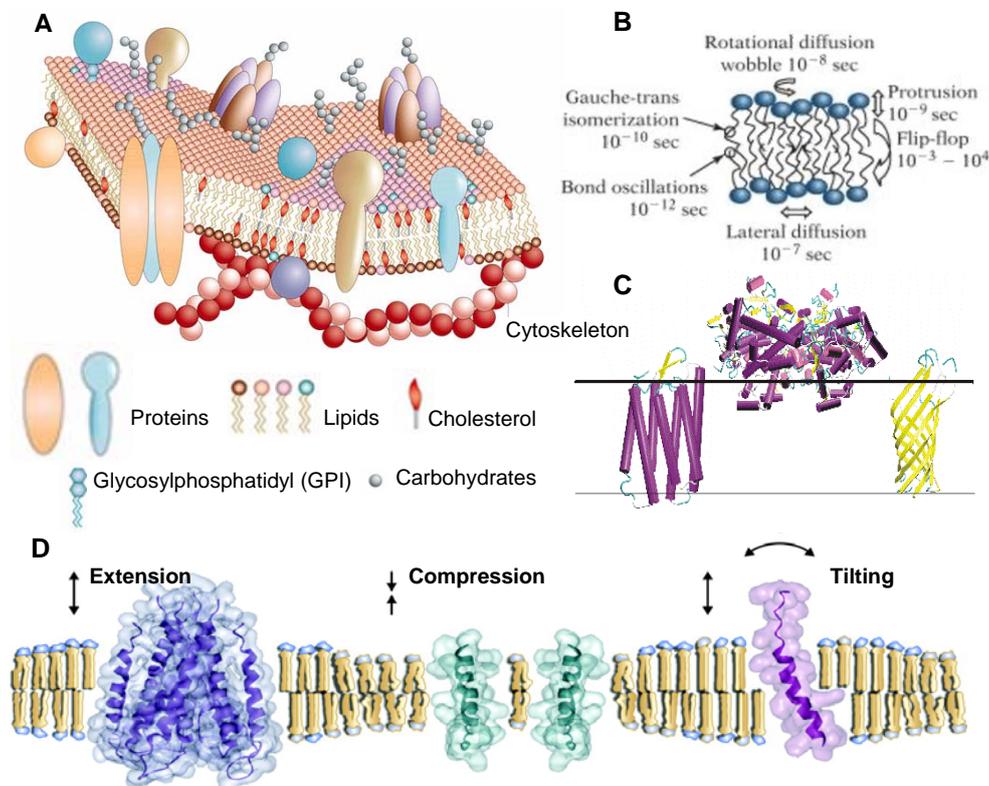


Figure 2. Biological membrane and membrane proteins. **A.** The Fluid-Mosaic-Model of the cell membrane ¹⁷. **B.** Lipid motions in biomembranes and their approximate correlation times ¹⁸. **C.** Diversity in membrane proteins (transmembrane and peripheral; α helical and β barrel) in the lipid bilayer. **D.** Modified view of the structure of a biological membrane. Because component lipids and proteins are not naturally matched in this membrane, they must strain (expend energy) to match each other hydrophobically, resulting in a high-energy membrane. Compensatory conformational changes include lipid acyl chain extension and transmembrane helix tilting when lipids surround a protein with a long transmembrane region, and lipid acyl chain compression when lipids surround a protein with a short transmembrane domain.

Not only do integral proteins perturb the lipids, but the physical state of the lipids does also actively influence protein function ¹⁹. According to the shell model ²⁰ the lipids form a shell (annulus) around the protein. Annular lipids are envisioned to be stably associated with proteins, though component lipids can exchange with the ‘bulk’ bilayer lipids ²¹. ESR ²²⁻²⁴, NMR ¹⁸ and optical spectroscopy have been used to studying annular lipids and protein-lipid interactions.

New models view the membrane as a complex, highly cooperative and heterogeneous system, which displays dynamic and structural properties on many length- and time scales ¹². It is also clear that there are strong interactions between lipids and proteins in the membrane ¹⁷.

Membrane proteins. Membrane proteins constitute by weight up to 80% of the biological membranes. Their common property is that part of their structure is buried in a lipid bilayer²⁵. The position of membrane proteins in the membrane depends on the polarity of the amino acid residues. Membrane proteins can pass through a membrane (integral or transmembrane proteins), or lie on top of a bilayer (peripheral or surface membrane proteins) (see Figure 2C). The most simple transmembrane proteins (e.g., the major coat protein of M13 introduced in section 1.4) consist of a single chain (usually α -helical), which spans the membrane. Larger proteins (e.g., protein complex CP29 described in section 1.5) consist of multiple segments. These are usually the chains of α -helices or strands of β -barrels connected with the loops (Figure 2C).

The topology of an integral membrane protein can be predicted from its amino acid sequence^{26,27}. A sequence of more than 20 hydrophobic amino acids indicates that this part of a protein traverses the lipid bilayer. A protein chain surrounded by lipid molecules without water molecules with which they could make hydrogen bonds, will tend to form α -helices and β -sheets, in which intrachain hydrogen bonding is maximized. If the side chains of all amino acids in a helix are nonpolar, hydrophobic interactions with the surrounding lipids further stabilize the helix¹. Polar amino acids (lysine, arginine, glutamic acid, and aspartic acid) are found exclusively in the aqueous phases. The side chains of tyrosine and tryptophan are often present in the interface between lipids and water²⁸, able to interact both with lipids and water, and serving as membrane interface anchors.

Effect of lipids on protein structure. Protein-lipid interactions are expected to play a prominent role in the membrane structure²⁹. Not only do integral proteins perturb the lipids, but the physical state of the lipids does also actively influence protein function¹⁹. Membranes are very dynamic structures with constant movements of lipids in the bilayer, both in the transverse direction across the bilayer and the lateral direction in the plane of this two-dimensional matrix (see Figure 2B). The movements in the lateral direction give rise to the fluid nature of the membrane and enable interactions among proteins and between proteins and lipids³⁰. It is thought that lipid dynamics profoundly influence the function of membrane proteins, not the least in dynamically differentiated and spatially separated in-plane membrane domains²³. On the other hand, there is evidence for stabilizing lipid-protein interactions: several hydrogen bonds and/or ion-pair interactions stabilize head group binding, whereas hydrophobic lipid side chains fit tightly into hydrophobic grooves at the protein surface and are stabilized by multiple nonpolar, van der Waals interactions with amino acid residues³¹.

1.2 Protein structure characterization

Around 90% of the protein structures available in the Protein Data Bank ^{32,33} has been determined by X-ray crystallography (see Table 1). This method allows the exact 3D coordinates of all the atoms in the protein to be determined to within a certain resolution. Roughly 9% of the known protein structures have been obtained by NMR techniques. Alternatively electron microscopy (EM), atomic force microscopy (AFM) can also be used to determine lower-resolution 3D structures ³⁴⁻³⁷. Certain aspects of the secondary structure as a whole can be determined via other biochemical techniques such as circular dichroism (CD) ³⁸⁻⁴⁰, small-angle X-ray scattering (SAXS) ^{41,42}, fluorescence and ESR spectroscopy ⁴³. Cryo-electron microscopy has recently become a means of determining protein structures at low resolution (less than 5Å) and is anticipated to increase in power as a tool for high resolution work in the next decade. In the past few years it has become possible for highly accurate physical molecular models to complement the *in silico* study of biological structures. These include various technologies of 3D Molecular Design and visualization, and Molecular Dynamics (MD) simulations, which are constantly being improved due to refinement of the models and due to continuous increase of the computational power allowing longer simulations or simulations of more complex systems.

Table 1. Protein database current holdings ^{32,33}.

Techniques	Proteins	Nucleic acids	Protein/NA complexes	Other	Total
X-ray	46071	1142	2118	17	49348
NMR	6844	850	144	7	7845
Electron Microscopy	163	16	59	0	238
Other	110	4	4	9	127
Total	53188	2012	2325	33	57558

1.2.1 SDSL-ESR spectroscopy

Site-directed spin labeling electron spin resonance spectroscopy ⁴⁴⁻⁴⁶ is a relatively new biophysical characterization technique that might to some extent function as an alternative to powerful methods as X-ray crystallography and NMR spectroscopy. SDSL-ESR provides both local structural and dynamic information on proteins ⁴⁷ and has been applied to membrane proteins ⁴³.

In SDSL-ESR a spin probe (nitroxide) is incorporated into the protein by attaching it to a cysteine side chain (see Figure 3). In the presence of paramagnetic species, e.g. a nitroxide, which contains an unpaired electron, ESR absorption is observed ²⁵.

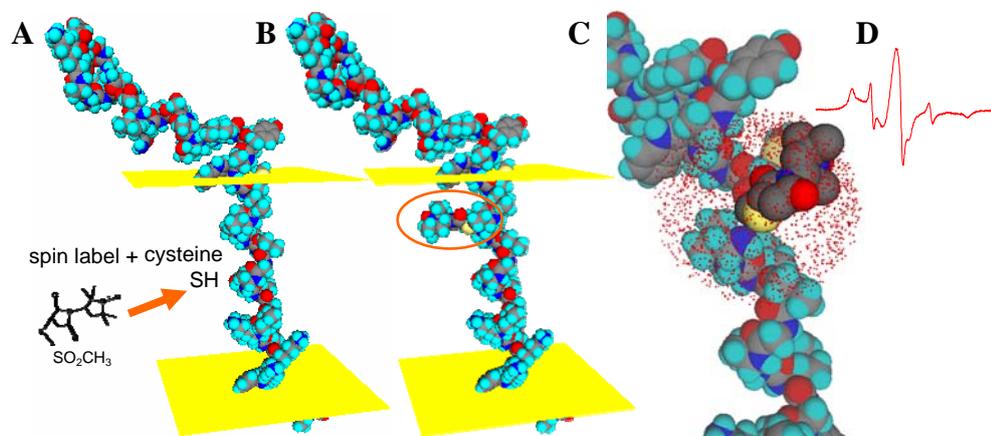


Figure 3. Site directed spin labeling ESR spectroscopy. Nitroxide spin label is attached to the cysteine (A), which replaces the original amino acid at the strategically chosen site (B). Anisotropy of the spin label side chain fast motion (C) is revealed by the line-shape of the ESR spectrum (D). Analysis of the ESR spectrum provides parameters, which describe the rate of the spin label dynamics and the conformational restrictions at a local protein site. In (C) small red points represent the single spin label conformations (rotamers).

One of the principal applications of spin label ESR is to study the mobility of nitroxide-labeled molecules⁴⁸. The dynamics, which can be extracted from an ESR spectrum, occurs on a picosecond-nanosecond time-scale in physiological conditions^{48,49}. The sensitivity of the conventional spin label ESR spectra to molecular motion is determined by the transverse relaxation process (T_2 process)⁴⁸ and it is limited by the spectral anisotropies of motions faster than $\approx 10^{-8} - 10^{-7}$ s⁵⁰.

For a spin label bound to a protein, the potential surface that determines its motion, or the rotational conformational space, is very complex, involving interactions with the protein backbone, the adjacent side chains, and collisions with solvent molecules⁵¹. Thus SDSL is a powerful tool to study the local structure in the proteins, monitor conformational changes in protein topology^{45,47}, and to determine backbone fluctuations at high temperature conditions^{47,52,53}.

With respect to structure determination, dual-probe-SDSL⁵⁴ enables also distance measurements in biological molecules^{55,56}. Depending on the particular experimental method, distances from 0.4 to 8.0 nm can be accurately measured⁵⁷⁻⁵⁹. The disadvantage of this method is that distance analysis by ESR can only be done at low temperature (200 K)⁴³ although attempts have been made to estimate inter-residue distances at physiological temperatures⁶⁰. The information about this method can be found elsewhere^{58,61-65}.

1.2.2 Multi-component ESR spectra analysis

To determine the picture of the actual heterogeneity within biomembranes and at specific sites of proteins, a special methodology should be applied including advanced spectral analysis and inverse-problem solving techniques⁶⁶. Such an analysis is based on mathematical modeling, spectrum fitting and parameters optimization^{67,68}. As a large amount of information evolves from such an approach a special method of solution condensation called GHOST was developed to facilitate the analysis and interpretation of the experimental data^{67,68}. It combines solution density filtering, χ^2 goodness filtering, solution-space slicing, and group determination, leading to a graphical presentation of the spectral parameters (see Figure 4).

Due to protein conformational variations, conformational transitions, and the complexity of protein dynamics on the time scales detectable by ESR spectroscopy the measured spectrum at a single mutant position is often a superposition of several components⁶⁸. In general, each component of an ESR spectrum can be simulated on the basis of different dynamic models⁶⁹. In order to accurately resolve spectroscopic parameters of each spectral component (so that the total multi-component simulated spectrum fits the experimental spectrum) a good optimization method needs to be applied. The advanced multi-component spectral analysis^{67,68} (see Figure 4) was applied in this thesis (it is publicly available online as EPRSIM-C: A Spectral Analysis Package⁷⁰).

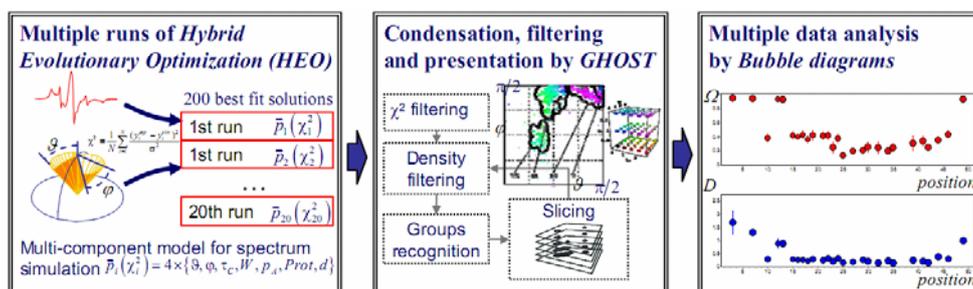


Figure 4. Overview of the method of ESR spectral analysis. Spectral analysis is based on multi-component spectral simulations and optimizations of spectral parameters (left), multiple solutions filtering, condensation and presentation (center), and multiple data analysis (right).

After the experimental data is measured and prepared (often a file conversion is needed when transferring the spectra from the spectrometer for further analysis) the analysis of ESR spectra can be organized into the following steps:

1. ESR spectral simulation and spectral parameters optimization. An appropriate simulation model has to be chosen for simulation of an experimental ESR spectrum. The maximal complexity, i.e., the number of spectral components, has to be defined. In addition optimization constants based on experimental parameters (ESR centre field and sweep, magnetization tensors, etc.) have to be defined as well as the initial spectral parameters. The

latter are then optimized with a hybrid evolutionary algorithm. The result of spectral simulations and optimization of spectral parameters is a population of multiple solutions, which fit the experimental spectrum (left box in Figure 4).

2. GHOST condensation. In order to make a relevant characterization based on spectral simulations the multiple solutions have to be filtered according to the quality of the fit and the solution density and the groups of solutions have to be recognized (central box in Figure 4) into so-called motional patterns. Detected motional patterns then have to be checked by studying corresponding contributions in the total spectrum line-shape. Sometimes a meaningless combination of spectral parameters may appear when the optimization algorithm tries to fit a spectral line-shape with artifacts. Such motional patterns are eliminated from further analysis, although they may be used to detect artifacts.

Initially, GHOST condensation and GHOST presentation algorithms (central box in Figure 4) were developed in the Mathematica environment, which was however too slow for high-throughput analysis. As a part of this thesis all the analysis algorithms were reprogrammed in a faster and more flexible independent software program called GHOSTMaker, a part of EPRSIM-C: A Spectral Analysis Package ⁷⁰.

3. Multiple data analysis. In multiple ESR data analysis (several mutant positions, temperature or concentration series) the first two steps are repeated for each spectrum of a series. In case of the temperature measurements, an additional check of the detected motional patterns may use the fact that the local dependence of a spectral parameter (correlation time, local motional restriction, polarity, etc) on the temperature has to be monotonous. The recent version of GHOSTMaker is capable of simultaneous presentation of multiple ESR data (right box in Figure 4), allowing also the comparison of several spectral series and data export to other analysis software packages.

The efficiency of the spectra optimization algorithms has been enhanced as described in Chapter 2. This analysis approach then was applied to interpret the ESR data from the light-harvesting complex CP29 (Chapter 3), and to develop a new SDSL-ESR-based method of protein structure characterization (Chapters 4 and 5).

1.3 Protein structure modeling and optimization

1.3.1 Protein structure modeling

Five choices have to be made when modeling a biomolecular system: 1) the scale of structure and dynamics have to be defined; 2) the degrees of freedom for the elementary particles (e.g. atoms, atom groups) that define the dimension of the conformational space have to be determined; 3) the force field (what interactions are taken into account) has to be chosen; 4) a sampling scheme of the conformational space has to be set up; and 5) boundary conditions (for dynamics simulations) have to be identified ⁷¹.

In order to produce efficient simulations that create calculation output in a short period of time, the model may split dynamics of the biomolecular system from its structure. Thus, the modeling of a static protein structure has no time scale and relies only on the data about the structures of amino acids, chemical bond lengths and bond angles. The backbone (main chain) of a protein is modeled by setting the secondary structure. Then amino-acid side chains are attached to the backbone. To define particular conformations of amino acid side chains the rotamer libraries can be used^{6,7}. At this step the structure of the protein can be compared with one in the Protein Database (if it is available) and/or checked by employing the potential energy calculations, while determining possible steric conflicts. More advanced methods of structure validation also include information about various structure stabilizing effects, e.g., sulfide bonds, hydrogen bonding, and other stabilizing weak interactions.

The identification of the interactions that stabilize protein structures has provided the framework for the development of computational models of protein structure and dynamics. To provide an accurate representation of the protein, these models include terms that reflect bond stretching, bending, and rotation. Although bond lengths and angles are formally determined by interactions of electrons and nuclei as described by quantum mechanics, these interactions can be treated by simple physical models. For example, the bond-stretching potential, $V(r)$, is determined by calculating the distance for each covalent bond, r , and comparing that distance to an ideal value, $r_{standard}$. A similar expression can be written for bending and also for bond rotation. All potentials are collected into a single expression, and each atom is uniquely identified by specific interactions with every other atom. The expectation for a protein is that the structure will adopt a conformation that represents the lowest-possible-energy state as given by the total potential energy⁴.

1.3.2 Comparison to molecular dynamics simulations

The application of MD simulations has started to make new and specific quantitative predictions about biological properties not yet reported from experiment⁷²⁻⁷⁴. The constant increase of modern computer computational power and further development of MD simulation packages including coarse-grained methods^{75,76} will gradually solve the current difficulties of short simulation time length and poor conformational sampling. In addition, certain work has to be done to verify simulations, force fields and simulation methods^{72,77}.

As compared to related papers on SDSL-ESR based on MD simulations^{51,78-85} our approach has the following advantages: 1) the simplicity of the underlying physical principles in the structural model; 2) the simultaneous analysis of multiple SDSL-ESR data from all available mutant positions; 3) no need for dynamics trajectories as compared to MD (our approach completely relies on conformational space modeling as described in the Chapter 4); 4) a 3-4 orders of magnitude faster speed as compared to calculations based on MD simulations.

1.3.3 Optimization methods

Optimization is an important tool in the analysis of complex systems. Mathematically speaking, optimization is the process of minimizing or maximizing of a function subject to constraints or variables⁸⁶. A good example is maximization of the efficiency in the design and operation in a manufacture production process, e.g. optimization of a gas turbine for the maximum efficiency and minimal costs.

In our work we apply optimization when solving an inverse-problem, i.e., searching for the parameters of a model that lead to a satisfactory description of the experimental data. First of all, this requires optimization of the simulated ESR spectra (minimization of the fitting criterion). This is the subject of Chapter 2. Secondly, this requires the optimization of the structure of the protein system (tuning the structure in order to improve the fit of the simulated local restrictions to the experimentally obtained restrictions), which is discussed in Chapter 5.

Optimization is a computationally demanding process, and that is why it is usually implemented for computer calculations in optimization algorithms. There are no universal optimization algorithms, and each particular problem may require a separate algorithm or at least a special modification of an existing algorithm. Several optimization methods (stochastic and deterministic) may be combined. For example, when searching for multiple optimal solutions two major characteristics should be satisfied: the diversity and the accuracy of multiple solutions. To provide both, an algorithm must be able to efficiently perform both two corresponding processes: exploration and exploitation of the search space.

In general, the optimization problem often can be simplified. For example, in case of correlated parameters the number of parameters can be reduced. The dimension of the search space can also be reduced by separating linear and nonlinear parameters as there is often a fast analytical solution for linear parameters optimization.

In both of our cases we deal with a multi-modal optimization (multiple local minima) of biophysical models of many parameters. We minimize the fitting function, which describes how well the simulated data is like the experimental data. The parameters are usually defined within certain intervals.

1.4 Photosynthetic minor antenna complex CP29

Oxygenic photosynthesis is one of the most fundamental processes sustaining life on Earth. Solar energy is converted into the chemical energy of an ATP molecule, and the reducing equivalents used for the conversion of CO₂ into carbohydrates (the building blocks of biomass) are generated. The first step in this process, light-driven charge separation, is conducted by Photosystem I and Photosystem II, two chlorophyll-binding protein complexes embedded in the thylakoid membranes of cyanobacteria, algae and plants^{87,88}.

In Photosystem II the outer antenna system consists of 4 complexes: LHCII, the major antenna subunit, present as a trimer in the membrane, and three minor monomeric complexes CP24, CP26, CP29^{89,90} (See Figure 5A). The minor complexes contain 15% of the pigments of PSII and they are located between LHCII^{91,92} and the core that contains the reaction centers where charge separation takes place.

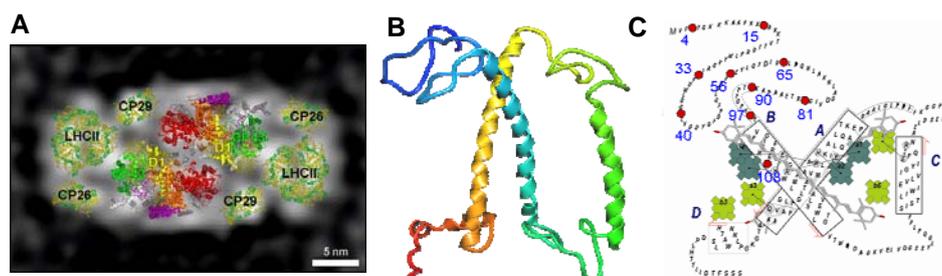


Figure 5. Photosynthetic minor antenna complex CP29. **A.** Overlay of X-ray structures of protein complexes on to the luminal top view of the 3D structure of the spinach LHCII–PSII supercomplex derived from cryo-EM and single particle analysis⁹⁰. **B.** 3D structure of LHCII (prepared in Argus Lab). **C.** Schematic presentation of the structure of CP29⁹³ with positions of Chl a (dark green), positions of Chl a or b (light green) and the two central carotenoids (grey)⁹⁴. The red dots and numbers indicate the position of the labeled residues used in this work.

Besides light-harvesting, minor antenna complexes are also involved in photoprotective mechanisms that are used by plant to dissipate excess of energy under high-light conditions^{95,96}. Structural changes may be responsible for a switch between light-harvesting state and the quenching state⁹⁶. It has been proposed that phosphorylated LHCII physically moves from PSII to PSI to balance the energy distribution between and optimize the rate of electron transfer through the two photosystems or induce cyclic electron flow around PSI⁹⁷⁻⁹⁹. In this state transition, CP29 may provide a functional link between a mobile LHCII antenna and the PSI core^{88,100}.

Structural information on the minor antenna complexes CP24, CP26 and CP29 is still lacking, but sequence analysis and site-selected mutagenesis have revealed that they share a high structure similarity with LHCII^{94,101,102}. CP29 is the largest of the outer antenna complexes of PSII, and it contains a long N-terminal domain (around 100 amino acids). Although the structure of CP29 has not been resolved, the high sequence homology, especially in transmembrane domain suggest an organization similar to that of LHCII, while the organization of the N-terminal domain differs from that of LHCII.

To obtain structural and dynamical information of CP29, in particular on its N-terminal domain picosecond fluorescence techniques have been used⁹³ providing distances between specific sites of the N-terminal domain and the chlorophyll molecules. We used an alternative method, SDSL-ESR spectroscopy in combination with multi-component spectral analysis to obtain additional insight into the structural organization and dynamical properties of the N-terminal domain of CP29 (see Chapter 3).

1.5 Major coat protein of bacteriophage M13

The major coat protein of bacteriophage M13 was chosen as a simple model system to develop and tune the protein structure optimization approach.

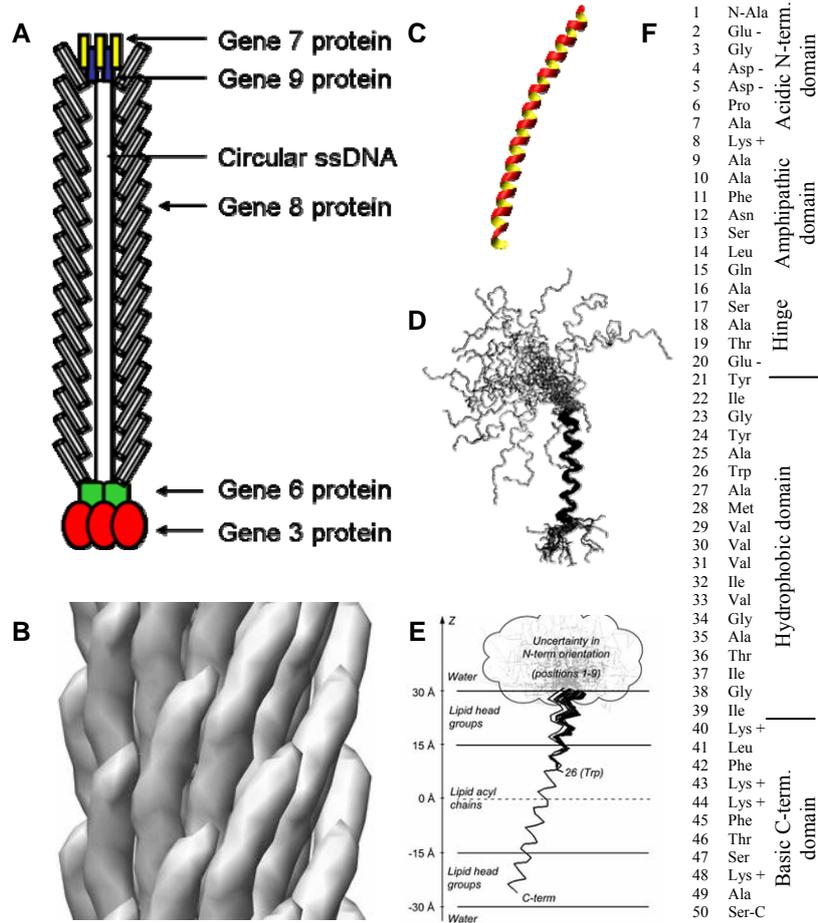


Figure 6. Major coat protein of bacteriophage M13. **A.** Schematic illustration of bacteriophage M13 filament. **B.** Arrangement of the coat protein subunits in M13¹⁰³. **C.** Curved α -helix model of M13 coat protein found by X-ray crystallography¹⁰⁴ deposited in the Protein Data Bank as PDB file 1IFJ. **D.** Best-fit superposition of 25 structures of M13 coat protein in SDS micelles¹⁰⁵ PDB file 2CPS. **E.** FRET data-based 100 best-fit structures of AEDANS-labeled M13 coat protein in DOPC/DOPG vesicles¹⁰⁶. **F.** Primary structure of M13 coat protein (gp8).

Bacteriophage M13 has been thoroughly studied by various biophysical techniques and the structure of the virion protein sheet was determined by X-ray fiber diffraction. The viral particle is composed of single-stranded circular DNA molecule that is encapsulated in a long cylindrical protein coat. The protein coat is composed of about 2800 copies of the

major coat protein (gp8). At both termini there are five copies of each of the two minor coat proteins, gp7 and gp9 at one end and gp3 and gp6 at the other end (see Figure 6) ^{107,108}. A detailed description of the filamentous bacteriophages has been reviewed ¹⁰⁸. In addition, extensive knowledge about bacteriophage M13 and recent technological advances has led to a successful application of M13 virus in Nanotechnology ^{109,110}.

The major coat protein is a small protein with a molecular weight of about 5240 Da. It forms a 1.5-2.0 nm thick flexible cylindrical shell (see Figure 6B). It is 50 amino acids long and it is composed of three specific domains: a hydrophobic core, an acidic N-terminal and a basic C-terminal domain (Figure 6F). The secondary structure is largely α -helical (also proven by primary sequence prediction) with several flexible positions in the N terminus ¹⁰⁸.

After integration into the lipid bilayer, the M13 coat protein adopts a transmembrane configuration. The structure of the protein, the dynamics, and protein embedment into the lipids were studied with X-ray crystallography ¹⁰⁴ (Figure 6C), NMR spectroscopy ^{105,111} (Figure 6D), site-directed labeling in combination with ESR ^{108,112-118} and fluorescence spectroscopy ^{106,119-122} (Figure 6E), circular dichroism spectroscopy ¹²³ and other methods.

According to X-ray crystallography the protein is a slightly curved α -helix extending from the N-terminus to the C-terminus ¹⁰⁴. Studies based on NMR spectroscopy ^{105,124,125} suggest that in micelles the protein contains two α -helical segments, residues 7-16 and 25-45. NMR spectroscopy in a dehydrated lipid bilayer ¹²⁶ resulted in a 3D structure where the first α -helical segment (residues 8-18) rests on the membrane surface, the transmembrane α -helix (residues 21-45) crosses the membrane at an angle of 26° up to residue Lys40, whereas the helix tilt changes to 16° , which was also observed previously by solid-state NMR ¹²⁷.

The results obtained with SDSL-ESR spectroscopy for the protein in different lipid bilayers (14:1PC-22:1PC) ^{113-115,117} were generally in accordance with two α -helical segments model, while the modeling based on the recent fluorescence data obtained for the protein in DOPC/DOPG vesicles suggested a model of two α -helical domains with an unstructured region (residues 1-9) and a general tilt (residues 12-46) of 18° with respect to the membrane normal. The state of art regarding M13 protein can be found in a recent publication ¹⁰².

We used M13 coat protein to test our method of calculating the restrictions of the conformational space (Chapter 4) and to develop a new protein structure optimization approach (Chapter 5).

1.6 About this thesis

In this thesis the problem of membrane protein structure characterization is tackled with a spectroscopic SDSL-ESR technique enhanced with computational methods, i.e., multi-component spectrum analysis, modeling of protein conformations and protein structure optimization.

Chapter 2 describes the details of multi-component ESR spectra analysis. A so-called GHOST condensation method is employed for filtering, grouping and presenting of the multi-solution results. This analysis is being enhanced to enable the handling of large sets of spectroscopic data (e.g., for the joint analysis of SDSL-ESR spectra from multiple sites of spin-labeled proteins).

Chapter 3 describes the application of this approach to analyze the ESR data from ten spin label positions at the N-terminal domain of the light-harvesting protein complex CP29. The multi-component ESR spectra analysis permit to trace the structural organization of the long N-terminal domain of CP29 leading to a structural model for its N-terminal domain.

Chapter 4 introduces a novel way to translate the local structural constraints gained by SDSL-ESR data into a low-resolution structure of a protein by simulating the restrictions of the local conformational spaces of the spin label attached at different protein sites along the primary structure of the membrane-embedded protein. The proposed structural model takes into account the restricting effect of the protein backbone, amino acid side chains and lipid environment.

In Chapter 5, the sensitivity of this approach is tested for artificial oligopeptides and then for membrane-embedded M13 major coat protein decorated with a limited number of strategically placed spin labels by employing high-throughput site-directed mutagenesis. This chapter introduces an optimization algorithm, which is used to optimize the parameters of the protein-lipid model by improving the fit of the simulation data to the experimental conformational space data. The result of the optimization, i.e., a family of best-fit structures of membrane-embedded M13 protein consistent with the available SDSL-ESR data is compared to a recent model based on site-directed fluorescence labeling.

Chapter 6 summarizes this thesis with an overview of our latest progress in the field of protein structure characterization based SDSL-ESR spectroscopy, protein conformational space modeling, and structural optimization.

References

- 1 Lehninger, A.L., Nelson, D.L., & Cox, M.M. *Lehninger principles of biochemistry*, 4th ed. (W.H. Freeman, New York, 2005).
- 2 Ramachandran, G.N. & Sasisekharan, V. *Adv. Protein Chem.* **23**, 283 (1968).
- 3 Capaldi, S. *et al. J. Biol. Chem.* **282**, 31008 (2007).
- 4 Allen, J.P. *Biophysical chemistry*. (Wiley-Blackwell Pub., Oxford; Hoboken, NJ, 2008).
- 5 Ramachandran, G.N., Ramakrishnan, C., & Sasisekharan, V. *J. Mol. Biol.* **7**, 95 (1963).
- 6 Dunbrack, J.R.L. & Karplus, M. *J. Mol. Biol.* **230**, 543 (1993).
- 7 Lovell, S.C., Word, J.M., Richardson, J.S., & Richardson, D.C. *Proteins* **40**, 389 (2000).
- 8 Henzler-Wildman, K. & Kern, D. *Nature* **450**, 964 (2007).
- 9 Ansari, A. *et al. Proc. Natl. Acad. Sci. U. S. A.* **82**, 5000 (1985).
- 10 Singer, S.J. & Nicolson, G.L. *Science* **175**, 720 (1972).
- 11 Edidin, M. *Nat. Rev. Mol. Cell Biol.* **4**, 414 (2003).
- 12 Venturoli, M., Maddalena Sperotto, M., Kranenburg, M., & Smit, B. *Phys. Rep.* **437**, (2006).
- 13 Israelachvili, J.N. *Biochim. Biophys. Acta* **469**, 221 (1977).
- 14 Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G., & Engelman, D.M. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4083 (2004).
- 15 Killian, J.A. & von Heijne, G. *Trends Biochem. Sci.* **25**, 429 (2000).
- 16 Killian, J.A. & Nyholm, T.K.M. *Curr. Opin. Struct. Biol.* **16**, 473 (2006).
- 17 Pietzsch, J. *Mind the membrane in A living frontier – exploring the dynamics of the cell membrane* (Palazzo Arzaga, Italy, 2004).
- 18 Gawrisch, K. The dynamics of membrane lipids in *The structure of biological membranes*, edited by Philip Yeagle (CRC Press, Boca Raton, Fla, 2005), pp. 147.
- 19 Jensen, M.Ø. & Mouritsen, O.G. *Biochim. Biophys. Acta* **1666**, 205 (2004).
- 20 Anderson, R.G.W. & Jacobson, K. *Science* **296**, 1821 (2002).
- 21 Jacobson, K., Mouritsen, O.G., & Anderson, R.G.W. *Nat. Cell Biol.* **9**, 7 (2007).
- 22 Marsh, D. *Methods* **46**, 83 (2008).
- 23 Marsh, D. *Biochim. Biophys. Acta* **1778**, 1545 (2008).
- 24 Marsh, D. & Pali, T. *Biochim. Biophys. Acta* **1666**, 118 (2004).
- 25 Torres, J., Stevens, T.J., & Samsø, M. *Trends Biochem. Sci.* **28**, 137 (2003).
- 26 Bernsel, A. *et al. Proc. Natl. Acad. Sci. U. S. A.* **105**, 7177 (2008).
- 27 von Heijne, G. *Nat. Rev. Mol. Cell Biol.* **7**, 909 (2006).
- 28 de Planque, M.R.R. *et al. Biochem.* **42**, 5341 (2003).
- 29 de Planque, M.R. & Killian, J.A. *Mol. Membr. Biol.* **20**, 271 (2003).
- 30 Luckey, M. *Membrane structural biology: with biochemical and biophysical foundations*. (Cambridge University Press, Cambridge ; New York, 2008).
- 31 Palsdottir, H. & Hunte, C. *Biochim. Biophys. Acta* **1666**, 2 (2004).
- 32 Bernstein, F.C. *et al. J. Mol. Biol.* **112**, 535 (1977).
- 33 Webpage. RCSB Protein Data Bank, available at <http://www.rcsb.org/pdb/Welcome.do>, (2009).
- 34 Henderson, R. *Q. Rev. Biophys.* **37**, 3 (2004).
- 35 Fleishman, S.J., Unger, V.M., & Ben-Tal, N. *Trends Biochem. Sci.* **31**, 106 (2006).
- 36 Muller, D.J. & Engel, A. *Curr. Opin. Colloid Interface Sci.* **13**, 338 (2008).

- 37 Pebay-Peyroula, E. *Biophysical analysis of membrane proteins: investigating structure and function*. (Wiley-VCH, Weinheim, 2008).
- 38 Kelly, S.M. & Price, N.C. *Curr. Protein Pept. Sci.* **1**, 349 (2000).
- 39 Fasman, G.D. *Circular dichroism and the conformational analysis of biomolecules*. (Plenum Press, New York, 1996).
- 40 Uversky, V.N. *Protein Sci.* **11**, 739 (2002).
- 41 Svergun, D.I. & Koch, M.H.J. *Rep. Prog. Phys.* **66**, 1735 (2003).
- 42 Petoukhov, M.V. & Svergun, D.I. *Biophys. J.* **89**, 1237 (2005).
- 43 Hemminga, M.A. Introduction and future of site-directed spin labeling of membrane proteins in *ESR spectroscopy in membrane biophysics*, edited by Marcus A. Hemminga & Lawrence Berliner (Springer, 2007), pp. 1.
- 44 Hubbell, W.L. & Altenbach, C. *Curr. Opin. Struct. Biol.* **4**, 566 (1994).
- 45 Hubbell, W.L., Cafiso, D.S., & Altenbach, C. *Nat. Struct. Biol.* **7**, 735 (2000).
- 46 Hubbell, W.L., McHaourab, H.S., Altenbach, C., & Lietzow, M.A. *Structure* **4**, 779 (1996).
- 47 Lietzow, M.A. & Hubbell, W.L. **43**, 3137 (2004).
- 48 Hoff, A.J. *Advanced EPR: applications in biology and biochemistry*. (Elsevier, Amsterdam New York, 1989).
- 49 Steinhoff, H.-J., Müller, M., Beier, C., & Pfeiffer, M. *J. Mol. Liq.* **84**, 17 (2000).
- 50 Marsh, D. & Horvath, L.I. Spin label studies of the structure and dynamics of lipids and proteins in membranes in *Advanced EPR: applications in biology and biochemistry*, edited by A.J. Hoff (Elsevier, 1989), pp. 707.
- 51 Steinhoff, H.J. & Hubbell, W.L. *Biophys. J.* **71**, 2201 (1996).
- 52 Columbus, L. & Hubbell, W.L. *Biochem.* **43**, 7273 (2004).
- 53 Columbus, L., Kalai, T., Jeko, J., Hideg, K., & Hubbell, W.L. *Biochem.* **40**, 3828 (2001).
- 54 Fanucci, G.E. & Cafiso, D.S. *Curr. Opin. Struct. Biol.* **16**, 644 (2006).
- 55 Borbat, P.P., McHaourab, H.S., & Freed, J.H. *J. Am. Chem. Soc.* **124**, 5304 (2002).
- 56 Hustedt, E.J. *et al. Biophys. J.* **90**, 340 (2006).
- 57 Sale, K., Song, L., Liu, Y.S., Perozo, E., & Fajer, P. *J. Am. Chem. Soc.* **127**, 9334 (2005).
- 58 Berliner, L.J., Eaton, G.R., & Eaton, S.S. *Distance measurements in biological systems by EPR*. (Kluwer Academic/Plenum Publishers, New York, 2000).
- 59 Tsvetkov, Y.D., Milov, A.D., & Maryasov, A.G. *Russ. Chem. Rev.* **77**, 487 (2008).
- 60 Altenbach, C., Oh, K.J., Trabanino, R.J., Hideg, K., & Hubbell, W.L. *Biochem.* **40**, 15471 (2001).
- 61 Jeschke, G., Bender, A., Paulsen, H., Zimmermann, H., & Godt, A. *J. Magn. Reson.* **169**, 1 (2004).
- 62 Jeschke, G., Wegener, C., Nietschke, M., Jung, H., & Steinhoff, H.J. *Biophys. J.* **86**, 2551 (2004).
- 63 Jeschke, G. & Polyhach, Y. *Phys. Chem. Chem. Phys.* **9**, 1895 (2007).
- 64 Borbat, P.P. & Freed, J.H. *Chem. Phys. Lett.* **313**, 145 (1999).
- 65 Schweiger, A. & Jeschke, G. *Principles of pulse electron paramagnetic resonance*. (Oxford University Press, Oxford, UK ; New York, 2001).
- 66 Štrancar, J., Koklic, T., & Arsov, Z. *J. Membr. Biol.* **196**, 135 (2003).
- 67 Štrancar, J. Advanced ESR spectroscopy in membrane biophysics in *ESR spectroscopy in membrane biophysics*, edited by Marcus A. Hemminga & Lawrence Berliner (Springer, 2007), pp. 49.

- 68 Štrancar, J. *et al. J. Chem. Inf. Mod.* **45**, 394 (2005).
- 69 Schneider, D.J. & Freed, J.H. Calculating slow motional magnetic resonance spectra: A user's guide in *Biological magnetic resonance: spin labeling, theory and applications*, edited by L. J. Berliner & J. Reuben (Plenum Press, New York, 1989), pp. 1–76.
- 70 Štrancar, J. EPRSIM-C: A spectral analysis package, available at http://www.ijs.si/ijs/dept/epr/EPRSIMC_overview.htm, (2009).
- 71 van Gunsteren, W.F. *et al. Angew. Chem. Int. Ed. Engl.* **45**, 4064 (2006).
- 72 Lindahl, E. & Sansom, M.S. *Curr. Opin. Struct. Biol.* **18**, 425 (2008).
- 73 Gumbart, J., Wang, Y., Aksimentiev, A., Tajkhorshid, E., & Schulten, K. *Curr. Opin. Struct. Biol.* **15**, 423 (2005).
- 74 Scott, K.A. *et al. Structure* **16**, 621 (2008).
- 75 Bond, P.J., Holyoake, J., Ivetac, A., Khalid, S., & Sansom, M.S.P. *J. Struct. Biol.* **157**, 593 (2007).
- 76 Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., & de Vries, A.H. *J. Phys. Chem. B* **111**, 7812 (2007).
- 77 Ulmschneider, M.B. & Ulmschneider, J.P. *J. Chem. Theory Comput.* **4**, 1807 (2008).
- 78 Håkansson, P., Westlund, P.O., Lindahl, E., & Edholm, O. *Phys. Chem. Chem. Phys.* **3**, 5311 (2001).
- 79 LaConte, L.E., Voelz, V., Nelson, W., Enz, M., & Thomas, D.D. *Biophys. J.* **83**, 1854 (2002).
- 80 Stoica, I. *J. Phys. Chem. B* **108**, 1771 (2004).
- 81 Pistolesi, S. *et al. Biophys. Chem.* **123**, 49 (2006).
- 82 Budil, D.E., Sale, K.L., Khairy, K.A., & Fajer, P.G. **110**, 3703 (2006).
- 83 Beier, C. & Steinhoff, H.-J. *Biophys. J.* **91**, 2647 (2006).
- 84 DeSensi, S.C., Rangel, D.P., Beth, A.H., Lybrand, T.P., & Hustedt, E.J. *Biophys. J.* **94**, 3798 (2008).
- 85 Sezer, D., Freed, J.H., & Roux, B. *J. Chem. Phys.* **128**, 165106 (2008).
- 86 Nocedal, J. & Wright, S.J. *Numerical optimization*. (Springer, New York, 1999).
- 87 Allen, J.F. & Martin, W. *Nature* **445**, 610 (2007).
- 88 Kargul, J. & Barber, J. *FEBS J.* **275**, 1056 (2008).
- 89 Dekker, J.P. & Boekema, E.J. *Biochim. Biophys. Acta* **1706**, 12 (2005).
- 90 Nield, J. & Barber, J. *Biochim. Biophys. Acta* **1757**, 353 (2006).
- 91 Boekema, E.J., van Roon, H., van Breemen, J.F.L., & Dekker, J.P. *Eur. J. Biochem.* **266**, 444 (1999).
- 92 Caffarri, S., Croce, R., Cattivelli, L., & Bassi, R. *Biochem.* **43**, 9467 (2004).
- 93 van Oort, B. *et al. Chemical Physics* **357**, 113 (2009).
- 94 Bassi, R., Croce, R., Cugini, D., & Sandonà, D. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10056 (1999).
- 95 Mozzo, M., Passarini, F., Bassi, R., van Amerongen, H., & Croce, R. *Biochim. Biophys. Acta* **1777**, 1263 (2008).
- 96 Ahn, T.K. *et al. Science* **320**, 794 (2008).
- 97 Allen, J.F. *Science* **299**, 1530 (2003).
- 98 Kruse, O. *Naturwissenschaften* **88**, 284 (2001).
- 99 Haldrup, A., Jensen, P.E., Lunde, C., & Scheller, H.V. *Trends Plant Sci.* **6**, 301 (2001).
- 100 Kargul, J. *et al. FEBS J.* **272**, 4797 (2005).

- 101 Gastaldelli, M., Canino, G., Croce, R., & Bassi, R. *J. Biol. Chem.* **278**, 19190 (2003).
- 102 Vos, W.L., Nazarov, P.V., Koehorst, R.B.M., Spruijt, R.B., & Hemminga, M.A. *Trends Biochem. Sci.* **34**, 249 (2009).
- 103 Marvin, D.A., Welsh, L.C., Symmons, M.F., Scott, W.R., & Straus, S.K. *J. Mol. Biol.* **355**, 294 (2006).
- 104 Marvin, D.A. *Curr. Opin. Struct. Biol.* **8**, 150 (1998).
- 105 Papavoine, C.H., Christiaans, B.E., Folmer, R.H., Konings, R.N., & Hilbers, C.W. *J. Mol. Biol.* **282**, 401 (1998).
- 106 Nazarov, P.V., Koehorst, R.B.M., Vos, W.L., Apanasovich, V.V., & Hemminga, M.A. *Biophys. J.* **92**, 1296 (2007).
- 107 Makowski, L. *J. Mol. Biol.* **228**, 885 (1992).
- 108 Stopar, D., Spruijt, R.B., Wolfs, C.J., & Hemminga, M.A. *Biochim. Biophys. Acta* **1611**, 5 (2003).
- 109 Lee, S.W. & Belcher, A.M. *Virus-based fabrication of micro- and nanofibers using electrospinning*. (2004), Vol. 4, pp. 387.
- 110 Mao, C. *et al. Science* **303**, 213 (2004).
- 111 Opella, S.J., Zeri, A.C., & Park, S.H. *Annu. Rev. Phys. Chem.* **59**, 635 (2008).
- 112 Spruijt, R.B., Wolfs, C.J., Verver, J.W., & Hemminga, M.A. *Biochem.* **35**, 10383 (1996).
- 113 Meijer, A.B., Spruijt, R.B., Wolfs, C.J., & Hemminga, M.A. *Biochem.* **40**, 8815 (2001).
- 114 Meijer, A.B., Spruijt, R.B., Wolfs, C.J., & Hemminga, M.A. *Biochem.* **40**, 5081 (2001).
- 115 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *J. Chem. Inf. Mod.* **45**, 1621 (2005).
- 116 Stopar, D., Spruijt, R.B., & Hemminga, M.A. *Chem. Phys. Lipids* **141**, 83 (2006).
- 117 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *Biophys. J.* **91**, 3341 (2006).
- 118 Štrancar, J., Kavalenka, A., Zihlerl, P., Stopar, D., & Hemminga, M.A. *J. Magn. Reson.* **197**, 245 (2009).
- 119 Koehorst, R.B.M., Spruijt, R.B., Vergeldt, F.J., & Hemminga, M.A. *Biophys. J.* **87**, 1445 (2004).
- 120 Fernandes, F. *et al. Biophys. J.* **87**, 344 (2004).
- 121 Nazarov, P.V., Koehorst, R.B.M., Vos, W.L., Apanasovich, V.V., & Hemminga, M.A. *Biophys. J.* **91**, 454 (2006).
- 122 Vos, W.L., Koehorst, R.B.M., Spruijt, R.B., & Hemminga, M.A. *J. Biol. Chem.* **280**, 38522 (2005).
- 123 Sanders, J.C., Haris, P.I., Chapman, D., Otto, C., & Hemminga, M.A. *Biochem.* **32**, 12446 (1993).
- 124 Almeida, F.C. & Opella, S.J. *J. Mol. Biol.* **270**, 481 (1997).
- 125 Papavoine, C.H., Konings, R.N., Hilbers, C.W., & van de Ven, F.J. *Biochem.* **33**, 12990 (1994).
- 126 Marassi, F.M. & Opella, S.J. *Protein Sci.* **12**, 403 (2003).
- 127 Cross, T.A. & Opella, S.J. *J. Mol. Biol.* **182**, 367 (1985).

Chapter 2

Speeding-up a genetic algorithm for ESR-based spin label characterization of biosystem complexity

based on:

Kavalenka, A.A., B. Filipic, M.A. Hemminga, J. Strancar. 2005. Speeding up a genetic algorithm for ESR-based spin label characterization of biosystem complexity. *J. Chem. Inf. Mod.* 45:1628-1635.

Abstract

Complexity of biological systems is one of the toughest problems for any experimental technique. Complex biochemical composition and a variety of biophysical interactions governing the evolution of a state of a biological system imply that the experimental response of the system would be superimposed of many different responses. To obtain a reliable characterization of such a system based on spin-label Electron Spin Resonance (ESR) spectroscopy, multiple Hybrid Evolutionary Optimization (HEO) combined with spectral simulation can be applied. Implemented as the GHOST algorithm this approach is capable of handling the huge solution space and provides an insight into the "quasi-continuous" distribution of parameters that describe the biophysical properties of an experimental system. However, the analysis procedure requires several hundreds of runs of the evolutionary optimization routine making this algorithm extremely computationally demanding. As only the best parameter sets from each run are assumed to contribute into the final solution, this algorithm appears far from being optimized. The goal of this study is to modify the optimization routine in a way that 20-40 runs would be enough to obtain qualitatively the same characterization. However, to keep the solution diversity throughout the HEO run, fitness sharing and newly developed shaking mechanisms are applied and tested on various test ESR spectra. In addition, other evolutionary optimization parameters such as population size and probability of genetic operators were also varied to tune the algorithm. According to the testing examples a speed-up factor of 5-7 was achieved.

2.1 Introduction

Complexity is one of the basic properties of natural biological systems. It qualitatively describes the number of (biochemical or biophysical) patterns/solutions that coexist in a system. In a pure system, only one solution can describe the entire system, whereas in complex systems distributions of solutions can exist (see Figure 1).

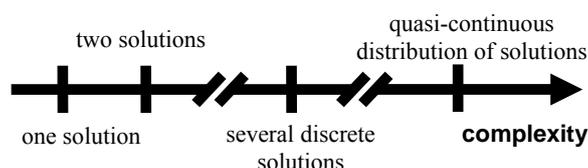


Figure 1. *Biosystem complexity axis of increasing complexity from simple single-solution to quasi-continuous distribution of solutions.*

The complexity of a biological membrane, for example, originates in its biochemical composition of a few hundred of lipids and many different proteins – channels and pumps, as well as membrane enzymes and receptors. In such a system, the constituents exhibit

different interactions to each other, from local steric and Van der Waals to more long-ranged Coulomb and dipolar interactions. The intensity and orientation of these interactions strongly depend on the type of interacting molecules as well as the potentials of the neighboring molecules. All these parameters make the biological membrane a very complex system in which many motional patterns can be found.

ESR spectroscopy in combination with nitroxide spin labeling (SL-ESR) has proven to be a powerful technique for the exploration of heterogeneity and motion in biological systems^{1,2}. The time scale of SL-ESR appears to be in the nanoseconds range, which is exactly the range needed to observe possible motional anisotropy of local rotational motions through motional averaging. The difference in anisotropy of rotational motion can be used to distinguish lateral domains together with other spectroscopic parameters such as the rate of motion, polarity, spin-spin broadening, *etc.* However, to determine the picture of the actual heterogeneity within biomembranes, a special methodology that includes advanced spectral analysis and inverse-problem solving techniques needs to be applied³. Such an analysis is based on mathematical modeling, spectrum fitting and spectral parameter optimization by means of evolutionary computation. A large amount of information evolves from such an approach. Therefore a special method of solution condensation called GHOST was developed¹. It incorporates solution density filtering, χ^2 goodness filtering, solution-space slicing, and domain determination, leading to a graphical presentation of the system parameters. This advanced approach named Hybrid Evolutionary Optimization (HEO) was shown to be powerful enough to study complex heterogeneous systems¹, but the computational demand appeared to be an obstacle for wider usage of the method.

The core of the problem lies in the optimization routine. To obtain a reliable result even in the case of quasi-continuous problems, the HEO procedure has to be executed at least 200 times. Each particular run consists of 100 generations with a population size of 300 candidate solutions that are exposed to various genetic operators. Since an average operator spends up to 10 spectrum calculations, HEO on average spends 60 million spectrum calculations. As a single spectrum calculation takes around 10 ms on a 1 GFLOPS processor, this results in 200 hours of computer time spent for a single characterization. Therefore, our aim was to enhance the HEO routine to speed up the approach to make it more applicable.

2.2 Theory and methodology

2.2.1 ESR spin labeling

ESR spectroscopy in combination with spin labeling can be applied to study the properties of biological membranes in a nondestructive way. In this approach spin-labeled analogs of different molecules are introduced into a system to report about their structural and motional properties. Since the nitroxide moiety is a small perturbation to the whole molecule, one can approximate that the description derived from spin probes is a reasonable approximation for the non-labeled molecules. This fact enables us to use ESR to explore biological systems *in vivo* so that there is no need for (bio)chemical extraction the subsystem of interest. In this way, various coexisting states of the system can be detected and characterized.

As was mentioned in the Introduction, ESR spin labeling inherits a unique sensitivity to the motional and polarity properties of the labeled molecules providing an opportunity to extracting information on structure and dynamics of the lipids and membrane proteins (*i.e.* restriction and rate of rotational motions, relative membrane locations, and oxygen profile). The complexity of such a system results in a large number of solutions superimposed in the ESR spectrum of such a labeled system (Figure 2).

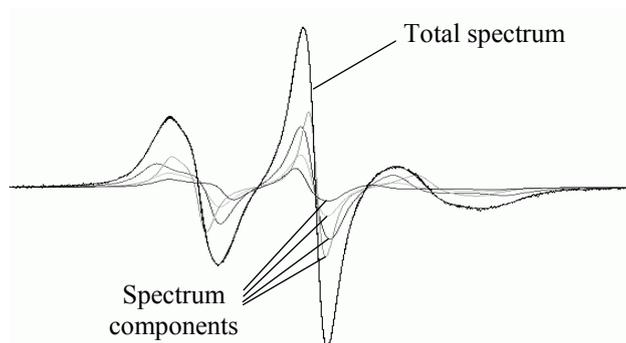


Figure 2. Superimposed four component ESR spectrum. The total ESR spectrum is a sum of four spectral components derived within described simulation model and determined by four sets of the following spectral parameters $\{\mathcal{G}, \varphi, \tau_c, W, p_A, prot, d\}$.

2.2.2 ESR spectrum modeling

Generally, to describe the ESR spectra of spin labels, the stochastic Liouville equation should be used⁴⁻⁶. However, under physiological condition the majority of the local rotational motions is fast with respect to the ESR time scale – as calculated by numerous molecular dynamics simulations – and therefore the fast motional approximation can be applied, reducing the computational demand by a factor of 100.

Since the basic approach has been already discussed elsewhere^{7,8}, we will emphasise only the physical background of the spectral parameters involved in our calculations. Firstly, one or two parameters are used in partial averaging of the rotational motion. While averaging the magnetic properties of the spin Hamiltonian for spin probes directed at every allowed direction with respect to the external magnetic field an order parameter S or opening cone angle \mathcal{G} (that defines the maximal tilt angle) and asymmetry cone angle φ (that describes the maximal restriction of spinning) will be applied. Secondly, the traces of the interaction tensors \mathbf{g} and \mathbf{A} are linearly corrected with p_A ⁹ and $Prot$ ¹⁰ parameters to take into account the effects of polarity and proticity, respectively. Thirdly, when calculating the convolution of the magnetic field distribution and basic lineshape, in addition two linewidth parameters are applied: a Lorentzian-type line is defined in the motional narrowing approximation¹¹ with a single (effective) rotational correlation time, τ_c and an additional broadening constant W . The latter arises primarily from unresolved hydrogen superhyperfine interactions and contributions of paramagnetic impurities (*e.g.* oxygen), external magnetic field inhomogeneities, field modulation effects, and spin-spin interaction.

To take into account the superposition of motional/polarity patterns, this basic set of six lineshape parameters \mathcal{G} , φ , τ_c , W , p_A and $Prot$ is expanded for the number of spectral components N_c . In addition there are N_c-1 weights d of these spectral components. Altogether, there are $7N_c-1$ spectral parameters, which have to be tuned by the optimization routine. Taking into account the resolution limit of SL-ESR which is around 30 parameters, this allows the usage of at most four spectral components.

2.2.3 Optimization

An optimization routine is used to find the set of spectral parameters that produces the best fit to the experimental spectrum. The goodness of fit (optimization objective function) was chosen to be the reduced χ^2 criteria:

$$\chi^2 = \frac{1}{N-p} \sum_{i=1}^N \frac{(y_i^{exp} - y_i^{sim})^2}{\sigma^2}, \quad (1)$$

where y^{exp} and y^{sim} are the experimental and simulated data, respectively, σ is the standard deviation of the experimental points, N is the number of spectral points, and p the number of model parameters.

For the optimization, HEO routine, a combination of the Genetic Algorithm (GA) with Downhill-Simplex local search was applied. Since the optimization scheme is presented elsewhere¹², we only briefly report on the implemented algorithm. The routine starts with a random initialization of solutions and continues with the tournament selection and application of genetic operators for 100 generations. The 3-point crossover with

probability of 0.7 and uniform mutation with probability of 0.01 are applied together with certain knowledge-based operators and local improvements (performed with Downhill-Simplex with probability of 0.002, see Figure 3)^{1,12}. The elite set (2% of the population size) is used to keep track of the best individuals found so far. One HEO consists of 100 generations with a population size of 300 individuals and provides the best parameter set found. In the 200 HEO runs a group of best parameter sets can be accumulated. This information is then filtered, grouped and graphically presented with a so-called GHOST condensation algorithm.

Taking only one best parameter set from each run can be a waste of computer time. In fact, HEO converges to the best solution region within 20-80 generations, thus creating a great number of similar solutions after 100 generations. Therefore, HEO was modified to increase the solution diversity within the population while preserving the same level of convergence rate. In such a case, it should be possible to include more than one parameter set into the final group of solutions and consequently rely on a smaller number of runs.

To maintain the population diversity throughout the GA generations and not to affect convergence, one should modify the selection scheme or add new operator(s) to keep the diversity within the population. In order to do that, one should clearly understand the HEO as well as the problem search space.

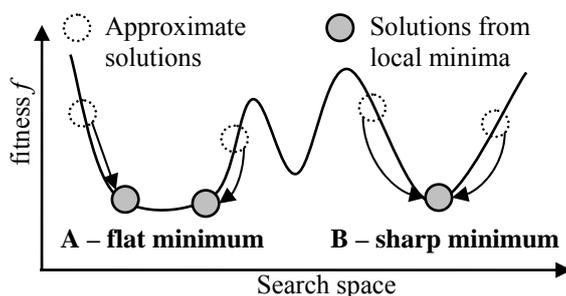


Figure 3. Schematic presentation of parameters search space and the effect of the local mutation procedure responsible for fine-tuning. Due to the noisy spectra and finite resolution of the local optimization routine starting approximations (white circles) are optimized into more accurate solutions (gray circles) according to the local phase-space landscape. **A.** In case of a flat valley (plateau in multidimensional space), the results of the local optimization routine strongly depend on the starting approximation. **B.** In case of sharply defined minimum, local optimization routine provides similar results independently of starting approximation unless starting approximation is too far from the local minimum.

2.2.4 Parameters search space

The optimization process should be thought of as searching for the minima in the landscape of the parameter search space (phase-space), which may contain both local and global minima. A powerful optimization routine should be able to find global minimum(a), which

can be of different types (Figure 3), i.e. well-defined minima (Figure 3B) or a flat minimum valley minima (Figure 3A). An optimization routine should therefore keep convergence to the minima of type **B** (discrete problems) and maintain the diversity to be able to reveal the minimum valleys (in continuous problems) already in a single run.

2.2.5 Population diversity in genetic algorithm

Genetic algorithms are general purpose global search algorithms that use principles of natural genetics. Simultaneously, a population of possible solutions is being optimized. A simple genetic algorithm (SGA) ¹³ is suitable for finding the optimum of a unimodal function in a bounded search space. However, both analysis and experiments show that the SGA cannot find multiple global maxima of a multimodal function ¹³⁻¹⁵ or a function with a flat global minimum, which is an extreme limit of the multimodal function. This limitation can be overcome by a mechanism that creates and maintains several subpopulations within the search space, referred to as “niching methods”. There exist sequential niching methods ^{16,17}; parallel niching methods (sharing ¹⁸, crowding ^{15,19} and clearing ¹⁴); speciation methods ²⁰⁻²² and clustering ^{23,24}; multi-population methods ²⁵ (island models ^{26,27} and migration models ²⁸).

Another way to find multiple optima is to make several runs of an ordinary GA. In each run the GA typically converges to a different optimum. Thus, several optima are found ²⁹. Exactly this strategy was used in the previous multiple HEO-based approach.

Since the methods that assume creating subpopulations do not match with our specific problem, we chose the sharing parallel niching method to maintain diversity within a single run together with a multiple run approach.

2.2.6 Maintaining population diversity: sharing and shaking operators

Sharing ^{15,18} requires that fitness is shared as a single resource among similar individuals in a population of solutions ³⁰. The fitness sharing method modifies the search landscape by changing the fitness function (2), i.e. the value of χ^2 , in densely-populated regions ³¹:

$$f'(j) = \frac{f(j)}{\sum_{i=1}^n \xi(d[i, j])}, \quad (2)$$

where the sharing function ξ is a function of distance $d[i, j]$ between two population elements and can be defined as.

$$\xi(j) = \begin{cases} 1 - \left(\frac{x}{\sigma_{share}} \right)^\alpha & ; \quad x < \sigma_{share} \\ 0 & ; \quad otherwise \end{cases} \quad (3)$$

It returns ‘1’ if the elements are identical and ‘0’ if they cross some threshold of dissimilarity, specified by constant σ_{share} . Here α is a constant, which regulates the shape of the sharing function. As a result of the sharing operator application, the population becomes better distributed in the search space which improves the population diversity (Figure 4A).

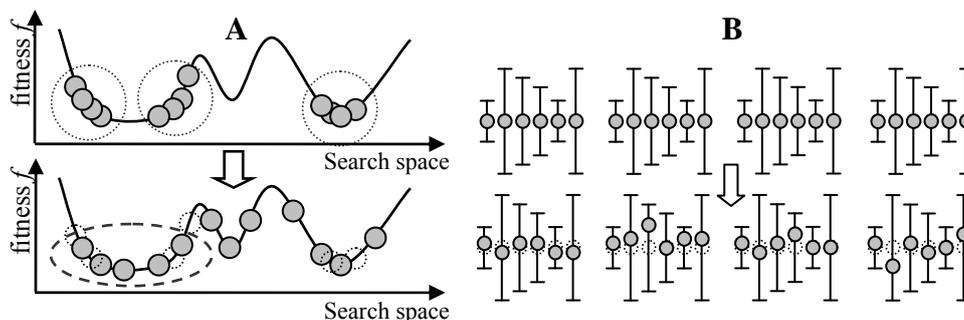


Figure 4. Schematic presentation of the fitness sharing operator function. **A.** Top: In a non-sharing routine crowding at the local minima is allowed, since there is no operator that would maintain diversity. Bottom: In a sharing-routine, fitness function is increased according to the density of solution, aiming to prevent crowding. **B.** Schematic presentation of the Gaussian shaking operator. Shaking operator implies a Gaussian random generator that provides a small deviation to the value of each parameter. The error bars indicate the width of Gaussian probability distribution of these deviations. The standard relative uncertainties of the spectral parameters $\{\vartheta, \varphi, \tau_c, W, p_A, prot, d\}$ are $\{0.02, 0.02, 0.04, 0.035, 0.035, 0.04, 0.02\}$, respectively, which follow average uncertainties that are found empirically for these parameters within the simulation model.

Shaking is a new operator that was developed to provide small Gaussian-like deviations to the spectral parameters (Figure 4B) before the crossover operator is applied. The shaking algorithm prevents “grid” formation and preserves the diversity in the solution population (for explanation of the grid problem see Discussion section).

2.2.7 Projection principle and GHOST condensation

The large amount of solutions resulting from the multiple HEO runs should be condensed and grouped together to construct a discrete or quasi-continuous description of the system. If the proposed model complexity (4 spectral components in our case) is sufficient to describe the system, the final description is also discrete. However, when the proposed complexity is lower than in reality, the model tries to describe the most important features of the system (ESR spectra in our work). In this case, the landscape at the point of the global minimum changes into a flat valley, and consequently, HEO needs to resolve the distribution of solutions describing this optimum region of the parameter search space. In this way, multiple-HEO approach incorporates the “projection principle” idea^{1,3}.

After solution filtering according to the local solution density and goodness of fit, performed in the same way as defined before¹, the GHOST condensed results are presented

in 2D cross-sections $\{S-\tau_c, S-W, S-p_A\}$ (Figure 5). The color of any solution point in the GHOST diagram is defined by RGB specification, where the intensity of each color component (red, green, blue) represents the relative value of the spectral parameters τ_c, W, p_A in their definition intervals $\{0 - 3 \text{ ns}\}$, $\{0 - 4 \text{ G}\}$, and $\{0.8 - 1.2\}$, respectively (Figure 5B). This technique enhances the possibility to distinguish groups of solutions, and to explore optimized values of model parameters.

The most important property of the GHOST algorithm is that there is no need to define the complexity (the number of different motional patterns) in advance – it comes out automatically from the GHOST condensation and graphical presentation.

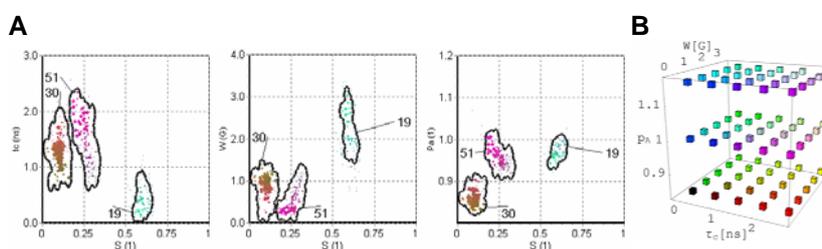


Figure 5. A. An example of the GHOST solution presentation of the spin labeled horse neutrophils taken from bronchoalveolar fluid (BAL) from horses suffering from the chronic obstructive pulmonary disease (COPD). Horses were sedated with medetomidine purchased from Domosedan (Turku, Finland). A 2.5-m long endoscope was introduced through the pre-cleaned and topically anesthetized nostril and advanced until it wedged in a bronchus. 300 ml of pre-warmed sterile physiological saline solution was infused through the biopsy channel into the bronchus and immediately re-aspirated into a sterile flask cooled in ice. Polymorphonuclear leukocytes were isolated from whole BAL samples, spin labeled with MeFASL(10,3), centrifuged, transferred to quartz capillary and measure at Bruker ELEXSYS E500 9.6 GHz spectrometer (field sweep of 10 mT; modulation: 0.15 mT, 100kHz; 5 scans of 40 s with 40 ms of time constant), fitted with ESRSIM BBW software and characterized using GHOST condensation procedure. **B.** Color legend. The RGB (red, green, and blue) color of any particular solution point codes the relative values of parameters τ_c, W , and p_A in their definition intervals.

2.3 Results and discussion

2.3.1 Evaluation criteria

To judge the success of the modification of the HEO algorithm the following criteria were selected: GHOST quality (solution diversity, solution domains determination, model parameters distribution); minimal fitness achieved in χ^2_{min} , and fitness deviation $\sigma(\chi^2)$, that is 40% of the best χ^2_{min} values; runs contribution histograms; and maximal detected solution density ρ_{max} . To check the generality of the new algorithm we analyzed two types of ESR spectra: experimental ones (from membranes and membrane proteins) and synthetic (discrete and continuous).

2.3.2 Multiple runs

Before making any implementation changes in the code, we simply reduced the number of HEO runs from 200 to 20 and increased the contribution of each run (more than one best parameter set). The results for a typical experimental spectrum are shown in the Figure 6 where the GHOST diagram (Figure 6B) and contribution histogram (Figure 6C) are compared with the original GHOST diagram based on the 200 runs (Figure 6A). It can be easily seen that this is not the right way to reduce the computational demand of the problem. With the modified approach, the GHOST diagram (Figure 6B) does not resemble the original one (Figure 6A). In addition it can be seen that only a few runs (such as the first, seventh, ninth and seventeenth) contribute to the GHOST presentation as it is shown by runs contribution histogram in the Figure 6C, whereas the other runs (*i.e.* the third, fourth, tenth, *etc.*) have no contribution at all. This causes the loss of solution diversity, a worse distribution of χ^2 (see minimum value and distribution width in “20 runs” column of Table 1 and a wrong solution domains determination (Figure 6B).

In addition one also can see a higher solution density as a consequence of the crowding in the search space. And even worse result is achieved when the modified “20 runs” approach is tested on a continuous problem: compare original “200 runs” (A) and “20 runs” (B) in the Figure 8. It can be easily seen that the results do not meet the original GHOST distribution. The bad GHOST picture arises from the fact that the contribution of the runs is extremely uneven (Figure 8B), originating in a solution crowding.

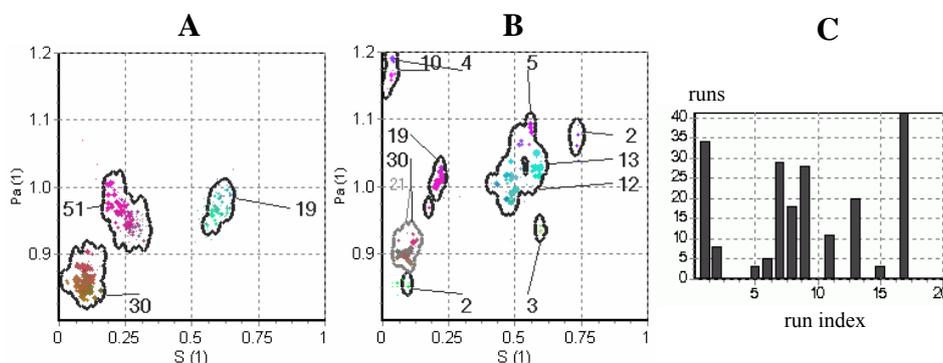


Figure 6. Typical characterization of spin labeled real membrane (see the caption to the Figure 5). **A.** GHOST as a result of 200 runs of HEO where only one solution is extracted from a single run. **B.** GHOST as a result of 20 runs of the same HEO algorithm where on average 10 solutions are taken from each run. **C.** Runs contribution histogram for the case of 20 runs where the number of runs is shown along the x-axis and number of solution (taken from particular run) along the y-axis.

According to the literature, the sharing implementation could change the result^{15,18}. To test the sharing approach the continuous problem was chosen (Figure 8A). The results of this test in terms of the runs contribution histogram and GHOST cross-section are shown in the Figures 8C. It can be seen that the GHOST representation better resembles the

original one, and also the runs contribution becomes more even. However, the distribution of χ^2 is worse (see the minimum value and the distribution width in “sharing” column of Table 2). This result was not good enough, even when we increased the population size from 300 to 600 (to keep convergence at the same level due to the sharing implementation).

Table 1. Optimization parameters after 200 and 20 runs for the real membrane spectrum (for the experimental preparation see the caption to the Figure 5).

Criteria	200 runs	20 runs
χ^2_{min}	3.4	4.09
$\sigma(\chi^2)$	2.04	1.87
ρ_{max}	64.2	71.5

2.3.3 Grid problem and shaking

By careful analysis of the parameters in the resulting solution distribution, we found the origin of the unsuccessful implementation of the sharing approach – the shortcoming of the three-point crossover, one of the most important operators in the GA algorithm. “Genetic material” related to good model parameters, spreads and copies among individuals in the population. After a few tens of generations the population forms a “grid” in the search space (Figure 7A) as a consequence of the rough action of the 3-point crossover operator. This leads to the loss of solution diversity.

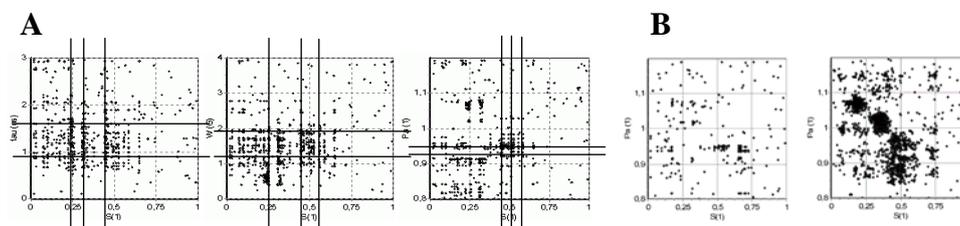


Figure 7: Schematic presentation of the “grid” problem (A) for three cross-sections of the phase-space and its solution (B) in single run. **A.** Due to the standard multipoint crossover, subgroups of parameters are “transferred” between generations untouched, resulting in a grid-like distribution of the GHOST solution (single run). The lines indicate very high vertical and horizontal densities of solutions that evolve from copying of parts of parameter sets within the optimization routine. **B.** Single run GHOSTs (with population size 600). Original version with crowding problem (**left**) – several solutions are crowded in many regions and the version with shaking that maintains diversity (**right**) – solutions crowded in each point previously now spread over the flat minima region with the help of shaking operator.

In the HEO algorithm only a local search operator is capable to restore the diversity and eliminate the “grid”, but due to the high computational cost and extremely high impact on the convergence to local minima the probability of the Downhill-Simplex local search operator

should be and is very low. Therefore the local search operator cannot be used to maintain population diversity. Instead, a new idea of “shaking” was introduced in our work keeping the standard crossover. As it was described in the Methods section, the shaking operator introduces a small deviation in parameters, thereby eliminates the effect of the “grid”.

Indeed, the implementation of the shaking operator allowed the algorithm to overcome the solution crowding and increased the population diversity already in a single run. This result is shown on the Figure 7B for a continuous problem that represents the most extreme case of the complexity.

The results of the implemented shaking operator are shown in the Figure 8. One can see that the shaking operator considerably improves the result of a single run as the GHOST pattern from 20 runs (Figure 8D) is very similar to the original one (Figure 8A), the runs histogram is very even Figure 8D), and finally the distribution of χ^2 is very good (Table 2).

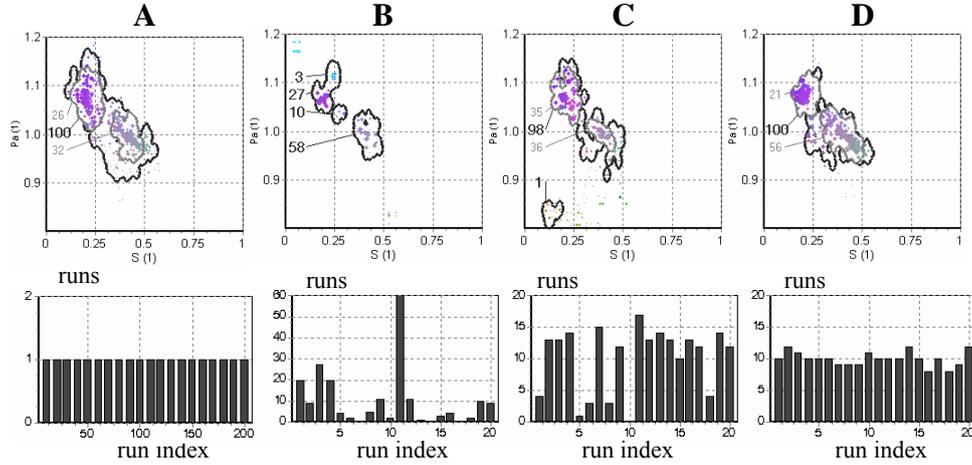


Figure 8: Comparison of the effectiveness of different multi-run HEO-GHOST approaches on the synthetic 15-component spectrum together with runs contribution histogram. **A.** GHOST and runs contribution as a result of 200 runs of original HEO routine. **B.** GHOST and runs contribution as a result of 20 runs of the original HEO routine. **C.** GHOST and runs contribution as a result of 20 runs of the modified HEO routine that includes sharing operator. **D.** GHOST and runs contribution as a result of 20 runs of the modified HEO routine that includes shaking operator as described in the text.

Table 2. Comparison of the χ^2 distributions and solution densities for the different multi-run HEO-GHOST approaches on the synthetic 15-component spectrum that simulates quasi-continuous distribution of spectral parameters (see also caption to the Figure 8).

Criteria	200 runs	20 runs	sharing	shaking
χ^2_{min}	1.17	1.22	1.65	1.24
$\sigma(\chi^2)$	0.9	0.4	1.29	0.9
ρ_{max0}	69.5	75.7	69	66.1

2.3.4 Testing of the modified algorithm

In further tests, the algorithm with the new shaking operator was also applied to several experimental and synthetic spectra in order to cover a wide range of possible systems related to discrete and continuous problems. The results of characterizations of four different examples are shown in the Figure 9, where the GHOST diagrams of different approaches are compared (original “200-runs” approach is compared against “shaking-20-runs” approach). The GHOST diagrams are very similar, confirmed also by the comparison of the averaged values and the distribution widths of the condensed parameters (table is not shown).

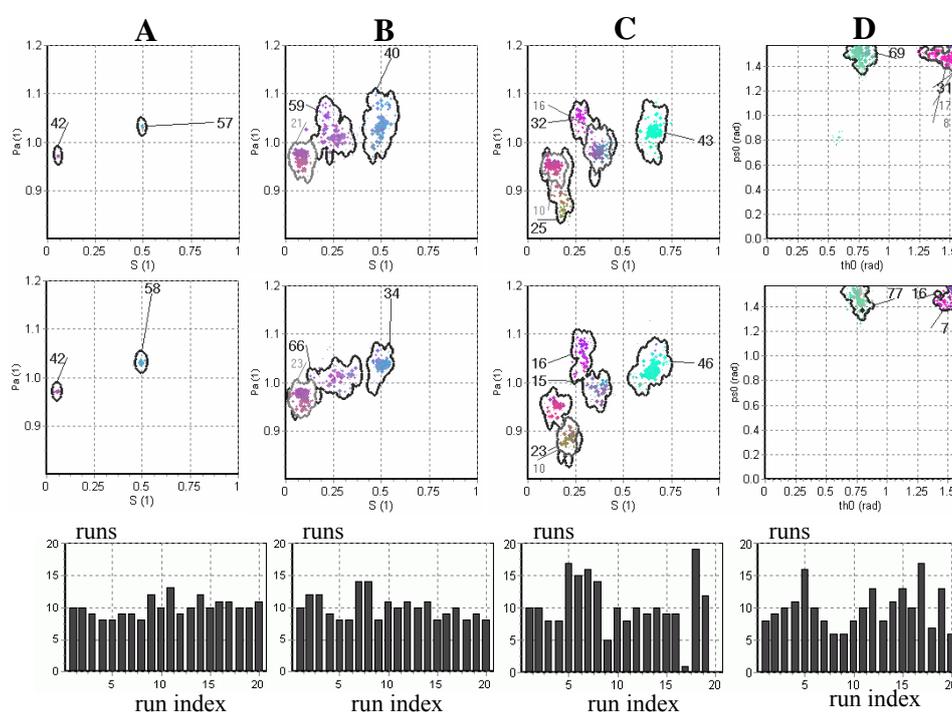


Figure 9: Comparison of GHOST plots of original-HEO approach versus shaking-modified-HEO together with runs contribution histogram for the shaking-modified-HEO based on 20 runs. The original-HEO approach with 200 runs (above) is compared versus modified-HEO (with shaking) based on 20 runs (below). **A.** GHOST plot and runs contribution of the synthetic discrete 2D spectrum that was constructed from two spectral components with the known parameter set and optimized as unknown one. **B** GHOST plot and runs contribution of the synthetic quasi-continuous spectrum (see the caption to the Figure 8). **C.** GHOST plot and runs contribution of the spectrum of the real membranes of breast cancer cells MT1 in the exponential phase of growth: MT1 breast cancer cells were seeded at approximately 10^6 cells in a culture flask with surface area of 75 cm^2 , spin labelled with the methyl ester of 5-doxy palmitate, MeFASL(10,3), and measured under the same conditions as the membranes of horse neutrophils (see the caption to the Figure 6). **D.** GHOST plot and runs contribution of the spectrum of the spin labelled (maleimide spin label) cystein mutant of major coat protein of bacteriophage M13 at amino acid position 46 reconstituted in dimyristoylphosphatidylcholine lipid bilayer³².

2.4 Conclusion

To reduce the computational demand of the original multiple HEO approach, we developed and implemented a novel shaking operator and carried out an extensive testing on various spectra that represent a wide range of possible applications. With the modified optimization algorithm we succeeded to keep the quality of the characterization, thereby considerably reducing the computational time of the ESR spectrum analysis by a factor of 5-7. With this successful modification the application of advanced ESR spectra analysis¹ to complex biosystems, such as biological membranes and membrane proteins, becomes more feasible. Further numerical calculations on both synthetic and experimental data should prove the advantages of the implemented modifications and hopefully find new possibilities to improve and speed-up ESR spectra analysis.

2.5 Acknowledgements

This work was carried out with the financial support of the Ministry of Higher Education, Science and Technology of the Republic of Slovenia through the Slovenian Research Agency. Initiation of the collaboration between the authors was enabled by the financial support from EC COST D22 Action.

References

- 1 Štrancar, J. *et al. J. Chem. Inf. Mod.* **45**, 394 (2005).
- 2 Columbus, L. & Hubbell, W.L. *Trends Biochem. Sci.* **27**, 288 (2002).
- 3 Štrancar, J., Koklic, T., & Arsov, Z. *J. Membr. Biol.* **196**, 135 (2003).
- 4 Budil, D.E., Sale, K.L., Khairy, K.A., & Fajer, P.G. **110**, 3703 (2006).
- 5 Robinson, B., Thomann, H., Beth, A., Fayer, P., & Dalton, L.R., The phenomenon of magnetic resonance: Theoretical considerations in *EPR and Advanced EPR Studies of Biological Systems*, edited by Larry R. Dalton (CRC Press, Boca Raton, Fla., 1985), pp. 11–110.
- 6 Schneider, D.J. & Freed, J.H., Calculating slow motional magnetic resonance spectra: A user's guide in *Biological magnetic resonance: spin labeling, theory and applications*, edited by L. J. Berliner & J. Reuben (Plenum Press, New York, 1989), pp. 1–76.
- 7 Schindler, H. & Seelig, J. *J. Chem. Phys.* **59**, 1841 (1973).
- 8 Štrancar, J., Sentjurc, M., & Schara, M. *J. Magn. Reson.* **142**, 254 (2000).
- 9 Marsh, D., Electron spin resonance: spin labels in *Membrane spectroscopy*, edited by E. Grell (Springer-Verlag, Berlin New York, 1981), pp. 51.
- 10 Steinhoff, H.-J. *et al. Biochim. Biophys. Acta* **1457**, 253 (2000).
- 11 Nordio, P.L., General magnetic resonance theory in *Spin labeling: theory and applications*, edited by L. J. Berliner (Academic Press, New York, 1976), pp. 5.
- 12 Filipič, B. & Štrancar, J. *Appl. Soft Comput.* **1**, 83 (2001).
- 13 Goldberg, D.E., *Genetic algorithms in search, optimization, and machine learning*. (Addison-Wesley Pub. Co., Reading, Mass., 1989).

- 14 Petrowski, A. A clearing procedure as a niching method for genetic algorithms in *Proceedings of the IEEE International Conference of Evolutionary Computation (ICEC '96)* (Nagoya, Japan, 1996), pp. 798.
- 15 Mahfoud, S.W., Ph.D. thesis, University of Illinois at Urbana-Champaign, 1995.
- 16 Beasley, D., Bull, D.R., & Martin, R.R. *Evol. Comput.* **1**, 101 (1993).
- 17 Glover, F. *ORSA J. Comput.* **1**, 190–206 (1989).
- 18 Goldberg, D.E. & Richardson, J. Genetic algorithms with sharing for multimodal function optimization in *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application* (L. Erlbaum Associates Inc., Cambridge, Massachusetts, United States, 1987).
- 19 De Jong, K.A., University of Michigan, 1975.
- 20 Li, J.-P., Balazs, M.E., Parks, G.T., & Clarkson, P.J. *Evol. Comput.* **10**, 207 (2002).
- 21 Spears, W. Simple subpopulation schemes in *Proc. Third Annual Conf. Evolutionary Programming (EP'94)*, edited by A. V. Sebald & L. J. Fogel (World Scientific, Singapore, 1994), pp. 296.
- 22 Deb, K., Master thesis, University of Alabama, 1989.
- 23 Streichert, F., Stein, G., Ulmer, H., & Zell, A. A clustering based niching method for evolutionary algorithms in *Genetic and Evolutionary Computation (GECCO 2003)*, edited by E. Cantú-Paz *et al.* (Springer, Berlin, 2003), Vol. 2723/2003, pp. 644.
- 24 Yin, X. & Gernay, N. A fast genetic algorithm with sharing scheme using cluster analysis methods in multimodal function optimization in *Interernational Conference on Artificial Neural Nets and Genetic Algorithms*, edited by R. F. Albrecht, C. R. Reeves, & N. C. Steele (1993), pp. 450.
- 25 Ursem, R.K. Multinational evolutionary algorithms in *Congress of Evolutionary Computation (CEC 1999)* (IEEE Press, Washington, DC, USA, 1999), Vol. 3, pp. 1633.
- 26 Gordon, V.S., Whitley, D., & Bohn, A., Dataflow parallelism in genetic algorithms in *Parallel problem solving from nature, 2 : Proceedings of the Second Conference on Parallel Problem Solving from Nature*, edited by Reinhard Männer & Bernard Manderick (Elsevier Science, Amsterdam, The Netherlands, 1992), pp. 533–542.
- 27 Bessaou, M., Petrowski, A., & Siarry, P. Island model cooperating with speciation for multimodal optimization in *Proceedings of the 6th International Conference on Parallel Problem Solving from Nature* (Springer-Verlag, 2000).
- 28 Martin, W.N., Lienig, J., Cohoon, J.P., Island (migration) models: Evolutionary algorithms based on punctuated equilibria in *Handbook of Evolutionary Computation*, edited by Thomas Bäck, David B. Fogel, & Zbigniew Michalewicz (Institute of Physics Publishing, Bristol, UK, 1998), pp. C6.3:1.
- 29 Darwen, P.J. & Yao, X. Every niching method has its niche: fitness sharing and implicit sharing compared in *Proceedings of the 4th International Conference on Parallel Problem Solving from Nature* (Springer-Verlag, 1996).
- 30 Mahfoud, S.W. Simple analytical models of genetic algorithms for multimodal function optimization in *Proceedings of the 5th International Conference on Genetic Algorithms* (Morgan Kaufmann Publishers Inc., 1993).
- 31 Sareni, B. & Krahenbuhl, L. *IEEE Transactions on Evolutionary Computation* **2**, 97 (1998).
- 32 Stopar, D., Spruijt, R.B., Wolfs, C.J., & Hemminga, M.A. *Biochim. Biophys. Acta* **1611**, 5 (2003).

Chapter 3

Site-directed spin labeling study of the light-harvesting complex CP29

based on:

Kavalenka, A.A., R.B. Spruijt, C.J.A.M. Wolfs, J. Štrancar, R. Croce, M.A. Hemminga, H. van Amerongen, 2009. Site-directed spin labeling study of the light-harvesting complex CP29. *Biophys. J.* 96:3620-3628.

Abstract

The topology of the long N-terminal domain (approx. 100 amino acid residues) of the photosynthetic light-harvesting complex CP29 was studied using electron spin resonance (ESR). Wild-type protein containing a single cysteine at position 108 and nine single cysteine mutants were produced, allowing to label different parts of the domain with a nitroxide spin label. In all cases the apoproteins were either solubilized in detergent or they were reconstituted with their native pigments (holoproteins) *in vitro*. The spin label ESR spectra were analyzed in terms of a multi-component spectral simulation approach, based on hybrid evolutionary optimization and solution condensation. These results permit to trace the structural organization of the long N-terminal domain of CP29. Amino acid residues 97 and 108 are located in the transmembrane pigment-containing protein body of the protein. Positions 65, 81 and 90 are located in a flexible loop that is proposed to extend out of the protein from the stromal surface. This loop also contains a phosphorylation site at Thr81, suggesting that the flexibility of this loop might play a role in the regulatory mechanisms of the light-harvesting process. Positions 4, 33, 40, and 56 are found to be located in a relatively rigid environment, close to the transmembrane protein body. On the other hand position 15 is located in a flexible region, relatively far away from the transmembrane domain.

3.1 Introduction

Photosynthesis in green plants and algae occurs in chloroplasts. Their highly-folded thylakoid membranes provide a home for the multisubunit protein complexes Photosystems I and II (PSI, PSII), that work in concert (linked by a cytochrome b6f complex) to convert sunlight energy into chemical energy ¹. The fourth major player is the ATP-synthase complex that uses the proton gradient across the thylakoid membrane, created by PSI/PSII, to convert ADP into ATP. PSI and PSII are supramolecular complexes composed of a core moiety, which contains all the cofactors of the electron transport chain and of an outer antenna system, the role of which is to collect light energy and to transfer it to the reaction centre where it can be used to drive charge separation. All antenna complexes of higher plants belong to the Lhc (Light-harvesting complex) multigenic family ². In particular six different gene products (Lhcb1-6) compose the outer antenna system of PSII. The major antenna complex of PSII is light-harvesting complex II (LHCII, the product of the genes Lhcb1-3 ³), harboring over 50% of the pigments and it is organized as trimers at the periphery of the PSII supramolecular complex ⁴. Three minor antenna complexes, CP29 (Lhcb4), CP26 (Lhcb5) and CP24 (Lhcb6) are located in between the LHCII trimers and the core complex and they are present as monomers. Recently, it has been proposed that the minor antenna complexes provide the sites of non-photochemical quenching, a mechanism

that protects PSII against photoinhibition⁵. In particular it has been shown that in CP29 a radical cation is formed on the zeaxanthin in the L2 site, which strongly interacts with Chl A5⁶ leading to the harmless dissipation of excess excitation energy.

The structure of LHCII has been resolved at 2.72 Å⁷ showing three transmembrane helices, two amphipathic helices on the luminal side of the membrane and the positions of 14 chlorophyll and 4 xanthophyll molecules per monomeric subunit. Structural information on the minor antenna complexes CP24, CP26 and CP29 is still lacking, but sequence analysis⁸ and site-selected mutagenesis have revealed that they share high structure similarity with LHCII, although they coordinate a smaller number of pigments^{9,10}.

CP29 is the largest member of the Lhc family and it is characterized by a long N-terminal domain (approx. 100 amino acid residues), which contains a phosphorylation site¹¹. Phosphorylation takes place, for instance, under cold stress and is accompanied by a structural change of the protein¹². It has been shown that there is a strong correlation between the presence of phosphorylated CP29 and the resistance of plants against cold stress, thus leading to the suggestion that the phosphorylation is involved in protective mechanisms¹³. However, details are lacking both on the structure and the structural changes.

CP29 belongs to the class of membrane proteins. In general, membrane proteins comprise almost one third of the total amount of proteins in an organism or cell. However, progress in determining their structures has been slow. Therefore, membrane proteins offer an enormous challenge in structural biology and there is an urgent need to develop and apply new biophysical methodologies that are able to generate detailed structural information. Among modern biophysical techniques, site-directed spin labeling electron spin resonance (SDSL-ESR) appears to show the highest potential to further develop the field¹⁴.

Recently, CP29 protein mutants reconstituted with plant pigments in detergent were selectively labeled at three positions in the N-terminal domain with a fluorescent dye TAMRA (6-carboxy-tetramethyl-rhodamine) and examined with picosecond fluorescence spectroscopy¹⁵. The results indicated that the N-terminus is folded back on the hydrophobic part of the protein, and suggested the presence of some structural heterogeneity in the N-terminal part.

The present paper focuses on the structure and dynamics of the N-terminal domain of CP29 in detergent systems with and without pigments. Site-directed mutagenesis was used to produce ten single-cysteine protein samples with cysteine positions equally distributed over the N-terminal domain. Following the approach of Stopar et al.¹⁶, single-cysteine protein samples were labeled with nitroxide spin labels. The ESR data allowed us to determine the free rotational space, local dynamics and polarity of the spin labeled sites that reflect the pigment-binding properties of CP29 and to arrive at a topological model for the N-terminal domain.

3.2 Materials and methods

3.2.1 Construction and isolation of overexpressed CP29 apoprotein

Lhcb4.1 cDNA of *A. thaliana* (from Arabidopsis Biological Resource Center DNA Stock Center) was subcloned into a pT7-7 expression vector. The construct contains the sequence of the mature CP29 protein with an additional methionine at the N-terminus and a 6 His-tag at the C-terminus. Mutations were introduced using the Stratagene Quick Change Site Directed Mutagenesis Kit. First, the natural occurring cysteine (position 108) was replaced by alanine. This mutant was also used to estimate the amount of nonspecific spin labeling. On this template, single cysteine residues were introduced in the N-terminal part at various positions resulting in the following mutants G4C, S15C, S33C, S40C, A56C, S65C, T81C, S90C and S97C. The constructs were checked by DNA-sequencing. The plasmids were amplified in the super competent *Escherichia coli* XL-1 Blue strain and the proteins overexpressed in the *E. coli* BL21 (*DE3*) strain. Inclusion bodies containing the CP29 apoprotein mutants were isolated as reported in ^{17,18} and stored in the presence of 10 mM dithiothreitol at -20 °C.

3.2.2 Pigment isolation, labeling and reconstitution of CP29-pigment complexes

Purified pigments were obtained from spinach. Concentrations of pigments were determined spectroscopically: chlorophylls as described by Porra ¹⁹ and carotenoids as described by Davies ²⁰. Just before labeling, inclusion bodies containing CP29 apoprotein were freshly purified from dithiothreitol and dissociated in lithium dodecyl sulfate (LDS) reconstitution buffer (2% LDS, 12.5% sucrose, 20 mM Na₂HPO₄ pH 7.6). CP29 apoproteins were labeled at room temperature for 3 hours with a 5 times molar excess of the spin label MTS-SL, (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate from TRC, Toronto, Canada). Excess spin label was removed using affinity chromatography on a His-Trap column. Prior to storage at -20 °C the excess of imidazole and NaCl from the elution buffer were removed by dialysis against LDS reconstitution buffer. Samples of CP29 apoprotein to be measured in β-D-maltoside (DM) buffer (0.03 % W/V + 10 mM Na₂HPO₄, pH 7.6) were prepared by using the detergent substitution procedure ²¹ followed by affinity chromatography on a His-Trap column to bring the apoprotein in DM buffer. Reconstitution and purification of protein-pigment complexes (holoproteins) were performed as reported in ²², but using a chlorophyll *a/b* ratio of 5.5. Solutions of the spin-labeled CP29 samples were washed and concentrated in sucrose-free DM buffer just before the ESR measurements. Integrity of the holoprotein samples was checked by fluorescence excitation and emission measurements, showing the complete absence of free chlorophylls and carotenoids in all preparations.

3.2.3 ESR measurements

All washed and concentrated spin-labeled CP29 preparations in DM buffer (final protein concentration between 0.07 and 0.2 mM) were transferred to 50 μ l capillaries up to 1 cm height and placed in a standard 4-mm quartz ESR tube. Spectra were measured on a X-band Bruker Elexsys E-500 ESR system equipped with a super-high-Q cavity ER 4122SHQE in combination with a SuperX X-Band Microwave Bridge type ER 049X. Temperature was controlled with a quartz variable-temperature Dewar insert. Spectra were recorded at 10 mT scan widths with a microwave power of 5 mW at 6 °C. To improve the signal-to-noise ratio, up to 100 scans were accumulated with a time constant of 20 ms, a modulation amplitude of 0.1 mT and a scan time of 82 s. Before analysis spectra were corrected for the background signal of the buffer.

3.2.4 ESR spectral simulation, optimization and solution condensation

The ESR spectra of spin-labeled CP29 samples were simulated with a multi-component model as described previously^{16,23}. The spectral parameters $\{\mathcal{G}, \varphi, \tau_c, W, p_A, prot\}$ of each component of the simulated spectra were simultaneously optimized with a multi-run hybrid evolutionary algorithm^{24,25}. Multiple solutions, which were obtained from optimization, were then filtered and grouped into domains with a GHOST condensation approach^{16,23,25,26}.

The simulation model for the ESR spectra employs a fast motional averaging approximation to describe the local motion of the spin label²⁵. The dynamics of the spin probe gives rise to a motion in a cone²⁷, which can be described with three parameters: a maximum opening cone angle \mathcal{G} , a cone asymmetry angle φ , and an effective correlation time τ_c . The magnetic interaction tensors \mathbf{g} and \mathbf{A} are linearly corrected with a polarity parameter p_A . Furthermore, a proticity parameter $prot$ is used that accounts for the effect of proton binding to the spin label on the \mathbf{g} tensor²⁸. It was found that the relative error for parameter $prot$ was quite large. Therefore this parameter will not be used in our further discussion²³. When calculating the convolution of the magnetic field distribution and the basic line shape, two line width parameters, τ_c and W , are applied. A Lorentzian line is used in the motional narrowing approximation with a single effective rotational correlation time τ_c ^{27,29}. The additional broadening of the spectral line arising from nonmotional effects is described by a constant W . This parameter arises from unresolved hydrogen superhyperfine interactions and contributions from paramagnetic impurities (e.g., oxygen), in addition to external magnetic field inhomogeneities, field modulation effects, and intermolecular spin-spin interactions if present and applicable.

To resolve coexisting motional patterns from the experimental ESR spectra, the simulated spectra were composed from four independent spectral components defined by four sets of spectral parameters $\{\mathcal{G}, \varphi, \tau_c, W, p_A, prot\}$ and four relative contributions following a previous approach²³. Typically, 20 runs of the population-based hybrid

evolutionary optimization (HEO) were used to produce 8000 (400 in each of 20 runs) solutions (spectral parameters and the weights of four spectral components)^{25,26}. The 200 best solutions were chosen (according to the quality of fit) and their four spectral components were separated into a pool of 800 parameter sets. Collected single spectral components are processed further with GHOST condensation, which filtered and then grouped the spectral components into domains^{25,26}. Each domain in a GHOST plot can be seen as a “motional pattern” of the spin label that is related to its local motional properties. Such motional patterns reflect the restrictions of the spin label arising from the local protein structure, i.e., local interactions between the spin label rotamers and neighboring amino acid side chains and the motional limitations imposed by the protein backbone. In addition, the motional patterns reflect different dynamical regimes of the spin probe, which may additionally include: a) dynamics inherited from the whole protein motion; b) protein backbone fluctuations²⁷. Also the spin label senses the accessibility of solvent molecules and adjacent acyl chain of the phospholipids in case it is in bilayer.

Filtering of the multiple solutions was done according to the fit quality of a particular solution and according to the density of the solution in the parameter space. The group recognition was done with a slicing method based on domain detection at several density levels³⁰. Visual analysis of the resulting GHOST plots, which present a combination of two parameters (φ and \mathcal{G} , τ_c and \mathcal{G} , p_A and \mathcal{G}), was used to revise the results of the automated group (motional patterns) recognition and to examine the distribution of the spectral characteristics within the groups. Candidate motional patterns were tested for their physical relevance by looking at the corresponding line shapes. Unusual line shapes resulting from abnormal combinations of parameters were omitted from further analysis. In this way the ESR experimental spectra are characterized in terms of multiple motional patterns, and the GHOST analysis provides the number of patterns, average parameters, and relative contribution of each pattern.

3.3 Results

3.3.1 CP29 reconstitution

Together with the wild type CP29 (WT-C108) nine cysteine-spin labeled CP29 apoproteins (G4C/C108A, S15C/C108A, S33C/C108A, S40C/C108A, A56C/C108A, S65C/C108A, T81C/C108A, S90C/C108A, and S97C/C108A) were obtained and reconstituted with pigments *in vitro*. All pigment-protein complexes were obtained in their monomeric state as assessed by sucrose gradient ultracentrifugation. The absorption spectra of the holoprotein mutants are identical to that of the wild-type construct and resemble the spectrum of the native CP29 complex, similar as in previous studies^{9,21,31}. This indicates that the mutations do not influence the pigment binding.

3.3.2 ESR experiments

The ESR spectra of the reconstituted holoprotein complexes and of the apoproteins in detergent solution are shown in Figure 1.

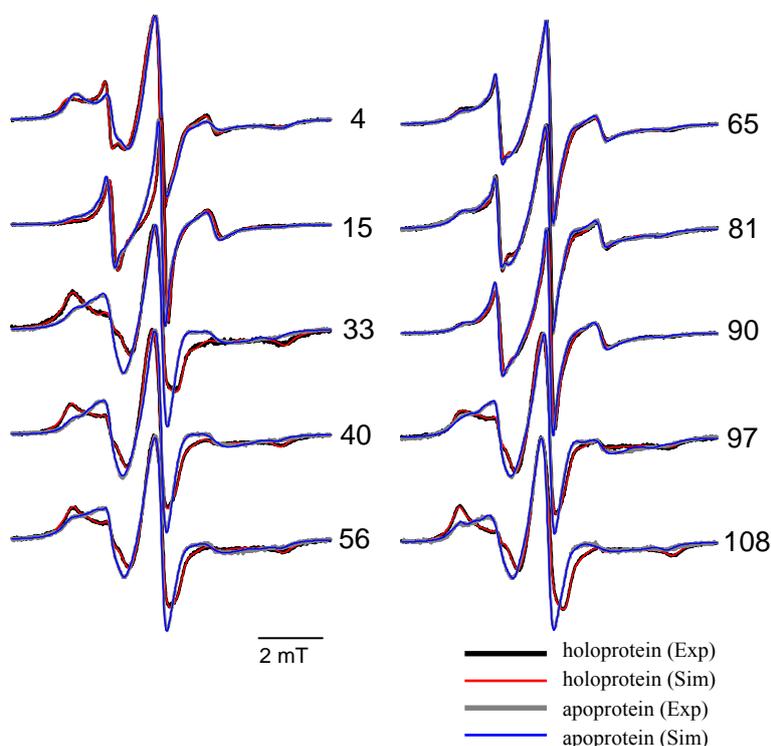


Figure 1. ESR spectra of MTS-SL spin-labeled CP29 protein samples at label positions 4, 15, 33, 40, 56, 65, 81, 90, 97 and 108 reconstituted in DM with (holoprotein, black line) and without (apoprotein, grey line) pigments. The total horizontal scan range is 10 mT. Spectral line heights are normalized to the same central line height (left peak). The simulated spectra are shown in red for holo- and blue for apoprotein samples.

In all cases the spectra have a multi-component character. As can be seen, the absence of the pigments has only a small effect on the spectra corresponding to positions 15, 65, 81 and 90 and for all these cases the ESR spectra show a strong sharp three-line component of mobile spin labels. In contrast, for positions 33, 40, 56, 97 and 108, there is a relatively large spectral difference between the holo- and apoproteins. At these positions the ESR spectrum has a typical immobile appearance, especially for the holoprotein in the presence of pigments. The ESR spectrum corresponding to position 4 shows a two-component spectrum with a strong immobile contribution. Close inspection of the ESR spectra corresponding to positions 4 and 15 reveals that there is a small increase of immobile component for the apoprotein.

To decompose the multi-component ESR spectra, we used a multi-component model of asymmetric motional restriction^{16,23} and optimized the fitted spectra employing a multi-run multi-solution hybrid evolutionary method²⁵. The goodness of fit was chosen to be the reduced χ^2 function:

$$\chi^2 = \frac{1}{N} \sum_{i=1}^N \frac{(y_i^{exp} - y_i^{sim})^2}{\sigma^2}, \quad (1)$$

where y^{exp} and y^{sim} are the experimental and simulated data, respectively, σ is the standard deviation of the experimental points, and N is the number of spectral points (in our case $N = 1024$). For all ten spin-labeled CP29 holo- and apoprotein samples the quality of the simulated ESR spectra is good (see Figure 1). For holoprotein spin labeled at positions 33, 56, 81 and 97 and apoprotein spin labeled at positions 4, 56, 81 and 108, the reduced χ^2 of the best fit solutions is between 1.6 and 3. For the other samples this is slightly higher, i.e., between 3 and 5. In general, χ^2 values below 5 can be considered to be very good.

The results from the simulations are summarized in so-called GHOST plots (such as a ϑ - φ GHOST shown for positions 65 and 108 in Figure 2). The GHOST methodology provides the motional patterns that characterize the spectrum, thus the GHOST plots provide the most significant and probable groups of solutions of spectral parameters. Each group corresponds to a particular motional pattern (e.g., mobile or immobile according to the rate of motion; restricted or unrestricted according to the extent of restrictions imposed by the local protein structure on free rotational space of the spin probe). The weight of the group represents the contribution of that particular component to the spectrum. For example, at position 65 (Figure 2) the rotational space of the component with 14% contribution is completely open (ϑ and φ around $\pi/2$), a component with a contribution of 57% is half-closed (ϑ around $\pi/4$) and still symmetric (φ around $\pi/2$), and a component with a contribution of 16% is very closed (ϑ around $\pi/6$ and φ close to 0), as suggested by the distribution of the cone angles of the spin label ϑ and φ (both angles can vary between 0 and $\pi/2$). On the other hand the rotational space for the spin label at position 108 of CP29 pigment-protein complex is very restricted as suggested by the major component with a contribution of 51% (ϑ around $\pi/6$ and φ close to 0) (Figure 2). In most cases the motional patterns in the GHOST plots (as shown in Figure 2) are represented in the parameter space by concentrated groups of solutions. Contrary, in the case of spin-labeled apoprotein mutants 40 and 90, there appear continuous patterns, which reflect smooth transitions between the spectral parameters. This may indicate a transition between structural conformations, or could represent a distribution of a local structure around the mutated residue. The samples having spectra with a relatively low signal-to-noise ratio turned out to be somewhat more problematic in terms of group recognition. Also the ESR spectra of mutants at positions 15 and 90 were more difficult to fit, and after group recognition, many spectral components were found distributed in the parameter space. Thus after group

recognition the final solution appeared to contain several motional patterns with a low contribution (Figure 4). This means an additional complexity of the corresponding spectra and consequently of the spin label motion at positions 15 and 90 relative to other spin-labeled positions. The four best-fitting spectral components in the simulated spectra (Figure 1) are presented in the GHOST plots with colored triangles (Figure 2). The size of a triangle is proportional to the contribution of the corresponding component in the spectrum.

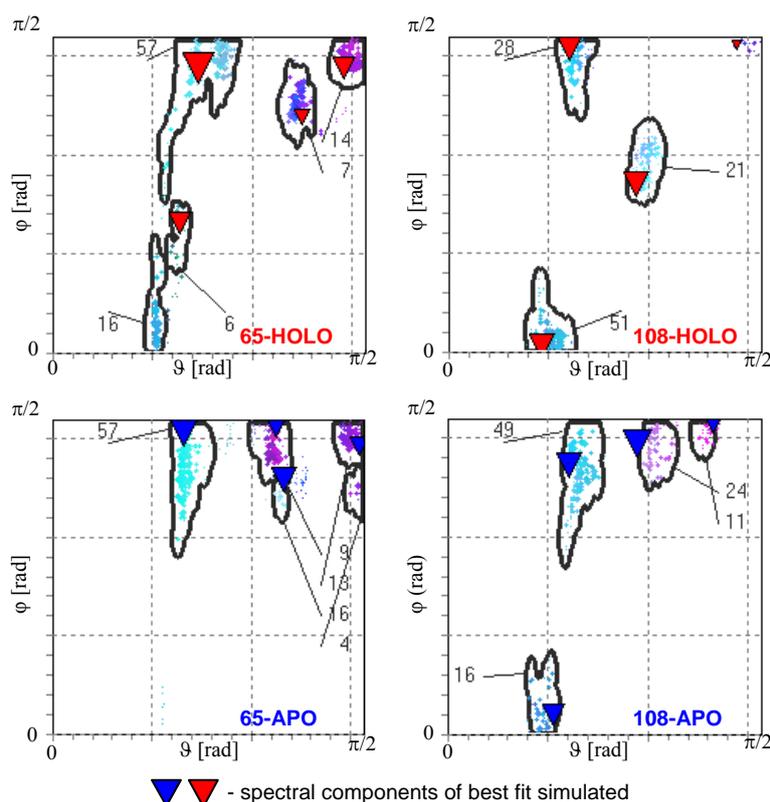


Figure 2. GHOST plot showing the optimized multiple solutions represented in a two-dimensional distribution of the angles ϑ and φ of MTS-SL spin-labeled CP29 protein samples at positions 65 and 108 reconstituted in DM with (top, holoprotein) and without (bottom, apoprotein) pigments. The components of each solution are represented with a point on the plot with a color, combined of red, green, and blue, which codes for the relative values of τ_c , W and p_A in their definition intervals $\{0 - 3 \text{ ns}\}$, $\{0 - 0.4 \text{ mT}\}$, and $\{0.8 - 1.2\}$, respectively. The closed black lines on the plot surround domains of the solutions grouped into motional patterns. The contribution of each pattern is shown in percents. Additionally, the four spectral components of the best fit solution are presented on the plot with red (top, holoprotein) and blue (bottom, apoprotein) triangles, while the area of each triangle is proportional to the relative contribution of the corresponding component in the simulated spectrum.

For further analysis (i.e., a more convenient comparison of multiple data between different spin label positions along the protein), the angles ϑ and φ are combined in a single parameter, Ω , which is defined as²³:

$$\Omega = \frac{\mathcal{I}\varphi}{(\pi/2)^2} \quad (2)$$

This parameter measures the space angle, i.e., the surface of the cone left for local spin label wobbling (free rotational space) and is shown for all ten spin-labeled CP29 holo- (Figure 3A) and apoprotein samples (Figure 3B). High values of Ω (between 0.7 and 1) correspond to nearly unrestricted motional patterns of the spin label (i.e., mobile spectral components), whereas low values (between 0 and 0.25) imply very high restrictions (i.e., immobile spectral components). In addition to the free rotational space Ω , the simulations provide the effective rotational correlation time τ_c ²⁹ and the polarity correction p_A for the magnetic interaction tensors \mathbf{g} and \mathbf{A} of the spin label^{16,28}. These parameters are presented in Figs. 3A and B as well. To elucidate the effect of pigment removal on the ESR data, we carried out a comparison of the most important motional patterns (with a contribution of more than 25%), as shown in Figure 3C. Figure 3D compares the weighted averages of the motional patterns of the spin-labeled holo- and apoproteins. In general, it can be seen in Figure 3 that high values of the free rotational space Ω correspond to high values of the effective rotational correlation time τ_c .

3.4 Discussion

The central issue in our research is related to the following questions: 1) what is the conformation of the unusually 100-residues long N-terminal domain of CP29 protein (which is much longer than for all other members of the Lhc family); 2) where is this domain located with respect to the membrane-embedded transmembrane protein body, and 3) what is the role of the pigments in determining the structure and dynamics of the N-terminal domain? To address these questions, we compared CP29 holo- and apoprotein by using ESR of spin labeled cysteine positions distributed over the N-terminal domain. In this respect, it should be noted that after reconstitution in the detergent DM the pigments provide a correctly folded transmembrane body domain of the protein, which can be considered as the native state of the protein^{21,32,33}. The detergent that is used for the reconstitution of CP29 protein with the pigments provides a good membrane-mimicking environment for CP29: DM it is not a very strong denaturing detergent providing a relatively compact protein-detergent complex³³. If the pigments are absent, the structure of CP29 protein is more loose and it may be partly unfolded³³. For LHCII in DM the spectroscopic properties are similar to those observed in the intact thylakoid membrane³⁴. Since LHCII and CP29 have a strong sequence homology in the transmembrane protein body⁹, this indicates that the structure of CP29 in DM may also be similar to the *in vivo* structure. Thus, the holo- and apo-states of CP29 provide a good starting point for a comparative spin-label ESR study addressing the questions given above.

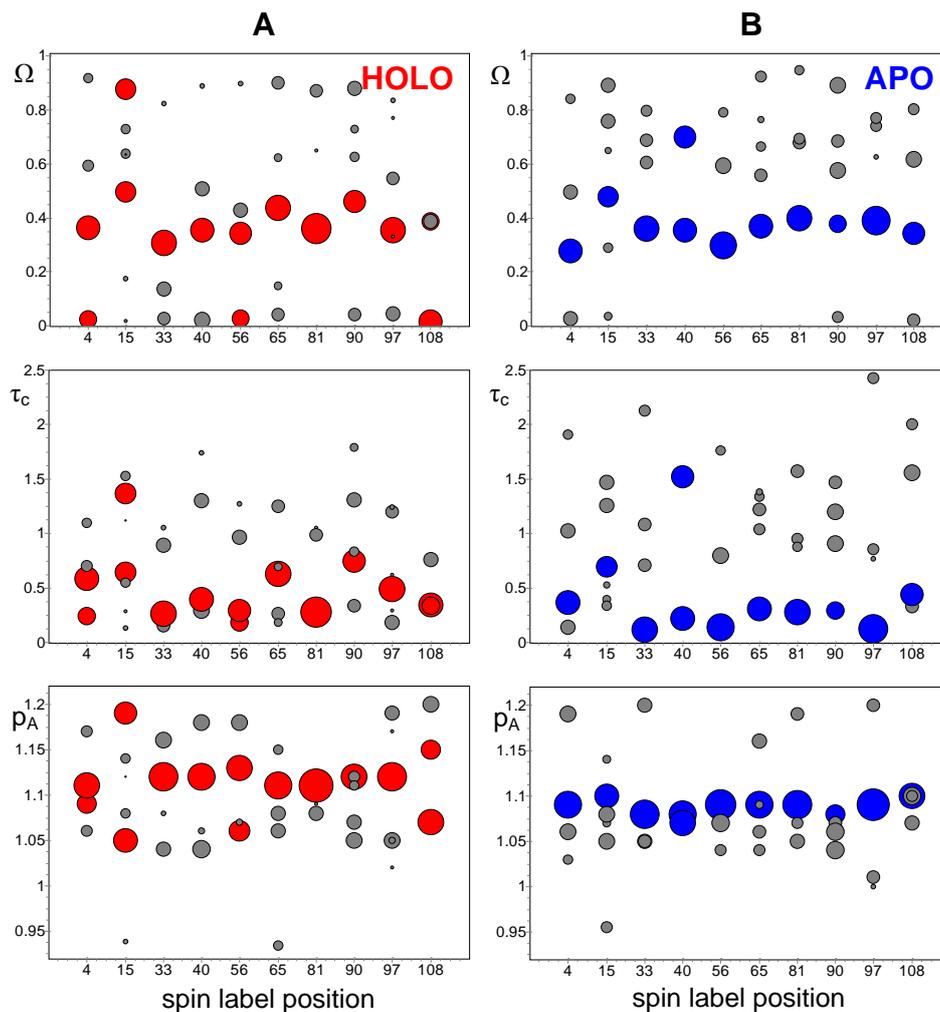
From a qualitative analysis of the ESR spectra (Figure 1) it follows that positions 33, 40, 56, 97 and 108 are located in protein domains that are strongly affected by pigment reconstitution of the CP29 complex. Positions 97 and 108 are located in the transmembrane protein body that contains the pigments¹¹. It is evident that these positions will be affected by the pigment reconstitution, bringing the protein from a relatively loose and partly unfolded structure without pigments into a native folded structure with pigments. Interestingly, positions 33, 40, and 56 follow the same trend. This indicates that this protein domain is located adjacent to the transmembrane protein body. Positions 65, 81 and 90 show a sharp mobile component indicating a relatively high degree of motion. Moreover, these positions are not affected by pigment reconstitution, suggesting that they are located far from the transmembrane region in a loop extending out from the stromal surface of the protein¹¹. Also positions 4 and 15 at the N-terminal end are just slightly affected by pigment reconstitution. Position 4 displays a clear two-component characteristics of a sharp mobile and a broad immobile component. Contrary, position 15 can be characterized only by a sharp mobile component and the broad immobile component is almost absent. This indicates that the spin label at position 4 is more restricted in its motion than the one at position 15. This finding is remarkable, since position 4 is close to the N-terminal end, where one would expect a large degree of motion due to fraying of the terminal amino acid residues. The ESR line shapes at positions 15, 65, 81 and 90 are roughly similar to each other.

To further analyze the multi-component ESR spectra, we carried out a spectral decomposition based on a multi-component model of asymmetric motional restriction^{16,23}, followed by a multi-run multi-solution hybrid evolutionary approach²⁵. The multi-component model turned out to be robust enough to cover many different combinations of coexisting local motional patterns. The multi-solution feature of the simulations provides the capability of determining the actual number of the spectral components related to spin probe motional patterns, the spectral parameters and the contribution of each component, without setting the number of the spectral components in advance. Due to practical considerations, we limited the maximum number of spectral components to four.

The main general advantages of our multiple-solution algorithm are: 1) determination of multiple components (motional patterns), since a single solution characterization may not be capable of revealing all components; 2) revealing a transition between spectral parameters, which could be very useful in the case of multiple protein conformations; 3) detecting defects in the line shape. Concerning line shape defects, a spectral component may arise in the optimization to simulate a particular feature of the line shape to improve the fit. In such a case checking of the parameter space via GHOST plots (such as shown in Figure 2) in combination with the line shape analysis helps to clarify the characterization results and to remove meaningless components, if needed^{23,25}. Also, the appearance of low-quality fits and an unusual distribution of the spectral parameters in the

parameter space may indicate artifacts in the spectra. In most cases we found high-quality fit solutions and well-defined two-dimensional GHOST patterns, indicating that the ESR spectra do not have artifacts and that the group recognition was carried out in a correct way.

As can be seen in Figs. 3A and B, the GHOST analysis results in a number of motional patterns.



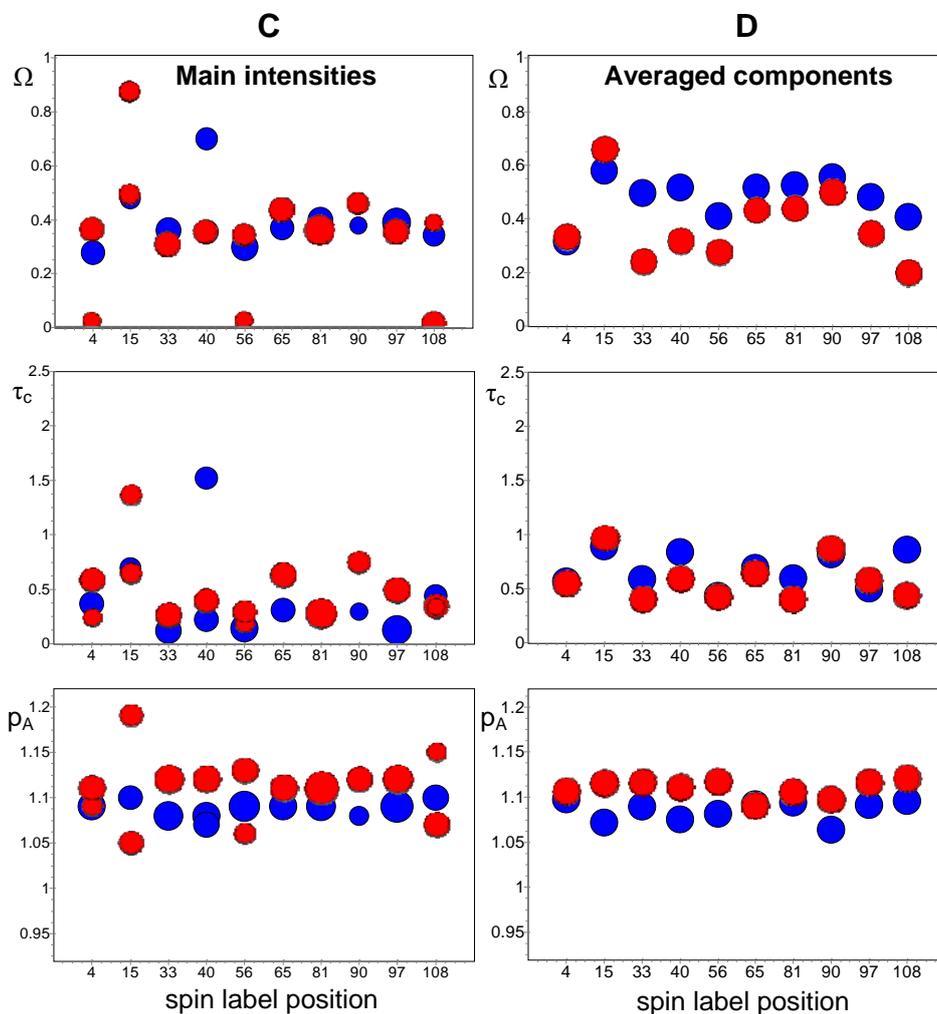


Figure 3. ESR data of MTS-SL spin-labeled CP29 protein samples reconstituted in DM with (A, holoprotein, red circles) and without (B, apoprotein, blue circles) pigments. Less pronounced motional patterns with a contribution below 25% are represented by grey circles. The horizontal axis indicates the spin label position, the vertical axes give Ω , τ_c and p_A . High values of Ω (between 0.7 and 1) correspond to (nearly) unrestricted motional patterns of the spin label (i.e., mobile spectral components), whereas low values (between 0 and 0.25) imply very high restrictions (i.e., immobile spectral components). (C) Comparison of the most important motional patterns (with a contribution of more than 25%) of spin-labeled CP29 protein samples with (holoprotein, red circles) and without pigments (apoprotein, blue circles). (D) Weighted averages of the motional patterns of spin-labeled CP29 protein samples with (holoprotein, red circles) and without pigments (apoprotein, blue circles). The area of the circles in A, B, and C is proportional to the relative contribution of the motional patterns to the multiple solution.

There are several factors that can contribute to a multi-component character: 1) differences in local structure around the spin label at the binding site; 2) various rotamers of the side chain of the spin label and interactions between certain rotamers with the local environment; 3) sample heterogeneity on the level of the micelles in which CP29 protein is incorporated, for example arising from differences in protein-to-detergent ratios and micellar sizes; and 4) nonspecific labeling. To estimate the amount of nonspecific labeling, we produced a mutant of wild-type CP29, in which the cysteine at position 108 is replaced by an alanine. Spin labeling of this mutant shows that the amount of nonspecific labeling is less than 5%. As can be seen in Figure 3A even by discarding motional patterns with small contributions (less than 10-20%), there is more than one component left in a majority of the cases.

Because the free rotational space Ω is very sensitive to the local environment of the spin label side chain (adjacent protein domains and/or solvent molecules), there are two different ways to handle multiple motional patterns:

1) Assign the motional patterns to one or two protein conformations and further use this result to interpret the effect of pigment binding on the conformation of the protein and locations of the pigments in the protein. In this case, we select the components with the highest intensity (above 25%) in the GHOST analysis (Figure 3C). The other motional patterns are then assigned to sample heterogeneities and minor structural components. Two or more components may manifest similarities, consistent changes of the model parameters and thus can be considered to be parts of a single major motional pattern. Such a pattern (prolonged in parameter space) with an evident transition of the model parameters then most likely represents the transition between conformational states. This enables an analysis of the results in terms of different protein conformations.

2) As we will concentrate on the effect of pigment binding of CP29 protein, we do not need to assign the various motional patterns, but we can focus on the differences in the results with and without pigments. Therefore another approach is to take the weighted average of all patterns (Figure 3D). When comparing the averaged data for the protein with and without pigment, the difference will be dominated by the effect of pigment binding.

In comparing the Ω values for the holo- and apoprotein in Figure 3A and B, it can be seen that for almost all spin label positions the range of values increases from low values to higher values. This is especially true for the motional patterns with $\Omega \approx 0$ in Figure 3A, in which the spin label motion is highly restricted. These motional patterns are almost gone in Figure 3B. In turn, in Figure 3B a larger range of motional patterns is observed for Ω values from 0.6 to 1.0, indicating local conformations with less restricted spin label motion. Since this effect is found throughout the whole N-terminal domain, it is assigned to partly unfolding of the protein on going from the holo- to the apo-state. As can be seen in the intensity-filtered data in Figure 3C, at several positions (4, 15, 40, 56 and 108) two values for Ω , τ_c and p_A can be identified. These positions appear to be spread over the entire

sequence of the N-terminal domain of CP29 protein. This effect is also related to a relatively loose and partly unfolded state of the apoprotein, as discussed above. However, no consistent pattern exists between the various values for Ω , complicating a detailed analysis of the data in terms of different conformations of the N-terminal protein domain. Although there appears to be a wealth of information in Figs. 3A and B, a full assignment of motional patterns is not possible without additional knowledge about the N-terminal domain and without having more amino acid residues systematically replaced in a certain protein domain.

This difficulty does not exist by taking the weighted average of all motional patterns (Figure 3D). These data represent the general trend, but details about the various components are lost. In Figure 3D, apart from information about the average free rotational space Ω , also information is available about the average effective rotational correlation time τ_c and local polarity p_A of the spin label attached to the protein. In Figure 3D it can be seen that in all cases (except for positions 4 and 15) the values for Ω for the pigment-free CP29 protein are above the values for the reconstituted protein. This indicates that the N-terminal part of the pigment-free apoprotein has a relatively loose and flexible structure in which the available space for the spin label is expected to be less restricted. Based on the polarity effect shown in Figure 3D (a high value for p_A reflects an increased local polarity²⁵), we can conclude that overall the spin-labeled sites in the apoprotein are more in an apolar environment as compared to the holoprotein. This could reflect an enhanced exposure to the acyl chains of the solubilizing detergent molecules, probably due to the relatively loose and partly unfolded state of the apoprotein.

The trend in the free rotational space Ω as shown in Figure 3D closely follows the qualitative interpretation of the ESR spectra in Figure 1, indicating that the “loop” positions 65, 81 and 90 are only slightly affected by the pigment binding to CP29. Also the observed differences between the holo- and apo-state of the protein on positions 33, 40, 56, 97 and 108 are consistent with the analysis of Figure 1. In the N-terminal domain, position 4 is slightly affected by the absence of pigment, however, its value for Ω is similar to the values for the positions in the more structured domains. This is remarkable for an N-terminal end position and could indicate a local structure that limits the free rotational space of the spin label. Alternatively, this N-terminus could interact with the transmembrane protein body, which is in agreement with recent fluorescence experiments with the fluorescent dye TAMRA (6-carboxy-tetramethyl-rhodamine) covalently attached to a cysteine at position 4¹⁵. In contrast, position 15 does not show a strong effect to pigment removal, but its value for Ω is at a high level, indicating rather unrestricted spin label motion at a location probably relatively far from the transmembrane protein body.

3.4.1 Summarizing model

CP29 has a strong sequential homology with LHCII, the major difference being an N-terminal insert from amino acid residue 56 to 98⁹. Also light-spectroscopic experiments have revealed a high degree of structural and functional similarity between CP29 and LHCII and demonstrate an unequivocally high similarity for the transmembrane protein bodies^{22,35-38}. Because of this strong sequence homology and spectroscopic similarities, we took the crystal structure of LHCII from spinach⁷ as a starting point for constructing a model for CP29 (Figure 4). In this figure, the extra N-terminal insert is shown as a red loop extruding from the main protein body. The amino acid residues 97 and 108 that were used for spin labeling, are located in the transmembrane protein body of the protein. Position 108 is situated on the putative transmembrane helix B of the protein, close to the center plane of the protein; position 97 is at the end of this helix, close to the stromal surface of the protein. These locations are consistent with the relatively strong difference between holo- and apoprotein and the relatively low values of Ω that are indicative for a restricted spin label motion (Figure 3). Positions 65, 81 and 90 are located in the extra N-terminal loop that is proposed to extend out of the protein in the stromal space, because for these sites the label has a large degree of freedom and is not influenced by pigment binding. This loop also contains a phosphorylation site at Thr81. This finding suggests that the flexibility of this loop could play a role in presumed regulatory functions of the phosphorylation.

Positions 33, 40, and 56 show far less rotational freedom and moreover, the corresponding ESR spectra are substantially affected by pigment reconstitution, indicating that the domain in which they are located should be close to the transmembrane protein body. Their relatively low values for Ω are similar to the values found for positions 97 and 108 (Figure 3). This observation is consistent with the crystal structure of LHCII, in which these positions are located in a folded protein domain at the stromal side of the protein⁷ (Figure 4). The next labeled position towards the N-terminal end, position 15, shows a high value for Ω suggesting rather unrestricted spin label motion. This indicates that this protein domain is in a flexible state. This is in agreement with the finding that the structure of the N-terminal amino acid residues 1 to 14 is not resolved in the crystal structure of LHCII. Finally, position 4 at the N-terminal end displays clear two-component characteristics of a broad immobile component in combination with a sharp mobile one (Figure 1, 3C). It is slightly affected by the absence of pigments, however, its value for Ω (Figure 3D) is similar to the values for the positions in the more motionally restricted domains (i.e., position 97). This suggests that the N-terminus interacts with the transmembrane protein body probably by folding back to it, however, without being strongly affected by the holo- or apo-state of the protein. This topology is in agreement with recent fluorescence experiments with the fluorescent dye TAMRA (6-carboxy-tetramethyl-rhodamine) covalently attached to a cysteine at position 4 that indicate that in about 80% of the cases the N-terminus is folded

back on the hydrophobic core¹⁵. Next to position 4 there are two phenylalanine residues. It could be hypothesized that this domain interacts with the hydrophobic amino acid residues that can be found in a groove on the stromal side of the transmembrane protein body.

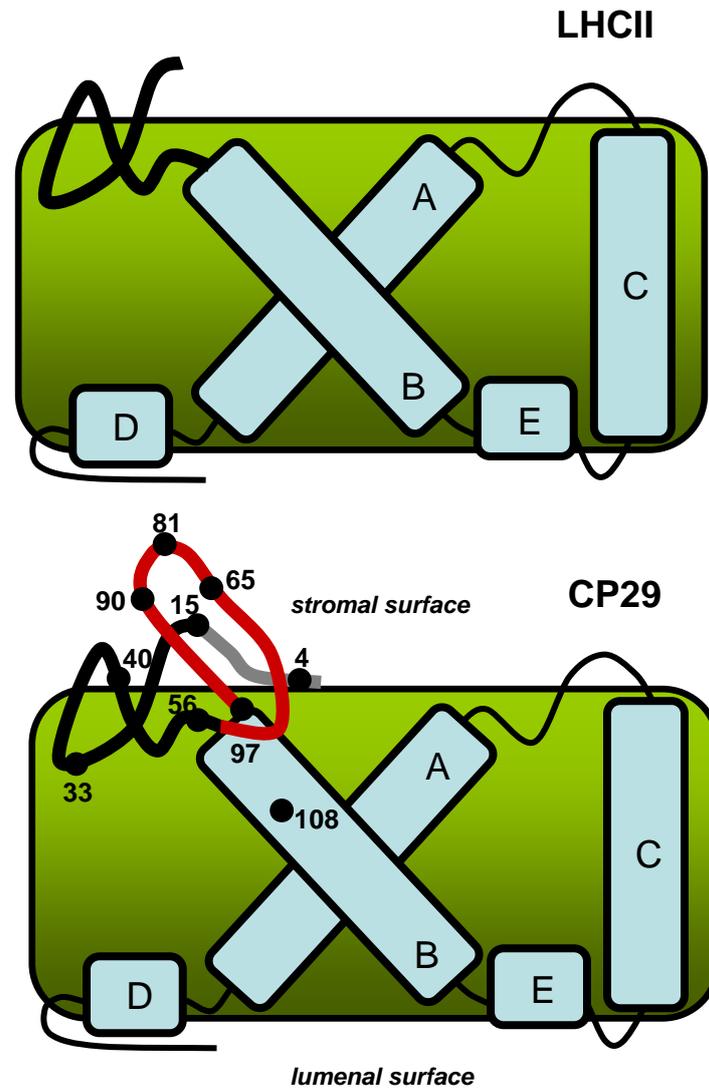


Figure 4. Schematic structural model of CP29 based on the crystal structure of LHCII from spinach⁷ (PDB ID: 1RWT). The main helical structures (A-E) of the transmembrane protein body are shown in light blue. The extra N-terminal insert of CP29 (as compared to LHCII) is shown as a red loop extruding from the main transmembrane protein body. The N-terminus from amino acid residue 1-14 is indicated in grey, as this part of the structure is not resolved in the crystal structure of LHCII. The numbers refer to the labeled positions (black dots).

Until so far, we have limited ourselves to analyze the ESR spectra of singly spin-labeled CP29 protein mutants. The main difficulty that we encountered was the limited number of available single-cysteine mutants, but this problem can be tackled by a high-throughput approach. In addition, a double-labeling approach can be applied that provides distances between spin labels placed in various domains of the protein, in a similar way as has been carried out for the major light-harvesting chlorophyll *a/b* protein (LHCIIb)³⁹. Therefore, site-directed spin labeling ESR spectroscopy is an attractive and powerful way to study the conformation and topology of the protein domains in CP29.

Acknowledgements

This work was supported by the “Stichting voor Fundamenteel Onderzoek der Materie (FOM)”, which is financially supported by the “Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)”. We thank Emilie Wientjes for making seminal contributions to this work.

References

- 1 Nelson, N. & Yocum, C.F. *Ann. Rev. Plant Biol.* **57**, 521 (2006).
- 2 Jansson, S. *Trends Plant Sci.* **4**, 236 (1999).
- 3 Caffarri, S., Croce, R., Cattivelli, L., & Bassi, R. *Biochem.* **43**, 9467 (2004).
- 4 Dekker, J.P. & Boekema, E.J. *Biochim. Biophys. Acta* **1706**, 12 (2005).
- 5 Avenson, T.J. *et al. J. Biol. Chem.* **283**, 3550 (2008).
- 6 Ahn, T.K. *et al. Science* **320**, 794 (2008).
- 7 Liu, Z. *et al. Nature* **428**, 287 (2004).
- 8 Green, B.R. & Dumford, D.G. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 685 (1996).
- 9 Bassi, R., Croce, R., Cugini, D., & Sandonà, D. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10056 (1999).
- 10 Sandonà, D., Croce, R., Pagano, A., Crimi, M., & Bassi, R. *Biochim. Biophys. Acta* **1365**, 207 (1998).
- 11 Testi, M.G., Croce, R., Laureto, P.P.-D., & Bassi, R. *FEBS Lett.* **399**, 245 (1996).
- 12 Croce, R., Breton, J., & Bassi, R. *Biochem.* **35**, 11142 (1996).
- 13 Mauro, S., Dainese, P., Lannoye, R., & Bassi, R. *Plant Physiol.* **115**, 171 (1997).
- 14 Hemminga, M.A. & Berliner, L.J., *ESR spectroscopy in membrane biophysics*. (Springer, New York, 2007).
- 15 Van Oort, B. *et al. Chem. Phys.*, In press (2008).
- 16 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *J. Chem. Inf. Mod.* **45**, 1621 (2005).
- 17 Nagai, K. & Thøgersen, H.C. *Methods Enzymol.* **153**, 461 (1987).
- 18 Paulsen, H., Rüdiger, U., & Rüdiger, W. *Planta* **181**, 204 (1990).
- 19 Porra, R.J., Thompson, W.A., & Kriedemann, P.E. *Biochim. Biophys. Acta* **975**, 384 (1989).

- 20 Davies, B., Analysis of carotenoid pigments in *Chemistry and biochemistry of plant pigments*, edited by T.W. Goodwin (Academy Press, New York, 1965), pp. 489.
- 21 Giuffra, E., Cugini, D., Croce, R., & Bassi, R. *Eur. J. Biochem.* **238**, 112 (1996).
- 22 Croce, R., Muller, M.G., Caffarri, S., Bassi, R., & Holzwarth, A.R. *Biophys. J.* **84**, 2517 (2003).
- 23 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *Biophys. J.* **91**, 3341 (2006).
- 24 Filipič, B. & Štrancar, J. *Appl. Soft Comput.* **1**, 83 (2001).
- 25 Štrancar, J. *et al. J. Chem. Inf. Mod.* **45**, 394 (2005).
- 26 Kavalenka, A.A., Filipič, B., Hemminga, M.A., & Štrancar, J. *J. Chem. Inf. Mod.* **45**, 1628 (2005).
- 27 Columbus, L. & Hubbell, W.L. *Biochem.* **43**, 7273 (2004).
- 28 Steinhoff, H.-J. *et al. Biochim. Biophys. Acta* **1457**, 253 (2000).
- 29 Štrancar, J., Sentjurc, M., & Schara, M. *J. Magn. Reson.* **142**, 254 (2000).
- 30 Štrancar, J., Koklic, T., & Arsov, Z. *J. Membr. Biol.* **196**, 135 (2003).
- 31 Caffarri, S., Passarini, F., Bassi, R., & Croce, R. *FEBS Lett.* **581**, 4704 (2007).
- 32 Paulsen, H., Finkenzeller, B., & Kühlein, N. *Eur. J. Biochem.* **215**, 809 (1993).
- 33 Horn, R., Grundmann, G., & Paulsen, H. *J. Mol. Biol.* **366**, 1045 (2007).
- 34 Hemelrijk, P.W., Kwa, S.L.S., van Grondelle, R., & Dekker, J.P. *Biochim. Biophys. Acta* **1098**, 159 (1992).
- 35 Gradinaru, C.C., van Stokkum, I.H.M., Pascal, A.A., van Grondelle, R., & van Amerongen, H. *J. Phys. Chem. B* **104**, 9330 (2000).
- 36 Pascal, A. *et al. Eur. J. Biochem.* **262**, 817 (1999).
- 37 Mozzo, M., Passarini, F., Bassi, R., van Amerongen, H., & Croce, R. *Biochim. Biophys. Acta* **1777**, 1263 (2008).
- 38 Mozzo, M., Dall'Osto, L., Hienerwadel, R., Bassi, R., & Croce, R. *J. Biol. Chem.* **283**, 6184 (2008).
- 39 Jeschke, G. *et al. J. Biol. Chem.* **280**, 18623 (2005).

Chapter 4

Analysis of side chain rotational restrictions of membrane-embedded proteins by spin label ESR spectroscopy

based on:

Štrancar, J., A. Kavalenka, P. Zihel, D. Stopar, M.A. Hemminga. 2009. Analysis of side chain rotational restrictions of membrane-embedded proteins by spin-label ESR spectroscopy. *J. Magn. Reson.* 197:245-248.

Abstract

Site-directed spin-labeling electron spin resonance (SDSL-ESR) is a promising tool for membrane protein structure determination. Here we propose a novel way to translate the local structural constraints gained by SDSL-ESR data into a low-resolution structure of a protein by simulating the restrictions of the local conformational spaces of the spin label attached at different protein sites along the primary structure of the membrane-embedded protein. We test the sensitivity of this approach for membrane-embedded M13 major coat protein decorated with a limited number of strategically placed spin labels employing high-throughput site-directed mutagenesis. We find a reasonably good agreement of the simulated and the experimental data taking a protein conformation close to the one determined by fluorescence resonance energy transfer analysis (Nazarov et al., *Biophys. J.* 92 (2007) 1296-1305).

4.1 Introduction

To identify the biological functions of proteins, it is imperative to know their three-dimensional structure. In this context, the least understood class of proteins are the integral membrane proteins^{1,2}. Although they represent 30-40% of all expressed sequences, they amount to less than 1% of proteins of known structure³. Thus membrane proteins remain an enormous challenge in structural biology.

The progress of high-resolution structural studies of membrane proteins using the two common techniques, NMR and X-ray diffraction, has been limited because both approaches are restricted by technical and practical difficulties⁴. As a result, there is an urgent need for new biophysical methodologies that can provide detailed structural information. Among the more modern biophysical techniques, site-directed spin-labeling electron spin resonance (SDSL-ESR) appears to show a high potential to further advance the field⁵⁻¹⁰.

The basis of this technique is high-throughput site-directed mutagenesis to introduce unique cysteine residues at desired locations in the protein. As site-directed mutagenesis is becoming an increasingly powerful tool in protein preparation, the usefulness of SDSL-ESR in membrane protein studies grows tremendously⁷. An additional advantage is that the membrane proteins can be examined in their native membrane environment, such as reconstituted lipid bilayer systems under their physiological conditions.

Our objective is to present the basic ideas of a new method tailored to transfer the SDSL-ESR data into structural information. To demonstrate the power of our analysis, we use the M13 major coat protein, a small reference membrane protein, and we decorate it

with a limited number of strategically placed spin labels. We extract the experimental free rotational space of the spin labels attached to the protein as published previously¹¹⁻¹³. Here we develop a molecular model to describe the conformations of the protein in a lipid environment in terms of the available free rotational space for the spin label, showing that our method provides a new advance for spin-label ESR spectroscopy in the determination of protein structures.

4.2 Methodology

For a membrane-embedded protein, the conformational space of a spin-labeled side chain is determined by three main factors: i) the local rotations of the spin-label side chain attached to the protein backbone, ii) the restrictions of the rotamers by the backbone and side chains of the neighboring amino acid residues, and iii) the restrictions imposed by the surrounding lipids. These effects are illustrated in Figure 1. In our conformational modeling, it is assumed that at room temperature the backbone motion is slow on the ESR time scale and significantly slower than the motion of the side chains¹⁴. Thus the protein fold on a timescale beyond several nanoseconds is defined by series of pairs of dihedral angles ϕ and ψ . Possible dihedral angle pairs are restricted due to steric clashes of the backbone atoms by taking into account the minimal interatomic distances (van der Waals distances, contact distances)^{15,16}. The bond lengths and angles are fixed to the values reported in the literature^{17,18}, as there is no need to resolve the individual conformation at the atomistic resolution. Instead, we want to detect the shape of the restricted conformational space that is experimentally measured by ESR. For each amino acid position including the spin-labeled cysteine side chain, the full conformational space of a side chain is generated by discrete rotations around the single bonds (Figures 1A and B). The torsion potentials are modeled by a discrete set of equiprobable but not equidistant rotational states, such that their density increases with the depth of the torsion potential at a given angle.

The statistical weight p_i of a certain conformation of spin label i is given by:

$$p_i = \begin{cases} 0, & \text{backbone overlap} \\ 1, & \text{no backbone overlap} \end{cases} \left[\prod_{k \in \text{neighboring amino acids}} \left(1 - \frac{N_{k-i \text{ overlap}}^k}{N_{all}^k} \right) \right] (1 - \sin \theta_i), \quad (1)$$

where the product in the central factor runs over all neighboring amino acid residues that share the space with the conformations of the spin label.

The first factor in Eq. (1) indicates that the conformations of the spin-labeled cysteine side chain, which overlap with the backbone are completely rejected (Figures 1C and D) as the motion of the backbone is much slower than the motion of the side chains. However, the overlap with the neighboring amino acid side chains (Figures 1E and F) is assumed to be “soft” rather than “hard”, as the wobbling of the side chains is fast on the

ESR time scale. This is taken into account by the second factor in Eq. (1), which describes the reduction of the statistical weight by the ratio of the number of overlapping conformations and the number of all possible conformations of the neighboring side chains that are allowed by the backbone overlap check.

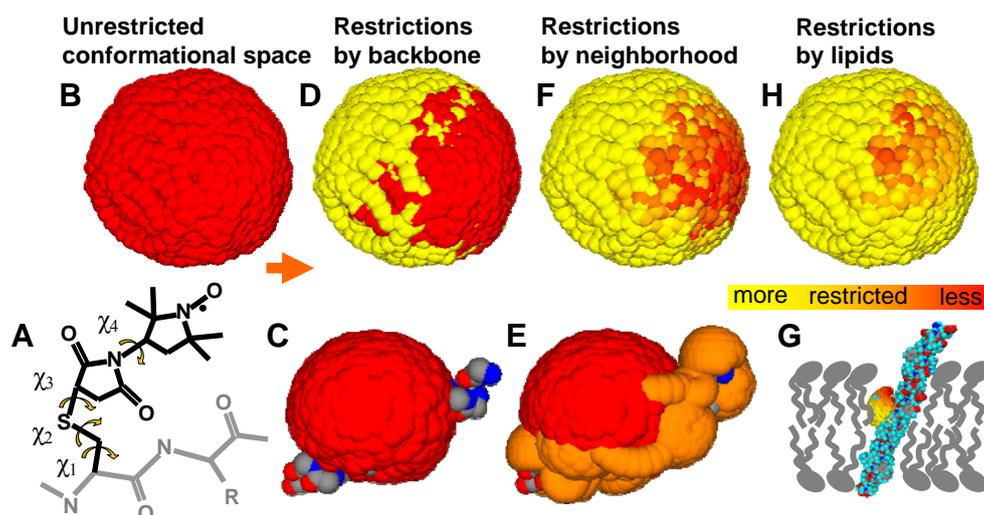


Figure 1. Schematic illustration of the conformational space of the spin-label side chain for membrane-embedded 3-maleimido proxyl spin-labeled M13 coat protein. For simplicity, the protein is assumed to be in a perfect α -helical conformation and embedded in a bilayer of 1,2-dierucoyl-sn-glycero-3-phosphocholine between amino acid position 9 and 46; the spin label is attached to a cysteine residue at position 25. **A.** The spin label attached to a cysteine residue has four free rotations (χ_1 , χ_2 , χ_3 , χ_4) around the four single bonds. **B.** Unrestricted spin label conformational space (shown in red) resulting from the free rotations of the side chain around the four single bonds. **C.** Steric overlap of the spin label with the protein backbone reduces the set of possible conformations. **D.** The available spin label conformational space after steric overlap with the protein backbone (forbidden conformations are shown in yellow). **E.** The wobbling spin label shares space with the wobbling side chains of the neighboring amino acid residues (indicated in orange). **F.** The available spin label conformational space after steric overlap with both the backbone and the side chains of the neighboring amino acid residues. The soft interaction with the neighboring amino acids is indicated by a continuous yellow-orange-red color scale (see inset). **G.** As the lipids tend to orient the amino acid side chains, conformations that are perpendicular to the membrane normal are highly restricted, which further reduces the set of allowed spin label conformations. **H.** The final available spin label conformational space subject to all three types of restrictions.

Finally, the statistical weight of each conformation of a side chain of the spin label is also decreased by restrictions due to adjacent lipids. The aligning effect of the lipids is parameterized by the angle ϑ between the membrane normal and the direction of the side chain of a particular conformation (which is defined as the direction from the $C\beta$ to the oxygen atom of the nitroxide) and in a first approximation described by $(1 - \sin \vartheta)$ (Figures 1G and H). This is provided by the third factor in Eq. (1). This factor is a simplification, based on the following requirements: (1) there are no restrictions in case of a

parallel orientation with respect to the lipids; (2) as soon as there is a non-zero angle, there should be a non-zero first derivative effect; (3) at a direction perpendicular to the lipids, the restriction should be strongest; (4) the derivative of this perpendicular effect should be zero again: there is not a very large difference whether lipid molecules are perfectly or nearly perpendicular to the side chains. The most simple and effective function that meets those criteria is the $(1 - \sin \vartheta)$ function. Since the side chain of the 3-maleimido proxyl spin label, which is used in the ESR experiments, is twice as large as compared to amino acid side chains, the aligning effect of the lipids on the other amino acid side chains can be ignored. Based on similar arguments we did not take into account the restrictive effects of amino acid side chains on one another.

Thus, the conformational space of a spin label at a specific site on a membrane-embedded protein will be sensitive to its local environment. For membrane-embedded M13 coat protein, the location of the protein relative to the lipid bilayer is defined by locking the positions of the amino acids that were experimentally determined to be at water-lipid interface^{19,20}. Note that this kind of description is proposed to describe the time-averaged SDSL-ESR experimental data and cannot be compared to the much more time-consuming molecular dynamics approach, which on the other hand would actually resolve the time evolution of the conformations.

Recently we have introduced a method of analysis of ESR spectra of site-directed labeled proteins, which provides information about the conformational space of the spin-labeled sites¹¹⁻¹³. The conformational space of a spin label is quantified by the normalized free rotational space Ω , which measures the effective solid angle of the cone left for spin label wobbling. This parameter can also be deduced from molecular modeling of the restriction in the rotational space of the side chains (Figure 1), by interpreting the results of the modeling in terms of a cone model^{12,13}. For this, we calculate the average direction of the nitroxide N-O bonds using the statistical weights of the conformations. The averages are converted into two cone angles ϑ_0 and φ_0 that characterize the anisotropy of the rotational space. From the cone angles we finally compute the simulated normalized free rotational space Ω as follows:

$$\Omega = \frac{\vartheta_0 \varphi_0}{(\pi/2)^2}, \quad (2)$$

which can then be compared to the experimental values of Ω ¹².

In summary, the free rotational space of a spin label is an attractive parameter to consider for protein structure analysis, as it will be affected by its local environment as given by the primary sequence, fold of the protein backbone, adjacent protein domains in a tertiary protein structure and, for membrane proteins, the phospholipids in which the protein is embedded. All computer models were realized as Delphi classes using the Borland Delphi 6.0 environment. The Pascal classes and the software are available from the authors upon request.

4.3 Results

The protein modeling was tested by comparing the simulated free rotational space of a membrane-bound M13 major coat protein to recently published experimental data¹² (Figure 2, red triangles). For this protein, consisting of 50 amino acid residues, 27 single cysteine mutants were available. They span the whole primary sequence of the protein and they cover almost the complete range of values of the free rotational space Ω of the 3-maleimido proxyl spin label for the protein reconstituted in phospholipid bilayers consisting of 1,2-dierucoyl-*sn*-glycero-3-phosphocholine^{11,20}.

The experimental free rotational space Ω was compared with the value of Ω obtained from the simulation of the restrictions of the side chain rotational spaces (Figure 2). For simplicity, we assumed a membrane-embedding of the protein based on a recently published model, using an α -helical protein with a tilt angle of 18° with respect to the membrane normal and with membrane crossing points at positions 9 and 47^{19,21,22}. To analyze the effect of protein conformation and membrane-embedding on the simulated free rotational space Ω , we generated a number of 5000 different helical structures of the protein with dihedral angles ϕ and ψ uniformly distributed around the values for an α -helix: $-57\pm 30^\circ$ and $-47\pm 30^\circ$, respectively. The Ω values related to the original α -helical protein model ($\phi = -57^\circ$ and $\psi = -47^\circ$) are indicated with white triangles in Figure 2. The observed variation in Ω values represents the effect of the various amino acid residues in the primary sequence of the protein. In one set of simulations, we left out the lipid effect in Eq. (1), showing the variation of Ω for a ‘free’ protein (Figure 2A). At all spin label positions along the primary sequence of the protein the simulated Ω values were summarized into frequency histograms (see the cyan-blue histograms of the relative frequency of a given value of Ω in Figure 2). As can be seen, the calculated restrictions from the simulated helical structures produce a wide range of Ω values that nicely cover the experimental data. In a second simulation approach, the effect of the lipids was included. In this case, there is a reasonably good agreement between the SDSL-ESR experimental data and the simulated data for all spin label positions (Figure 2B). The deviating positions 25-29 most likely indicate that the simulated structure did not produce locally a secondary structure motif that would sufficiently restrict the conformational space of the spin label. We will address this problem by introducing an optimization procedure in our calculation, which would tune the backbone dihedral angles and in fact eventually would produce an optimized ensemble of best-fitting structures.

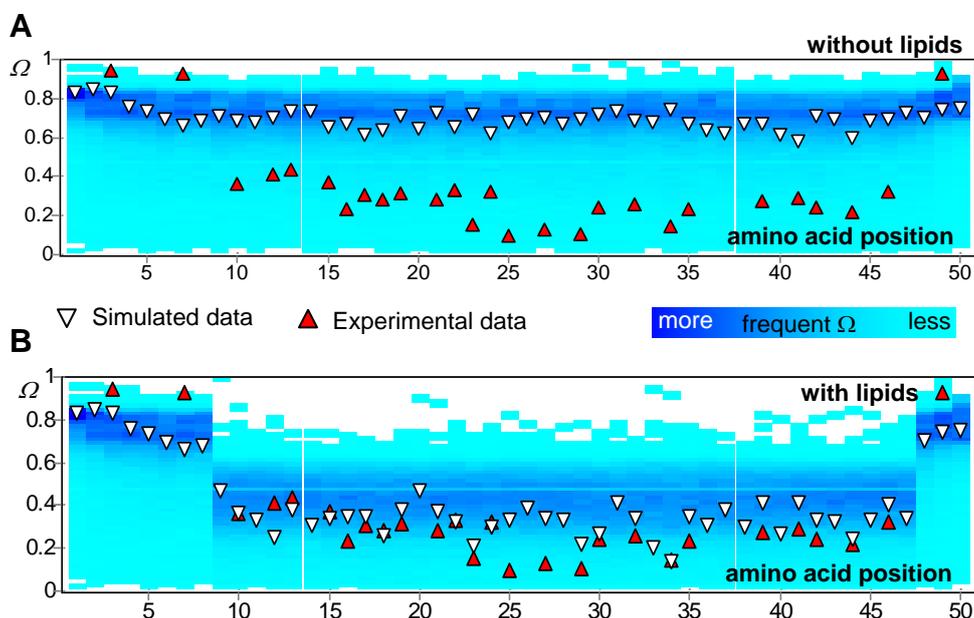


Figure 2. Sensitivity of the free rotational space to the primary sequence and variations of the protein secondary structure and the effect of the lipids for membrane-embedded spin-labeled M13 coat protein. The histograms of the relative frequency of a given value of Ω (color-coded by continuous shades of blue, such that cyan is the lowest frequency and dark blue is the highest frequency within the set of 5000 modeled near helical structures; see text) at an amino acid position along the primary sequence are plotted both for the ‘free’ protein (A) and for the protein in a lipid environment (B). The red triangles correspond to the experimental values of Ω . The white triangles indicate the Ω values related to the original α -helical protein model ($\varphi = -57^\circ$ and $\psi = -47^\circ$) as defined in ^{19,21,22}.

4.4 Conclusion

The key factor to the efficiency of our computational approach is the adjusted spatial and temporal resolution of the molecular modeling guided by the characteristic scales of the spin-label ESR experiments. The ESR experiment is insensitive to the exact atomic coordinates, but it enables us to track the rotational conformations of the amino acid side chains. With this in view, the simulation algorithm is designed to optimize its runtime without compromising the level of detail of the analysis. Since ESR spectroscopy is very sensitive to the available space of the fast rotational motion of the spin label attached to the protein, the rotational conformational space of the side chain can be taken as the most strategic unit in our protein modeling. The proposed search of the conformational space for each spin-labeled protein mutant requires a new approach in the modeling strategy as the standard modeling techniques and molecular dynamics simulations are not ideally suited

for such an ESR data analysis and consequently are not realistically applicable within the computer time frames possible.

The next step will be to set up an optimization algorithm that will enable to find the best possible structures of the protein based on the Ω data. In this respect, the backbone dihedral angles will be continuously changed and the local restrictions will be recalculated, thereby optimizing the secondary structure of the protein. The goodness of fit to the experimental data will guide the optimization procedure through the search space towards more favorite structures. At the end of the optimization, more than one structure can produce equally good fits to the experimental Ω data, indicating a set of allowed global protein conformations. Such a method is comparable to the distance geometry approach employed in two-dimensional solution NMR spectroscopy that also results in a family of structures²³. Based on our experience with evolutionary optimization methods^{24,25}, the estimated time frame for the structure optimization for a protein of size of 150 amino acid residues and 50 different single cysteine mutants will be about four weeks using twenty 4-GFLOP processors, which makes this approach highly competitive compared to other high-resolution methodologies.

As compared to well-established ESR tools of structure determination, such as accessibilities and distance constraints, our method provides an alternative approach. Our method has the advantage of providing direct information about the local secondary structure at physiological temperatures (i.e., room temperature) with singly labeled protein samples, without changing the sample conditions. In the case of accessibility experiments relaxation agents, such as Ni²⁺ ions or oxygen, need to be added to the sample. To determine distance constraints, two spin labels need to be engineered at the protein and for spin echo ESR experiments the sample has to be cooled to a low temperature (around 50 K)⁷.

Acknowledgements

This work was supported in part by contract no. QLG-CT-2000-01801 of the European Commission (MIVase – New Therapeutic Approaches to Osteoporosis: targeting the osteoclast V-ATPase) as well as by the Slovenian Research Agency (programs P1-0060 and P1-0055 and project J1-6581).

References

- 1 Arora, A. & Tamm, L.K. *Curr. Opin. Struct. Biol.* **11**, 540 (2001).
- 2 Lacapere, J.-J., Pebay-Peyroula, E., Neumann, J.-M., & Etchebest, C. *Trends Biochem. Sci.* **32**, 259 (2007).
- 3 White, S., Membrane proteins of known 3D structure, Available at http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html, (2009).
- 4 Torres, J., Stevens, T.J., & Samso, M. *Trends Biochem. Sci.* **28**, 137 (2003).
- 5 Hubbell, W.L., Cafiso, D.S., & Altenbach, C. *Nat. Struct. Biol.* **7**, 735 (2000).
- 6 Columbus, L. & Hubbell, W.L. *Trends Biochem. Sci.* **27**, 288 (2002).
- 7 Hemminga, M.A. & Berliner, L.J., *ESR spectroscopy in membrane biophysics*. (Springer, New York, 2007).
- 8 Beier, C. & Steinhoff, H.-J. *Biophys. J.* **91**, 2647 (2006).
- 9 Budil, D.E., Sale, K.L., Khairy, K.A., & Fajer, P.G. **110**, 3703 (2006).
- 10 LaConte, L.E., Voelz, V., Nelson, W., Enz, M., & Thomas, D.D. *Biophys. J.* **83**, 1854 (2002).
- 11 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *J. Chem. Inf. Mod.* **45**, 1621 (2005).
- 12 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *Biophys. J.* **91**, 3341 (2006).
- 13 Štrancar, J. *et al. J. Chem. Inf. Mod.* **45**, 394 (2005).
- 14 Karplus, M. & McCammon, J.A. *CRC Crit. Rev. Biochem.* **9**, 293 (1981).
- 15 Ho, B.K. & Brasseur, R. *BMC Struct. Biol.* **5**, 14 (2005).
- 16 Ho, B.K., Thomas, A., & Brasseur, R. *Protein Sci.* **12**, 2508 (2003).
- 17 Word, J.M. *et al. J. Mol. Biol.* **285**, 1711 (1999).
- 18 Xiang, Z. & Honig, B. *J. Mol. Biol.* **311**, 421 (2001).
- 19 Koehorst, R.B.M., Spruijt, R.B., Vergeldt, F.J., & Hemminga, M.A. *Biophys. J.* **87**, 1445 (2004).
- 20 Stopar, D., Spruijt, R.B., & Hemminga, M.A. *Chem. Phys. Lipids* **141**, 83 (2006).
- 21 Nazarov, P.V., Koehorst, R.B.M., Vos, W.L., Apanasovich, V.V., & Hemminga, M.A. *Biophys. J.* **91**, 454 (2006).
- 22 Nazarov, P.V., Koehorst, R.B.M., Vos, W.L., Apanasovich, V.V., & Hemminga, M.A. *Biophys. J.* **92**, 1296 (2007).
- 23 Bax, A. *Annu. Rev. Biochem.* **58**, 223 (1989).
- 24 Filipič, B. & Štrancar, J., 2003.
- 25 Kavalenka, A.A., Filipič, B., Hemminga, M.A., & Štrancar, J. *J. Chem. Inf. Mod.* **45**, 1628 (2005).

Chapter 5

Optimization of membrane protein structure based on SDSL-ESR constraints and conformational space modeling

based on:

Kavalenka, A., M. A. Hemminga, and J. Štrancar, submitted. Optimization of membrane protein structure based on SDSL-ESR constraints and conformational space modeling.

Abstract

The problem of structure determination of membrane proteins is addressed with a new combination of site-directed spin labeling (SDSL) electron spin resonance (ESR) spectroscopy and structure modeling of a membrane-embedded protein. A structural model is developed to simulate the free rotational space of a spin label attached to a membrane protein, taking into account the restricting effect of the protein backbone, amino acid side chains and lipid environment. To validate our model, we compared the simulation data of the conformational space of 3-maleimido proxyl spin label with experimental data obtained from SDSL-ESR spectra of 27 mutants of M13 coat protein reconstituted in phospholipid bilayers. By using an optimization algorithm we optimized the parameters of the protein-lipid model by improving the fit of simulation data to experimental conformational space data. The outcome of the optimization is a family of best-fit structures of membrane-embedded M13 protein, which not only agree with the available SDSL-ESR data, but also are consistent with a recent model based on site-directed fluorescence labeling.

5.1 Introduction

One of the most challenging fields of structural biology and structural proteomics is the structure determination of membrane proteins ^{1,2}. Although one third of all proteins are membrane proteins, less than 1% of the known protein structures correspond to membrane proteins ³. Difficulties in the application of standard high-resolution methods for three-dimensional protein structure determination, i.e., X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, have stimulated the development of alternative approaches. One of such techniques is site-directed spin labeling (SDSL) electron spin resonance (ESR). This technique provides both a structural and dynamical characterization of the local conformations of a membrane protein in its native environment, and therefore evolves into a very useful method for the structure analysis of membrane proteins ⁴⁻⁸. However, the problem of transforming local structural information from different labeled protein sites into a global protein structure remained unsolved.

In the present work a structural model is developed to simulate the free rotational space of a membrane protein specifically spin labeled at different sites along the primary sequence. This is carried out by calculating the motional restrictions of the spin label due to the primary and secondary protein structure, as well as due to the membrane lipids ⁹. A comparison of the experimental and theoretical values of the free rotational space profiles and optimization of parameters of the protein-lipid model reveals the structure and membrane embedment of the protein. Optimization of the protein-lipid model implies tuning the secondary structure of the protein, and optimizing the relative position, tilt and orientation of the protein in the membrane. At the end of the optimization, several

structures may produce equally good fits to the experimental data, indicating a set of allowed global protein conformations. Such a method is comparable to the distance geometry approach employed in two-dimensional solution NMR spectroscopy that also results in a family of structures¹⁰.

In a recent SDSL-ESR study of membrane-embedded M13 protein, the spectra of 27 spin-labeled single-cysteine protein mutants were measured and analyzed in terms of the free rotational space of the spin labels attached to the protein, i.e., the effective space within which the attached spin labels may wobble^{6,7}. In the present work, this set of data is used as an input in the structure calculations guided by the free rotational space. The resulting family of best-fit structures of membrane-embedded M13 protein agrees well with the available SDSL-ESR data⁷, but is also consistent with a recent model based on site-directed fluorescence labeling¹¹⁻¹⁴.

5.2 Theory and methodology

5.2.1 Modeling approach

For SDSL-ESR spectroscopy a protein is labeled at a specific site with a spin label of a size comparable to the size of the amino acid residues. This makes the rotational conformational space of the spin label sensitive to the local protein structure. In addition lipids surrounding the spin-labeled site at the membrane protein provide restrictive forces that will also limit the rotational conformational space of the spin label. There are three interesting features of ESR spectroscopy that can facilitate efficient modeling of the conformational space of the spin label:

1. The motion of protein side chains at physiological conditions (i.e., room temperature) is fast on the nanosecond ESR time scale⁴. For example, if the temperature is decreased, or a protein is put in an extremely rigid environment, the side chain motion of the spin label will be significantly slowed down or the spin label could become immobilized due to stabilizing interactions¹⁵, making it insensitive to the space restrictions imposed by molecular groups of the protein and lipids, in which the protein is embedded. Such cases, however, are easily recognized experimentally by rigid-like ESR lineshapes and will be avoided in the approach discussed here.

2. The backbone motion is much slower than the nanosecond ESR time scale^{16,17}, or at least slower than the side chain motion. This is especially valid for proteins embedded into membranes or in large multi-chain protein complexes¹⁸. However, backbone atoms near the terminal ends can move to a larger extent, but such a case can be easily recognized in the ESR spectra and treated separately during the modeling.

3. The ESR experiment is insensitive to the exact atomic coordinates. Therefore there is no need for a precise calculation of an individual side chain conformation. This means that in the calculations only the average effects of all wobbling chains (amino acid side chains of the protein and acyl chains of the lipids) needs to be taken into account.

These characteristic aspects of ESR spectroscopy offer us the opportunity to develop an approach that is computationally manageable and which provides a sufficiently high resolution to match the SDSL experiment. Under these conditions modeling of the conformational space of a spin-labeled side chain and its restrictions will be based on the following steps (see Figure 1) ⁹:

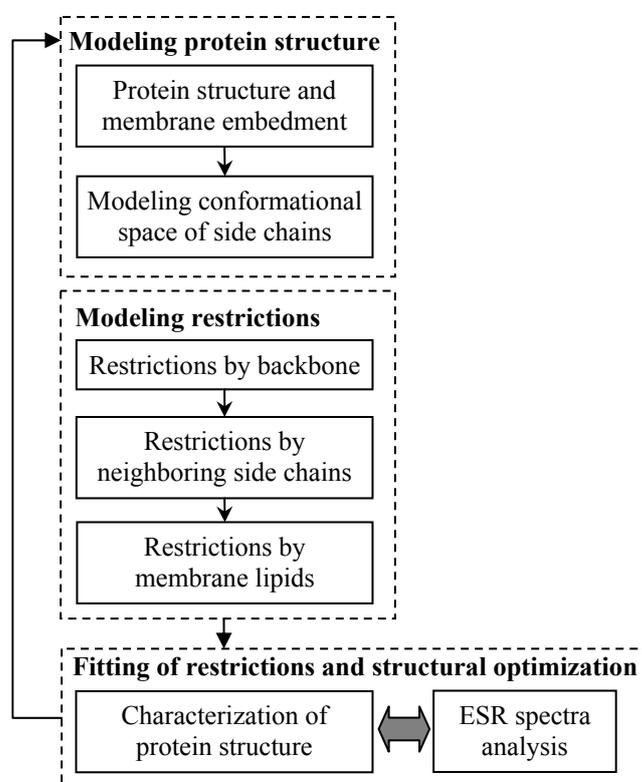


Figure 1. Overview of the approach of membrane protein structure modeling, calculation of the restrictions, and structure optimization based on local structural constraints from SDSL-ESR.

1. Modeling of the membrane protein structure including all local rotations of the side chain of the amino acid residues and spin label, i.e., the conformational space.

2. Modeling of the restrictions of the conformational space of the spin label by the protein backbone, the side chains of neighboring amino acid residues, and the surrounding lipids.

3. Characterization of the restricted conformational space of the spin label in terms of the so-called normalized free rotational space Ω , to enable a) comparison with the

experimental SDSL-ESR data; b) optimization of the parameters of the structural model to provide a simultaneous fit of the modeled restrictions to the experimental restrictions; c) characterization of the structure of the protein and its embedment in the membrane.

5.2.2 Modeling of membrane-embedded protein

Protein modeling. To derive the atomic coordinates of a protein molecule, first its backbone structure is modeled, based on the backbone dihedral angles φ_i and ψ_i for each i -th amino acid residue. This is followed by the attachment of the amino acid side chains to the backbone. Since an atomistic resolution of modeling is not necessary for our case, the atomic structures of the amino acid residues and spin label are constructed using an approximation of fixed bonds lengths and bonds angles¹⁹. The bond lengths and bond angles that we use are based on previously reported values²⁰⁻²². The so-called Ramachandran plot²³, which contains the allowed distributions of the backbone dihedral angles φ_i and ψ_i , was calculated with our model and compared with previously published plots²³⁻²⁶. When optimizing the protein structure a pre-calculated Ramachandran plot was used to speed up the calculations by excluding forbidden secondary structures from the search. In the description of the protein structure two coordinate systems, absolute and relative, are simultaneously used. In an absolute Cartesian coordinate system, the structure of the protein and the coordinates of the amino acid side chains are stored. A relative coordinate system is used when constructing the side chain conformations²⁷⁻²⁹. The details of the protein modeling are described in the Appendix A.

Protein embedment in a membrane. For our analysis, we use membrane-embedded M13 major coat protein, which is 50-residues long and almost α -helical³⁰. This protein was selected, because it has been subject of a large number of biophysical studies in bilayers, making it ideally suited as a reference protein for a review, see ref³¹. In addition, an excellent set of experimental spin label data is available of membrane-embedded M13 protein⁷ that can be used to test the performance of our approach. Since this protein has a single transmembrane domain, it is virtually placed in a lipid bilayer by setting its initial start and end point of the transmembrane region using the information from previous work^{7,11,12,14,32}. The tilt angle of the protein is derived from the effective length of the transmembrane region and the steric thickness³³ of the lipid bilayer. Also the initial orientation angle, which defines the protein rotation around the symmetry axis of the helix, is taken from previous work^{11,12,14}. The effect of the lipids is modeled as a restrictive potential along the transmembrane region as the lipids tend to orient the side chains of the amino acid residues of the protein parallel to the acyl chains of the lipids. As discussed in the next section, we will assume that this restrictive potential is also effective in the head group region of the membrane.

5.2.3 Modeling side chain restrictions

Sampling side chain conformations. Different combinations of torsion angles around single bonds of a side chain result in a set of rotamers, or side chain conformations, i.e., the conformational space. For sampling the conformational space, we use a so-called Residue-Parts-Groups mechanism, which links neighboring side chain conformations and considerably speeds up the conformational sampling when checking the overlap between neighboring side chains. In this approach, the residue is split into parts according to the number of free bond rotations, so that all atoms within one part preserve their relative positions (see Appendix A, Figure A1D). Each large part is split into atom groups, while each group contains one heavy atom (C, O, N or S) with hydrogen atoms, if there are any. Thus, the largest parts are the aromatic rings of tryptophan, tyrosine, and phenylalanine. Glycine and proline are considered as an exception. The side chain of glycine is just a hydrogen atom, which makes it very flexible, so that almost all dihedral angles are possible. The side chain of proline has a cyclic structure, which gives it conformational rigidity by locking one of the dihedral angles. When calculating the conformational spaces, we will mainly focus on the repulsive van der Waals interactions that dominate close interatomic distances^{16,34,35}. For simplicity, interactions stabilizing the tertiary structure, such as sulfide bridges and hydrogen bonding, and helix-helix interactions due to effects of macrodipoles are not included.

In determining the overlap between two side chain conformations, the steric contacts between atoms are checked only if the conformations are close enough in space: the distance is checked on the level of residues, parts, groups and then finally atoms. When determining the steric contacts between any two atoms, the distance between the atoms is compared with the sum of the original van der Waals radii, assuming that there are no interactions between the atoms that would allow any closer contacts. If a conformation of one side chain overlaps with another neighboring side chain, automatically all conformations, which partially repeat the current conformation, inherit this overlap result.

Unrestricted conformational space. When modeling the conformational space of an amino acid side chain, we use equidistant rotational states with an optimized grid step varying from 10 to 45°⁹. A torsion potential (similar as the “three staggered potential”) is implemented via fixed orientations of two subsequent bonds. However, the orientation of the second subsequent bond relative to the first one is much more poorly defined through molecular orbital effects, and it rather depends on van der Waals steric clashes, as was shown elsewhere^{36,37} (see Appendix A). The van der Waals interaction is approximated only by a repulsion part (hard sphere exclusion volume) that takes care of the steric clashes in accordance with the high-temperature approximation that is used in our approach. Note, that in this case the effective van der Waals radii are reduced standard van der Waals radii,

arising from nonbonding electrostatic interactions in O-H, C=O and CH groups, as well as the anisotropy of C atom electron shell and nonspherical shape of the electron shell of C atoms. This enables a closer contact between the atoms than allowed on basis of the original van der Waals radii. The effective van der Waals radii are calculated from an analysis of protein structure data in the PDB data bank ²⁴ (see Table A4). Conformations that have internal overlap are eliminated. In general, a too large number of conformations becomes a computational bottleneck in determining the restrictions due to overlap between neighboring side chains. In the course of our work, we found a compromise between modeling accuracy and computational costs by taking approximately 3000 conformations for the spin label and at most 1000 for the amino acid side chains. For simplicity, we assume that the initial probability $P^{initial}$ of the different conformations in the unrestricted conformational space is equal ⁹. Lists of allowed conformations, i.e., the unrestricted conformational space, calculated for each amino acid residue of the protein (see Figure 2A) are stored in memory and then used later for calculating the restrictions.

Restricted conformational space. The statistical weight of side chain conformations not restricted by the backbone is further reduced by restrictions from adjacent amino acid side chains. If two residues are close and their side chains are large enough, their conformational spaces will overlap (see Figure 2A). The extent of overlap depends on the relative position of these two residues in the protein, i.e., on the local secondary structure. By this effect, the statistical weights of the overlapping conformations of both residues are reduced, i.e., the statistical weight of the i -th conformation that shares space with $N_k^{overlaps}$ conformations of the neighboring k -th residue is reduced by a factor ⁹:

$$F_k^i = \frac{N_k^{all} - N_k^{overlaps}}{N_k^{all}}, \quad (1)$$

where N_k^{all} is the number of all conformations of the neighboring residue.

For a spin label attached to the protein, a chosen conformation of its side chain may overlap with many conformations of other neighboring amino acid side chains. Therefore, the probabilities for each of the overlapping pairs of conformations should be factorized. As a result the statistical weight of a particular spin label conformation will be reduced to a value between 0 and the initial value $P^{initial}$, depending on the extent of overlap. Thus the combined restriction of the conformational space of the spin label from the neighboring residues is a product of factors ⁹:

$$P_i = P_i^{initial} \prod_k^n F_k^i, \quad (2)$$

where P_i and $P_i^{initial}$ are the statistical weights (probabilities) of the restricted and unrestricted i -th side chain conformation, respectively. The factors F_k^i are given by Eqn. 1. This effect is illustrated in Figure 2B.

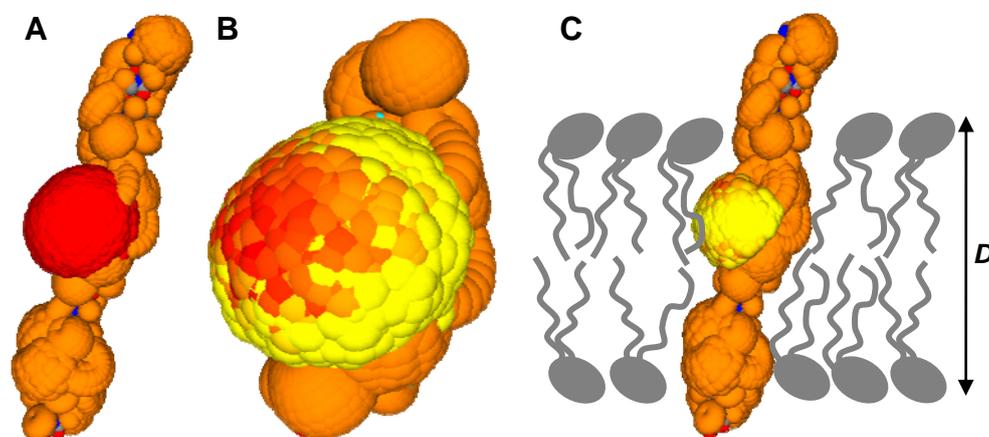


Figure 2. Modeling of the allowed conformational space of M13 protein labeled with 3-maleimido proxyl spin label at position 25. **A.** Initial unrestricted conformational space of the spin label in which each small sphere represents one side chain conformation, as given by the coordinates of the oxygen atom of the nitroxide group (red), and the rotational conformational space of the amino acid residues, as given by the coordinates of the most distant heavy atom (orange) of the residue. **B.** The conformational space of the spin label is restricted by the protein backbone and neighboring amino acid side chains. The statistical weight of the conformations is coded with a red-yellow color gradient in which the red and yellow color correspond to unrestricted and restricted conformations, respectively. **C.** Restrictive effect of the lipids for M13 protein embedded in 14:1PC bilayers^{6,7}. The steric thickness of the phospholipid bilayer D includes both the hydrophobic and head group regions of the membrane

Finally, if the spin-labeled protein site is in a transmembrane region the statistical weight of the spin label conformations is further reduced by the lipid molecules. The side chains of the amino acid residues as well as of the spin label that are surrounded by the lipid environment will feel the fluctuating lipid acyl chains as well as a restrictive effect of lipid head groups. It is assumed that this lipid effect arises from a lateral pressure profile that is present in a membrane³⁸. For a spin label at each single mutant position in the lipid bilayer we hypothesize that the lateral pressure is constant for all its conformations. Thus only the direction of the conformation relative to membrane normal makes a difference in the restrictive effect of the lipids. The same applies for the side chains of the amino acid residues of the protein. Such a restrictive lipid effect is in agreement with the finding that the fluctuations of the lipid molecules are on the time scale from ps to ns³⁹ and that the lipid-protein interactions are just slightly more favorable than lipid-lipid interactions⁴⁰. Recent molecular dynamics simulations show that side chains from aromatic, polar and charged amino acid residues tend to orient along the membrane normal^{41,42}, supporting our model. This leads to the statistical weight P_i of the i -th conformation given by⁹:

$$P_i = P_i (1 - \sin \vartheta_i), \quad (3)$$

where ϑ_i is the angle between the membrane normal and side chain direction of a i -th conformation (which is the vector from the β -carbon atom to nitroxide oxygen atom of the spin label side chain).

In combining the contributions for the restrictions of the conformational space, the statistical weight P_i of a certain conformation on a single i -th side chain conformation can be summarized as ⁹:

$$P_i = P_i^{initial} \begin{cases} 0, & \text{backbone overlap} \\ 1, & \text{no backbone overlap} \end{cases} \prod_k^n \left(1 - \frac{N_k^{overlaps}}{N_k^{all}} \right) (1 - \sin \vartheta_i). \quad (4)$$

The effect of the lipids on the conformational space of membrane-embedded spin-labeled M13 coat protein is shown in Figure 2C.

Characterization of the restricted conformational space of the spin label. ESR spectroscopy is sensitive to the orientation of the spin label nitroxide group relative to the external magnetic field, thus the distribution of orientations is reflected in the measured ESR spectrum. In general the orientation of the NO group does not coincide with the geometrical orientation of the side chain of the spin label. Both orientations depend on a combination of free rotations of the side chain. The difference between geometrical orientations of the side chain and the orientation of the nitroxide group varies from one conformation (rotamer) to another. This is taken into account in our model: conformational space restrictions are calculated for geometrically defined conformations, while the calculation of the restrictions for the conformational space is based on nitroxide group orientations.

To characterize the restricted conformational space of the spin label, we refer to the cone model that is also used in the analysis of experimental ESR spectra ^{7,43,44}. The cone model is parameterized with the angles ϑ_0 and φ_0 (Figure 3A), which describe the amplitude and the anisotropy of the spin label rotational motion within a cone, respectively. Parameters ϑ_0 and φ_0 , available from ESR spectra analysis, are connected with the averages $\overline{\cos^2(\vartheta)}$ and $\overline{\sin^2(\varphi)}$:

$$\overline{\cos^2(\vartheta)} = \frac{1}{3} [\cos^2(\vartheta_0) + \cos(\vartheta_0) + 1] \quad (5)$$

$$\overline{\sin^2(\varphi)} = 1 - \overline{\cos^2(\varphi)} = \frac{1}{2} \left(1 - \frac{\sin(2\varphi_0)}{2\varphi_0} \right) \quad (6)$$

The averages $\overline{\cos^2(\vartheta)}$ and $\overline{\sin^2(\varphi)}$ can be calculated numerically from the modeled restriction of the conformational space of the spin label (Figure 3B-D). The average $\overline{\cos^2(\vartheta)}$ characterizes the opening of the simulated spin label conformational space (a larger value indicates a higher restriction of the conformational space). The average

$\overline{\sin^2(\varphi)}$ characterizes the asymmetry of the simulated spin label conformational space (a smaller value corresponds to a more asymmetric conformational space). The average $\overline{\cos^2(\vartheta)}$ is calculated by:

$$\overline{\cos^2(\vartheta)} = \frac{\sum_i^N \left(\overline{\xi_i^{NO}} \cdot \overline{\xi_{avg}^{NO}} \right)^2 p_i \rho_i}{\sum_i^N p_i \rho_i}, \quad (7)$$

where $\overline{\xi_i^{NO}}$ is the normalized nitroxide NO bond direction of the i -th conformation, $\overline{\xi_{avg}^{NO}}$ is the normalized average nitroxide NO bond direction, and p_i and ρ_i are the probability (statistical weight) and local density of the i -th conformation, respectively. The average $\overline{\sin^2(\varphi)}$ is calculated by:

$$\overline{\xi_i^X} = \overline{\xi_i^{NO}} - \left(\overline{\xi_i^{NO}} \cdot \overline{\xi_{avg}^{NO}} \right) \overline{\xi_{avg}^{NO}}, \quad (8)$$

$$\sin^2(\varphi_i) = 1 - \left(\overline{\xi_i^X} \cdot \overline{\xi_{ref}^X} \right)^2 \quad (9)$$

$$\overline{\sin^2(\varphi)} = \frac{\sum_i^N \left(1 - \left(\overline{\xi_i^X} \cdot \overline{\xi_{ref}^X} \right)^2 \right) p_i \rho_i \sqrt{1 - \left(\overline{\xi_i^{NO}} \cdot \overline{\xi_{avg}^{NO}} \right)^2}}{\sum_i^N p_i \rho_i \sqrt{1 - \left(\overline{\xi_i^{NO}} \cdot \overline{\xi_{avg}^{NO}} \right)^2}} \quad (10)$$

where $\overline{\xi_i^X}$ is a projection of $\overline{\xi_i^{NO}}$ on a plane perpendicular to $\overline{\xi_{avg}^{NO}}$; $\overline{\xi_{ref}^X}$ is a normalized reference direction for the calculation of the asymmetry of the conformational space that corresponds to the highest radial density of $\overline{\xi_i^X}$ directions.

Both angles ϑ_0 and φ_0 are combined into one characteristic parameter, the so-called normalized free rotational space:

$$\Omega = \frac{\vartheta_0 \varphi_0}{(\pi/2)^2}, \quad (11)$$

which can be compared to the normalized free rotational space values extracted from SDSL-ESR experimental data^{7,9}. The results of testing of the sensitivity of the normalized free rotational space are presented and discussed in the Appendix B.

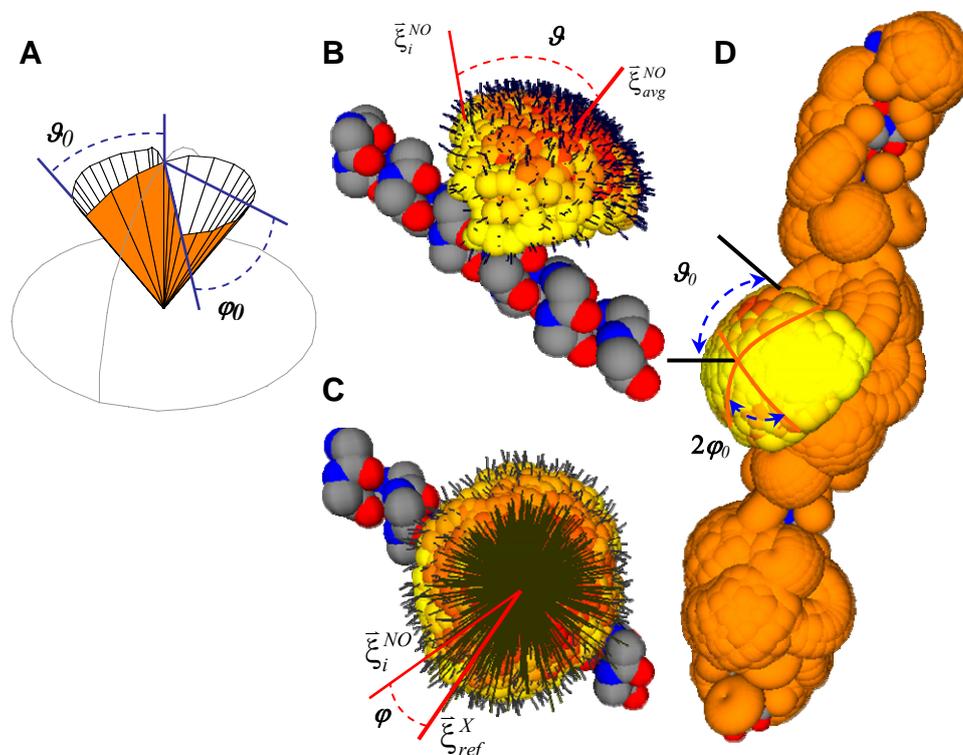


Figure 3. Characterization of the conformational space of M13 protein labeled with 3-maleimido proxyl spin label at position 25 in terms of a cone model. **A.** Cone model parameterized with angles ϑ_0 and φ_0 . **B.** The opening of the conformational space is characterized by the average $\cos^2(\vartheta)$ (Eqn. 7). The angle ϑ is between the NO bond direction $\bar{\xi}_i^{NO}$ in the i -th conformation and the average direction of the NO bond $\bar{\xi}_{avg}^{NO}$. The NO vectors at different conformations are represented with small black direction lines. **C.** The asymmetry of the conformational space is characterized by $\sin^2(\varphi)$ (Eqn. 10). The angle φ is between the X direction $\bar{\xi}_i^{NO,X}$ (a projection of $\bar{\xi}_i^{NO}$ on a plane perpendicular to the average NO direction $\bar{\xi}_{avg}^{NO}$) of the i -th conformation and a reference X direction $\bar{\xi}_{ref}^{NO,X}$ (direction that corresponds to the highest radial density of the $\bar{\xi}_i^{NO}$ directions). **D.** Schematic illustration of the cone angles ϑ_0 and φ_0 on the conformational space of the nitroxide spin label.

Protein structure optimization. The goal of our work is to use the free rotational space that is experimentally obtained for a spin-labeled membrane protein along its primary sequence, as a constraint in optimizing its three-dimensional structure and membrane-embedding. For this, we use a stochastic optimization algorithm to tune the secondary structure of the protein and the relative position of the protein in the membrane, so that the calculated local restrictions would correspond to the characteristics extracted from the experimental ESR data. The parameters that are optimized are listed in Table 1.

Table 1. Optimization parameters of protein structure and membrane embedment for a single membrane-spanning transmembrane M13 major coat protein³⁰.

Parameter	Unit	Description
$\{\varphi_i, \psi_i\}$	°	2x50 pairs of dihedral angles (the first and last angles, φ_1 and ψ_{50} , are not defined)
n_{start}	-	Starting position of the transmembrane region of the protein
n_{end}	-	End position of the transmembrane region of the protein
D	Å	Steric thickness of the membrane (Figure 2C)
θ	°	Tilt angle of the protein with respect to the membrane normal
φ	°	Rotational angle (rotation of the protein relative to the bilayer around the long axis of the protein)
$dshift$	Å	Shift of the protein in the bilayer along the membrane normal (used for the fine-tuning of the transmembrane position of the protein)

The optimization module is based on a stochastic algorithm of the Metropolis Monte Carlo family⁴⁵ with several elements of the Evolutionary Optimization (mutation operator, replacement operator and elite)^{46,47}. Unlike conventional Evolutionary Optimization, each optimization run in our algorithm tunes a single structure. One run counts for 200 generations. At each generation the current structure of the protein is modified by mutation (change of backbone dihedral angles) and crossover (introduction of the secondary structure motifs achieved in previous generations and stored in elite) operators. The parameters that describe the protein-lipid model (Table 1) are tuned simultaneously: the steric thickness³³ has to match protein tilt angle and transmembrane length, in addition the relative orientation of the protein in the membrane is tuned by a vertical shift and the rotation of the protein around its long axis. The quality of the fit at each generation is evaluated by the goodness of fit as follows:

$$\chi^2 = \frac{1}{N} \sum_i \left(\frac{\Omega_{exp,i} - \Omega_{sim,i}}{\Delta\Omega_{exp,i}} \right)^2, \quad (12)$$

where N is the number of spin-labeled mutants, $\Omega_{exp,i}$ and $\Omega_{sim,i}$ correspond to experimentally derived and simulated free rotational space values at i -th mutant position, while $\Delta\Omega_{exp,i}$ represents the inaccuracy of the experimental free rotational space. The goodness of fit χ^2 guides the optimization routine determining whether a current structure is accepted for the next generation of structural evolution.

After multiple runs of optimization, many final structures will have the same goodness of fit, given by different Ω profiles. This provides a family of structures that characterizes the low-resolution structure of the protein. Such a method is comparable to the distance geometry approach employed in two-dimensional solution NMR spectroscopy that also results in a family of structures^{10,48,49}.

5.3 Results

The secondary structure of membrane-embedded M13 protein, the thickness of the lipid bilayer and the position of the protein relative to the membrane normal were optimized with a multi-run optimization algorithm. The simulation data allow a comparison with the experimental data obtained from SDSL-ESR spectra of 27 mutants of M13 protein reconstituted in 14:1PC bilayers^{6,7}. This protein served as a reference membrane protein to test the basic ideas of our approach. Initially the secondary structure of the protein was set to an α -helical conformation ($\varphi = -57^\circ$ and $\psi = -47^\circ$). The lipid effect was defined for the transmembrane region between amino acid positions 14 and 46 according to the experimental profiles for the free rotational space Ω and rotational dynamics⁷. The initial steric thickness of the bilayer was set to 40 Å resulting in an initial protein tilt of about 35° in accordance with the fluorescence-based protein model^{11,12,14}. The multi-run evolutionary optimization was repeated for 1000 times. Each run contained 200 generations. At each generation a new structural conformation of the protein was obtained by modifying stochastically the dihedral angles of the main chain, by tuning the parameters of lipid bilayer, and by optimizing the relative position and orientation of the protein in the lipids. For each new structure the corresponding local structural restrictions were calculated. Thus altogether about 200,000 different global structural conformations were checked.

The result of structural optimization is presented as a family of most successful (in terms of goodness of fit) global protein conformations together with the summarized simulated restrictions (Figure 4A). The goodness of fit χ^2 of this assembly of 50 structures is in the range from 2.7 to 3.7. These structures are mainly α -helical, some of them demonstrate a tiny kink, i.e., as can be seen for the structure in Figure 4B. The tilt angle varies between 10 and 40° with a mean value around 26°. This mean tilt angle is in excellent agreement with the experimental data ($23^\circ \pm 4$, as determined from quantitative fluorescence site-directed analysis¹⁴). The steric bilayer thickness of these structures is in the range from 37 to 44 Å with a mean value around 41 Å. This value is quite reasonable as compared to the typical steric thicknesses of a phospholipids bilayer around 44 Å^{33,50}. The transmembrane region was found between amino acid positions 12-17 and 46-47. In most cases the transmembrane region starts at position 14 and ends at position 46.

One of the best-fit structures is shown in Figure 4B. The corresponding goodness of fit of this structure is $\chi^2 = 3.2$. For this topology, the transmembrane region is between amino acid residues 14 and 46 as indicated by a drop of the free rotational space Ω in the simulated data below 0.5. The tilt angle is about 25° and the resulting steric bilayer thickness turns out to be 42 Å. The local restrictions of the same structure simulated without including the lipid effect (Figure 4C) provide a goodness of fit χ^2 of 33.0, which is much worse than for the membrane-embedded protein. This indicates that the presence of lipid restrictions is required to get a good fit to the experimental data.

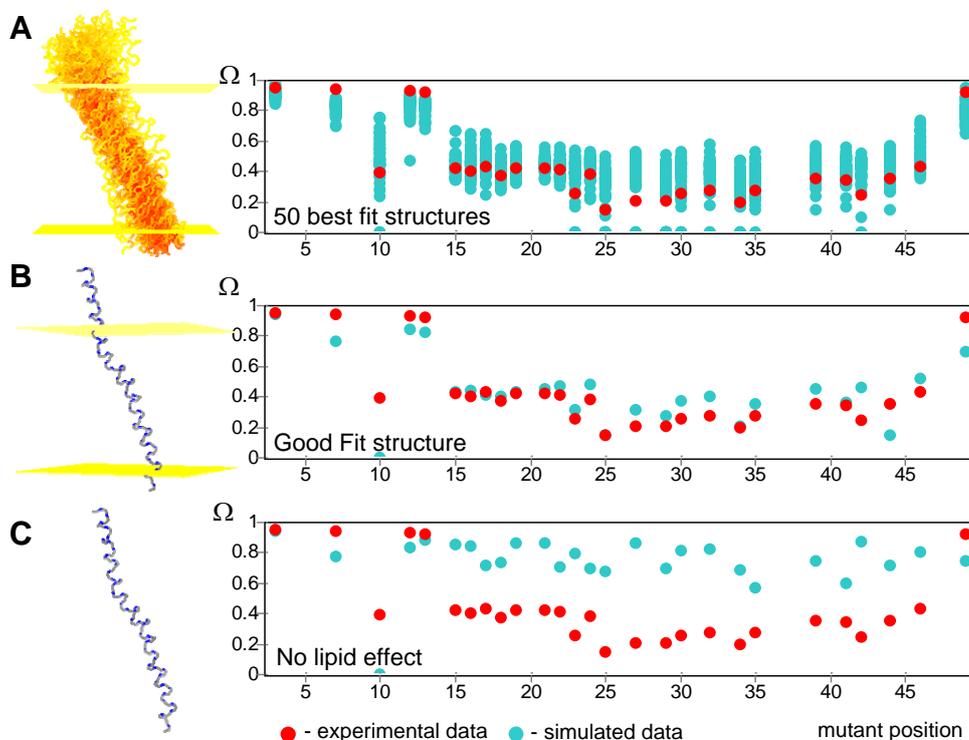


Figure 4. Optimization of the structure and membrane-embedding of M13 protein by fitting the calculated local restrictions (blue circles) to the restrictions profile obtained from SDSL-ESR (red circles). **A.** A population of best-fit structures (50 structures with a goodness of fit χ^2 in the range from 2.7 to 3.7) of M13 protein in 14:1PC lipid bilayers. The yellow-red color gradient represents the structural density in this population of structures. The steric thickness of lipid bilayer (including hydrophobic and head group regions) is represented with yellow planes. **B.** A single structure from the population of best-fit structures from (A) with $\chi^2 = 3.2$. **C.** The free rotational space for same secondary structure of M13 protein calculated without taking into account the lipid effect ($\chi^2 = 33.0$).

5.4 Discussion

Extensive testing of the model (see Appendix B) shows that the free rotational space is sensitive to the protein primary sequence, secondary structure, and to the position and orientation of the protein in the lipid system. In addition, calculated restrictions from 5000 randomly simulated structures of M13 protein reconstituted in phospholipids produce a wide range of Ω values that cover all 27 points of the experimental data⁹. The optimization routine that is implemented in our model is capable to considerably improve the fits (starting with an initial α -helical structure). Multiple structures (Figure 4A) obtained in 1000 runs of optimization are in excellent agreement with a recently proposed fluorescence-based M13 protein model^{11,12,14}. This nice accordance implies that our

modeling and optimization approach is fundamentally sound and that the simplifications we have made in our model are acceptable.

As compared to related papers on SDSL-ESR that employ molecular dynamics simulations^{51-55,56,57,58} our approach has the following advantages: 1. The simplicity of the underlying physical principles in the structural model; 2. The simultaneous analysis of multiple SDSL-ESR data from all available spin-labeled mutant positions; 3. There is no need for dynamics trajectories; 4. As a consequence, our calculations are 3-4 orders of magnitude faster than calculations based on molecular dynamics simulations. In the present case, extending the computation time will impose a severe limit to the calculations, as the optimization of the 200,000 structures resulting in Figure 4A already takes four weeks of CPU time on a 20-cores computer cluster (6×Opteron Double Core 2.4 GHz and 8×Athlon Single Core 2.13 GHz).

Analysis of best-fit single structures (like one in Figure 4B) indicates that the optimization algorithm was successful to provide accurate fits for different parts of the protein picking up the main trends of the experimental data. Some experimental points may not be fitting well (e.g., mutant positions 10, 42, 44 for the structure in Figure 4B). This results in ranges of calculated Ω values for the family of best-fit structures (Figure 4A). The observed discrepancy between simulated and experimental data could be either due to the simplifications we assume in calculating the lipid effect, or to an incorrectly determined local motif of the secondary structure, related to simplifications in the protein structure determination. Although, the Ω trend remains correct for short subsequences, still some single mutant positions may experience different lipid effect most likely due to different local orientations of the spin label with respect to the membrane normal.

The present paper focuses on a small protein embedded in a lipid bilayer. However, our method can also be used for larger proteins, consisting of several protein domains (e.g., helices) and it is not necessary for the applicability of the method that the domains are transmembrane or helical. In such a more complicated case, the position and orientation of each domain relative to the rest of the complex have to be parameterized, which generally will require some approximate information about the folding of the protein domains. Another limitation of the method is that SDSL-ESR data should be available from a sufficient number of spin-labeled sites and that, according to assumptions in the model, only data can be used that are obtained at physiological temperature (i.e., room temperature).

Fitting the simulated to the experimental data remains a challenging task. The number of parameters even for a small protein is already high (for a 50-amino acid residue long protein we have more than 100 parameters). On the other hand, in our case there are 27 experimental points to be fitted simultaneously. For such a task an efficient optimization algorithm is needed that would be capable to efficiently handle a high number of optimization parameters. A good candidate is a hybrid evolutionary algorithm that could

optimize a population of structures simultaneously⁵⁹. In that case, to make the optimization more efficient, the information about known secondary structure motifs, tertiary structure interactions as well as information about protein-lipid interaction could be implemented in specialized genetic operators.

5.5 Conclusion

The proposed method allows fitting of the simulated restrictions of the free rotational space to the experimental data obtained from SDSL-ESR spectra. For a reference membrane protein (membrane-embedded M13 protein reconstituted in 14:1PC bilayers), a multi-run optimization results in a family of favorable protein structures, which not only agree with the available SDSL-ESR data, but also are consistent with the previously published model based on site-directed fluorescence labeling^{11,12,14,30}.

Our simulations show that the simultaneous analysis of available SDSL-ESR data based on structural modeling provides information about membrane protein structure and structural characteristics. Structural modeling allows using additional data from primary structure analysis (secondary structure predictions, hydrophathy index calculation) as well as any additional information potentially available from other experimental techniques (X-ray high-resolution structures, information from NMR spectroscopy, fluorescence spectroscopy, infrared spectroscopy, and circular dichroism) could be combined to provide optimization constraints for structure modeling. In this perspective, structural modeling is thought to be a connecting link, which could transfer multiple data into a high-resolution structure or structural characterization and reveal the functional properties of membrane proteins. The present method provides a challenging starting point for the development of a powerful methodology for the structure characterization of membrane protein, as an alternative to conventional techniques.

Acknowledgements

We acknowledge the Slovenian Research Agency for funding this work within the research program P1-0060 (to J.Š.) and the financial support from Ad Futura (Science and Education Foundation of the Republic of Slovenia, Public Fund) (to A.A.K.). The authors thank David Stopar and Primož Ziherl for helpful and stimulating discussions.

5.6 Appendix A. Protein structure modeling. Tuning interatomic contact distances with Ramachandran plot calculation

The modeling of the protein structure makes use of previously reported fixed bond lengths and bond angles²⁰⁻²² that are presented in Tables A1 and A2. The secondary structure of a protein is parameterized with pairs of backbone dihedral angles φ_i and ψ_i (Figure A1A). Due to steric restrictions between backbone atoms not all angle pairs are allowed. For a three-residue fragment the allowed combinations of the angles φ_i and ψ_i for the central amino acid residue is presented in a so-called Ramachandran plot²³. To derive the φ_i - ψ_i distribution, a list of backbone atom pairs has to be checked for steric interatomic contacts²⁴. In our case, the sterically allowed effective minimum distances between the backbone atoms (sums of atom van der Waals radii) were tuned by computing Ramachandran plots and comparing them with previously published plots²³⁻²⁶. As a consequence the van der Waals radii of the atoms (Table A3) needed to be reduced in agreement with the literature^{24,36,60,61}. Note that the reduction of the “apparent” van der Waals radii simulates the effect of atomic interactions when checking steric overlaps^{16,34,35}. For example, nonbonding electrostatic interactions between O-H, C=O and CH groups, as well as the anisotropy of the C atom electron shell contribute to the reduction of the average distance at which those groups can be found. In our case, calculation of the Ramachandran plot suggested a reduction of the minimally allowed initial interatomic distances between backbone atoms by 3-16% (Table A3). The atom pairs in Table A3 are split into three groups according to the restrictive effect that they impose on the distribution of backbone dihedral angles φ and ψ . Some discrepancy can be found between the Ramachandran plot calculated with our model (Figure A1B) and the distribution of the angles φ and ψ based on PDB structures analysis²⁵ (Figure A1C). This difference arises from the fact that in our calculations of the Ramachandran plot we used a three-alanine peptide, whereas φ and ψ angle distribution of the reference²³⁻²⁶ was obtained by analysis of different amino acid residues of structures deposited in the PDB data bank. For the conformational sampling within the conformational space, the residue is split into parts according to the number of free bond rotations, so that all atoms within one part preserve their relative positions (blue ovals in Figure A1D). Each complex part is split into atom groups, while each group contains one heavy atom (C, O, N or S) with hydrogen atoms, if there are any. The result of the tuning of the contact distances is presented in Table A4 with the reduced van der Waals atoms radii.

Optimization of membrane protein structure based on SDSL-ESR constraints and conformational space modelling

Table A1. Values of the chemical bond lengths used in the modeling of the protein structure. These values are based on previously reported constants²⁰⁻²².

Bond	Bond length (Å)
N-C ^α (backbone)	1.46
C ^α -C (backbone)	1.53
C-N (backbone)	1.33
C-C	1.53
C-S, S-C	1.80
C-O	1.42
C=O	1.24
C-N (Lys)	1.50
C-N, N-C (Arg, Asn, Gln)	1.32
C-H, O-H, N-H	1.00
S-H	1.30
C-C (in ring)	1.35
C-C (spin label)	1.45
N-O (nitroxide)	1.40

Table A2. Values of the chemical bond angles used in the modeling of the protein structure. These values are based on previously reported constants²⁰⁻²².

Bond	Bond angle (°)
- C -	109.5
- C =	120.0
- N -	120.0
O-H	104.5
S-H	104.5

Table A3. Computational results of the interatomic distances between the backbone atoms tuned by computing Ramachandran plots. Tuned values suggest a reduction of the minimally allowed interatomic distances, which is in accordance with the literature^{24,36,61}.

Atoms pairs	Original van der Waals interatomic distances (Å)	Tuned interatomic distances (Å)
Restricting ϕ		
C ^β , O _{i-1}	3.15	2.70
O _{i-1} , C	3.05	2.55
Restricting ψ		
C ^β , O	3.15	2.70
C ^β , N _{i+1}	3.30	2.95
N, H _{i+1}	2.72	2.30
C ^β , H _{i+1}	2.92	2.65
Restricting both ϕ and ψ		
O _{i-1} , O	2.80	2.70
O _{i-1} , N _{i+1}	2.95	2.70
O _{i-1} , H _{i+1}	2.57	2.50
H, H _{i+1}	2.34	2.10

Table A4. Reduced van der Waals atoms radii used in the modeling of the conformational space of the amino acid side chains in accordance with^{34,60,61}, compared with the original values.

Atom, atom group	Original van der Waals radius (Å)	Reduced van der Waals radius (Å)
C	1.75	1.25
C (carboxyl)	1.65	1.15
C (aromatic)	1.65	1.65
N	1.55	1.25
O	1.40	1.20
S	1.80	1.50
H	1.17	1.17
H (polar, aromatic)	1.00	1.00
CH, CH ₂ , CH ₃	2.27	1.50
SH, OH, NH, NH ₂ , NH ₃	2.17	1.55

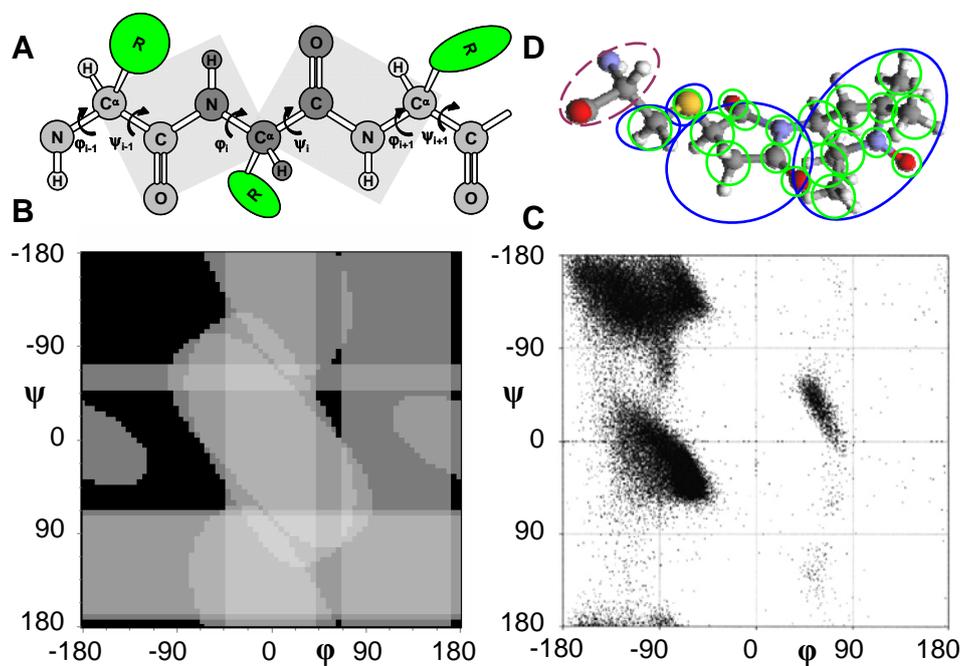


Figure A1. Protein secondary structure parameterization. **A.** Definition of the backbone dihedral angles. Available backbone free rotations and corresponding dihedral angles are shown on a three-residue model. Amino acid side chains are schematically presented with the green ovals marked with 'R'. **B.** Ramachandran plot: the distribution of allowed ϕ and ψ backbone dihedral angles (islands colored with black) calculated with our model for a three-alanine peptide. Nine backbone atom pairs were analyzed. Different grey scale regions represent ϕ and ψ combinations that are forbidden due to steric clashes between atoms pairs. **C.** Distribution of ϕ and ψ dihedral angles obtained by analysis of 240 protein structures from the Protein Data Bank²⁵. **D.** Residue side chain presented with a so-called Residue-Parts-Groups mechanism. The residue is split into side chain parts (blue ovals) including backbone zero-part (brown oval). Each part is composed of atom groups (green circles) which contain one of the heavy atoms (C, O, N or S) and often a few hydrogen atoms.

5.7 Appendix B. Contribution of different restrictive factors to conformational space restriction

The sum statistical weight of the conformational space, P_{SUM} , which is the sum of the statistical weights of the single conformations P_i , shows the total effect of the different restrictive factors that contributes to a reduction of the statistical weight. This is given by:

$$P_{SUM} = P_{SUM}^{initial} (1 - R_B - R_S - R_L),$$

where $P_{SUM}^{initial}$ is the sum statistical weights of the unrestricted conformational space; R_B , R_S , and R_L characterize the reduction of the sum statistical weight due to overlap with the backbone and neighboring side chains, and due to the lipids, respectively.

A relative analysis of the different restrictive factors that contribute to the restriction of the conformational space of the spin label shows that approximately half of the restrictions arise from side chain conformations that overlap with the backbone (Table B1). The conformational overlap with neighboring side chains and lipids contributes to the restrictions with about 30 and 10%, respectively. This means that half of the conformations become fully restricted due to ‘hard’ overlap with the backbone, the other half of the conformations considerably lose their statistical weights and finally the occurrence probability of still allowed conformations is redistributed. These results indicate that the major restrictive factor that defines the conformational space of the side chains is the secondary structure of the protein. The side chains compete for the available space with each other and also with the surroundings (e.g., the lipid acyl chains).

Table B1. Computational results of the relative comparison of the restricting factors that contribute to the reduction of the conformations statistical weights and restrict the conformational space of the spin label.

Restricting factor	Notation	Reduction of $P_{SUM}^{initial}$
Overlap with the backbone	R_B	~ 50-60 %
Side chain neighborhood	R_S	~ 30 %
Lipids	R_L	~ 10 %
Reduced sum probability	P_{SUM}	~ 1 %

5.7.1 Spin label conformational space sensitivity to primary structure

To test the sensitivity of the conformational space of the spin label to the primary structure of a protein, the normalized free rotational space Ω was calculated for the 3-maleimido proxyl spin label attached to the central (10th) cysteine residue of a number of artificially-designed 19-residue peptides in an α -helical conformation ($\varphi = -57^\circ$ and $\psi = -47^\circ$). Since the secondary structure in tests was set to be uniform among all oligopeptides, the resulting

differences in Ω (Figure B1A) can be assigned to the properties of the amino acid residues, i.e., to the primary structure. As can be seen in Figure B1A, the restrictive effect of the primary structure depends on the flexibility and on the size of the amino acid side chains.

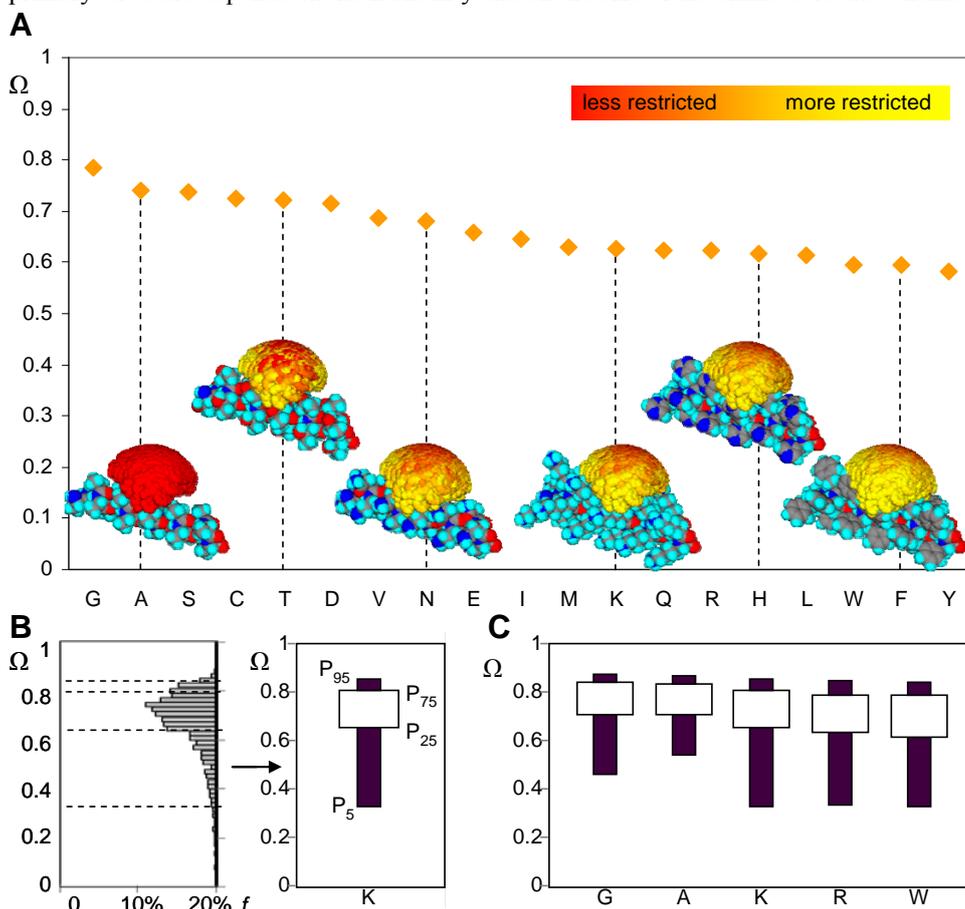


Figure B1. Spin label conformational space sensitivity to primary structure. **A.** Normalized free rotational space Ω (orange data points) for artificially-designed 19-residue peptide homopolymers in an α -helical conformation ($\varphi = -57^\circ$ and $\psi = -47^\circ$) with the 3-maleimido proxyl spin label attached to the central (10^{th}) cysteine residue. For a few oligopeptides the conformational space of the spin label is illustrated in a molecular model with the yellow color corresponding to the restricted and red to the unrestricted conformations. **B.** Frequency (f) histogram of the Ω distribution for several thousands of conformations of an oligolysine peptide (K) (left) and a simplified representation showing its inter-quartile ranges given by the 5th, 25th, 75th, and 95th percentile (right). **C.** Inter-quartile ranges of typical peptides: oligoglycine – G, oligoalanine – A, oligolysine – K, oligoarginine – R and oligotryptophan – W. In B and C all backbone dihedral angle pairs are the same along the oligopeptide.

The long and flexible side chains of lysine and arginine as well as the bulky side chains of tryptophan, phenylalanine, tyrosine and histidine show the strongest restrictions.

On the contrary, the side chains of glycine, alanine, and serine are smaller resulting in a less restricted conformational space for the spin label (Figure B1A).

To examine the accumulated effect of primary structure for the different secondary structure motifs, the backbone dihedral angles along the oligopeptides were varied stochastically within the allowed regions of the Ramachandran plot. Since the secondary structure was uniform along the peptide, the changes in the calculated normalized free rotational space Ω for the different oligopeptides and for the specific secondary structure are directly related to the properties of the particular amino acid side chains, i.e., to the primary structure. To illustrate this effect, the Ω values for an oligolysine peptide were sorted in a histogram (Figure B1B). This Ω distribution was simplified by plotting the inter-quartile ranges that correspond to 5th, 25th, 75th, and 95th percentiles of the Ω distribution. For a number of typical oligopeptides the inter-quartile ranges are plotted in Figure B1C. Amino acid residues with small side chains (i.e., glycine and alanine) are not very restrictive, as indicated by a relatively large median and narrow distribution of the conformational space of the spin label (Figure B1C, columns G and A). In this case, the spin label will be less sensitive to different elements of secondary structure. However, amino acid residues with bulky side chains (tryptophan) or long and flexible side chains (arginine and lysine) more strongly confine the conformational space of the spin label (see columns W, R and K in Figure B1C).

Among the tested structures the probability that the value of the normalized conformational space Ω of the spin label in the oligolysine peptide exceeds 0.77 is 25%, and the probability that the spin label is immobilized with an Ω value below 0.63 is also 25% (Figure B1C, column K). It is even more striking that the probabilities of the spin label being very unrestricted with Ω above 0.84 and very restricted with Ω below 0.39 are both equal to 5%. The same applies to the oligoarginine and oligotryptophan peptides with virtually identical probability levels (Figure B1C, columns R and W). This indicates that the spin label is more sensitive to the secondary structure rather than to the primary structure.

5.7.2 Spin label conformational space sensitivity to secondary structure

To explore the sensitivity of the conformational space of the spin label to the secondary structure, the backbone restrictions R_B (i.e., the reduction of the sum statistical weight due to overlap with the backbone) were calculated for an artificially-designed oligoalanine peptide. By using alanine as a small amino acid residue, we minimized the effect of primary structure. The dihedral angles ϕ and ψ at the position of the spin label were systematically varied within the allowed regions of the Ramachandran plot with a grid step of 5°, resulting in approximately 1000 different secondary structures. The remaining part of the oligopeptide was fixed to an α -helix. The calculations show that R_B varies from 40 to 100% (Figure B2A), indicating that most restrictions arise from an overlap of the conformational

space with the backbone. Consequently Ω varies from 0.3 to 0.9 (data is not shown). Thus changing the secondary structure of a protein locally at the spin label position considerably affects its conformational space.

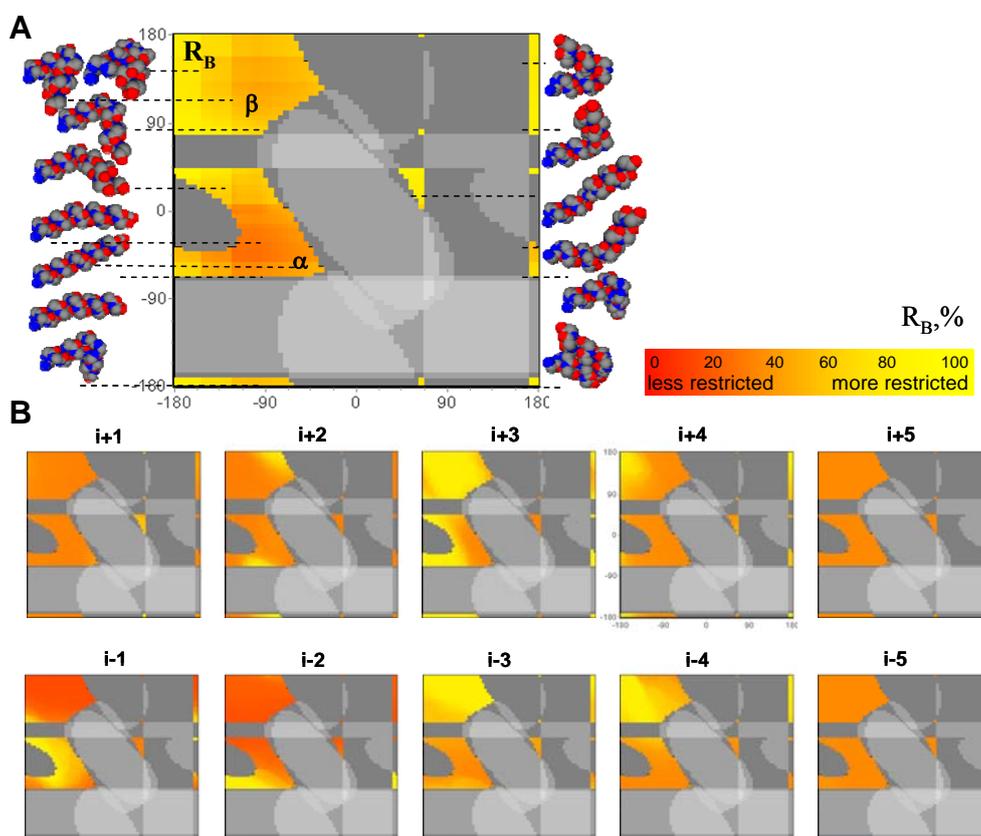


Figure B2. Sensitivity of the conformational space of the spin label to the secondary structure. The 3-maleimido proxyl spin label is attached to the central (10^{th}) cysteine residue of a number of artificially-designed 19-residue oligoalanine peptide in an α -helical conformation ($\varphi = -57^\circ$ and $\psi = -47^\circ$). **A.** Backbone restrictions R_B (i.e., the reduction of the sum statistical weight due to overlap with the backbone) at the spin-labeled site. The dihedral angles φ and ψ at the position of the spin label were systematically varied within the allowed regions of the Ramachandran plot with a grid step of 5° , resulting in approximately 1000 different secondary structures. The remaining part of the oligopeptide was fixed to an α -helix. Each allowed conformation is represented with a colored dot in the Ramachandran plot with the coordinates corresponding to the angles φ and ψ at the spin label position (the red-yellow color gradient encodes R_B from 0 to 100%). Molecular models of typical secondary structures conformations (indicated on the left and right sides of the figure) point to the corresponding regions of the plot. The regions for α -helix and β -sheet conformations are indicated as well. **B.** Backbone restrictions at the spin-labeled site $i=10$ arising from adjacent amino acid positions up to five to the N and C-terminal end (i.e., $i\pm 1$, $i\pm 2$, $i\pm 3$, $i\pm 4$, $i\pm 5$). The remaining secondary structure was taken as α -helix. For labeling of the axes and for the other details, see (A).

To investigate what effect of secondary structure at neighboring amino acid positions on the conformational space of the 3-maleimido-proxyl spin label, we repeated the calculation of R_B by changing the dihedral angles further away from the i -th labeled site. The result is shown in Figure B2B for up to five amino acid positions towards the N and C-terminal ends. The remaining secondary structure was fixed to an α -helix. In all cases, the conformational space of the spin label is affected by this effect of secondary structure, the range of backbone restrictions varies from 30 to 100% (Figure B2B). Based on this finding it may not be necessary to have all positions labeled. Instead, sites could be labeled alternately, reducing the number of mutants by a factor of two. The restrictive effect is most strong, as expected, at one helical winding up or down to the spin label site (i.e., at $i\pm 3$ and $i\pm 4$). For positions up to the C-terminal end (i.e., at $i+3$, the restrictive effect is slightly larger as compared to positions down to the N-terminal end. This may be related to the fact that in α -helices the amino acid side chains have the tendency to slightly tilt toward the N-terminal end of the helix ⁶².

5.7.3 Spin label conformational space sensitivity to lipid environment

The tendency of lipids to reduce the conformational space of amino acid side chains was introduced in the model as an additional restrictive factor that limits the conformational space of the spin label. This effect is demonstrated on exploring the normalized free rotational space Ω of membrane-embedded spin-labeled M13 protein. We assume that the protein is in an α -helical conformation ¹¹⁻¹⁴. To enable comparison with experimental spin-label ESR data, we assume that the protein is reconstituted in 1,2-dierucoyl-*sn*-glycero-3-phosphocholine (22:1PC) phospholipid bilayers ⁷ with the transmembrane region defined between amino acid positions 9 and 47 and a steric bilayer thickness of 55 Å. This protein-lipid model implies that the restrictive lipid effect extends into the phospholipid headgroup region. In this way, the tilt angle of the protein in the membrane turns out to be around 20°, in good agreement with a protein model based on site-directed fluorescence labeling ^{11,12,14}. For comparison we also examine the protein without a lipid environment.

For almost all amino acid positions in the transmembrane region of M13 protein, the normalized free rotational space Ω is reduced by 20-40% due to the lipid effect (Figure B3A). Also, the trend of the values for Ω is slightly changed by the lipids. This is due to the fact that at each spin-labeled site the lipid effect strongly depends on the relative orientation of the conformational space of the spin label in relation to the membrane normal (i.e., the restrictive effect of the lipids depends on the angle between the lipids acyl chains and the spin label side chain conformations).

The effect of the orientation of M13 protein in the membrane (i.e., rotation about the helical axis) on the conformational space of the spin label was studied by systematically changing the orientation angle from 0 to 360° by 1°. Then the values for Ω were calculated

for the different spin label positions along the protein. The secondary structure of the protein, the transmembrane region, and the tilt angle were set to the same values as in the previous calculation. From these Ω data the frequency histograms were determined (Figure B3B) and for each amino acid position collected in Figure B3C. As can be seen, this approach results in quite wide ranges for Ω in the transmembrane region. This effect is due to the change of the relative orientation of the conformational space of the spin label in relation to the membrane normal. Thus simultaneous analysis of the experimental Ω values from SDSL-ESR data combined with modeling of the membrane-embedded protein may reveal the correct orientation angle of the protein in the membrane.

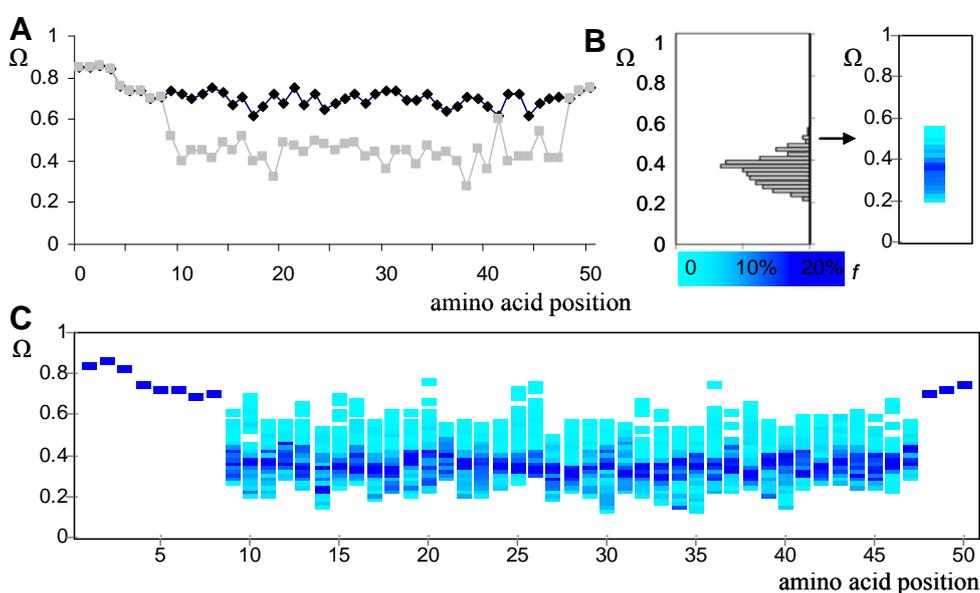


Figure B3. Sensitivity of the conformational space of the spin label to the lipid environment. The 3-maleimido proxyl spin label is attached to membrane-embedded M13 protein in an α -helical conformation ($\varphi = -57^\circ$ and $\psi = -47^\circ$). The transmembrane region is defined between amino acid positions 9 and 47, corresponding to a tilt angle of 20° . **A.** Normalized free rotational space Ω of the spin label at different amino acid positions on the protein in the presence of lipids (grey curve) and without lipids (black curve). **B.** Effect of orientation of M13 protein in the membrane by systematically changing the orientation angle from 0 to 360° in steps of 1°, shown in a frequency histogram (left) and using a blue color coding (right). The spin label is at position 23. **C.** Amino acid position dependence of the Ω data as determined in (B).

References

- 1 Lacapere, J.-J., Pebay-Peyroula, E., Neumann, J.-M., & Etchebest, C. *Trends Biochem. Sci.* **32**, 259 (2007).
- 2 Torres, J., Stevens, T.J., & Samso, M. *Trends Biochem. Sci.* **28**, 137 (2003).
- 3 White, S., Membrane proteins of known 3D structure, Available at http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html, (2009).
- 4 Hemminga, M.A. & Berliner, L.J., *ESR spectroscopy in membrane biophysics*. (Springer, New York, 2007).
- 5 Fanucci, G.E. & Cafiso, D.S. *Curr. Opin. Struct. Biol.* **16**, 644 (2006).
- 6 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *J. Chem. Inf. Mod.* **45**, 1621 (2005).
- 7 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *Biophys. J.* **91**, 3341 (2006).
- 8 Bashtovyy, D., Marsh, D., Hemminga, M.A., & Pali, T. *Protein Sci.* **10**, 979 (2001).
- 9 Štrancar, J., Kavalenka, A., Zihlerl, P., Stopar, D., & Hemminga, M.A. *J. Magn. Reson.* **197**, 245 (2009).
- 10 Bax, A. *Annu. Rev. Biochem.* **58**, 223 (1989).
- 11 Nazarov, P.V., Koehorst, R.B.M., Vos, W.L., Apanasovich, V.V., & Hemminga, M.A. *Biophys. J.* **92**, 1296 (2007).
- 12 Nazarov, P.V., Koehorst, R.B.M., Vos, W.L., Apanasovich, V.V., & Hemminga, M.A. *Biophys. J.* **91**, 454 (2006).
- 13 Vos, W.L., Koehorst, R.B.M., Spruijt, R.B., & Hemminga, M.A. *J. Biol. Chem.* **280**, 38522 (2005).
- 14 Koehorst, R.B.M., Spruijt, R.B., Vergeldt, F.J., & Hemminga, M.A. *Biophys. J.* **87**, 1445 (2004).
- 15 Sale, K., Sar, C., Sharp, K.A., Hideg, K., & Fajer, P.G. *J. Magn. Reson.* **156**, 104 (2002).
- 16 Vasquez, M. *Curr. Opin. Struct. Biol.* **6**, 217 (1996).
- 17 Karplus, M. & McCammon, J.A. *CRC Crit. Rev. Biochem.* **9**, 293 (1981).
- 18 Taylor, R.D., Jewsbury, P.J., & Essex, J.W. *J. Comput. Chem.* **24**, 1637 (2003).
- 19 Xiang, Z. & Honig, B. *J. Mol. Biol.* **311**, 421 (2001).
- 20 Engh, R.A. & Huber, R. *Acta Crystallogr. A.* **47**, 392 (1991).
- 21 MacKerell, A.D. *et al. J. Phys. Chem. B* **102**, 3586 (1998).
- 22 Word, J.M. *et al. J. Mol. Biol.* **285**, 1711 (1999).
- 23 Ramachandran, G.N., Ramakrishnan, C., & Sasisekharan, V. *J. Mol. Biol.* **7**, 95 (1963).
- 24 Ho, B.K., Thomas, A., & Brasseur, R. *Protein Sci.* **12**, 2508 (2003).
- 25 Lovell, S.C. *et al. Proteins* **50**, 437 (2003).
- 26 Hovmöller, S., Zhou, T., & Ohlson, T. *Acta Crystallogr. D.* **58**, 768 (2002).
- 27 Choi, V. *J. Chem. Inf. Mod.* **46**, 438 (2006).
- 28 Rainey, J.K. & Goh, M.C. *J. Chem. Inf. Comput. Sci.* **44**, 817 (2004).
- 29 Zhang, M. & Kavraki, L.E. *J. Chem. Inf. Comput. Sci.* **42**, 64 (2002).
- 30 Vos, W.L., Nazarov, P.V., Koehorst, R.B.M., Spruijt, R.B., & Hemminga, M.A. *Trends Biochem. Sci.* **34**, 249 (2009).
- 31 Stopar, D., Spruijt, R.B., & Hemminga, M.A. *Chem. Phys. Lipids* **141**, 83 (2006).

- 32 Stopar, D., Spruijt, R.B., Wolfs, C.J., & Hemminga, M.A. *Biochim. Biophys. Acta* **1611**, 5 (2003).
- 33 Nagle, J.F. & Tristram-Nagle, S. *Biochim. Biophys. Acta* **1469**, 159 (2000).
- 34 Bower, M.J., Cohen, F.E., & Dunbrack, R.L., Jr. *J. Mol. Biol.* **267**, 1268 (1997).
- 35 Vasquez, M. *Biopolymers* **36**, 53 (1995).
- 36 Tombolato, F., Ferrarini, A., & Freed, J.H. *J. Phys. Chem. B* **110**, 26248 (2006).
- 37 Tombolato, F., Ferrarini, A., & Freed, J.H. *J. Phys. Chem. B* **110**, 26260 (2006).
- 38 Marsh, D. *Biochim. Biophys. Acta* **1778**, 1545 (2008).
- 39 Gawrisch, K., The dynamics of membrane lipids in *The structure of biological membranes*, edited by Philip Yeagle (CRC Press, Boca Raton, Fla, 2005), pp. 147.
- 40 Lee, A.G. *Biochim. Biophys. Acta* **1612**, 1 (2003).
- 41 Johansson, A.C. & Lindahl, E. *Biophys. J.* **91**, 4450 (2006).
- 42 Johansson, A.C.V. & Lindahl, E. *Proteins* **70**, 1332 (2008).
- 43 Štrancar, J. *et al. J. Chem. Inf. Mod.* **45**, 394 (2005).
- 44 Kavalenka, A.A., Filipič, B., Hemminga, M.A., & Štrancar, J. *J. Chem. Inf. Mod.* **45**, 1628 (2005).
- 45 Kirkpatrick, S., Gelatt, C.D., Jr., & Vecchi, M.P. *Science* **220**, 671 (1983).
- 46 Eiben, A.E. & Smith, J.E., *Introduction to evolutionary computing*. (Springer, New York, 2003).
- 47 Fogel, D.B., Bäck, T., & Michalewicz, Z., *Evolutionary computation*. (Institute of Physics Publishing, Bristol; Philadelphia, 2000).
- 48 Castellani, F. *et al. Nature* **420**, 98 (2002).
- 49 Papavoine, C.H., Aelen, J.M., Konings, R.N., Hilbers, C.W., & Van de Ven, F.J. *Eur. J. Biochem.* **232**, 490 (1995).
- 50 Balgavý, P. *et al. Biochim. Biophys. Acta* **1512**, 40 (2001).
- 51 Håkansson, P., Westlund, P.O., Lindahl, E., & Edholm, O. *Phys. Chem. Chem. Phys.* **3**, 5311 (2001).
- 52 LaConte, L.E., Voelz, V., Nelson, W., Enz, M., & Thomas, D.D. *Biophys. J.* **83**, 1854 (2002).
- 53 Stoica, I. *J. Phys. Chem. B* **108**, 1771 (2004).
- 54 Pistolesi, S. *et al. Biophys. Chem.* **123**, 49 (2006).
- 55 Budil, D.E., Sale, K.L., Khairy, K.A., & Fajer, P.G. **110**, 3703 (2006).
- 56 Beier, C. & Steinhoff, H.-J. *Biophys. J.* **91**, 2647 (2006).
- 57 DeSensi, S.C., Rangel, D.P., Beth, A.H., Lybrand, T.P., & Hustedt, E.J. *Biophys. J.* **94**, 3798 (2008).
- 58 Sezer, D., Freed, J.H., & Roux, B. *J. Chem. Phys.* **128**, 165106 (2008).
- 59 Filipič, B. & Štrancar, J., 2003.
- 60 Shetty, R.P., De Bakker, P.I., DePristo, M.A., & Blundell, T.L. *Protein Eng.* **16**, 963 (2003).
- 61 Grigoryan, G., Ochoa, A., & Keating, A.E. *Proteins* **68**, 863 (2007).
- 62 Creighton, T.E., *Proteins: Structures and molecular properties*, 2nd ed. (W.H. Freeman, New York, 1993).

Chapter 6

Summarizing discussion

6.1 Introduction

Revealing the function of a protein requires high-resolution structural information, information about protein dynamics, and biological data on this particular protein. As long as a protein may be crystallized, high-resolution structural information can be obtained by using X-ray crystallography. Currently there is no better method for obtaining high-resolution structural information. Nevertheless, for proteins that are difficult to crystallize, or to concentrate, only very limited structural information is available. Therefore, it is not surprising that the structure determination of membrane protein is one of the most challenging fields of structural biology and structural proteomics^{1,2}. Due to very heterogeneous environment in which they are found, the classical methods have difficulties with the structure determination of membrane proteins³.

The structure of membrane proteins has been already tackled with X-ray crystallography and NMR. However, when studying membrane proteins, the protein has to be incorporated in a membrane-like environment, which can create special conditions that could affect its structure. Obtained structures then need to be checked for correspondence with the functional protein conformation. And an even more difficult task is to obtain information about the dynamics of membrane proteins. Therefore the development of alternative structure determination methodologies is important to complement the structural picture provided by the well-established high-resolution techniques.

Difficulties in the application of standard high-resolution methods for the characterization of the three-dimensional structure of membrane proteins, therefore call for the development of alternative approaches. Low-resolution structural data can be obtained with small angle X-ray scattering (SAXS)^{4,5}, circular dichroism (CD)⁶⁻⁸ and atomic force microscopy (AFM)^{9,10}. Molecular dynamics simulations and other computational techniques reinforce alternative experimental methods, such as NMR spectroscopy¹¹⁻¹⁵ and electron microscopy (EM)^{16,17}. One of the alternative techniques is site-directed spin labeling (SDSL) electron spin resonance (ESR). This technique provides both a structural and dynamical characterization of the local conformations of a membrane protein (or any other protein) in its native environment^{3,18-21}. Site-directed spin labeling at multiple sites of proteins has been widely applied for characterization of protein structures²²⁻²⁶.

In this perspective my thesis presents a novel combination of site directed spin labeling ESR spectroscopy and molecular modeling that both describe restrictions to side-chain conformational spaces, stable within the nanosecond time window of ESR spectroscopy. In this approach, multiple local data are treated simultaneously: the restrictions along protein sequence are used to navigate the optimization of a protein backbone conformation, which finally provides a family of equally-good global conformations of the protein chains (Figure 1). In this final chapter an overview will be provided of our latest progress in this field.

6.2 Combining experimental methods and modeling for protein structure characterization

Already since the 50s, X-ray crystallography has been used to determine the three-dimensional structure of proteins. In this approach, diffraction patterns are obtained that are converted into an electron density map, which serves as a basis to generate a model for the protein structure. In general, such an approach includes the following steps: a) acquisition of experimental data; b) conversion of data into structural constraints; c) building a structural model; d) optimization of the model, and e) estimation of the accuracy of the resulting structures. Coupling a high-resolution structure with data from techniques that describe dynamics and the incorporation of biological data may further help in understanding of protein function²⁷.

In our case, the experimental data consist of a series of ESR spectra, which contain information about local structural restrictions and dynamical characteristics along the protein backbone. With the help of advanced spectrum analysis based on spectral modeling (simulation) and optimization of spectral model parameters, all the spectra are analyzed and the corresponding structural constraints are determined. The structural model of a protein, which has an atomic resolution, is parameterized with a set of backbone dihedral angle pairs, and in case of a membrane protein additional parameters describe the position and orientation of the protein with respect to the lipids. An optimization algorithm is then used to maximize the fit of the simulated restrictions to the experimentally determined constraints. Finally, the resulting family of favorable three-dimensional structures of the protein has to be checked with published data, or data about similar systems. In addition, the accuracy of the resulting structures may be estimated.

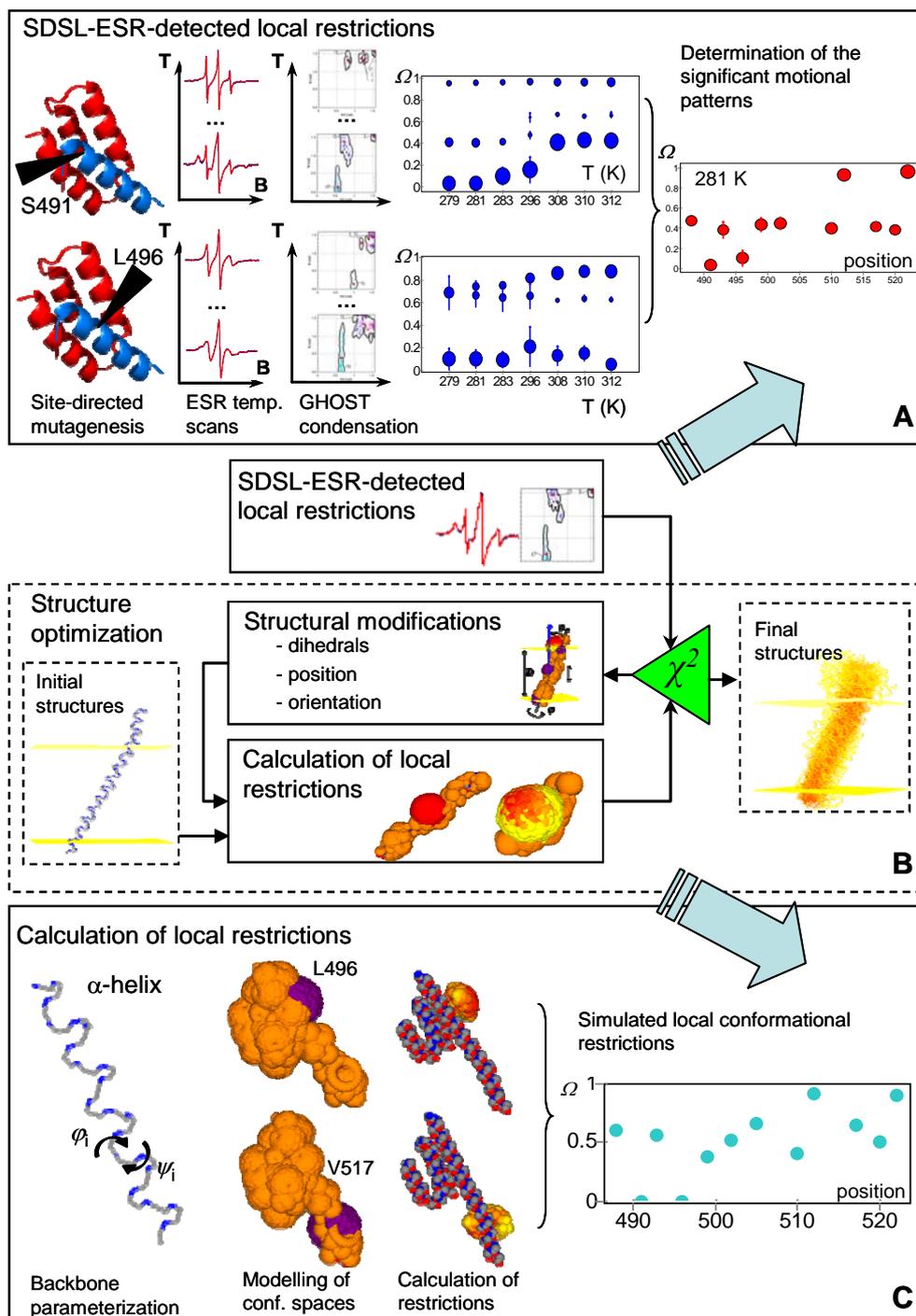


Figure 1. Overview of the SDSL-ESR approach for protein structure determination. **A.** Detection of the local restrictions from SDSL-ESR spectroscopic data. The method is illustrated for NTAIL-XD protein complex spin labeled at two typical mutant positions S491 and L496. The method is based on site-directed mutagenesis, measurements of ESR spectra at different temperatures (the experimental spectra are shown in blue, the simulated spectra are in red), GHOST condensation, and determination of the significant motional patterns for each mutant position. **B.** The structure determination approach is based on modeling of the conformational space of the amino acid side chains. This modeling is coupled to SDSL-ESR spectroscopy, and enhanced by structural optimization. Optimization starts with an initial structure, and finally comes up with a family of favorite structures. **C.** The determination of local restrictions is based on modeling of the conformational space. This method includes modeling of the protein structure (parameterized by pairs of backbone dihedral angles φ and ψ), modeling of conformational spaces of the side chains, and calculation of the conformational space restrictions. The method is illustrated for the NTAIL-XD protein complex spin labeled at two typical mutant positions L496 (at the interacting part of NTAIL-XD complex) and V517 (at the disordered part of NTAIL) ²⁸.

6.3 Pushing the limits of SDSL-ESR spectroscopy

Although the ultimate goal of this work is protein structure determination, a strong bias of the proposed method is towards SDSL-ESR spectroscopy. The modeling of the structure, the simulation of the conformational spaces and the structural optimization were developed to mimic SDLS-ESR, and to employ the SDLS-ESR experimental data as structural constraints. SDSL is an alternative biophysical technique that when compared to NMR and X-ray crystallography methods, provides a lower resolution and more qualitative structural data. It supplements the NMR technique with additional information about the dynamics of the protein backbone and side chains. It may also be capable of enhancing the ‘resolution’ as determined with X-ray crystallography. The spin label ESR experiments can be done at physiological conditions, at which the protein reveals its functionality, and in a native environment, what is especially important for the class of membrane proteins. Therefore, the method provides a unique opportunity to track conformational changes of proteins. In addition, the experiments could be conducted at low costs due to the small amounts of material needed ¹⁸.

In contrast to isotope labeling studies that are used in NMR spectroscopy, which are essentially non-disturbing, site-directed labeling can introduce a protein modification when molecular probes are covalently added; thus this approach can bring about unwanted effects on the structure and function of the protein. For this reason, when using site-directed labeling approaches, it is preferable to avoid replacing crucial amino acid residues, for instance by selecting solvent-exposed loop regions of membrane proteins where structural perturbations due to the presence of labels is minimized (see for instance ²⁹). In addition, site-directed labeling enables the use of a wide variety of probes, including environmental probes that allow monitoring further away from the protein backbone, structural probes for monitoring near the backbone, fluorescent, and spin probes. New labeling strategies based

on expanding the genetic code are being developed which will allow a position-specific incorporation of 'unnatural' labeled amino acids³⁰⁻³³. Such an approach will permit the specific insertion of small tailor-made probes and will offer the maximum flexibility in the site-directed labeling methodology³⁴.

In general, it is always preferable to use a range of sites for site-directed labeling via high-throughput mutagenic approaches^{20,21,35} and, if possible, conduct mutant measurements at a series of temperatures²⁸. More complete spectroscopic information obviously leads to a better understanding of the structural principles and helps to explain the heterogeneity, typical for complex ESR spectra from proteins.

6.3.1 Multi-component analysis and interpretation of SDSL-ESR data

The understanding of the spectroscopic data from complex biological systems is also enhanced by applying a multi-component analysis of ESR spectra³⁶. In spin-labeled proteins there are several factors that can contribute to a multi-component character: 1) differences in local protein structure; 2) difference in local protein dynamics; 3) sample heterogeneity; 4) various dynamical regimes of spin label, possible immobilization due to interactions with the environment; 5) non-specific labeling.

ESR spectra, composed of up to four spectral components can be simulated and the spectroscopic parameters can be extracted by spectral optimization (fitting of simulated spectra to the simulated one). Determination of the complexity of the system (the number of spectral components) as well as the structural and dynamical characteristics of each system component also require multiple spectral simulations as usually many combinations of component parameters may produce equally good results³⁶.

Diversity is an important issue in approaches for multi-solution spectral simulation. Multiple diverse solutions provide a more complete description of a complex system, revealing possible conformational transitions. In our methodology, diversity and accuracy of the solutions are achieved with a hybrid evolutionary optimization algorithm (Chapter 2).

The number of spectral components and the characteristics of the components are determined through filtering of multiple solutions and grouping them by recognition of motional patterns³⁶. A powerful analysis tool requires adequate background knowledge, so that unwanted effects can be eliminated (e.g., the amount of nonspecific labeling should be estimated). Thus, the wealth of ESR spectroscopic information requires a deep knowledge about the studied system to be able to assign and interpret each component in terms of structure and dynamics.

6.3.2 Detection of the local restrictions by SDSL-ESR spectroscopy

In the experimental part of the methodology, site-directed mutagenesis is used to replace a strategically chosen amino acid residue with a cysteine. This cysteine is then targeted by a spin label^{26,37}. After protein purification, concentration, and – if needed – reconstitution into the membrane, the spin-labeled protein sample is prepared for ESR experiments (see Figure 1A). Taking into account that the spectral lineshape is highly sensitive to the motional properties of the spin label, the temperature has to be chosen in such a way that these motional properties will depend primarily on the protein structure, and not on the internal label properties. This condition imposes that the conformational space of a spin label attached to a protein should be fully exploited to be restricted by structural elements, such as the protein backbone, rotational space of neighboring amino acids, and lipids (in case of a membrane protein).

At low temperature the conformational space is degenerated into a small number of low-energy conformational states (rotamers), which depend mostly on the minimization of the internal energy in accordance with the angular potentials of the rotamer. However, when the temperature increases, the side chains start to exploit their full rotational spaces, which at the same time become restricted due to steric overlap with the more rigid backbone and due to sharing physical space with the side chains of the neighboring amino acids. If the temperature is increased even further, the backbone can lose its stable conformation. This state of the backbone will lead to an undesired situation for our methodology and should be avoided. Therefore, the lifetime of the protein backbone should be long in comparison with the nanosecond time scale of the ESR experiment. Under this condition, slow backbone motions will not be reflected in the ESR spectral lineshape, which is then mainly determined by the fast motions of the spin label side chain. If the lifetime of a protein conformation is shorter than the ESR time window, both backbone and side chains will contribute to the ESR spectra, preventing the extraction of useful structural information from the lineshape. In such a case, the protein backbone dynamics might be slowed down by increasing the viscosity of the environment, e.g., with addition of sucrose, or by lowering the temperature.

6.3.3 Cleaning of the motional patterns based on temperature series of SDSL-ESR measurements

In any case, it is advantageous to measure ESR spectra at several temperatures. After acquiring a temperature series of the ESR spectra at each mutant position, spectral simulations and optimizations are used to extract the appropriate motional patterns (see Figure 1A) (Chapters 2 and 3). Spectral simulation and automatic optimization of the spectral parameters help to characterize the site-specific motional properties in a well-

defined and high-throughput manner. However, the detected motional patterns should be checked for reliability before interpretation, or further usage for protein modeling and structural optimization. As ESR spectra are always noisy, it is impossible to analyze a single spectrum precisely. Thus, to increase the reliability of the analysis, ideally a suitable series of ESR spectra has to be measured, analyzed and interpreted jointly. Depending on the subject of research, an experimental series of different spin labels, various environments and/or chemical concentrations can be applied.

However, performing measurements and comparing results at different temperatures is the most straightforward approach to clear artifacts in the spectral analysis. This cleaning can be easily done in terms of a so-called “bubble diagram” (Figure 1A and 2). In this diagram the average characteristic values of the chosen parameter (such as the free rotational space Ω , rotational diffusion D , etc.) are plotted against the parameter of the series (e.g., temperature, or mutant position). The bubble size is related to the spectral weight of the motional pattern, while the vertical bar at each bubble represents the second moment of a distribution of that particular motional pattern in the phase space. Such a presentation allows the determination of general trends of significant patterns in the data series, as well as recognizing the numerical and computational artifacts by applying the following criteria. Reliable domains should appear regularly and consistently in the series of the external variable, e.g., by having a locally monotonous temperature dependence of Ω . Further on, isolated solutions and solutions with a spectral weight below a certain threshold of a few percent are discarded. In addition, spectrally irrelevant solutions (e.g., that describe none of the spectral features) are deleted. In the final presentation only the patterns that meet all mentioned conditions are kept.

To illustrate this “cleaning” strategy, a check of a typical temperature-dependent series of ESR measurements at a chosen mutant position, resulting in a series of motional patterns, is shown in Figure 2. Irregular temperature behavior (at 279, 281, and 283 K) indicates a deviation due to inappropriate filtering of the optimization results. It is expected that the temperature dependence of any parameter of our system is monotonous, unless it feels a major structural rearrangement, such as a phase transition. It is therefore expected that motional patterns evolve smoothly in some small temperature range. Therefore, deviating solutions are very likely to be caused by numerical artifacts. For the same reason, isolated solutions originating from either insignificant motional patterns with a small spectral weight, or inappropriate spectral components fitting noisy spectral details (e.g. small-weight patterns at temperatures 308, 310, 312 K), are also irrelevant (see Figure 2A). To increase the accuracy of the structure determination, all these inappropriate solutions should be systematically removed.

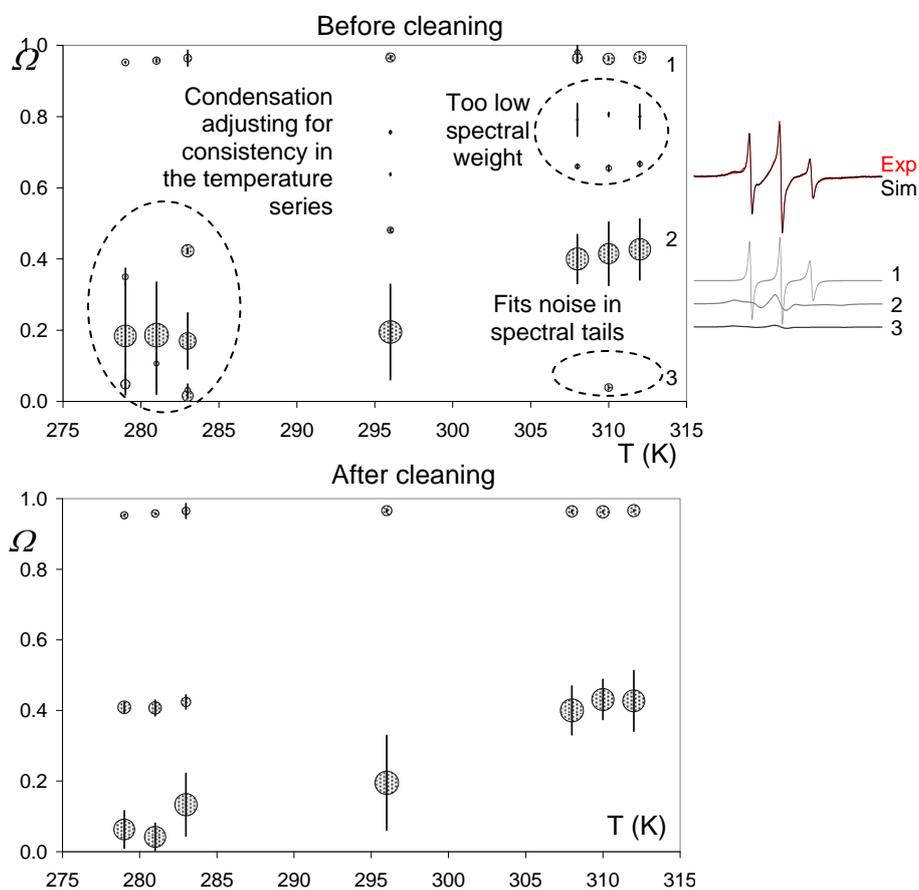


Figure 2. An example of motional pattern cleaning. A temperature-dependent series of bubble diagrams at each spin-labeled protein site (an example of a bubble diagram is shown for the NTAIL-*XD* protein complex spin labeled at position S491²⁸) is used to detect insignificant and/or false solutions. High values of Ω (between 0.7 and 1) correspond to (nearly) unrestricted motional patterns of the spin label, whereas low values (between 0 and 0.25) imply very high restrictions. Adjustment of the condensation procedure allows achieving consistency of motional patterns in the temperature series (see the motional patterns at 279, 281, 283 K marked with a dashed oval). Removing spectral components with a low intensity, or components that fit noise in the tails in the ESR spectrum (see motional patterns at 308, 310, 312 K marked with dashed ovals) allows focusing on the most important meaningful motional patterns. For illustration, the motional patterns at 308, 310, 312 K are numbered 1-3 and the corresponding ESR spectral components of the simulated 310 K spectrum are presented on the right.

Temperature dependencies can also be used to verify the main assumption in modeling the conformational space, i.e., that the backbone motion should be slow on the ESR time scale, whereas the motion of the side chains should be fast. This effect can be diagnosed by a sudden transition in the free rotational space as a function of temperature. In such a case, ESR

experiments at several temperatures are needed to identify whether the protein is in a permanent disordered state, or if the lifetime of the backbone conformation is too short.

Finally, the complexity of the motional patterns should also be taken into account. More than one reliable motional pattern at a site obviously means that the spin label feels different restrictions to its wobbling space. For example, a completely unrestricted motional pattern could indicate nonspecific labeling at unrestricted positions in the protein (i.e., the terminal ends). However, any situation where more than one motional pattern is revealed with a significant contribution means that there are coexisting local protein conformations. In the protein modeling any of these patterns can be used, and appropriate structures can be derived accordingly.

In this thesis, the SDSL-ESR method was used in combination with multi-component analysis to study light-harvesting membrane protein complex CP29. The results of multi-component analysis of ESR data permit to trace the structural organization of the long N-terminal domain of CP29 (Chapter 3).

6.4 Enhancement of conformational space modeling for the detection of the local restrictions

The protein backbone structure is parameterized by the dihedral angles φ_i and ψ_i at each i -th amino acid residue, following by the attachment of the amino acid side chains to the backbone (Figure 1C) (Chapter 4). It is assumed that atomic structures of the spin label and amino acid residues can be constructed using an approximation of fixed bonds lengths and bonds angles³⁸, based on previously reported values³⁹⁻⁴¹. In the modeling, the unrestricted conformational spaces of all amino acid side chains are attached to their respective backbone C^α atoms, which are assumed to be fixed in space. The unrestricted conformational space is a result of the rotations of the side chain of an amino acid, or of a spin label around single bonds. The side chain is rotated around its single bonds in different steps in accordance with the type of amino acid residue (or spin label). The steps in these rotations are derived by taking into account that the most computationally demanding step in conformational space analysis is the calculation of the restrictions of the conformational space. This calculation involves checking of the overlap between side chains, which quadratically depends on the number of rotamers in the conformational space. Therefore, it is clear that the number of rotamers has to be reduced as much as possible, however, by maintaining a certain degree of accuracy.

6.4.1 Sampling of side chain conformational rotamers

Since ESR spectroscopy is very sensitive to the available space of the fast rotational motion of the spin label attached to the protein, the rotational conformational space of the side chain can be taken as the most strategic unit in our protein modeling (see Chapter 4). In SDSL-ESR spectroscopy a protein is labeled at a specific site with a spin label of a size slightly larger than the size of the largest amino acid residues. Together with high-temperature conditions, this guarantees that the fully-exploited conformational space of the amino acid side chain becomes restricted due to steric overlap with the local backbone conformation, and due to the conformational spaces of the neighboring amino acids. In addition, the surrounding phospholipids (for membrane proteins) affect the conformational space of the spin label. To employ these restrictions for protein structure determination, the conformational space has to be measured experimentally and simulated at the same time and then compared. In this respect, it is important to note that the ESR experiment is insensitive to the exact atomic coordinates, but very sensitive to the motional anisotropy of the nitroxide group. Therefore there is no need for a precise calculation of a side chain conformation. Instead, the relative probability of side chain conformations in the conformational space of the spin label has to be determined.

6.4.2 Restrictive effect of the lipids on side chain conformations

A calculation of the restrictive effects should also take into account the average space-sharing effects of all surrounding wobbling chains from the neighboring amino acid side chains of the protein(s) and alkyl chains of the lipids. Two possible views about lipid effect on protein side chains: 1) high mobility of lipids (characterized by fluidity/viscosity) and conformational space sharing; against 2) lipid rigidity according to shell model and severe immobilization of the lipid and protein chains. From one hand, lipids fluidity is in favor of explaining protein-protein interactions between integral proteins. On the other hand, the contact between integral membrane proteins and lipids must be very tight to maintain the seal of the membrane as a permeability barrier⁴².

On the contrary to the restriction calculations that arise from the backbone and side chains, in case of a membrane environment the steric overlapping effect cannot be derived explicitly by calculating the overlap of atoms and groups. This arises, because the position of the atoms of the lipids is not precisely known. Therefore, it is clear that the lipid effect has to be introduced in the calculations in a more phenomenological way. In the simplest approximation the effect of the alkyl chains of the phospholipids should take into account following assumptions⁴³: a) side chain conformations, which stretch out from the main body of the protein perpendicular to the lipids, should be restricted by the highest extent; b) there are minimal restrictions in case of a parallel alignment to the membrane normal; c) the

lipids ordering is effective as soon as there is any non-zero angle between the side chain of a spin label and a lipid alkyl chain, meaning that the derivative of the lipid effect should be linear when ϑ angle approaches zero; d) perpendicular and near-perpendicular conformations should be restricted by approximately a similar extent, meaning that the derivative of the lipid effect should be zero, when the angle approaches $\pi/2$; e) the amplitude of lipid effect on the conformational space of the side chains can be deduced from the effect of electron density profile (static restriction, highest at the membrane surface, ⁴⁴) as well as from the effect of lipid chain rotational conformational space (dynamic restrictions, increasing towards the centre of membrane ⁴⁵). All these approximations can be merged into a probability function for the lipid effect (Chapter 5, Eq. 3). This description of the lipid effect is in agreement with results of recent molecular dynamics simulations studies, which shows that aromatic, polar and charged amino acid side chains tend to orient along the membrane normal ^{46,47}. As can be seen from Eq. 3 (Chapter 5), we assume that the lipid effect is depth independent. This is reasonable, as the two restrictive effects from virtually fixed headgroups and more flexible tails sum almost to a constant effect at different membrane depths ^{44,45}.

6.5 Structural optimization

The comparison of the simulated values for Ω_{sim} with the normalized free rotational space Ω_{exp} extracted from SDSL-ESR experimental data ^{21,43} is used to govern an optimization algorithm, which tunes the secondary structure of the protein and the parameters of its relative orientation and position (Figure 1B) (see also Chapter 5 for details).

6.5.1 Structural optimization algorithm

A single run of optimization (Figure 3), which counts for 200 generations, starts with the initialization of a protein structure (setting the pairs of backbone dihedral angles $\{\varphi_i, \psi_i\}$ and the relative orientation and position of the protein in the system) (Table 1, Figure 4), as well as the initialization of the optimization parameters and constants (such as selection, mutation, crossover, elite, and shaking). In each generation, the optimization parameters and the dihedral angles of the protein backbone are updated first (these parameters usually change with run generation number N_{gen}). Then the current structure of the protein system is modified by *internal operators* (modification of the secondary structure of a protein via a mutation and crossover operators applied to the backbone dihedral angles) and *external operators* (modification of the position and orientation of the protein towards the membrane, or towards the other protein). The external operators also include a rotation of the protein around its long axis, given by angle φ (relevant for helical chains, see Figure 4).

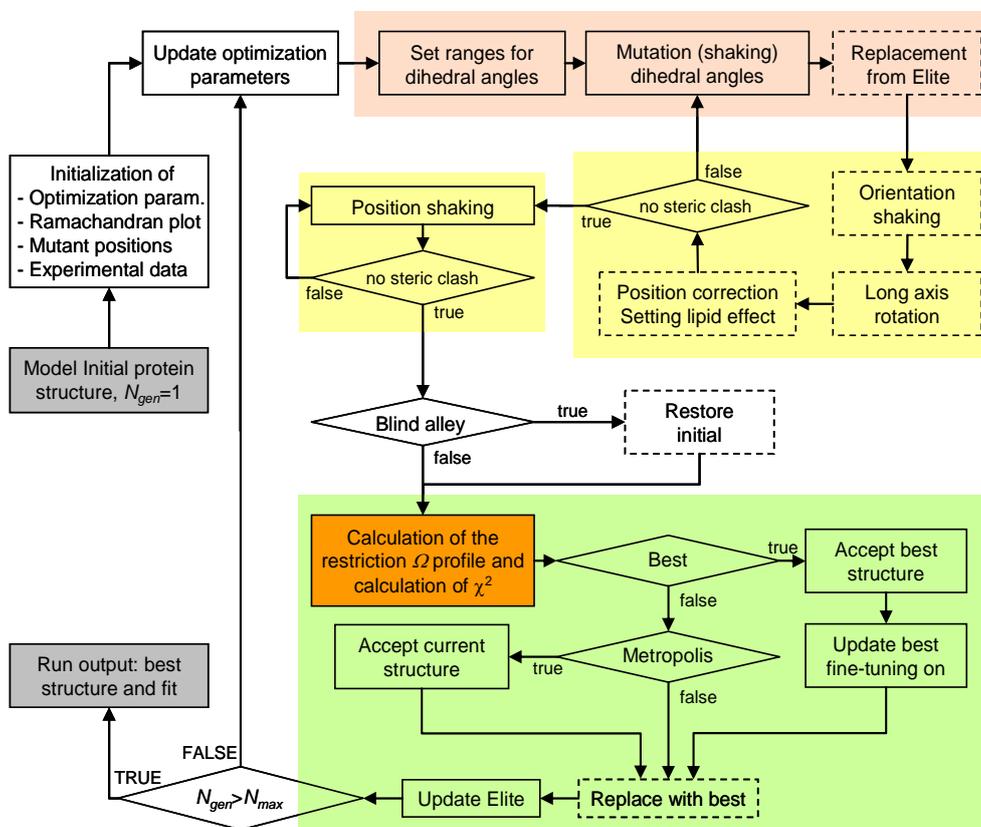


Figure 3. Scheme of a single run of the algorithm for protein structure optimization. The algorithm is split into several functional parts: internal structure optimization operators (red box), external structure optimization operators (yellow box), restrictions calculation and data fitting (orange box), decision making part (green box). The algorithm includes modeling of the lipid effect, as is needed in case of membrane proteins⁴⁸. The run generation number is N_{gen} . N_{max} is the maximum number of generations, typically 100.

Table 1 gives the parameters of the protein-lipid model, or protein complex model that can be optimized. At several stages in the optimization procedure, the protein structure in the system is checked for steric clashes. In case of internal steric overlap, the algorithm returns and makes another try with the current operator. There is a maximum number of clashed structures allowed in one generation. If this number is achieved the current problematic structure is replaced with the initial structure to protect the algorithm from going into a dead end. After the new structure is generated, the local restrictions at the mutant positions are calculated, and the obtained restriction profile Ω_{sim} is compared with the experimental restriction profile Ω_{exp} by a goodness of fit (Eq. 12, Chapter 5).

Table 1. Internal and external optimization parameters for different protein systems.

Parameter	Unit	Description
Internal parameters		
$\{\varphi_i, \psi_i\}$	$^\circ$	Pairs of dihedral angles (the first and last angles, φ_1 and ψ_N , are not defined)
φ	$^\circ$	Rotational angle (rotation of the protein around the long axis)
External parameters – membrane proteins		
tm_{start}	-	Starting position of the transmembrane region of the protein
tm_{end}	-	End position of the transmembrane region of the protein
n_{ref}	-	Reference residue usually in the centre of the protein
D	\AA	Steric thickness of the membrane
θ	$^\circ$	Tilt angle of the protein with respect to the membrane normal
d_{shift}	\AA	Shift of the protein in the bilayer along the membrane normal (used for the fine-tuning of the transmembrane position of the protein)
External parameters – protein complex		
Δx	\AA	Displacement vector of the protein relative to the partner
A	$^\circ$	Orientation tensor of the protein relative to the partner

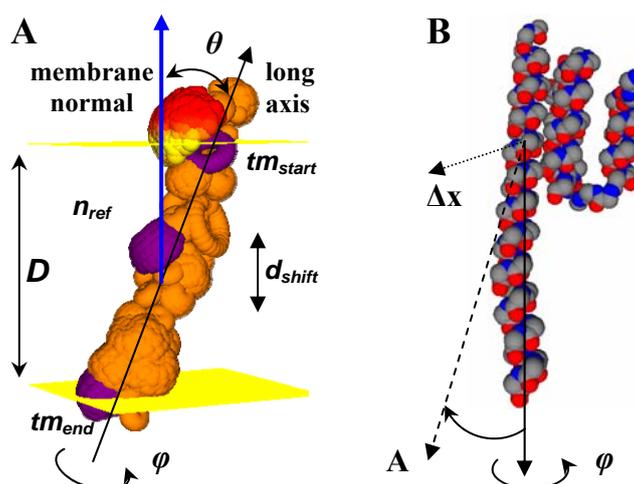


Figure 4. Parameters for the protein structure optimization. **A.** Relative position and orientation of membrane-embedded M13 coat protein⁴⁸. The protein is shown with the conformational spaces of the amino acid side chains and spin label. The starting, tm_{start} and ending, tm_{end} residues of the transmembrane part of the protein as well as a reference n_{ref} residue in the centre of the transmembrane domain are highlighted. The yellow planes indicate the restrictive region of the lipid bilayer. **B.** NTAIL protein, presented by backbone atoms relative to the partner protein XD²⁸. Both protein systems are parameterized according to Table 1.

The minimum value of χ^2 in a single run is defined as χ^2_{best} and it corresponds to a protein structure, which produces restrictions for the conformational space with the best fit to experimental data. If the goodness of fit χ^2 of a current structure is better than χ^2_{best} , then χ^2 replaces χ^2_{best} and the current structure becomes the parent for the structure in the next generation and the fine-tuning mutation mode is turned on. Even if χ^2 of the new structure

does not outperform the χ^2_{best} , it may still become the parent for the new generation, if the Metropolis criterion⁴⁹ is satisfied. In case the new structure is rejected, the parent stays the same as in the previous generation. However, with some probability (depending on the number of consequent unsuccessful generations) the current structure is replaced with the one that corresponds to χ^2_{best} to let the algorithm to produce good-fit structures still within the current generation. Finally, each structure with calculated restrictions is split into short subsequences with the corresponding goodness of fit (fitting corresponding part of the experimental data). Any successful subsequence updates the elite database, which is used later for elitist crossover. The algorithm repeats this main loop until the maximal number of generations is reached. The outcome of the optimization is always a family of best-fit protein structures, which agree with the experimental SDSL-ESR data. Note that in the case of a membrane-embedded protein the protein structures are found relative to the membrane⁴⁸.

In case of larger proteins, or membrane proteins that span the membrane several times, the protein backbone can be split into a set of protein domains, separated by loosely structured loops. The assignment of loosely structured loops and terminal ends could be carried out via the normalized rotational diffusion being much higher in the unstructured parts as compared to the structured parts, such as transmembrane domains^{20,21,48}.

6.5.2 Future improvement of structural optimization

Our algorithm of structural optimization, which is based on modification of the dihedral angles, blindly samples a high-dimensional conformational space of the protein. Similar to protein folding in nature, where the protein naturally finds its way through the conformational space, the optimization should also be equipped with mechanisms that help to reduce the huge number of conformational possibilities.

To be able to get deeper insights into the three-dimensional protein structure, one might take into account any available information about the evolution of the protein from a linear sequence of amino acid residues to folding and post-folding modifications. In the first step, the local secondary structure is formed, which is guided by hydrogen bonding. At the same time protein folding starts (spontaneous collapse into a compact state) mediated by hydrophobic interactions among nonpolar residues. This is followed by longer-range interactions, which stabilize the tertiary structure bring together structure chains, e.g., two α -helices that come together, β -strands organize into a β -sheet or β -barrel. In addition some proteins need the assistance from specialized proteins (chaperones) while folding. Others require post folding modifications to obtain the native functionality. This information can be used when improving the protein structure optimization^{50,51}.

Thermodynamically, the folding process can be viewed as a kind of free-energy funnel^{52,53}. The unfolded states are characterized by a high degree of conformational

entropy and relatively high free energy. As folding proceeds, the narrowing of the funnel represents a decrease in the number of conformations. In protein folding the virtual “surface” of the energy landscape makes it possible that the alpine skier coming from the top of the summit (an initial disordered structural conformation of the proteins) finds his way through the mountain terrain and comes to the finish at the proper place (native functional structure of the protein).

From one hand, an energy-based calculation of the intermediate conformational states, including checking for possible stabilizing interactions (i.e., hydrogen bonding), could help to reject unfavorable conformations. From the other hand, the efficiency of the optimization search may be improved by the introduction of known secondary structure motifs (α -helix, 3_{10} helix, β -sheet, turns and loops etc.) based on primary sequence analysis and secondary structure predictions. According to primary sequence analysis, the protein could be split into regions of certain secondary structures and each region could be described by a set of parameters (e.g., thickness of the helix/residues per turn, the pitch of the helix⁵⁴) including the angles, which would describe the relative orientation of these protein parts one with respect to each other. This would end up with a smaller number of optimization parameters than in case of structure parameterization via dihedral angles. The dihedral angles could be anyway recalculated from the structure. Such a reduced number of optimization parameters has been implemented already, e.g., in a FRET-data-based model⁵⁵⁻⁵⁷.

6.6 Computational demands

Our method requires solving of the inverse problem, both in the GHOST analysis, as well as in the protein structure optimization. Generally, this means that these procedures are very time consuming. Therefore, in our work a lot of effort was spent on speeding up the numeric calculations. Since the computational demands of the modeling of the conformational space strongly depends on the number of dihedral rotations, we optimized the generation of the conformational space by discretizing the experimentally derived probability histograms of the side chain rotamer angles⁵⁸⁻⁶³ of each single bond rotation for each type of amino acid side chain. The optimization efficiency was further increased especially by the introduction of a special operator that keeps track of successful structural segments (that successfully fit the corresponding segment of the SDSL-ESR-based restriction profile), as well as by the introduction of an operator for local structural tuning. Currently the characterization of a 50-amino acid membrane protein with 27 spin label positions⁴⁸ takes about 5 weeks of CPU time on a small 20-core (2.3 GHz) computer cluster. One week was needed to analyze SDSL-ESR data and to extract the corresponding motional GHOST patterns. The GHOST analysis was then used in protein structure optimization, which took another four weeks of CPU time to obtain 1000 best-fit structures

by evolving through approximately 200,000 protein structures. The computational demand is roughly linearly dependent on the protein size. Similarly, the methodology also allows a linear decrease of computational time by increasing the computational power. It should be noted that protein structure optimization is practically inaccessible, if molecular dynamics simulations would be applied to derive the restrictions of the conformational space even with a much larger computer cluster size. This impracticability arises, because the molecular dynamics simulations should reach an ESR averaging time of a few nanoseconds for each of the structure scanned.

6.7 Complementarity to other methods

Our method is comparable to the distance geometry approach employed in two-dimensional solution NMR spectroscopy that also results in a family of structures^{13,64}. However, the number of restraints used in the modeling in our method (free rotational space and rotational diffusion) depends on the number of available spin-labeled protein mutants. Such a data set is smaller than the data set available from NMR spectroscopy (nuclear Overhauser effect, one- and three-bond J-coupling, carbon and proton chemical shift, and rotational diffusion anisotropy)⁶⁵. The number of restraints could be increased by producing more spin-labeled protein mutants, or by using additional structural restraints from other low-resolution methods (e.g., distance measurements by fluorescence spectroscopy or dual-spin label SDSL-ESR, global conformation constraints by small angle X-ray scattering (SAXS) and circular dichroism (CD)).

Even though ESR spectroscopy is insensitive to the exact atomic coordinates, the sensitivity to the anisotropy of the local conformational space of the spin label and the corresponding modeling of the conformational space allow the determination of the backbone fold with almost atomic resolution. However, there is no structural information about individual side chain conformations. Instead, the protein can be represented as a space needed for all the amino acid side chains to wobble. The ability to track the protein structure and dynamics in a native environment and at physiological temperature is one of the great advantages of the proposed method. Another advantage is the higher sensitivity of ESR when compared to NMR, which means that much lower concentrations of protein samples are needed to perform the experiment³.

6.8 General results of the application of SDSL-ESR, structural modeling and optimization to proteins

Although a limited number of single mutant spin label positions is able to provide structure information about a protein, a better approach (like one that employs protein structure

modeling) will be needed to fully characterize the protein structure. Our simulation tests on synthetic oligopeptides (Chapter 5) as well as the structural optimization of M13 coat protein in lipid bilayers (Chapter 5) show that a typical nitroxide spin label is capable to feel local conformational effects up to five neighboring amino acids, in case of a helical conformation of the protein. However, in general a large number of spin-labeled protein mutants might be desired, to narrow the uncertainty in the structure calculations. As a rough estimate, every two or three amino acid positions should be sufficient for characterization a high-quality secondary structure.

What could contribute and enhance the approach:

- Calibrate the approach of the calculation of the local restrictions: carry out SDSL-ESR experiments and modeling on known water-soluble proteins, of which a high-resolution structure is known, or on single α -helical transmembrane proteins in lipid systems.
- Employ more structural constraints from SDSL-ESR: use the spin label motional cone angles ϑ and φ separately (not combined in Ω).
- Optimization can be enhanced by taking into account structure-stabilizing weak interactions, by calculation of the free energy of conformations, and removing unfavorable global conformations from the search.
- Using known secondary structure motifs, and in general, make use of parameter sets that are better than a list of backbone dihedral angles pairs.
- Implementation of the new knowledge-based optimization operators.

References

- 1 Lacapere, J.-J., Pebay-Peyroula, E., Neumann, J.-M., & Etchebest, C. *Trends Biochem. Sci.* **32**, 259 (2007).
- 2 Torres, J., Stevens, T.J., & Samso, M. *Trends Biochem. Sci.* **28**, 137 (2003).
- 3 Hemminga, M.A., Introduction and future of site-directed spin labeling of membrane proteins in *ESR Spectroscopy in Membrane Biophysics*, edited by Marcus A. Hemminga & Lawrence Berliner (Springer, 2007), pp. 1.
- 4 Svergun, D.I. & Koch, M.H.J. *Rep. Prog. Phys.* **66**, 1735 (2003).
- 5 Petoukhov, M.V. & Svergun, D.I. *Biophys. J.* **89**, 1237 (2005).
- 6 Kelly, S.M. & Price, N.C. *Curr. Protein Pept. Sci.* **1**, 349 (2000).
- 7 Fasman, G.D., *Circular dichroism and the conformational analysis of biomolecules*. (Plenum Press, New York, 1996).
- 8 Uversky, V.N. *Protein Sci.* **11**, 739 (2002).
- 9 Muller, D.J. & Engel, A. *Curr. Opin. Colloid Interface Sci.* **13**, 338 (2008).
- 10 Pebay-Peyroula, E., *Biophysical analysis of membrane proteins: investigating structure and function*. (Wiley-VCH, Weinheim, 2008).
- 11 Watts, A. *et al. Methods Mol. Biol.* **278**, 403 (2004).
- 12 Dominguez, C., Boelens, R., & Bonvin, A.M.J.J. *J. Am. Chem. Soc.* **125**, 1731 (2003).

- 13 Castellani, F. *et al. Nature* **420**, 98 (2002).
- 14 Arora, A. & Tamm, L.K. *Curr. Opin. Struct. Biol.* **11**, 540 (2001).
- 15 Wüthrich, K., *NMR of proteins and nucleic acids.* (Wiley, New York, 1986).
- 16 Henderson, R. *Q. Rev. Biophys.* **37**, 3 (2004).
- 17 Fleishman, S.J., Unger, V.M., & Ben-Tal, N. *Trends Biochem. Sci.* **31**, 106 (2006).
- 18 Fanucci, G.E. & Cafiso, D.S. *Curr. Opin. Struct. Biol.* **16**, 644 (2006).
- 19 Alexander, N., Al-Mestarihi, A., Bortolus, M., McHaourab, H., & Meiler, J. *Structure* **16**, 181 (2008).
- 20 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *J. Chem. Inf. Mod.* **45**, 1621 (2005).
- 21 Stopar, D., Štrancar, J., Spruijt, R.B., & Hemminga, M.A. *Biophys. J.* **91**, 3341 (2006).
- 22 Li, Q. & Fung, L.W.-M. *Biochem.* **48**, 206 (2009).
- 23 Huang, H. & Cafiso, D.S. *Biochem.* **47**, 12380 (2008).
- 24 Jao, C.C., Hegde, B.G., Chen, J., Haworth, I.S., & Langen, R. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19666 (2008).
- 25 Pistolesi, S. *et al. Biophys. Chem.* **123**, 49 (2006).
- 26 Hubbell, W.L., Gross, A., Langen, R., & Lietzow, M.A. *Curr. Opin. Struct. Biol.* **8**, 649 (1998).
- 27 Cowieson, N.P., Kobe, B., & Martin, J.L. *Curr. Opin. Struct. Biol.* **18**, 617 (2008).
- 28 Kavalenka, A. *et al. Biophys. J.*, submitted (2009).
- 29 Altenbach, C., Kusnetzow, A.K., Ernst, O.P., Hofmann, K.P., & Hubbell, W.L. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7439 (2008).
- 30 Becker, C.F. *et al. Magn Reson Chem* **43**, S34 (2005).
- 31 Dougherty, D.A. *Curr. Opin. Chem. Biol.* **4**, 645 (2000).
- 32 Hendrickson, T.L., CrΓ©cy-Lagard, V.r.d., & Schimmel, P. *Annu. Rev. Biochem.* **73**, 147 (2004).
- 33 Wang, L., Xie, J., & Schultz, P.G. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225 (2006).
- 34 Vos, W.L., Nazarov, P.V., Koehorst, R.B.M., Spruijt, R.B., & Hemminga, M.A. *Trends Biochem. Sci.* **34**, 249 (2009).
- 35 Kavalenka, A.A. *et al. Biophys. J.* **96**, 3620 (2009).
- 36 Štrancar, J. *et al. J. Chem. Inf. Mod.* **45**, 394 (2005).
- 37 Hubbell, W.L., Cafiso, D.S., & Altenbach, C. *Nat. Struct. Biol.* **7**, 735 (2000).
- 38 Xiang, Z. & Honig, B. *J. Mol. Biol.* **311**, 421 (2001).
- 39 Engh, R.A. & Huber, R. *Acta Crystallogr. A.* **47**, 392 (1991).
- 40 MacKerell, A.D. *et al. J. Phys. Chem. B* **102**, 3586 (1998).
- 41 Word, J.M. *et al. J. Mol. Biol.* **285**, 1711 (1999).
- 42 Luckey, M., *Membrane structural biology: with biochemical and biophysical foundations.* (Cambridge University Press, Cambridge ; New York, 2008).
- 43 Štrancar, J., Kavalenka, A., Zihlerl, P., Stopar, D., & Hemminga, M.A. *J. Magn. Reson.* **197**, 245 (2009).
- 44 Wiener, M.C. & White, S.H. *Biophys. J.* **61**, 434 (1992).
- 45 Vermeer, L.S., de Groot, B.L., Reat, V., Milon, A., & Czaplicki, J. *Eur. Biophys. J.* **36**, 919 (2007).
- 46 Johansson, A.C. & Lindahl, E. *Biophys. J.* **91**, 4450 (2006).
- 47 MacCallum, J.L., Bennett, W.F.D., & Tieleman, D.P. *Biophys. J.* **94**, 3393 (2008).

- 48 Kavalenka, A., Hemminga, M.A., & Štrancar, J. *Biophys. J.*, submitted (2009).
- 49 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. *J. Chem. Phys.* **21**, 1087 (1953).
- 50 Dill, K.A., Ozkan, S.B., Shell, M.S., & Weikl, T.R. *Annual Review of Biophysics* **37**, 289 (2008).
- 51 St-Pierre, J.F., Mousseau, N., & Derreumaux, P. *J. Chem. Phys.* **128**, 045101 (2008).
- 52 Dill, K.A. & Chan, H.S. *Nat Struct Mol Biol* **4**, 10 (1997).
- 53 Bryngelson, J.D., Onuchic, J.N., Socci, N.D., & Wolynes, P.G. *Proteins: Structure, Function, and Genetics* **21**, 167 (1995).
- 54 Berg, J.M., Tymoczko, J.L., & Stryer, L., *Biochemistry*, 5th ed. (W.H. Freeman, New York, 2002).
- 55 Koehorst, R.B.M., Spruijt, R.B., Vergeldt, F.J., & Hemminga, M.A. *Biophys. J.* **87**, 1445 (2004).
- 56 Nazarov, P.V., Koehorst, R.B.M., Vos, W.L., Apanasovich, V.V., & Hemminga, M.A. *Biophys. J.* **91**, 454 (2006).
- 57 Nazarov, P.V., Koehorst, R.B.M., Vos, W.L., Apanasovich, V.V., & Hemminga, M.A. *Biophys. J.* **92**, 1296 (2007).
- 58 Dunbrack, J.R.L. *Curr. Opin. Struct. Biol.* **12**, 431 (2002).
- 59 Lovell, S.C., Word, J.M., Richardson, J.S., & Richardson, D.C. *Proteins* **40**, 389 (2000).
- 60 Vasquez, M. *Curr. Opin. Struct. Biol.* **6**, 217 (1996).
- 61 Dunbrack, J.R.L. & Cohen, F.E. *Protein Sci.* **6**, 1661 (1997).
- 62 Dunbrack, J.R.L. & Karplus, M. *J. Mol. Biol.* **230**, 543 (1993).
- 63 Ramya Bhargavi, G., Sheik, S.S., Velmurugan, D., & Sekar, K. *J. Struct. Biol.* **143**, 181 (2003).
- 64 Bax, A. *Annu. Rev. Biochem.* **58**, 223 (1989).
- 65 Brunger, A.T. *et al. Acta Crystallogr. D. Biol. Crystallogr.* **54**, 905 (1998).

Summary

Site-directed spin labeling (SDSL) electron spin resonance (ESR) spectroscopy is a relatively new biophysical tool for obtaining structural information about proteins. This thesis presents a novel approach, based on powerful spectral analysis techniques (multi-component spectral simulations and evolutionary optimizations of ESR spectra) and modeling of the protein structure by calculating the restrictions of the conformational space of the attached spin label.

First, the feasibility of the ESR spectral analysis was enhanced by speeding-up the spectrum optimization and by automation of the analysis routines to enable the handling of large sets of spectroscopic data (e.g., for the joint analysis of SDSL-ESR spectra from multiple sites of a spin-labeled protein). According to the testing examples a speed-up factor of 5-7 was achieved.

Secondly, SDSL-ESR was used to study the topology of the long N-terminal domain of the photosynthetic light-harvesting complex CP29. Wild-type protein containing a single cysteine at position 108 and nine single cysteine mutants were produced, allowing to label different parts of the domain with a nitroxide spin label. In all cases the apoproteins were either solubilized in detergent, or they were reconstituted with their native pigments *in vitro*. The spin label ESR spectra were analyzed in terms of a multi-component spectral simulation approach. These results permit to trace the structural organization of the long N-terminal domain of CP29 leading to a structural model for its N-terminal domain.

Thirdly, we proposed a novel way to translate the local structural constraints gained by SDSL-ESR data into a low-resolution structure of a protein by simulating the restrictions of the local conformational spaces of the spin label attached at different protein sites along the primary structure of the membrane-embedded protein. The proposed structural model takes into account the restricting effect of the protein backbone, amino acid side chains and lipid environment. We tested the sensitivity of this approach for artificial oligopeptides and then for membrane-embedded M13 major coat protein decorated with a limited number of strategically placed spin labels by employing high-throughput site-directed mutagenesis. We found a reasonably good agreement of the simulated and the experimental data taking a protein conformation close to an α -helix.

Finally, by using an optimization algorithm we optimized the parameters of the protein-lipid model by improving the fit of the simulation data to the experimental conformational space data. The outcome of the optimization was a family of best-fit structures of membrane-embedded M13 protein, which not only agree with the available SDSL-ESR data, but also was consistent with a recent model based on site-directed fluorescence labeling.

Therefore, the present method provides a challenging starting point for the development of a powerful methodology for the protein structure characterization, an alternative approach to conventional techniques.

Acknowledgements

Marcus, thank you! It was a great opportunity to learn a lot new in your group! Science, biophysics, human relationships, self organization, assertiveness and personal competence... these are very important things you deliver to your students and colleagues. Thank you for your constant guidance and assistance with my PhD studies, paper writings, and thesis finalization.

Thank you to Prof. Dr. Herbert van Amerongen, the head of the laboratory of Biophysics for supporting and stimulating our research work and for the help with the finalization of the thesis! Thank you, Netty, Cor, Ruud, Elena for the help and support with work and during my stays in Wageningen. Special thanks to Douwe, Bart, and Rob for the assistance with finalization of my PhD! Thanks to the students with whom we had the Advanced Spectroscopy graduate course in Wageningen in 2004: Marjolein, Saskia, Ed, Marcus, Daphne, and Justin.

I would like to thank my Slovenian supervisor Dr. Janez Štrancar! Janez, your principal ideas and your constant guidance in the last three years lead us to the excellent scientific results and helped me to approach the successful end of my PhD. You have been showing how one has to be persistent, hard-working and devoted to his work till the very end. In addition, many nice trips around Slovenia made me fall in love with the county, its beautiful nature, mountains, forests, rivers, lakes and the sea.

I would like to thank our Slovenian collaborators: Dr. Bogdan Filipič from the Department of Intelligent Systems at the Jožef Stefan Institute for the help with the developing of the optimization algorithms. With your help we speeded-up our calculations for the advanced EPR spectral analysis and got a nice paper published. I would like to thank Dr. Primož Zihel and Dr. David Stopar, – without your contribution we wouldn't be able to cross the line in our epic journey with the manuscript on local structural restrictions analysis. David, your SDSL-EPR data on M13 coat protein played a crucial role in the developing of our new approach to protein structure characterization based on conformational spaces modelling and employing local restriction extracted from SDSL-EPR data.

Thanks, Zoran, Tilen, Iztok, Zrinka, Marjana, Sandra, Maja, Jana, Marjeta, Milan, and Slavko for nice working environment and fruitful discussion. Thanks, Ajasja and Jan for the help on the development of the approach of conformational space modelling! Thanks, Iztok, Dejan, Maral, Dilek, Michal for the valuable advices on GHOST condensation software.

Also thanks to my friends in Ljubljana, who supported me all this time: Saša Knezevic, Vladimir Minja, Zoran and Milada, Zoran and Urška, Stanimir and Zora, Alen and Inna, Wadie, Uroš, Mladen, Ljubica, Nina and Jurij... as well as Igor, Iztok, Špela, Franci, Miha and all other colleagues for joint sport activities that helped us to go forward with our research work. Thanks to my Belarusian mentors with whom I started my scientific work: Prof. Dr. Vladimir Apanasovich (Minsk), and Dr. Mikalai Yatskou (Luxemburg), and who introduced and supervised me in the field of spectroscopy data analysis, modelling and optimization. Thanks to my Belarusian colleagues Dr. Peter Nazarov (Luxemburg), Vladimir Lutkovski (Minsk), Genadz Astapenko (Minsk), Ekaterina Makarova (Warsaw), Marina Repich (Trento, Italy), Dr. Sergey Laptенок (Wageningen, The Netherlands), Alexander Golovati (Luxemburg), Dr. Victor Skakun (Minsk), Dr. Anatoly Digris (Minsk), and all colleagues from the department of systems analysis for our joint work, support and help during my postgraduate studies and studies at the faculty.

Special thanks to Prof. Dr. Arkadi Vernikov and also to Irina, Elena, Silouan, and Alexander Suharev for the spiritual support and for the friendship!

List of publications

In reviewed journals:

- Kavalenka, A.A.; Filipič, B.; Hemminga, M.A.; Štrancar, J. Speeding up a genetic algorithm for EPR-based spin label characterization of biosystem complexity. *Journal of Chemical Information and Modelling* **45**:1628-1635 (2005).
- Štrancar, J.; Kavalenka, A.; Zihlerl, P.; Stopar, D.; Hemminga, M.A. Analysis of side chain rotational restrictions of membrane-embedded proteins by spin-label ESR spectroscopy. *Journal of Magnetic Resonance* **197** (2), 245-248 (2009).
- Kavalenka, A.; Spruijt, R.B.; Wolfs, C.J.A.M.; Štrancar, J.; Croce, R.; Hemminga, M.A.; van Amerongen, H. Site-directed spin labelling study of the light-harvesting complex CP29. *Biophysical Journal* **96** (9), 3620-3628 (2009).

Submitted to reviewed journals:

- Kavalenka, A.; Hemminga, M.A.; Štrancar, J. Optimization of membrane protein structure based on SDSL-ESR constraints and conformational space modelling, submitted for publication.
- Kavalenka, A.; Urbančič, I.; Belle, V.; Rouger, S.; Costanzo, S.; Kure, S.; Fournel, A.; Longhi, S.; Guigliarelli, B.; Štrancar, J. Conformational analysis of the partially disordered measles virus NTAIL-XD complex explored by SDSL EPR spectroscopy, submitted for publication.
- Štrancar, J.; Kavalenka, A.; Urbančič, I.; Ljubetič, A.; Hemminga, M.A.; SDSL-ESR-based protein structure characterization, submitted for publication.

In international conferences proceedings:

- Kavalenka, A.A.; Štrancar, J. Maintaining solution diversity in a hybrid EA for EPR-based spin label Characterization of Biosystem Complexity. In: Bio Inspired Optimization Methods and their Application (BIOMA 2006). 147-156 (Ljubljana, Slovenia, 2006).
- Kavalenka, A.; Yatskou, M.M.; Apanasovich, V.V. Simultaneous analysis of multi-dimensional data by simulation modelling. In: Young researches in science 2005: physics and mathematics (Minsk, Belarus, 2005).
- Kavalenka, A.A.; Štrancar, J.; Apanasovich, V.V. Speeding-up complex EPR spectra analysis for biosystem complexity characterization. In: 8th International Conference on Pattern Recognition and Information Processing (PRIP'05). 48-51 (Minsk, Belarus, 2005)
- Yatskou, M.M.; Kavalenka, A.; Apanasovich, V.V.; Calzaferri, G. Principles of Monte Carlo simulations in physical chemistry: luminescence of organized dye molecules. In: Krasnoprosin, V. & Aluja, J.G. (ed.) MS'2004 - International Conference on Modelling and Simulation. 368-372 (Minsk, Belarus, 2004).
- Nazarov, P.V.; Kavalenka, A.; Makarava, K.U.; Lutkovski, V.M.; Apanasovich, V.V. Neural network based algorithm of preliminary data analysis: application to fluorescence and EPR spectroscopy. In: Krasnoprosin, V. & Aluja, J.G. (ed.) MS'2004 - International Conference on Modelling and Simulation. 130-134 (Minsk, Belarus, 2004).

Curriculum vitae

Aleh (*Rus.* Олег) Kavalenka was born in Minsk, Belarus (Soviet Union), on July 14th 1980 in the family of Lubov and Alexander Kavalenka. In 1998 after he graduated from the *Belarusian State University's lyceum* he entered the *faculty of Radio Physics* (Minsk, Belarus). He received his master degree and started his postgraduate studies at the *Department of Systems Analysis at Belarusian State University* under supervision of Prof. Dr. Vladimir Aranasovich and Dr. Mikalay Yatskou in 2003. In 2003 and 2004 he visited the *Laboratory of Biophysics at Wageningen University* (Wageningen, the Netherlands) where he was introduced to the field of Biophysics of Membrane Proteins and also took special courses, including a graduate course on Advanced Spectroscopy, under supervision of Dr. Marcus Hemminga. Since 2004 he is working at the *Laboratory of Biophysics and EPR Centre of the Jožef Stefan Institute* (Ljubljana, Slovenia) in collaboration with Dr. Janez Štrancar on EPR data analysis, and protein structure characterization. Since September 2006 he has been taking additional courses at the *Jožef Stefan International Postgraduate School* as well. His scientific work is related to membrane proteins structure determination, spectroscopic data analysis, protein structure modelling, and evolutionary optimization. In September 2008 he was granted admission to the doctorate (PhD Programme) at *Wageningen University*.

**Education Statement of the Graduate School
Experimental Plant Sciences**

The Graduate School
**EXPERIMENTAL
PLANT
SCIENCES**

Issued to: **Aleh Kavalenka**
Date: **30 September 2009**
Group: **Laboratory of Biophysics, Wageningen University**

1) Start-up phase	<i>date</i>
▶ First presentation of your project Introductory presentation, Wageningen, The Netherlands	Apr 2003
▶ Writing or rewriting a project proposal Development of an alternative approach to membrane protein structure determination based on the local rotational conformational spaces analysis	Apr 2007
▶ Writing a review or book chapter	
▶ MSc courses Advanced Spectroscopy course (BIP-31806)	Mar-Apr, 2004
▶ Laboratory use of isotopes	
<i>Subtotal Start-up Phase</i>	<i>9.5 credits*</i>
2) Scientific Exposure	<i>date</i>
▶ EPS PhD student days EPS PhD Student's Day 2009	Feb 26, 2009
▶ EPS theme symposia EPS Theme 3 Symposium 'Metabolism and Adaptation'	Feb 18, 2009
▶ NWO Lunteren days and other National Platforms Annual Meetings of the Dutch Lipids and Biomembranes group, Lunteren Annual Meetings of the Dutch Lipids and Proteins and Nucleic Acids group, Lunteren Annual Meetings of the Dutch Lipids and Biomembranes group, Lunteren	Mar 16-18, 2004 Dec 10-12, 2007 Mar 10-11, 2008
▶ Seminars (series), workshops and symposia Attending Biophysical seminars in Wageningen & Ljubljana	2004-2009
▶ Seminar plus	
▶ International symposia and congresses Bio Inspired Optimization Methods and their Application BIOMA 2008, Ljubljana, Slovenia. Regional Biophysics Conference, Balatonfured, Hungary. 15th International Symposium "Spectroscopy in theory and practice", University of Nova Gorica, Slovenia. COST D27 Final Evaluation Conference, Dubrovnik, Croatia. Bio Inspired Optimization Methods and their Application BIOMA 2006, Ljubljana, Slovenia. Conference Young researchers in science: physics and mathematics, Minsk, Belarus. PRIP'05 8th International Conference on Pattern Recognition and Information Processing, Minsk, Belarus. Bio Inspired Optimization Methods and their Application BIOMA 2004, Ljubljana, Slovenia. Modelling and Simulation International Conference, Minsk, Belarus. MIVASE EU project workshop "Molecular Inhibitors of Vacuolar ATPase", Wageningen, The Netherlands.	Oct 13-14, 2008 Aug 21-25, 2007 Apr 18-21, 2007 May 11-13, 2007 Oct 9-10, 2006 Nov 14-18, 2005 May 18-20, 2005 Oct 11-12, 2004 Apr 27-29 2004 Apr 15-16, 2004
▶ Presentations Poster: EPS PhD Student Day, "Site-directed spin labeling study of the light-harvesting complex CP29", University of Leiden, The Netherlands Poster: Conference in Sienna, Italy "Using conformational space modelling for easier identification of sensitive SDSL positions" Oral: Scientific Day at Institute (IJS), "Optimization of protein structure with constraints from SDSL-EPR spectroscopy", Bistra, Slovenia Oral: Lunteren meeting, "Modelling side chain rotational restrictions of spin-labelled membrane proteins", Lunteren, The Netherlands Oral: Research Seminar at Postgraduate School, "Conformational space modeling for protein structure determination", Ljubljana, Slovenia Oral: Regional Biophysics Conference, Balatonfured, Hungary, "From local rotational conformational spaces to the protein structure" Poster: COST D27 Evaluation Conference, Dubrovnik, Croatia, "Speeding-up the HEO-GHOST Algorithm for EPR-based Biosystem Characterization" Oral: Symposium "Spectroscopy in theory and practice", Nova Gorica, Slovenia, "SDSL EPR – GHOST – SICS approach to protein structures" Oral: - seminar at Postgraduate School, "Proteins Structure", Ljubljana, Slovenia Oral: - seminar at Postgraduate School, "Nanophysics of Proteins", Ljubljana, Slovenia Oral: Conference BIOMA 2006, "Solution diversity in a hybrid EA for SL EPR-based characterization of biosystem complexity", Ljubljana, Slovenia Poster: 39th Annual International Meeting, "From SDSL EPR experiments to membrane protein structure", Edinburgh, UK Oral: Conference "Young researchers in science", "Simultaneous analysis of multi-dimensional data by simulation modelling", Minsk, Belarus Oral: 8th International Conference on Pattern Recognition and Information Processing (PRIP'05), "Speeding-up complex EPR Spectra Analysis for Biosystem Complexity Characterization", Minsk, Belarus Poster: Regional Biophysics Meeting, "Speeding-up the HEO-GHOST Algorithm for EPR-based Biosystem Characterization", Zrece, Slovenia Oral: Conference Modelling and Simulation, "Principles of Monte Carlo simulations in physical chemistry: luminescence of organized dye molecules", Mir Belarus Oral: Conference Modelling and Simulation, "Neural network based algorithm of preliminary data analysis: application to fluorescence and ESR spectroscop Minsk, Belarus Oral: - Research seminar, "Speeding-up Spin Label EPR-based Characterization of Biosystem Complexity", Ljubljana, Slovenia Oral: - Research seminar, "Global analysis of time-resolved fluorescence spectra", Minsk, Belarus Oral: Conference of young scientists, "Global analysis of time-resolved fluorescence spectra" (in russian), Minsk, Belarus	Feb 26 2009 Sep 24, 2008 Sep 17, 2008 Mar 10-11, 2008 Sep 25, 2007 Aug 21-25, 2007 May 11-13, 2007 Apr 18-21, 2007 Mar 25 2007 Feb 07, 2007 Oct 9-10, 2006 Apr 2-5, 2006 Nov 14-18, 2005 May 18-20, 2005 Mar 16-20, 2005 Apr 27-29 2004 Apr 27-29 2004 Dec 14, 2004 Feb 2004 Sep 2003
▶ IAB interview	
▶ Excursions	
<i>Subtotal Scientific Exposure</i>	<i>25.3 credits*</i>
3) In-Depth Studies	<i>date</i>
▶ EPS courses or other PhD courses 3d European School in Materials Science: "Complex Metallic Alloys: Surfaces and Coatings", Ljubljana, Slovenia. 2nd European School in Materials Science: "Properties and Application of Complex Metallic Alloys", Ljubljana, Slovenia. Stochastic Optimization of Complex Systems, lectures course, Jozef Stefan Postgraduate school, Ljubljana, Slovenia Selected Chapters from Nanosciences and Nanotechnologies, lectures course, Jozef Stefan Postgraduate school, Ljubljana, Slovenia Physics of Nanomaterials, course with research seminar, Jozef Stefan Postgraduate school, Ljubljana, Slovenia Chemistry of Materials, lectures course with several research seminars, Jozef Stefan Postgraduate school, Ljubljana, Slovenia Stochastic Optimization of Complex Systems, lectures course, Jozef Stefan Postgraduate school, Ljubljana, Slovenia EPS Graduate Course "Advanced ESR Spectroscopy in Membrane Biophysics", Wageningen, The Netherlands	May 26-31, 2008 May 21-26, 2007 2006 2006-2007 2006-2007 2006-2007 2006 Mar-Apr, 2004
▶ Journal club	
▶ Individual research training "Molecular Dynamics simulations of the protein-lipids system", Chemical faculty of the Moscow State University, Russia	May 2004
<i>Subtotal In-Depth Studies</i>	<i>13.2 credits*</i>
4) Personal development	<i>date</i>
▶ Skill training courses Speed Reading, Ljubljana, Slovenia (1 week) Course on "Writing Research Project", Jozef Stefan Postgraduate school, Ljubljana, Slovenia (3 days, 3 lect) Pedagogics, Belarusian State University, Minsk, Belarus Psychology, Belarusian State University, Minsk, Belarus Informational Technologies, Belarusian State University, Minsk, Belarus English, two semester course with seminar & examination, Belarusian State University, Minsk, Belarus Philosophy, two semesters course with essay & examination, Belarusian State University, Minsk, Belarus Personal Competence, Assertiveness, Time Management & Project Planning course, Wageningen, The Netherlands Advanced English course (4 weeks), Oxford House college, London, UK	year 2007 spring, 2006 year 2005 year 2005 year 2005 year 2005 2004-2005 2004-2005 Mar 16-18, 2004 year 2002
▶ Organisation of PhD students day, course or conference	
▶ Membership of Board, Committee or PhD council	
<i>Subtotal Personal Development</i>	<i>14.5 credits*</i>
TOTAL NUMBER OF CREDIT POINTS*	62.5

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 credits

* A credit represents a normative study load of 28 hours of study