MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

ALEH KAVALENKA

DEVELOPMENT OF AN ALTERNATIVE APPROACH TO MEMBRANE PROTEIN STRUCTURE DETERMINATION BASED ON THE ANALYSIS OF LOCAL ROTATIONAL CONFORMATIONAL SPACES

DOCTORAL DISSERTATION

LJUBLJANA, MAY 2009

DEVELOPMENT OF AN ALTERNATIVE APPROACH TO MEMBRANE PROTEIN STRUCTURE DETERMINATION BASED ON THE ANALYSIS OF LOCAL ROTATIONAL CONFORMATIONAL SPACES

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia, May 2009

Supervisor: *Assist. Prof. Dr. Janez Štrancar*

Evaluation Board:

Assist. Prof. Dr. Igor Serša, Chairman, Jožef Stefan Institute, Slovenia
Prof. Dr. Marcus Hemminga, Member, Wageningen University, The Netherlands
Assist. Prof. Dr. Bogdan Filipič, Member, Jožef Stefan Institute, Slovenia

Aleh Kavalenka

DEVELOPMENT OF AN ALTERNATIVE APPROACH TO MEMBRANE PROTEIN STRUCTURE DETERMINATION BASED ON THE ANALYSIS OF LOCAL ROTATIONAL CONFORMATIONAL SPACES

Doctoral Dissertation

RAZVOJ NOVEGA PRISTOPA K DOLOČANJU STRUKTURE MEMBRANSKIH PROTEINOV NA OSNOVI ANALIZE LOKALNIH ROTACIJSKO KONFORMACIJSKIH PROSTOROV

Doktorska disertacija

Supervisor: Assist. Prof. Dr. Janez Štrancar

May 2009

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL
Ljubljana, Slovenia

Index

Abstract

The problem of structure determination of membrane proteins is addressed with a new combination of site-directed spin labelling (SDSL) electron paramagnetic resonance (EPR) spectroscopy and structure modelling of a protein and its conformational spaces. This new approach is aimed at structural characterization of membrane proteins and intrinsically disordered proteins.

In the first part of the thesis the advanced analysis of EPR spectra based on spectral simulations and condensation of multiple solutions was enhanced facilitating its application for protein structure characterization. In the second part a novel approach was developed to simulate the free rotational space of a spin label attached to a protein, taking into account the restricting effect of the protein backbone, amino acid side chains and lipid environment in case of membrane proteins. In the third part this simulations were coupled with protein backbone modelling and an optimization algorithm to optimize the secondary structure of the protein as well as the parameters of relative position and orientation in a protein-lipid or protein-protein systems. The outcome of the optimization is a family of best-fit structures used for characterization of the global conformation of a protein. Finally, the method was applied to study the structure of the membrane-embedded major coat protein of bacteriophage M13 as well as the intrinsically disordered N_{TAIL} protein in N_{TAIL}-XD protein complex of the measles virus.

Thus, the present method represents a challenging starting point for the development of an alternative powerful methodology for the structure characterization of proteins, particularly intrinsically disordered or membrane proteins as the combination of EPR and structural modelling provides a valuable complement to the time and spatial windows of the conventional techniques used in protein structure determination.

Povzetek

Disertacija se z novo kombinacijo metod mestno specifičnega označevanja (SDSL), elektronske paramagnetne resonance (EPR) in modeliranja strukture proteina ter njegovih konformacijskih prostorov loteva problema določevanja struktur proteinov. Nov pristop je namenjen predvsem strukturni karakterizaciji membranskih in intrinzično neurejenih proteinov. V prvem delu disertacije so predstavljene izboljšave analize EPR spektrov s pomočjo GHOST kondenzacije, ki naredijo metodo primernejšo za strukturno karakterizacijo proteinov. V drugem delu je predstavljen nov pristop za simulacijo prostih rotacijskih prostorov spinskega označevalca, kovalentno pritrjenega na protein, in izračun omejitev rotacijskih prostorov zaradi proteinske hrbtenice, stranskih skupin sosednjih aminokislinskih ostankov in (v primeru membranskih proteinov) tudi lipidov. V tretjem delu so simulacije rotacijskih prostorov uporabljene za optimizacijo sekundarne proteinske strukture in relativne orientacije proteinov v sistemih protein-protein ali protein-lipidi. Naloga optimizacije je poiskati čim boljši približek lokalnim restrikcijam, dobljenih iz GHOST kondenzacije eksperimentalnih EPR spektrov. Rezultat optimizacije je družina struktur z najboljšim ujemanjem z eksperimentalnimi podatki, ki nam opisuje globalne konformacije proteina. Na koncu smo novo metodo preizkusili za karakterizacijo strukture dveh proteinskih sistemov. Prvi je glavni protein virusne kapside bakteriofaga M13, vstavljen v membrano, drugi sistem pa je intrinzično neurejen NTAIL protein v NTAIL-XD proteinskem kompleksu virusa ošpic. S tem smo predstavili dobro izhodišče za razvoj alternativne metode za strukturno karakterizacijo proteinov, predvsem membranskih in intrinzično neurejenih, ki jo nam podaja kombinacija EPR in strukturnega modeliranja na časovni skali, ki je nedosegljiva ustaljenim metodam za določevanje struktur proteinov.

Abbreviations

IDP	=	intrinsically disordered protein
NMR	=	nuclear magnetic resonance
SDSL	=	site-directed spin-labelling
EPR (ESR)	=	electron paramagnetic (spin) resonance
FRET	=	Förster (or fluorescence) resonance energy transfer
M13	=	filamentous bacteriophage
PC	=	phosphatidylcholine
14:1 PC	=	1,2-dimyristoleoyl-sn-glycero-3-phosphocholine
MV	=	measles virus
NTAIL	=	nucleoprotein
P	=	phosphoprotein
XD	=	X domain of P
MTSL	=	methanethiosulfonate spin label
MD	=	molecular dynamics
SL	=	spin label
GA	=	genetic algorithm
SGA	=	simple genetic algorithm
HEO	=	hybrid evolutionary algorithm
S/N	=	signal-to-noise ratio
GHOST	=	condensation algorithm that filters and groups multiple solutions found during optimization of the simulated spectra
RGB	=	red green blue colour model

1 Introduction

Proteins are the key molecules in cells of living organisms, including human beings. The knowledge about protein structure and function provides important insights and practical applications in medicine, agriculture, nutrition, and industry [100]. And what is more, our ultimate concern in proteins is with the wonder of life itself. All constructed out of amino acids, the proteins are very diverse in their structure and function. A class of proteins called membrane proteins is one of the most challenging fields of structural biology and structural proteomics [96,187]. Although one third of all proteins are membrane proteins, less than 1% of the known protein structures correspond to membrane proteins [201], pointing out to the difficulties with membrane proteins structure determination by the existing methods. The later are also less successful in structure determination of the so-called intrinsically disordered or intrinsically unstructured proteins (IDP) [42]. Difficulties in the application of standard high-resolution methods for determination of three-dimensional structure of these classes of proteins, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, therefore, stimulate the development of alternative approaches. One of such techniques is site-directed spin labelling (SDSL) electron paramagnetic resonance (EPR). This technique provides both a structural and dynamical characterization of the local conformations or a membrane protein in its native environment, and therefore evolves into a very useful method for the structure analysis of membrane proteins [8,46,63,171,172].

1.1 Proteins

Interestingly, the whole wealth and diversity of proteins and protein functions (see Figure 1) is related to their enormously diverse 3D structures which primary originates in a series of 20 different basic amino acids, sequentially linked into one-dimensional biopolymer. Particular combination of amino acids and the specific environment govern the folding of the primary protein sequence into an astonishing 3D structure. Thus, special structural organization with specific dynamical properties define the functionally of the proteins.

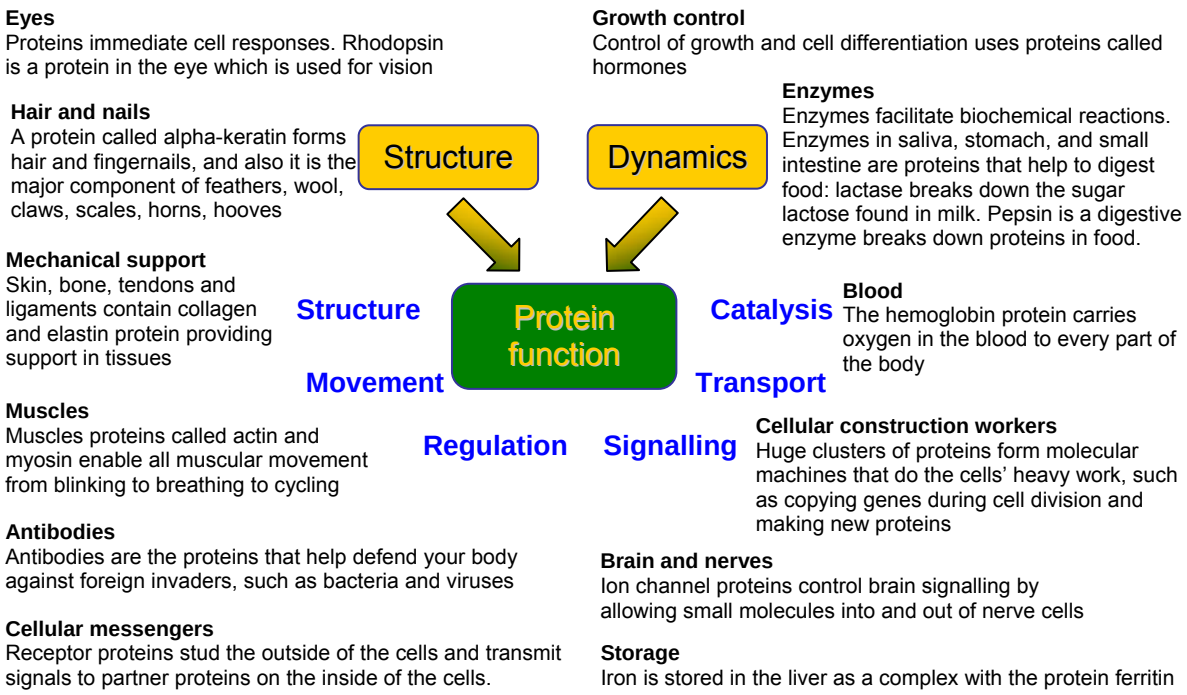

Figure 1: *Multitude of protein functions in the living organism.*

Proteins with their structural and functional variety could be compared in a way to a wealthy human language with its diverse vocabulary, and the set of amino acids to the letters of alphabet. The words of language dialects are unique due to unique environment: physical location, professional sphere, or social class. The proteins are also unique due to unique environment. And as the language can be rich to express human thoughts and ideas in our society, similarly the created nature is rich and diverse when expressing with the help of proteins the whole multitude of biological functionality in cells of living organisms. Nevertheless, when the structure and the function of proteins are affected it causes human diseases [90]. Actually, due to their functional importance proteins are the main targets for the majority of modern drugs, as it has evolved in our world to be the most popular way to cope with diseases.

1.1.1 Protein structure and dynamics

Mainly the primary sequence of a protein eventually defines its 3D structure. Ordering of the amino acid chain into a particular pattern (α-helices, β-strands forming β-sheets, and turns can be found on protein in Figure 2A) is referred to the secondary structure. Complete folding of the secondary structure into the three-dimensional order is then called the tertiary structure. And, finally, several folded proteins organized into a functional protein complex form the quaternary structure (see Figure 2A).

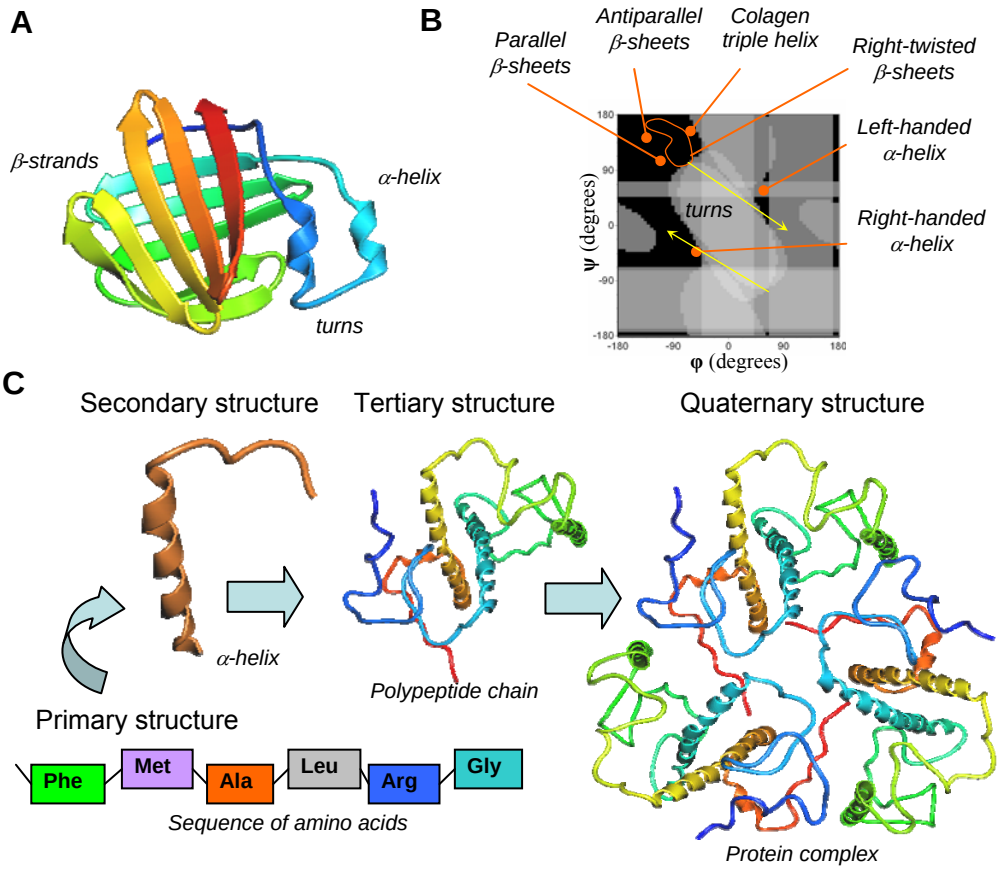

Figure 2: *Protein structure*. **A**. The structure of zebrafish liver bile acid-binding protein rich in β-sheets, also having two short α-helices and several turns (loops) [25]. **B**. Parameterization of the secondary structure of the protein with Ramachandran plot, the distribution of φ and ψ backbone dihedral angles. Black areas indicate the allowed combinations of φ and ψ dihedral angles. Secondary structure motifs, α-helix and β-sheets as well as turns are marked with orange lines and yellow arrows respectively. For turns one of two consequent residues has to be glycine (this smallest amino acid may have combinations of φ and ψ in the regions forbidden for other amino acids). **C**. The levels of protein structure: the primary structure consists of a sequence of amino acids linked together by peptide bonds. The resulting polypeptide can be coiled into units of the secondary structure, such as an α-helix. The helix is a part of the tertiary structure of the folded polypeptide, which is itself one of the subunits that make up the quaternary structure of the multisubunit protein complex. The monomer representing the tertiary structure organization is a light-harvesting complex II [164]. 3D structures for (A) and (C) were obtained with ArgusLab 4.0 molecular modelling software [183].

In proteins, the amino acids are covalently linked in linear sequences, 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 structure [2]. 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 [146,147], which also identifies regions of different motifs of the secondary structure, e.g. α-helix and β-sheets (see Figure 2B). 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.

Interestingly, interactions that control secondary, tertiary and quaternary structure are much weaker than the covalent bond. Secondary structure motifs, α-helix and β-sheets, are entirely governed by hydrogen interaction. In case of transmembrane proteins secondary structure is stabilized additionally by hydrophobic interaction and van der Waals interactions imposed by membrane lipids. Different turns appear in protein structure due to main chain flexibility and the lack of secondary structure stabilizing interactions. Hydrogen bonding, hydrophobic forces, electrostatic forces and van der Waals forces are the weak interactions that cooperatively govern protein folding and stabilize the tertiary and the quaternary structures.

In order to unravel the functionality of the protein, the information about protein tree-dimensional structure has to be combined with the information about protein dynamics [58,66]. Protein dynamics includes both equilibrium fluctuations and non-equilibrium effects such as structural transitions. It is thought that exactly the fluctuations observed at equilibrium seem to govern biological function in processes both near and far from equilibrium [66].

Proteins function-dynamics is rooted in the multi-dimensional energy landscape (Figure 3A) that defines the relative probabilities (thermodynamics) of the conformational states and the energy barriers between them (kinetics). Protein dynamics is characterized by timescale and the amplitude of the fluctuations (as suggested by energy landscape) as well as by the directionality of the fluctuations. It should be noted, that particular energy landscape is very much tied to an individual set of external parameters: temperature, pressure and solvent (environment) conditions. Thus variation of these conditions changes the relative populations of the protein conformational states and the kinetics of conversion between them.

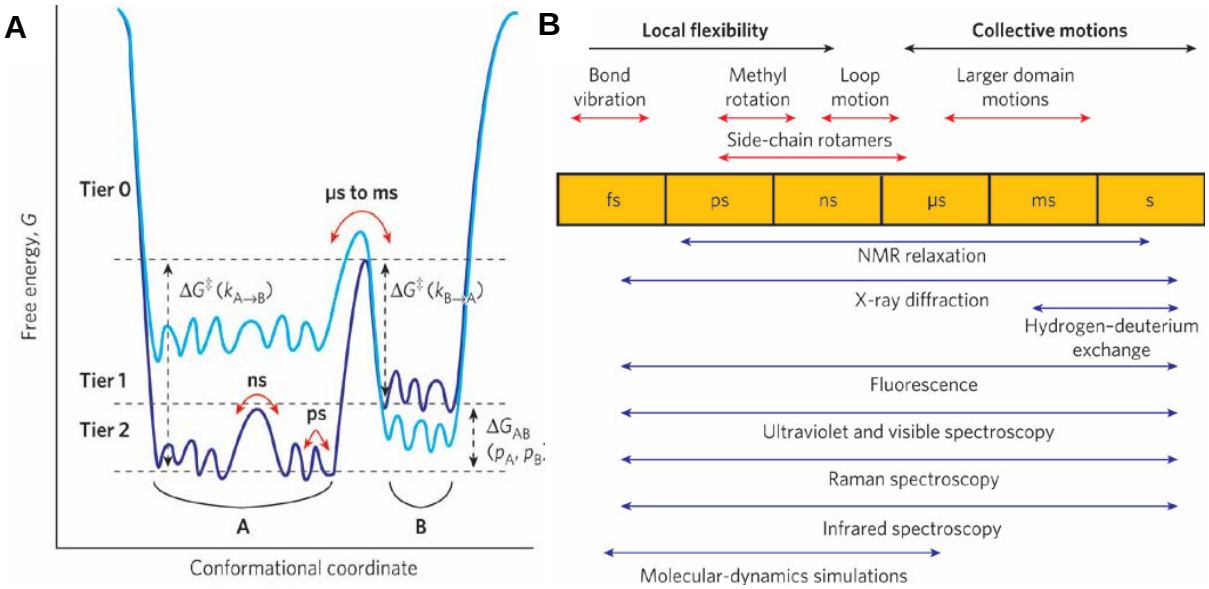

Figure 3: *Timescale of protein motions*. **A**. One-dimensional cross-section through the high-dimensional energy landscape of a protein showing the hierarchy of protein dynamics and the energy barriers. Each tier is classified following the description introduced by Frauenfelder and co-workers [6]. A state is defined as a minimum in the energy surface, whereas a transition state is the maximum between the wells. The populations of the tier-0 states A and B (p_A, p_B) are defined in accordance with Boltzmann distributions based on their difference in free energy (ΔG_{AB}). The barrier between these states ($\Delta G^{\ddagger}$) determines the rate of interconversion (k). Lower tiers describe faster fluctuations between a large number of closely related substrates within each tier-0 state. A change in the system will alter the energy landscape (e.g. from dark blue to light blue, or vice versa). For example, ligand binding, protein mutation and changes in external conditions shift the equilibrium between states. **B.** Comparison of the timescale of dynamic processes in proteins and the time windows of various experimental methods.

Dynamics in proteins at physiological temperatures is divided into three regimes (Figure 3A). The slowest, above-microsecond time-scale corresponds to the transitions between kinetically distinct states separated by energy barriers of several kT (the product of the Boltzmann constant and the absolute temperature). Typically, these are larger-amplitude collective motions between relatively small numbers of states. The probability of transitions between tier-0 states is very low. Many biological processes including enzyme catalysis, signal transduction and protein-protein interactions occur on this timescale. Faster picosecond-to-nanosecond timescale dynamics (tier-1 and tier-2) in contrast correspond to the fluctuations in a large ensemble of structurally similar states that are separated by energy barriers of less than kT, resulting in more-local, small-amplitude fluctuations at physiological temperature (Figure 3).

Although there are many experimental methods (Figure 3B) to study proteins dynamics at different time scales, different methods are often combined. Computational methods (molecular dynamics simulations) are currently more applicable to study the fast dynamics, although a large variety of approaches that simplify the force fields help to extend the simulation trajectory lengths and allow studying slower time scales processes [66].

1.1.2 Membrane proteins

A membrane protein is a protein molecule that is attached to, or associated with one of the cell membranes. The later define the periphery of the cell, separating its content from the surrounding. Membranes are composed of lipids and protein molecules that form a thin hydrophobic barrier around the cell. 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 by triggering the signaling cascades; and membrane enzymes participate in various reaction pathways. Because the individual lipids and membrane proteins are not covalently linked, the entire structure is remarkably flexible, allowing changes in the shape and size of the cell [100].

Complex structure of the biological membrane was initially described by the fluid mosaic model of Singer and Nicholson [43,159]. In this model the fatty alkyl chains in the interior of the membrane form a fluid hydrophobic region. Integral proteins float in lipids, held by hydrophobic interactions with their nonpolar amino acid side chains (see Figure 4). Both proteins and lipids are free to move laterally (at the rates about 1-2 micrometer per second [143]) in the plane of the bilayer, but movement of either from one face of the bilayer to the other is restricted. The refinements of the fluid mosaic model were suggested, based on the results of experimental and theoretical studies [196]. A concept of 'hydrophobic matching' [74,88,89,127] suggests that proteins and lipids need to 'adjust' to each other (Figure 4).

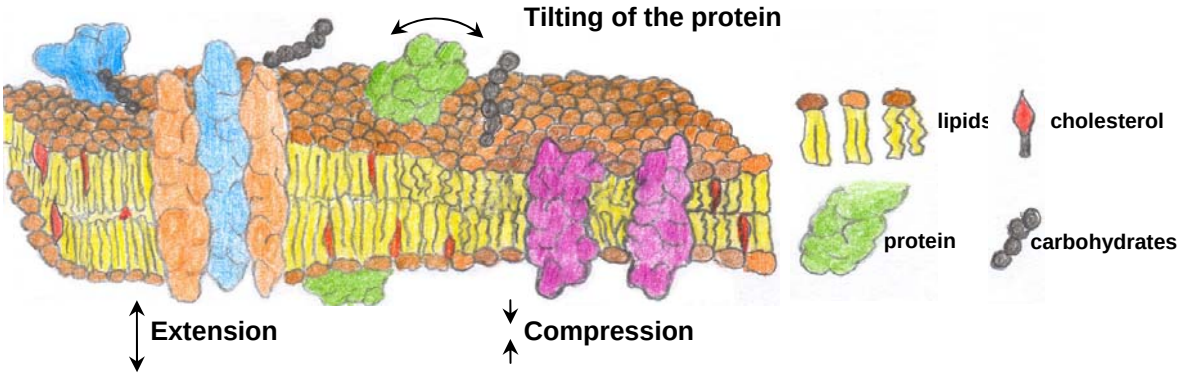

Figure 4: *Biological membrane and membrane proteins*. **A**. The fluid mosaic model of the cell membrane. Because component lipids and proteins are not naturally matched in this membrane, they must strain (expend energy) to match each other hydrophobically, resulting in a high-energy membrane. Compensatory conformational changes include lipid alkyl chain extension and transmembrane helix tilting when lipids surround a protein with a long transmembrane region, and lipid alkyl chain compression when lipids surround a protein with a short transmembrane domain.

Small proteins of one or few α-helical chains are likely to accommodate the bilayer thickness by helix tilting, however, larger proteins or less flexible proteins induce changes in bilayer thickness. More advanced but still discussed picture of the fluid mosaic model contained patches of lipids, the composition of which differed from the average for the bilayer [43]. This lipid domains, 'rafts', emerging and dissolving [143] are thought to have crucial biological function [61,75]. In a shell model [5] the lipids form a shell

(annulus) around the protein. Annular lipids are stably associated with proteins, though individual lipid molecules remain in the annular shell around a protein for only a short period of time [75,98]. EPR [116,117,119,120], NMR [53,199] and optical spectroscopy have been used to studying annular lipids and protein-lipid interactions. To conclude, the 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 [196]. It is also clear that that there are strong interactions between lipids and proteins in the membrane [143]. Thus, when modelling membrane protein structural restrictions, the effect of lipids in the transmembrane regions is very significant.

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 bilayer [187]. 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). Most simple transmembrane proteins (like the major coat protein of M13 introduced in section 3.4) consist of a single chain (usually α-helical), which spans across the membrane. The larger proteins consist of multiple segments. These are usually the chains of α-helices or strands of β-barrels connected with the loops.

The topology of an integral membrane protein can be partially predicted from its primary sequence. If the later is longer than 20 hydrophobic amino acids, such a part of protein will traverse the lipid bilayer. A protein chain surrounded by lipids lacks hydrogen-bonds which are otherwise formed with water molecules. Consequently, it tends to form α-helices and β-sheets, where intra-chain hydrogen bonding minimizes the chain free energy. If the side chains of all amino acids in a helix are nonpolar, hydrophobic interactions with the surrounding lipids further stabilize the helix [100]. Polar amino acids (lysine, arginine, glutamate, and asparagate) are found exclusively in the aqueous phases, which can, however, be found also on the polar side of the amphiphatic transmembrane helices that form a pore. The side chains of tyrosine and tryptophan are often presented in the interface between lipids and water [36], able to interact both with the lipids and with water, and serving as membrane interface anchors.

Protein-lipid interactions are expected to play a prominent role in the membrane structure [35]. Not only do integral proteins perturb the lipids, but the physical state of the lipids does also actively influence protein function [76]. Membranes are very dynamic structures with constant movements of lipids in the bilayer, both in the transverse direction across bilayer and the lateral direction in the plane of this two-dimensional matrix. The movements in the lateral direction give rise to the fluid nature of the membrane and enable interactions among proteins and between proteins and lipids [107]. It is though that lipid dynamics profoundly influence the function of membrane proteins, not least in dynamically differentiated and spatially separated in-plane membrane domains [117]. 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 residues [136].

1.1.3 Intrinsically unstructured proteins

Intrinsically disordered proteins (IDPs) are functional proteins that do not fold into well-defined, unique three-dimensional structures under physiological conditions [40,42,50,186,192]. IDPs show an extremely wide diversity in their structural properties. Indeed they can attain extended conformations (random-coil-like) or remain globally collapsed (molten-globule-like), where the latter possess regions of fluctuating secondary structure [128]. Although there are IDPs that carry out their function while remaining permanently disordered, many of them undergo induced folding, i.e. a disorder-to-order transition upon binding to their physiological partners [12,126] (see Figure 5). The functional role of intrinsically disordered proteins in crucial areas, such as transcriptional regulation, translation and cellular signal transduction, has only recently been recognized [42]. Many of the disordered regions and most if not all of the completely disordered proteins are involved in cell signalling or regulation. Qualitatively, it seems reasonable that highly flexible proteins would provide a better basis for responding to changes in the environment than rigid ones. [41] It can be additionally expected that the structural transition in case of IDPs are much faster than for other proteins since the interactions in the protein under physiological conditions are usually not strong enough to stabilize its structure without strong constraint of the partner proteins.

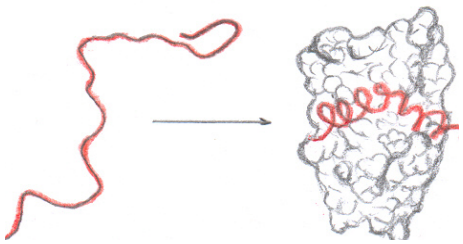

Figure 5: *Schematic illustration of folding of IDP upon binding*. The phosphorylated kinase-inducible domain (pKID) of the transcription factor cyclic-AMP-response-element-binding protein (CREB) is unstructured when it is free in solution but it folds on forming a complex with the KID-binding (KIX) domain of CREB-binding protein (CBP).

Spectroscopic methods such as NMR have now advanced in sensitivity and resolution, to the point at which the structural propensities and dynamics of sizeable disordered proteins in solution can be thoroughly characterized [42]. Site-directed spin labelling EPR has been used to detect local structural characteristics of disordered proteins [11,128].

1.2 Protein structure characterization

Around 90% of the protein structures available in the Protein Data Bank [14,200] have 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 Nuclear Magnetic Resonance (NMR) techniques. Alternatively electron microscopy (EM), atomic force microscopy (AFM) can also be used to determine 3D structure [51,65,129,140]. Certain aspects of the secondary structure as whole can be determined via other biochemical techniques such as circular dichroism (CD) [47,87,192], small-angle X-ray scattering (SAXS) [141,181], fluorescence and electron paramagnetic resonance (EPR) spectroscopy [63]. Cryo-electron microscopy has recently become a means of determining protein structures to 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 being constantly enhanced 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* [14,200].

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 Modelling of protein structure

The main goal of the protein structure characterization is to determine the protein function. Since both the structure and function of a protein strongly depend on the structure and dynamics of the environment, a modelling of a biomolecular system should in general include to the whole system and model the structure and the dynamics of all the system components. Five choices have to be made when modelling a bimolecular system: 1) scale of structure and dynamics have to be defined; 2) 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) force field (what interactions are taken into account) have to be revealed; 4) sampling

scheme of the conformational space have to be setup; and 5) boundary conditions (for dynamics simulations) have to be identified. When identifying these five choices, four problems arise: 1) force-field accuracy; 2) conformational sampling (search) efficiency; 3) ensemble sampling scheme (usually only a small part of the experimental system is modelled); 4) experimental conditions are not well defined (whether enough data is available, whether a proper comparison between simulated and experimental data is made) [193].

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

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 total potential energy [2].

One of the most efficient simulation tools to study the structure and the dynamics in proteins is the molecular dynamics (MD) simulations. It has started to make new and specific quantitative predictions about biological properties not yet reported from experiment [59,104,156]. Constant increase of modern computational power and further development of MD protocols including coarse-grained methods [16,114] gradually extend the simulation time range and improve poor conformational sampling. In addition, certain work has to be done to verify simulations, force fields and simulation methods [104,189]. These are the most important reasons why the search for new alternative protein structure characterization methods is one of the frontiers in biophysics.

2 Aims and Hypothesis

This work is aimed at developing a new approach for protein structure characterization based on modelling of local conformational spaces coupled to the structural constraints extracted from the spin label EPR spectroscopic data. To present the applicability of the new method, we focused on membrane proteins and intrinsically disordered proteins as these two classes of proteins have been very problematic in obtaining structural information. Either the problem originates in specific membrane environment, which prevents solubilization of the protein for NMR spectroscopy and crystallization required for X-ray crystallography, or the fast dynamic characteristics, especially important for the functionality of intrinsically disordered proteins, lead to the ensemble of protein conformations, which can not be resolved with the conventional methods. With this in mind, we split our work into three stages (see Figure 6):

1. EPR spectroscopy and data analysis (Figure 6A). An advanced analysis of EPR data based on spectral simulations, optimization of spectral parameters and GHOST condensation of multiple solutions is required, so that the molecular modelling can employ experimental data about local restrictions in proteins obtained with site-directed spin labelling EPR spectroscopy. The feasibility of spectral analysis is achieved by speeding-up the spectrum optimization algorithm (see sections 3.2 and 4.1) as well as by development and application of a new data condensation technique (section 3.1).

2. Conformational space modelling and restrictions calculations (Figure 6B). Since EPR 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 is taken as the most strategic unit in our approach. We develop a model to calculate the local restrictions to the conformational space of the spin label (section 3.3) and we test the sensitivity of this approach (section 4.2). The multiple SDSL-EPR data from a set of spin-labelled protein mutants, describing local restrictions along protein sequence, is modelled simultaneously (Figure 6B) allowing the comparison between experimental and simulated restriction profiles.

3. Structural optimization of proteins (Figure 6C). We set up a structural optimization (sections 3.5) that enables to find the best possible structures of the protein based on fitting simulated restrictions to the experimentally obtained restrictions. In this respect, the backbone dihedral angles are continuously changed and the local restrictions are recalculated, thereby optimizing the 3D structure of the protein. The goodness of fit to the experimental data guides the optimization procedure through the search space towards more favourite structures (Figure 6C). At the end the optimization provides a family of favourite global protein conformations (Figure 6D).. The whole procedure was then applied to the M13 major coat protein (sections 3.4.1, 3.5.2, and 4.3), a membrane protein, and to the measles virus nucleoprotein, an intrinsically disordered protein (sections 3.4.2, 3.5.3, and 4.4) of N_{TAIL}-XD complex.

According to the aims our proposed hypotheses are the following:

1. Restrictions of the conformational space of a spin-labelled protein can be modelled by means of reducing the probabilities of the side chain conformations due to overlap by backbone, neighbouring side chains and surrounding lipids (in case of membrane proteins).

2. The global analysis of the measured and modelled restrictions of the series of the spin labelled mutants can be used for 3D structure characterization of membrane proteins and other proteins.

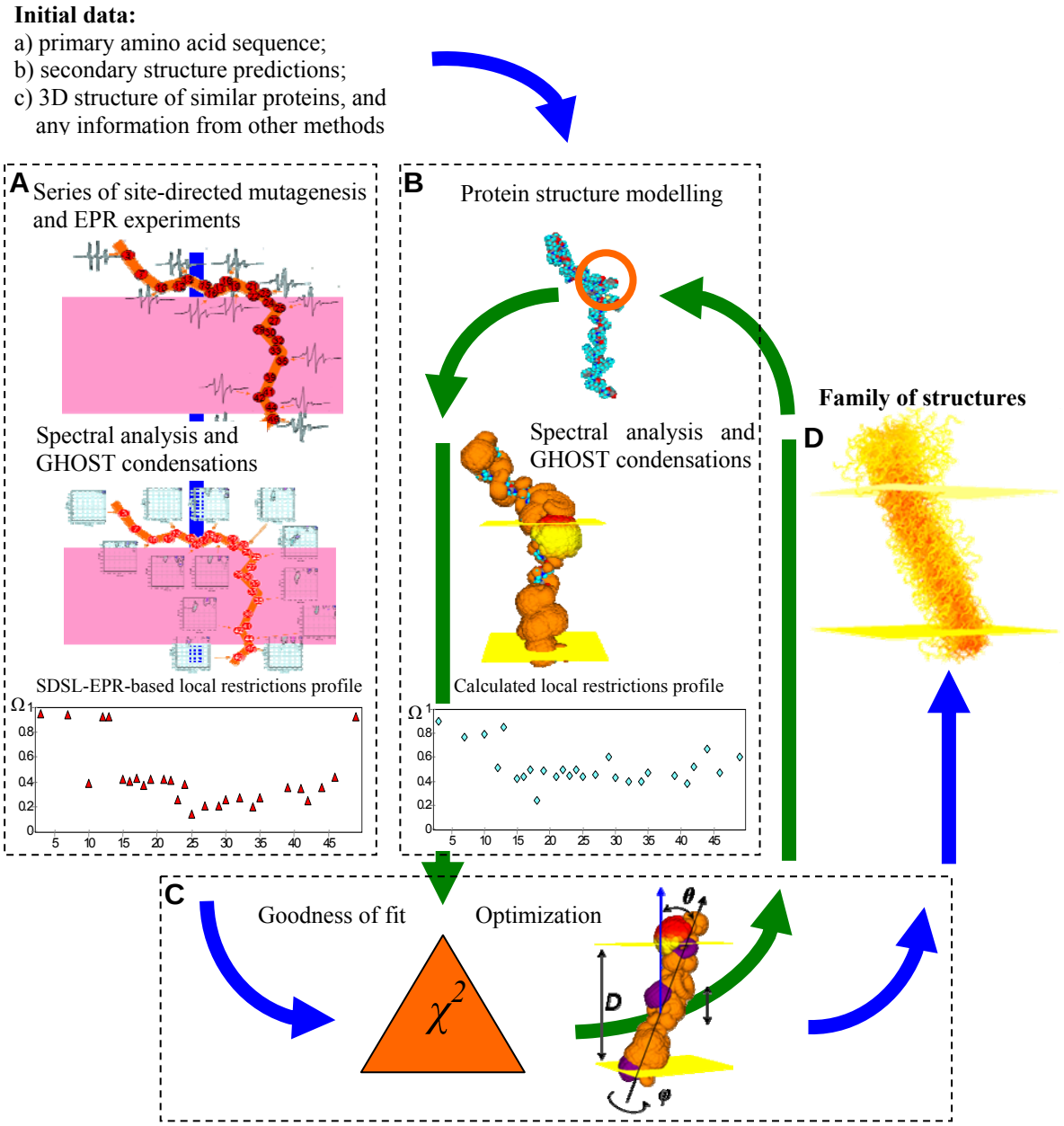

Figure 6: *Aims and hypothesis: overview of the SDSL-ESR approach for protein structure determination*. Analysis of a series of SDSL-EPR data in terms of free rotational space Ω (A) is combined with protein conformational space modelling and local restrictions simulation (B); fitting of the restrictions and structural optimization (C) allows to obtain a population of favourite structures (D).

3 Materials, Methods and Experiments

3.1 Determination of free rotational space through SDSL-EPR spectra

Te advanced analysis of SDSL-EPR spectra based on spectral simulations, optimization of spectral parameters and GHOST condensation of multiple solutions is required in order to determine the local conformational restrictions in proteins, which are employed for protein structure characterization.

3.1.1 Biosystem complexity

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 7). 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 neighbouring molecules. All these parameters make the biological membrane a very complex system in which many motional patterns can be found. Similarly, a high complexity of local motions can be found also in proteins, as they, in addition to the local heterogeneity, may be found in various static global conformations or be involved in slow passages through many different conformational states.

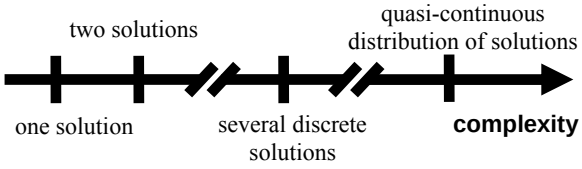

Figure 7: *Biosystem complexity axis.* Complexity is increasing from simple single-solution to quasi-continuous distribution of solutions.

EPR spectroscopy in combination with nitroxide spin labelling (SL-EPR) has proven to be a powerful technique for the exploration of heterogeneity and motion in biological systems [29,178]. The time scale of SL-EPR appears to be in the nanoseconds range [118], 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 in biological membranes or local motional patterns on proteins.

3.1.2 Site-directed spin labelling EPR spectroscopy

Site-directed spin labelling (SDSL) electron paramagnetic resonance (EPR) spectroscopy [70-72] is among new characterization techniques of biophysics, which is alternative to such powerful methods of protein structure determination as X-ray crystallography and NMR spectroscopy. SDSL-EPR provides both local structural and dynamic information on proteins [103] and has been widely applied to membrane proteins [63].

In SDSL-EPR a spin probe (nitroxide) is incorporated into the protein by attaching it to a cysteine side chain (see Figure 8). In the presence of paramagnetic species, e.g. a nitroxide, which contains an unpaired electron, EPR absorption is observed [187].

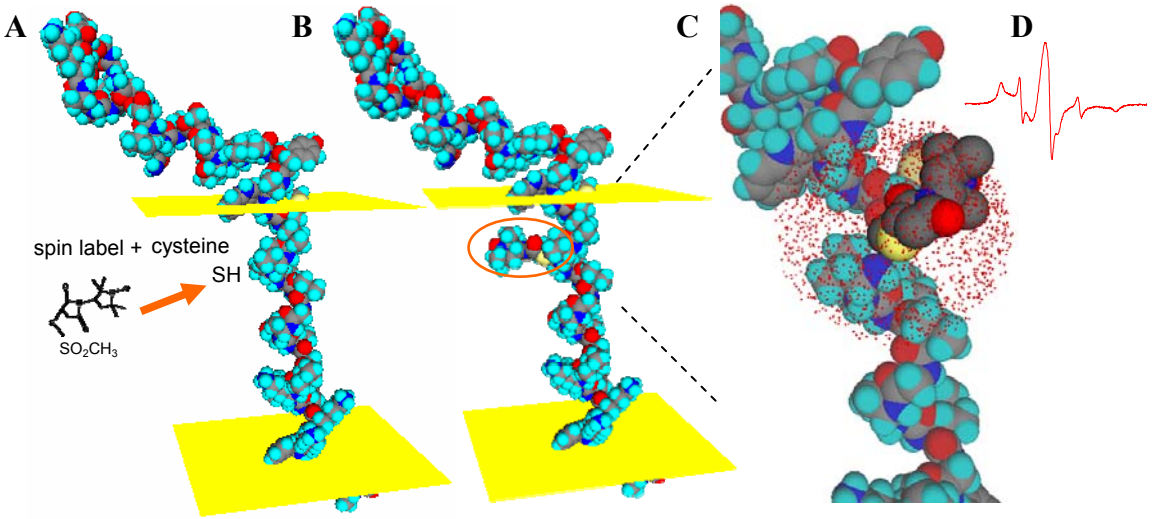

Figure 8: *Site directed spin labelling EPR spectroscopy*. Nitroxide spin label is attached to the cysteine (A), which replaces the original amino acid at the strategically chosen site (B). Anisotropy of the spin label side chain fast motion (C) is revealed in the line-shape of EPR spectrum (D). Analysis of EPR spectrum provides parameters, which describe the rate of the spin label dynamics and the conformational restrictions at local protein site. In (C) small red points represent the single spin label conformations (rotamers).

One of the principal uses of spin label EPR is to study the mobility of nitroxide-labelled molecules [68]. The dynamics, which can be detected in EPR spectrum in physiological conditions, is in picosecond-nanosecond time-scale [68,165]. The sensitivity of the conventional spin label EPR spectra to molecular motion is determined by transverse relaxation process (T_2 process) [68] and is limited by the spectral anisotropies to motions faster than $\approx 10^{-8} - 10^{-7}$ s [118]. SDSL often enhances resolution of data obtained with other methods although it provides lower resolution and more qualitative structural data compared to NMR and X-ray crystallography [46].

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 molecules [167]. Thus SDSL is a powerful tool to study the local structure in the proteins, monitor conformational changes in protein topology [71,103], and to determine backbone fluctuations at high temperature conditions [30,31,103].

In respect to structure determination, dual-probe-SDSL [46] enables also distance measurement in biological molecules [18,73]. Depending on the particular experimental method, distances from 0.4 to 8.0 nm can be accurately measured [13,150,188]. The disadvantage of this method is that distance analysis by EPR can only be done at low temperature (200 K) [63] although the attempts to estimate of inter-residue distances at physiological temperatures have been made [4]. The information about this method can be found elsewhere [13,17,77-79,155].

3.1.3 Characterization of SDSL-EPR spectra

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 techniques [177]. Such an analysis is based on mathematical modelling, spectrum fitting and parameters optimization [173,178]. 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 [173,178]. It combines solution density filtering, χ^2 goodness filtering, solution-space slicing, and group determination, leading to a graphical presentation of the system parameters (see Figure 9).

Due to protein conformational variations, conformational transitions, and the complexity of protein dynamics at time scale detectable by EPR spectroscopy the measured spectrum at a single mutant position is often a superposition of several components [178]. In general, each component of EPR spectra can be simulated on the basis of different dynamic models [154]. In order to accurately resolve spectroscopic parameters of each spectral component (so that the total multi-component simulated spectrum would fit the experimental spectrum) a good optimization method needs to be applied. The advanced multi-component spectral analysis [173,178] (see Figure 9) was applied in this thesis (it is publicly available online as EPRSIM-C: A Spectral Analysis Package [175], see also Appendix A).

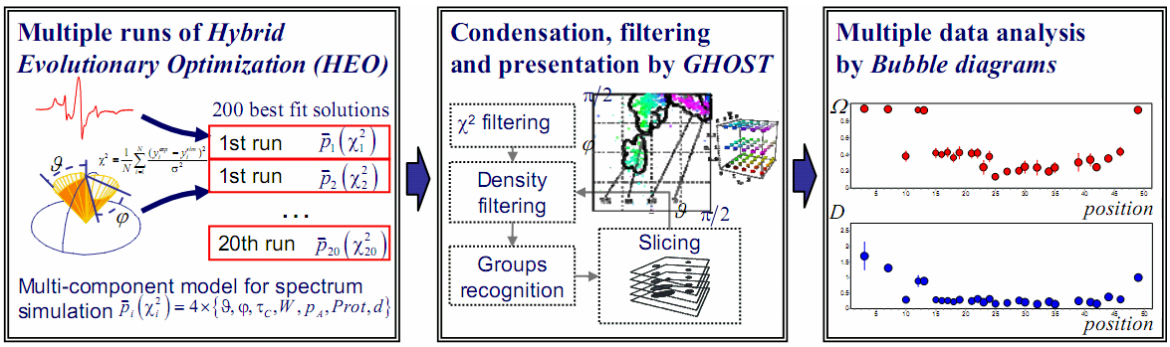

Figure 9: *Overview of the method of EPR spectral analysis*. Spectral analysis is based on multi-component spectral simulations and optimizations of spectral parameters (left), multiple solutions filtering, condensation and presentation (center), and multiple data analysis (right).

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 EPR spectra can be organized into the following steps:

1. EPR spectral simulation and spectral parameters optimization. An appropriate simulation models has to be chosen for simulation of experimental EPR spectrum. The maximal complexity, i.e. the number of spectral components, has to be defined. In addition optimization constants based on experimental parameters (EPR centre field and sweep, magnetization tensors, etc.) has to be defined as well as the initial spectral parameters. The later are then optimized with hybrid evolutionary algorithm. The result of spectral simulations the optimization of spectral parameters is a population of multiple solutions, which fit the experimental spectrum (left box in Figure 9).

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 recognized into groups of solutions (central box in Figure 9). Initially, GHOST condensation and GHOST presentation algorithms (central box in Figure 9) were developed in 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 [175] (see Appendix A).

3. Multiple data analysis. In multiple EPR data analysis (several mutant positions, temperature or concentration series) the first two steps are repeated for each spectrum of a series. However, motional patterns have to be checked against artefacts in spectral line-shape analysis, which can appear as solutions with low contribution and strange spectral parameters combination. In case the temperature measurements are available, additional check can rely on that the local temperature dependence has to be monotonous. The recent version of the GHOSTMaker (see Appendix A) is capable of simultaneous presentation of multiple EPR data (right box in Figure 9), allowing also the comparison of several spectral series and data export to other analysis software packages.

The abovementioned stages of the SL EPR spectral analysis are discussed in more details in the next sections.

3.1.3.1 EPR spectral simulation

Generally, to describe the EPR spectra of spin labels, the stochastic Liouville equation should be used [24,148,154]. However, under physiological condition the majority of the local rotational motions are fast with respect to the EPR time scale and therefore the fast motional approximation can be applied, reducing the computational demand by a factor of 100 [179].

Since the basic approach has been already discussed elsewhere [153,179], we will emphasize only the physical background of the spectral parameters involved in the calculations. Firstly, one or two parameters are used to parameterize the 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 – one order parameter S or opening cone angle ϑ (that defines the maximal tilt angle) and asymmetry cone angle φ (that describes the maximal angle allowed in long-axial-rotation). Secondly, the traces of the interaction tensors **g** and **A** are linearly corrected with p_A [115] and *Prot* [166], the parameters that take into account the effects of polarity and proticity, respectively. Thirdly, when calculating the convolution of the magnetic field distribution and basic lineshape, two linewidth parameters are additionally applied: a single (effective) rotational correlation time, τ_c, and an additional broadening

constant W. The first defines a Lorentzian-type line in the motional narrowing approximation [133], while the latter arises primarily from unresolved hydrogen superhyperfine interactions and contributions from 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 parameters ϑ, 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 resolved by the optimization routine. Taking into account the resolution limit of SL-EPR to be around 30 parameters, this allows the usage of at most 4 spectral components.

3.1.3.2 Multi-run spectrum optimization

An optimization routine is used to find the set of spectral parameters that produces the best fit to the experimental spectrum. To guide the optimization for solving an inverse problem, a common fitness function i.e., the reduced χ^2 (Eq. 1), is used. This function is calculated from the sum of the squared residuals between the experimental and simulated spectral points divided by the squared standard deviation of the experimental points and by the number of points in the experimental spectrum:

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

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 elsewhere [49], we only briefly report on the implemented algorithm.

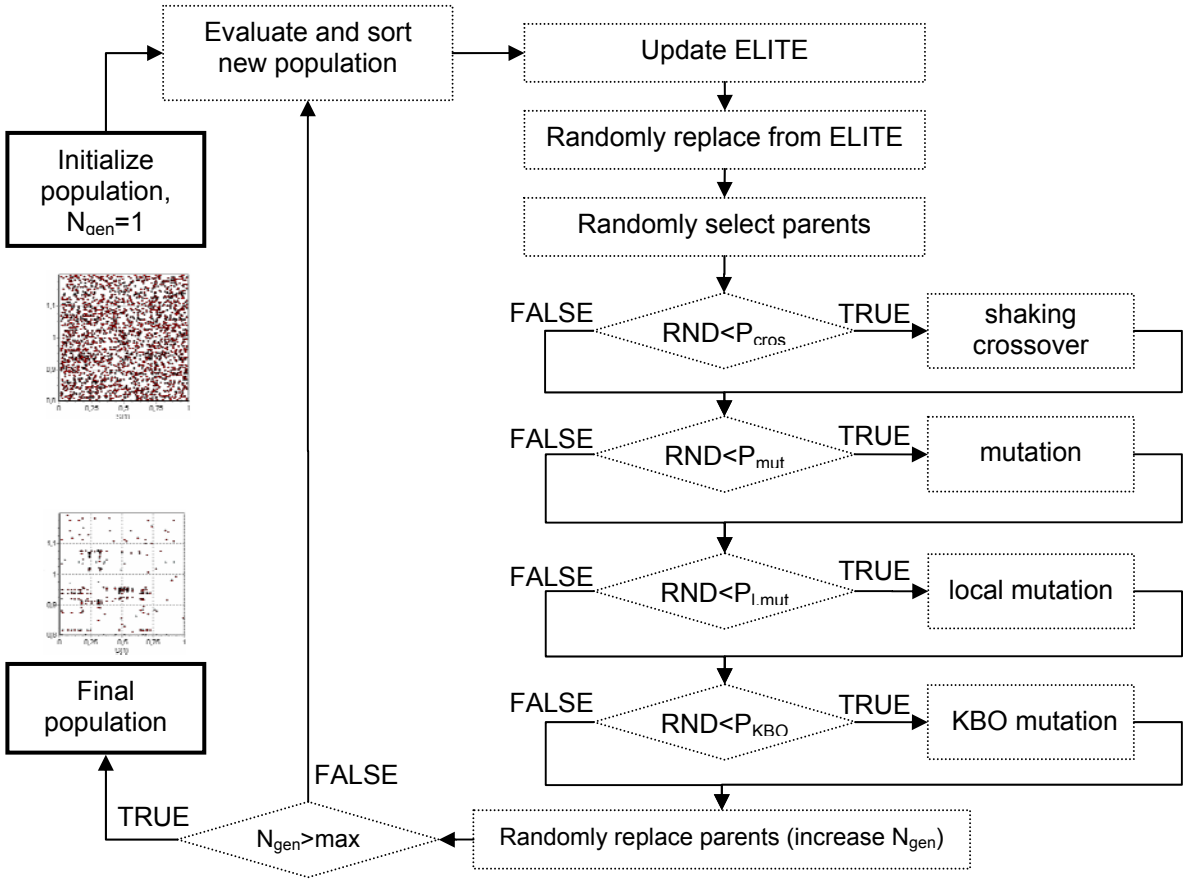

Figure 10: *The scheme of a single optimization run of the population-based HEO algorithm*. The blocks on the right part of the scheme represent the genetic operators, which modify the population of solutions (sets of simulated spectral parameters) at each generation.

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) [49,178]. 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 optimizing a population of 300 individuals. In the initial version of the algorithm, only the best parameter sets of each of 200 HEO runs were accumulated. A new version of HEO with the implemented shaking operator (shaking maintains diversity already within a single HEO run [85], see section 3.2.2.1), however, requires only 20 HEO runs. The latter is enough to accumulate the final set of 200 sets of parameters of spectral components, which are then filtered, grouped and graphically presented with a so-called GHOST condensation algorithm, described in the following section.

Efficiency of spectral simulations and optimization of spectral parameters depends a) on the quality of the measured data (high signal-to-noise ratio, absence of artefacts in the line-shape); b) on the correct choice of the spectral simulation model; c) on the correct set of the parameters determining the experimental conditions (e.g. central field, magnetic field sweep interval, magnetic properties of paramagnetic molecules in a reference environment, i.e., tensors) [173]. The most straightforward indication of successful optimization run is the low values of the fitness function χ^2 (Eq. 1) (acceptable values of χ^2 depend on S/N; in general, χ^2 should be below 10). However, another important indicators of the successful optimization procedure are the equal contribution of different runs into the final solution (measured in terms of run flatness parameter, which should be above 70%), and the absence of unusual combinations of spectroscopic parameters.

3.1.3.3 Projection principle and data 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 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 (EPR 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 [177,178].

After solution filtering according to the local solution density and goodness of fit, which is performed in accordance with [178], the GHOST condensation results are presented in 2D cross-sections $\{\vartheta - \varphi,\ \vartheta - \tau_c,\ \vartheta - W,\ \vartheta - p_A\}$ (Figure 11). This GHOST presentation technique helps to distinguish the 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.

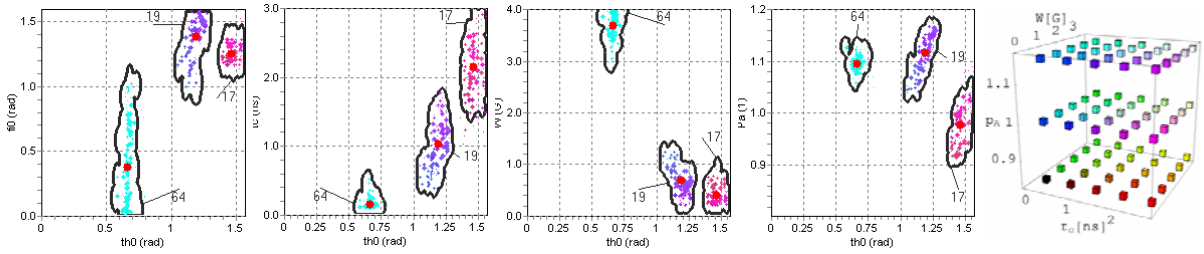

Figure 11: *Presentation of the GHOST condensation of multiple solutions*. An example of the GHOST condensed multiple solutions represented in two-dimensional distributions of the parameters ϑ and φ, ϑ and τ_c, ϑ and W, ϑ and p_A, of MTSL-SL spin labelled at position 496 N_{TAIL} protein sample in 30% sucrose at 281K in complex with XD (section 3.4.2). The components of each single solution are represented with a point on the plot with a colour, 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 ns}, {0 - 0.4 mT}, and {0.8 - 1.2}, respectively, according to colour legend (*right*). 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 average spectroscopic parameters of the detected motional patterns are presented on the plot with the red solid circles.

3.1.3.4 Multiple EPR data analysis

As a single EPR spectrum is always biased with small noise it is impossible to interpret it perfectly. Therefore a series of EPR spectra has to be measured, analyzed and interpreted jointly. For that purpose experimental series of different spin probes, different environments, different concentrations and/or different temperatures can be applied [190]. For protein structural characterization the site-directed spin labelling EPR measurements are often performed at several mutant positions [12,172]. Note that the measurements at different temperatures is the most straightforward approach to clear the artefacts in spectral analysis (see Figure 12A). Herein we present an extension of the approach of the multi-run multi-component EPR spectral simulations, which is aimed at multiple experimental data series analysis. All the possibilities of that approach are realized in GHOSTMaker software (see Appendix A).

GHOST condensation results in terms of motional patterns (see section 3.1.3.3) obtained for all the spectra in a series (Figure 12A) can be simultaneously presented in a three-dimensional plot (one dimension is presented with the size of the drawn bubbles) called bubble diagram (Figure 12B). In this diagram the vertical axis encodes one of the chosen characteristic (free rotational space Ω (Eq. 15), rotational diffusion D, order parameter S, correlation time τ_c, and etc.), and the horizontal axis encodes the parameter of the series (e.g. temperature, concentration, pH, type of spin label, mutant position). For each value of an external parameter many motional patterns can be found through GHOST condensation resulting in a vertical row of bubbles (Figure 12B), where a single bubble represents a motional pattern with the position of the bubble corresponding to the average characteristic value of that motional pattern and the bubble size being proportional to the contribution of that particular motional pattern in the total spectrum. The vertical bar at each bubble represents the second moment of a distribution of that particular motional pattern.

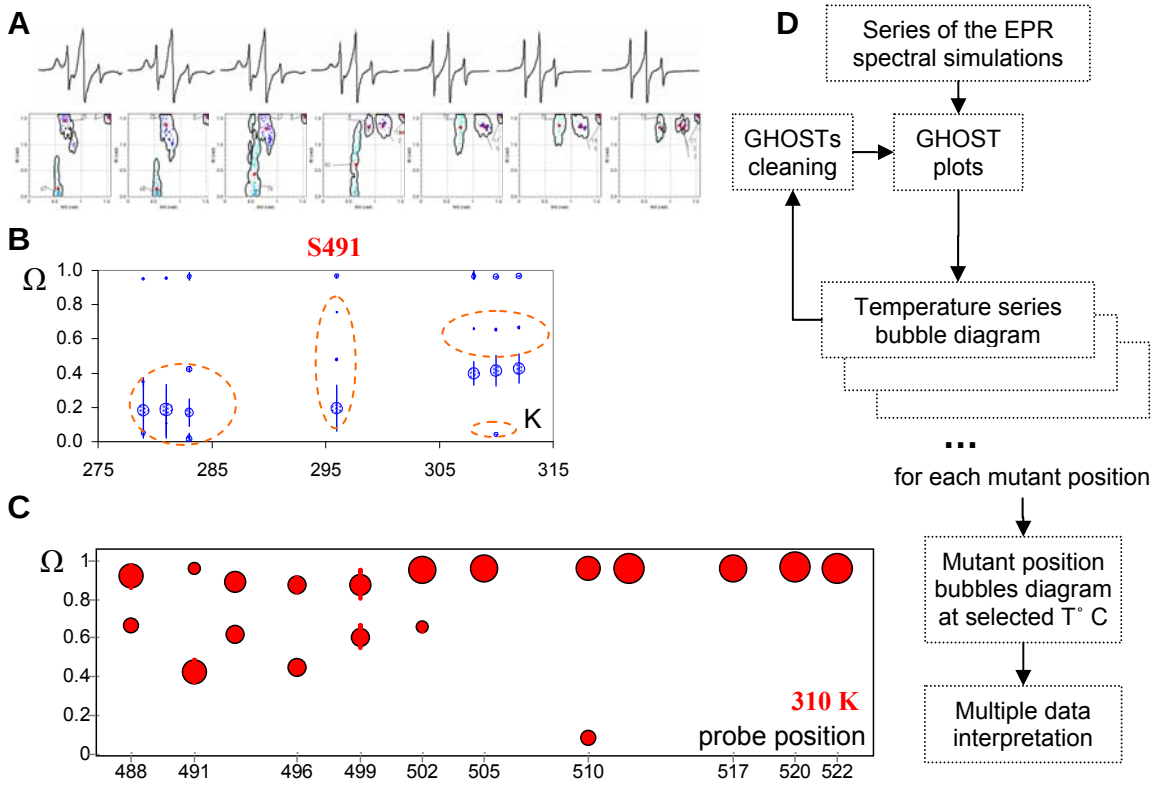

Figure 12: *Interpretation of multiple EPR data with bubble diagram*. **A.** Temperature series of SDSL-EPR spectra: experimental spectra (grey) are fitted with multi-component simulated spectra (black). The SDSL-EPR experiments for the mutant 491 of N_{TAIL}-XD complex in sucrose (see section 3.4.2) at different temperatures. Below are the corresponding series of ϑ-φ GHOST plots presenting the determined motional patterns. **B.** ϑ-φ GHOST plots condensed into the temperature-dependent Ω-bubble diagram (encircled region indicate the pattern, which need to be checked/cleaned). **C.** Temperature-dependent Ω-bubble diagram for different protein mutants condensed into a single bubble diagram for the selected temperature (310 K). In (B) and (C) each bubble represent a motional pattern, the size of bubble is proportional to the relative contribution of the particular pattern in the total spectrum. The vertical bar at each bubble represents the second moment of the motional pattern and proportional to its size in a GHOST plot. **D.** The overall scheme of multiple data analysis.

Bubble diagram allows tracking general trends in data series. It helps to determine the number of motional patterns, conformational transitions via distributed motional patterns. However, bubble diagram does not replace the GHOST plots, it is an additional presentation of the multiple data, and if presented just alone it actually reduces the information available in GHOST plots. In short, GHOST plots never have to be disregarded, but on the contrary, GHOST plots, line-shape analysis (with the Wizard application [175]), and bubble diagram have to be combined for the analysis and correct interpretation of the experimental data. Motional patterns determined with GHOST condensation (Figure 12A) are further condensed into the bubble diagram (Figure 12B). The latter enables to check and to clean the small and meaningless motional patterns (Figure 12E). The cleaning of motional patterns is based on checking the consistency of the data (e.g. the temperature dependence should be locally monotonous; in addition, the combinations of the parameters that describe motional pattern should be meaningful).

Cleaning and tuning of motional patterns has to be done while checking GHOST plots itself. By controlling the density level one has to achieve compact shape of the groups of motional patterns. If density level is decreased, i.e. fewer solutions are taken into account and the corresponding threshold density in density filtering is increased. In such a case, the border of the motional pattern group will contract towards the centre of such a group and often appearing (with the default density level) "flower" pattern will disappear guaranteeing the right density level. The last can be recognized also at the density level histogram, where the threshold density level indicator has to be positioned significantly higher than the background noise. Generally, the motional patterns with a small contribution (less than 10-20%) can be discarded from the analysis unless they represent an important recognized effect. Final criterion for accepting the small solutions can always be a simple visual check of the results of spectral simulations (the components of the simulated line-shape) – any spectral component that does not clearly describe at least one spectral feature can be disregarded.

The simulation of experimental spectra could be also repeated with the different number of spectral components, especially reduced number of spectral components in case of low S/N ratio. Note, that if a spectrum is too simple, signal-to-noise too low and the predefined complexity too high (for example using four components where there is obviously only one component), the optimization have to ignore too many parameters. The latter puts a lot of numerical stress on the GHOST condensation algorithm, which can therefore create non-existing patterns. To even further help with the analysis some components may be locked to specific parameter values in order to fit a specific feature in the spectra (like very low-populated water-soluble free probe spectral component, however, significant in amplitude).

To demonstrate the approach, the GHOST condensation results for the series of EPR spectra were collected for the same mutant position of N_{TAIL}-XD protein complex (section 3.4.2) measured at different temperatures (Figure 12C). Temperature data helped to identify the motional patterns and to condense then for further analysis and interpretation of the EPR data at different mutant positions for the selected representative temperatures (e.g. 310 K in Figure 12D).

3.2 Speeding-up SL EPR-based characterization of biosystem complexity

The most computational demanding part of SL EPR-based characterization approach was the optimization of the simulated EPR spectrum. To obtain a reliable result even in the case of quasi-continuous problems, the HEO procedure initially had to be executed at least 200 times. Each particular run consisted of 100 generations with a population size of 300 candidate solutions. At each generation after application of various genetic operators the spectrum was reevaluated. Tacking into account multiple spectrum calculations during several local optimizations, the singe HEO run was forced to do up to 150 thousand spectrum calculations. As a single spectrum optimization took ~ 40 min on a 1 GFlops processor (year 2004), the whole 200-runs optimization resulted in 130 hours of computer time. Since that was too long for a single spectrum characterization, one of the first goals was to reduce the computational time by enhancing the HEO routine.

Taking only one best parameter set from each run (see section 3.1.3.2) was 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. This provided the key idea for speeding up the spectrum optimization procedure. Accordingly, HEO was modified to increase the solution diversity within the population while preserving the same level of convergence rate. Thus, it became possible to include more than one solution into the final group of solutions and consequently rely on a smaller number of runs.

3.2.1 Parameter search space

The optimization process should be thought as searching for the minima in the landscape of the parameter search space (phase-space), which may contain both local minima and global minimum. A powerful optimization routine should be able to find global minimum(a), which can be of different types (Figure 13) – a well-defined minimum (Figure 13B) or a flat minimum valley (Figure 13A). 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 of type **A** already in a single run, *i.e.* to find distributed solutions (continuous problems).

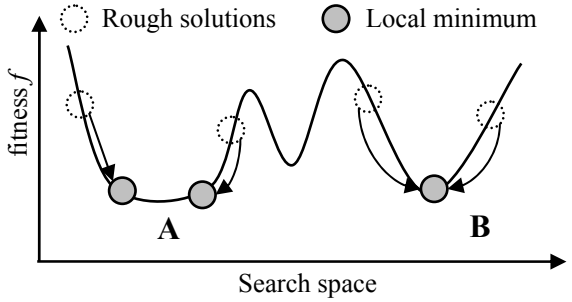

Figure 13: *Schematic presentation of parameter search space and the effect of the local mutation procedure responsible for fine-tuning*. Due to the noisy spectra and finite resolution of the local optimization routine starting approximations (white circles) are optimized into more accurate solutions (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.

3.2.2 Population diversity in genetic algorithm

A simple genetic algorithm (SGA) [54] 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 [54,110,142] 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 [55]; parallel niching methods (sharing [55], crowding [142] and clearing [142]); speciation methods [37,101,160] and clustering [180,205]; multi-population methods [191] (island models [15,56]and migration models [121]).

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 [34]. 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 looked for the method to maintain diversity within a single run together with a multiple run approach. First candidate was a sharing parallel niching method.

3.2.2.1 Maintaining population diversity: sharing and shaking operators

Sharing [55,110] requires that fitness is shared as a single resource among similar individuals in a population of solutions [109]. The fitness sharing method modifies the search landscape (Figure 14A) by changing the fitness function (Eq. 2), i.e. the value of χ^2, in densely-populated regions [152]. As a result the solutions population becomes better distributed in the search space improving the population diversity

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

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

$$sh(x) = \begin{cases} 1 - \left(\dfrac{x}{\sigma_{share}} \right)^{\alpha} & (x < \sigma_{share}) \\ 0 & otherwise \end{cases}. \quad (3)$$

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 14A). Alternative to sharing, shaking is a new operator that includes small Gaussian-like deviations to the spectral parameters (Figure 14B) before the crossover is applied. The shaking algorithm prevents "grid" formation and preserves the diversity in the solution population.

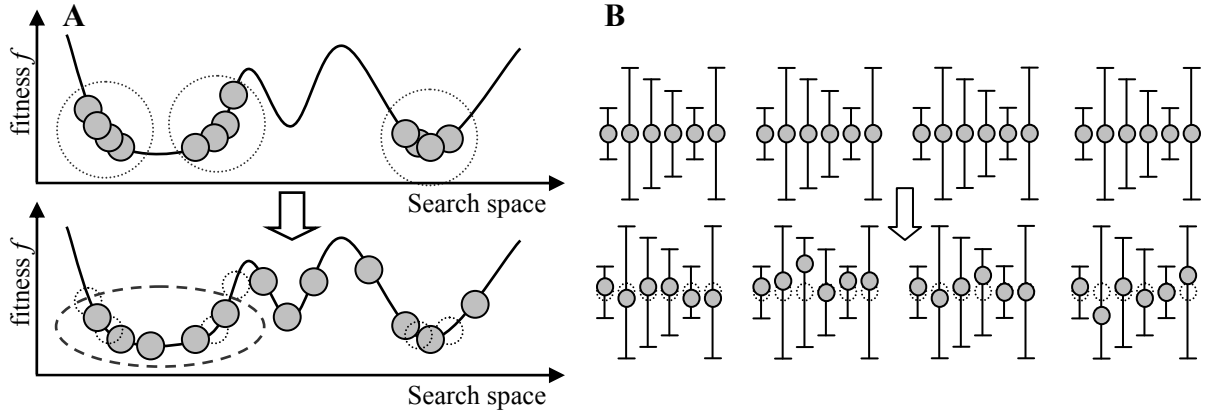

Figure 14: *Schematic presentation of the fitness sharing and Gaussian shaking operators*. **A.** Top: In a non-sharing routine crowding at the local minima is allowed, since there is no operator that would maintain diversity. Bottom: In a sharing-routine, fitness function is increased according to the density of solution, aiming to prevent crowding. **B.** 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.

3.3 Modelling protein structure and conformational space restrictions

3.3.1 Modelling approach overview

In SDSL-EPR spectroscopy a protein is labelled at a specific site with a spin label of a size slightly larger than the size of the largest amino acid residues. This makes the rotational conformational space of the spin label sensitive to the local protein structure, i.e. local backbone conformation and the conformational spaces occupied by the neighbouring amino acids, as well as by the surrounding lipids (for membrane proteins). To employ the properties of rotational conformational space for protein structure determination, the rotational conformational space has to be both accurately measured experimentally and calculated by modelling. The latter can be facilitated by the following features, specific for EPR spectroscopy:

1. The motion of protein side chains at physiological conditions (i.e., room temperature) is fast on the nanosecond EPR time scale [64]. However, if the temperature is decreased significantly, the side chain motion of the spin probe will be slowed down and/or immobilized due to stabilizing interactions [149], making it insensitive to the space restrictions imposed by the other side chains, backbone and lipids. In such a case the analysis of the rotational conformational spaces cannot be used for structure determination, and for the applicability of the approach the temperature had to be raised until the rigid-like EPR lineshape disappears.

2. The backbone motion is much slower than the nanosecond EPR time scale [83,195], or at least slower than the side chain motion. This is especially valid for proteins embedded into membranes or in large multi-chain protein complexes [182]. However, backbone atoms near the terminal ends can move to a larger extent, but such a case can be easily recognized in the EPR spectra and treated separately during the modelling.

3. The EPR experiment is insensitive to the exact atomic coordinates. On the contrary, it is very sensitive to the motional anisotropy of the nitroxide group. Therefore there is no need for a precise calculation of a conformation of an individual side chain; instead, the probability of all possible side chain conformations has to be determined. A calculation should also take into account the average effects of all the surrounding wobbling chains (the amino acid side chains of the protein(s) and alkyl chains of the lipids when present).

Overview of the method. These characteristic EPR aspects 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 modelling of the conformational space of a spin-labelled side chain and its restrictions will be based on the following steps (see Figure 15):

1. Modelling of membrane protein structure including all local rotations of the amino acid side chains and the spin-label side chain, i.e., the conformational space of the spin label;

2. Modelling of the restrictions of the spin label conformations by the protein backbone, the side chains of other neighbouring amino acid residues, and the restrictions imposed by 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 experimental SDSL-EPR data; b) optimization of the parameters of the structural model to provide a simultaneous fit of the modelled restrictions to the experimental restrictions; c) characterization of the structure of the protein and its embedment in the membrane.

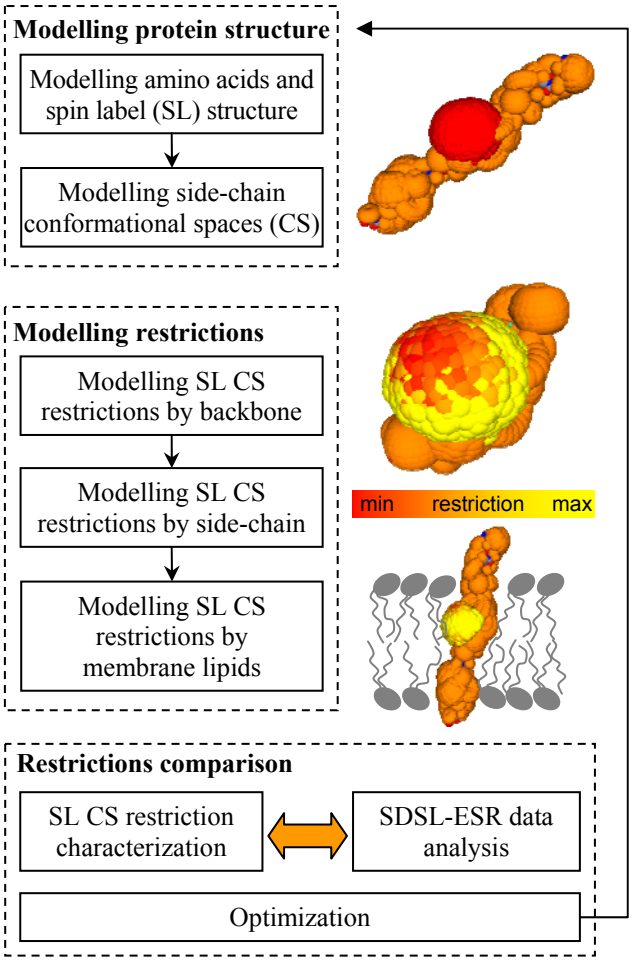

Figure 15: *Overview of the SDSL-ESR approach for protein structure determination.*

When calculating the conformational space restrictions, we will focus exclusively on the repulsive part of the van der Waals interactions that dominate close interatomic distances [23,194,195]. This approximation is mainly explained by the high-temperature condition where the entropic term becomes at least as important as the energy term of the free energy if not prevailing, and consequently the conformational spaces are restricted due to space sharing and not due to energy-minimized side chain rotameric conformations. It is therefore logical, that for simplicity, tertiary structure stabilizing interactions including sulfide bridges, hydrogen

bonding as well as helix-helix interactions due to effects of macrodipoles are not included. It should be stressed that the dynamics of the protein backbone is assumed to be much slower than the dynamics of the side chains [83] and also slow on the nanosecond EPR time scale, so that it is not need to be taken into account when simulating the spin label conformational space and modelling the local conformational restrictions at the spin label positions

3.3.2 Membrane-embedded protein structure modelling

Protein modelling. To derive the atomic coordinates, first the protein backbone structure is parameterized by 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. Since EPR spectroscopy is sensitive only to the shape of the conformational space of the spin label and not to the exact positions of the atoms of various conformations, an atomistic resolution of modelling is not required. Therefore the atomic structures of the amino acid residues and spin labels are constructed using an approximation of fixed bonds lengths and bonds angles [204], based on previously reported values [45,108,203]. The so-called Ramachandran plot [146], 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 plots [67,69,106,146]. 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, its orientation towards the membrane or other proteins and the coordinates of the amino acid side chains are stored. A relative coordinate system is used when constructing the side chain conformations [28,145,206]. The details of the protein modelling and the numerical values are presented in Appendix B.

Protein embedment in a membrane. In case of membrane protein, the embedment of protein into the lipids and the restrictive effect of lipids have to be modelled. When developing the methodology, we use the membrane-embedded M13 major coat protein, which is 50-residues long and almost α-helical [169,198]. 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 [93,131,132,170,172]. The tilt angle of the protein is derived from the effective length of the transmembrane region and the steric thickness [130] 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 [93,131,132]. The effect of the lipids is modelled as a restrictive potential along the transmembrane region as the lipids tend to restrict the side chain conformations which stick out from the protein in a perpendicular direction to the alkyl chains of the lipids.

3.3.3 Modelling of side chain conformational space restrictions

Sampling side chain conformations. Different combinations of torsion angles around single bonds of a side chain result in a set of rotamers, i.e. side chain conformations, which all together constitute the conformational space. For the conformational sampling within the conformational space, we propose a so-called Residue-Parts-Groups mechanism, which links neighbouring side chain conformations and considerably speeds-up the conformational sampling when checking the overlaps between neighbouring side chains. 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. 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. Thus, the largest parts are the aromatic rings of tryptophan, tyrosine, and phenylalanine. The glycine and proline residues, both known to be helix-breaking residues [27], are considered as an exception. The side chain of glycine is just a hydrogen atom. This makes the glycine residue very flexible to adopt most of the dihedral angles of a Ramachandran plot. Contrary, the side chain of proline has a cyclic structure, which imposes a certain conformational rigidity by locking one of the dihedral angles.

When checking 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 levels 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 neighbouring side chain, automatically all the conformations, which partially repeat the current conformation, inherit this overlap result.

Unrestricted conformational space. When modelling the conformational space of the side chains, we assume the backbone to be fixed. This approximation is based on the findings that the side chain dynamics is in the nanosecond time scale while the backbone motion is much slower, i.e., in the range of several tens or hundreds of nanoseconds [1,83], and the approximation is even more valid in case of protein surrounded by lipids or in viscose environment.

After carbon beta as a side-chain origin is fixed, the side-chain is rotated around its single bonds in order to derive all possible rotational conformations. Equidistant rotational states with an optimized grid step varying from 10 to 45° is used according to the type of an amino acid or a spin label and rotation level χ_1, χ_2 etc. (Figure 16A). A torsion potential (similar as the "three staggered potential") is implemented via fixed orientations of two subsequent bonds. Since the orientation of the second subsequent bond relative to the first one is much more poorly defined and the steric effects start to prevail [184,185], rotational conformation are restricted only via a repulsion part (hard sphere exclusion volume, in accordance with the high-temperature approximation). Note, that in this case the effective van der Waals radii are reduced standard van der Waals radii due to nonbonding electrostatic interactions in O-H, C=O and CH groups, as well as the anisotropy of C atom electron shell and non-spherical shape of the electron shell of C atoms. These enable closer contacts between the atoms as 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 [67] (see Appendix B, Table 11). Finally, conformations that have internal overlaps are eliminated.

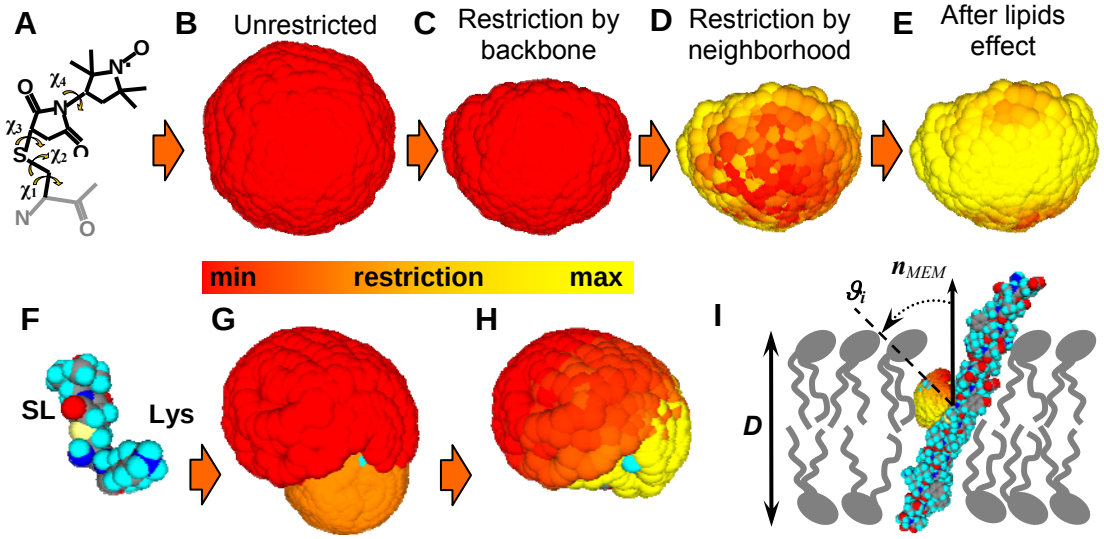

Figure 16: *Spin label conformational space restrictions modelling*. The statistical weight of the conformations is coded with the red-orange-yellow colour gradient: the red and yellow colour correspond to unrestricted and restricted conformations, respectively. **A.** Side chain free rotations (χ_1, χ_2, χ_3, χ_4) of the 3-maleimido proxyl spin label. **B.** Initial unrestricted spin label conformational space (each red sphere represents one side chain conformation; the coordinates of the sphere represent the coordinates of the oxygen atom of the nitroxide group). **C**. Spin label conformational space restricted by the protein backbone. **D.** Spin label conformational space restricted due to sharing space with neighbouring amino-acids side chains. **E.** Superposition of the restrictions due to local structure and restrictive effects of the lipids. **F.** Schematic illustration of the overlap of the conformational spaces of two neighbouring residues. Two modelled residues linked with a peptide bond: cysteine with 3-maleimido proxyl spin label attached (SL) and lysine (Lys). The backbone dihedral angles correspond to an α-helix. **G.** The overlap of the conformational spaces of the spin label (red) and the lysine residue (orange). **H.** Spin label conformational space restriction due to the overlap with the neighbouring lysine residue. **I.** Model illustrating the calculation of the restrictive effect of lipids: schematic representation of the M13 protein [171,172] labelled at position 25 and reconstituted into a phospholipid bilayer. The steric thickness of the lipid bilayer *D* includes both the hydrophobic and head group regions of the membrane. The restrictions arising from the lipids are schematically presented on the conformational space of the spin label. The colour gradient encodes the magnitude of the restriction effect, which depends on the orientation ϑ_i of each particular *i*-th conformation relative to the membrane normal n_{MEM}: The yellow colour corresponds to maximum restrictions; red to conformations that are not restricted.

In general, a too large number of conformations arises in rotational conformational space creation, becoming a computational bottleneck in determining the restrictions due to overlap between neighbouring side chains. In the course of our work, we found a compromise between modelling accuracy and computational costs by adjusting the rotational step resulting in approximately 3000 allowed 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 these different conformations in the unrestricted conformational space (Figure 16B) is equal. Lists of allowed conformations, i.e., unrestricted conformational spaces, calculated for each residue of the protein are stored in memory and then used later for the calculation of the restrictions.

Restricted conformational space. The restriction of the conformational space of the spin label side chain is calculated by checking the restrictions for each particular conformation due to: a) overlap with the backbone (Figure 16C); b) overlap with side chains of neighbouring amino acid residues (Figure 16D); and c) restrictions imposed by the surrounding membrane lipids that tend to suppress conformations perpendicular to the lipids alkyl chains, i.e., perpendicular to the membrane normal (Figure 16E).

In the first step the overlap with the backbone is determined. Sequentially we go through the conformational space of the spin label and check each individual conformation for overlap with the backbone. Since the backbone motion is assumed to be much slower than the side chain motions, the statistical weight of a conformation that overlaps with the backbone is diminishes, i.e., all such conformations are forbidden (removed conformations in Figure 16C).

The conformations that do not overlap with the backbone could be still restricted by the side chains of neighbouring amino acid residues. The statistical weight of the allowed spin label conformations 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 (Figure 16F-H). The extent of overlap depends on the relative position and orientation 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 neighbouring *k*-th residue is reduced by a factor:

$$F_k^i = \frac{N_k^{all} - N_k^{overlaps}}{N_k^{all}}, \tag{4}$$

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

A chosen conformation of the side chain of a spin label overlaps with many conformations of other neighbouring side chains, and each of these conformations may spend a certain time in the space, which is occupied by a chosen spin label conformation. If there is more than one overlapping neighbouring side chain, the overlaps can happen only in different fractions of time. Therefore, the probabilities for each of the overlapping pairs of conformations should be factorized. As the side chains wobble fast, the overlaps cannot occur over the full (EPR averaging) time period and thus as a result the statistical weight of a particular spin label conformation will be reduced to a value between 0 and initial value $P^{initial}$, depending on the extent of overlap. Thus the combined restriction of the conformational space of the spin label from the neighbouring residues is a product of factors:

$$P_i = P_i^{initial} \prod_k^n F_k^i \ , \tag{5}$$

where P_i and $P_i^{initial}$ are the statistical weights (probabilities) of the restricted and unrestricted *i*-th side chain conformation. The factors F_k^i are given by Eq. 4. Note, that this part of calculations is not limited just to side chains of the same protein chain, as two side chains of two different proteins could be in the close contact in the folded complex, and thus could restrict each other.

Finally, if the spin-labelled protein site is in a transmembrane region the statistical weight of the spin label conformations is further reduced. The side chains of the amino acid residues including the spin label that are surrounded by the lipid environment will feel the fluctuating lipid alkyl chains as well as the restrictive effect of lipid head groups. Since the latter expose lateral pressure [117] and occupy certain amount of space, the conformational spaces of side chains will be further restricted. However, for a spin label at each single mutant position we assume that the amplitude of lipid effect is constant for all conformations and only the direction of the conformation relative to membrane normal makes a difference in the restrictive effect of the lipids. 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 ns [53] and that the lipid-protein interactions are just slightly more favourable than lipid-lipid interactions [98]. Recent molecular dynamics simulations show that side chains from aromatic, polar and charged amino acid residues tend to orient along the membrane normal [80], supporting our model. With this in mind, the simplest approximation of the lipid effect should take into account the following issues [176]: 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 (Figure 16I); b) there are no restrictions in case of a parallel alignment to the membrane normal as the disruption to lipid packing is minimized; c) the effect of the lipids should be effective as soon as there is any non-zero angle ϑ 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 the 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 $\pi/2$; and e) the occupancy of the space around a particular conformation is reflected by the extent of reduction of the statistical weight of the *i*-th conformation due to sharing space with the neighbouring amino acid side chains. In accordance with all these constraints a simple empirical $(1 - \sin \vartheta)$ function is used, so that the statistical weight P_i of the *i*-th conformation is modified in the following way:

$$P_i = P_i\left(1-\sin\vartheta_i\right), \tag{6}$$

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).

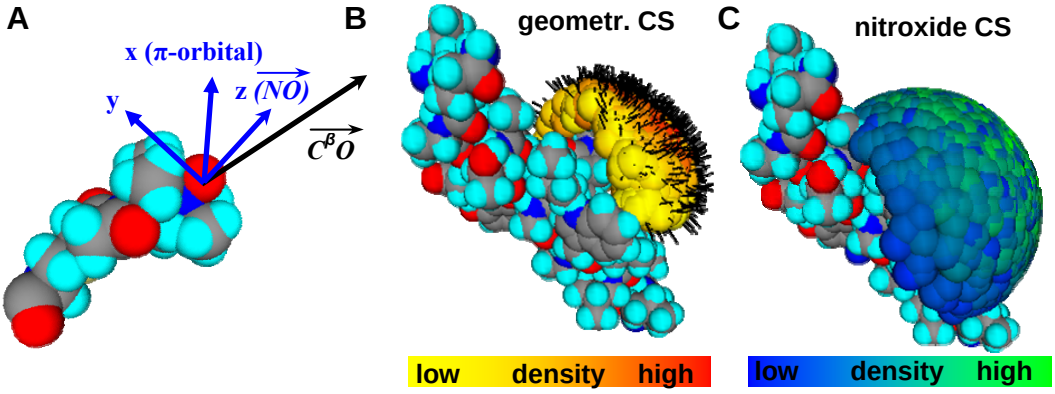

Figure 17: *Relation between the geometrical and nitroxide NO conformational spaces*. **A**. NO group axis, NO bond direction and $C^{\beta}O$ directions of the spin label side chain conformation. **B**. Unrestricted initial geometrical and nitroxide conformational spaces. Small spheres represent single conformations. For the nitroxide conformational space sphere coordinates represents NO vector direction. The colours of the conformations encode the local conformational density. **C**. Restricted geometrical and nitroxide conformational spaces. NO vectors at different conformation are represented with small black direction lines. The spin label is attached to the M13 protein at position 25. Only the amino acid residues that contribute to the restriction of the conformational space of the spin label are presented.

The backbone restrictions, superposition of the neighbouring amino acids restrictions, and lipid effect on a single *i*-th side chain conformation can be summarized [176] as:

$$P_i = P_i^{initial}\begin{Bmatrix}0, \text{ backbone overlap}\\ 1, \text{ no backbone overlap}\end{Bmatrix}\prod_k^n\left(1-\frac{N_k^{overlaps}}{N_k^{all}}\right)\left(1-\sin\vartheta_i\right). \tag{7}$$

Characterization of spin label conformational space restrictions. EPR 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 EPR spectrum. In general the orientation of the NO group does not coincide with the geometrical orientation of the side chain of the spin label (Figure 17A). 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. Note, that conformational space restrictions have to be calculated for geometrically defined conformations (Figure 17B), while the calculation of the restrictions for the conformational space is based on nitroxide group orientations (Figure 17C).

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 EPR spectra [85,172,178]. The cone model is parameterized with the angles ϑ_0 and φ_0 (Figure 18A), 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 the EPR spectral analysis, are connected with the directional averages $\overline{\cos^2(\vartheta)}$ and $\overline{\sin^2(\varphi)}$:

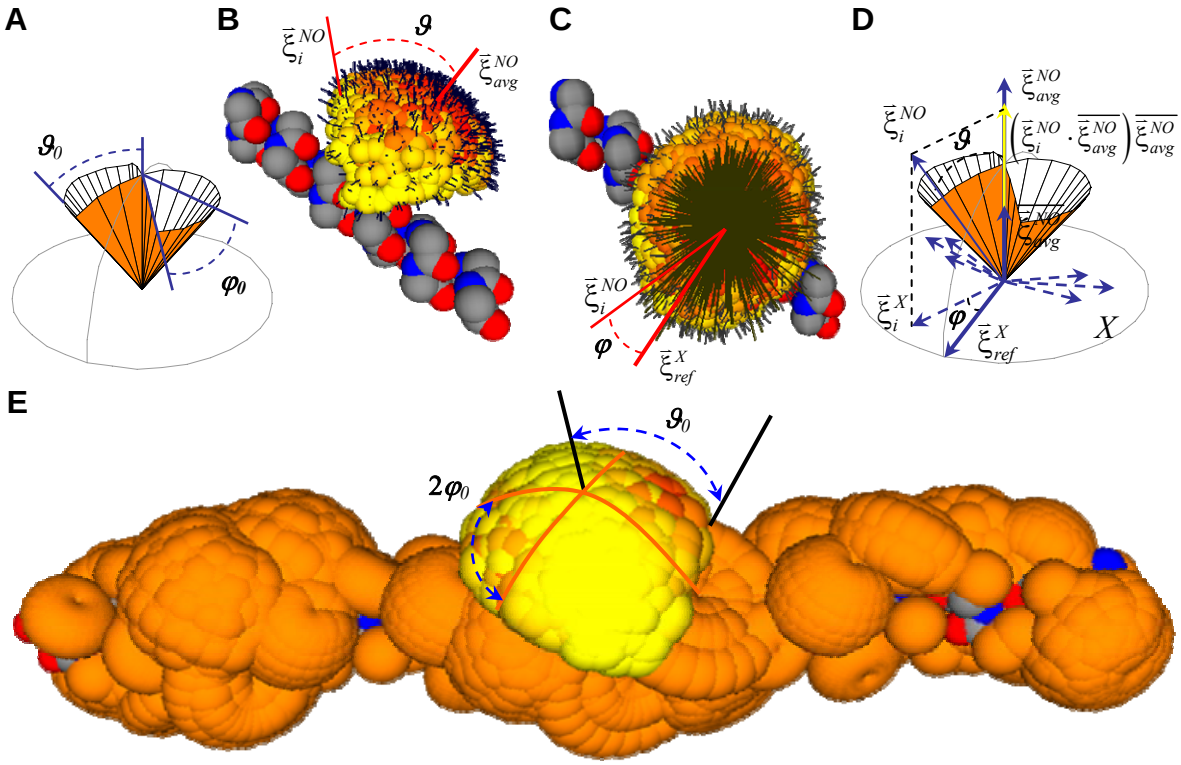

Figure 18: *Characterization of spin label conformational space restriction*. **A**. Definition of the opening ϑ_0 and the asymmetry φ_0 angles of the cone model. **B.** Conformational space opening characterized by the average $\overline{\cos^2(\vartheta)}$. The angle ϑ is between the NO bond direction $\vec{\xi}_i^{NO}$ in the *i*-th conformation and the average direction of the NO bond $\vec{\xi}_{avg}^{NO}$. **C**. Conformational space asymmetry characterized by $\overline{\sin^2(\varphi)}$. The angle φ is between the X direction, $\vec{\xi}_i^{X}$ (a projection of $\vec{\xi}_i^{NO}$ on a plane perpendicular to the average NO direction $\vec{\xi}_{avg}^{NO}$) of the *i*-th conformation and a reference X direction $\vec{\xi}_{ref}^{X}$ (direction that corresponds to the highest radial density of the $\vec{\xi}_i^{NO}$ directions). **D.** Definition of the vectors for characterization of the conformational space asymmetry φ. **E.** Characterization of the conformational space of spin label restricted by protein backbone, due to sharing the space with the neighbouring amino acids, and due to perpendicular restrictive effect of lipids (the yellow colour corresponds to maximum restrictions; red to conformations that are not restricted).

$$\overline{\cos^2(\vartheta)} = \frac{1}{3}\left[\cos^2(\vartheta_0) + \cos(\vartheta_0) + 1\right] \tag{8}$$

$$\overline{\sin^2(\varphi)} = 1 - \overline{\cos^2(\varphi)} = \frac{1}{2}\left(1 - \frac{\sin(2\varphi_0)}{2\varphi_0}\right) \tag{9}$$

The averages $\overline{\cos^2(\vartheta)}$ and $\overline{\sin^2(\varphi)}$ can be calculated numerically from the modelled restriction of the conformational space of the spin label (Figure 18B-C). 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:

$$\overline{\cos^2(\vartheta)} = \frac{\sum_i^N \left(\vec{\xi}_i^{NO} \cdot \overline{\vec{\xi}_{avg}^{NO}}\right)^2 p_i \rho_i}{\sum_i^N p_i \rho_i}, \tag{10}$$

where $\overline{\vec{\xi}_i^{NO}}$ is the normalized nitroxide NO bond direction of the *i*-th conformation, $\overline{\vec{\xi}_{avg}^{NO}}$ 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 (see Figure 18D). The average $\overline{\sin^2(\varphi)}$ is calculated by:

$$\vec{\xi}_i^{\,X} = \vec{\xi}_i^{\,NO} - \overline{\vec{\xi}_{avg}^{\,NO}}\cos(\vartheta)\left|\vec{\xi}_i^{\,NO}\right| = \vec{\xi}_i^{\,NO} - \overline{\vec{\xi}_{avg}^{\,NO}}\frac{\vec{\xi}_i^{\,NO}\cdot\overline{\vec{\xi}_{avg}^{\,NO}}}{\left|\vec{\xi}_i^{\,NO}\right|}\left|\vec{\xi}_i^{\,NO}\right| = \vec{\xi}_i^{\,NO} - \overline{\vec{\xi}_{avg}^{\,NO}}\left(\vec{\xi}_i^{\,NO}\cdot\overline{\vec{\xi}_{avg}^{\,NO}}\right), \tag{11}$$

$$F_i = p_i\rho\sin(\vartheta) = p_i\rho_i\sqrt{1-\left(\overline{\vec{\xi}_i^{\,NO}}\cdot\overline{\vec{\xi}_{avg}^{\,NO}}\right)^2}\,, \tag{12}$$

$$\sin^2(\varphi_i) = 1-\left(\overline{\vec{\xi}_i^{\,X}}\cdot\overline{\vec{\xi}_{ref}^{\,X}}\right)^2, \tag{13}$$

$$\overline{\sin^2(\varphi)} = \frac{\sum_i^N \sin^2(\varphi_i)F_i}{\sum_i^N F_i} = \frac{\sum_i^N\left(1-\left(\overline{\vec{\xi}_i^{\,X}}\cdot\overline{\vec{\xi}_{ref}^{\,X}}\right)^2\right)p_i\rho_i\sqrt{1-\left(\overline{\vec{\xi}_i^{\,NO}}\cdot\overline{\vec{\xi}_{avg}^{\,NO}}\right)^2}}{\sum_i^N p_i\rho_i\sqrt{1-\left(\overline{\vec{\xi}_i^{\,NO}}\cdot\overline{\vec{\xi}_{avg}^{\,NO}}\right)^2}}, \tag{14}$$

where $\vec{\xi}_i^{\,X}$ is a projection of $\overline{\vec{\xi}_i^{\,NO}}$ on a plane perpendicular to $\overline{\vec{\xi}_{avg}^{\,NO}}$ (Figure 18D); $\overline{\vec{\xi}_{ref}^{\,X}}$ is a normalized reference direction for the calculation of the asymmetry of the conformational space that corresponds to the highest radial density of $\vec{\xi}_i^{\,X}$ directions (Figure 18D).

Both angles ϑ_0 and φ_0 can be combined into the so-called normalized free rotational space [172]:

$$\Omega = \frac{\vartheta_0\varphi_0}{(\pi/2)^2}, \tag{15}$$

which is compared to the normalized free rotational space values extracted from SDSL-EPR experimental data [172,176]. The results of testing of the sensitivity of the normalized free rotational space are presented and discussed in Section 4.2.

3.4 Materials

The M13 coat protein was used to test our new method of calculation the restrictions of the conformational spaces and to develop a new protein structure optimization approach. In addition, we apply our approach to translate multiple SDSL EPR data into structural characterization of the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (N_{TAIL}, aa 401-525) alone or in a complex with and the C-terminal X domain (XD, aa 459-507) of the phosphoprotein (Section 3.5.3 and 4.4).

3.4.1 Major coat protein of bacteriophage M13

Bacteriophage M13 is thoroughly studied by various biophysical techniques, the structure of the virion protein sheet was determined by X-ray fibre 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 19) [111,170]. The detailed description of the filamentous bacteriophages can be found in the reviews [170]. In addition, extensive knowledge about bacteriophage M13 and recent technological advances lead to successful application of M13 virus in Nanotechnology [99,112].

The major coat protein is a small protein with molecular weight of about 5240 Da. It forms a 1.5-2.0 nm thick flexible cylindrical shell (see Figure 19B). It is 50 amino acids long and it is composed of three specific domains: a hydrophobic core, an acidic N-terminal and basic C-terminal domains. Secondary structure is largely α-helical (also proven by primary sequence prediction) with several flexible positions in the N terminus [170].

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 [122], NMR spectroscopy [134,138], site-directed labelling in combination with EPR

[124,125,163,169-172,176] and fluorescence spectroscopy [48,93,131,132,197], circular dichroism spectroscopy [151] and other methods.

According to X-ray crystallography the protein is slightly curved α-helix extending from the N-terminus to the C-terminus [122]. Studies based on NMR spectroscopy [3,138,139] suggest that in micelles the protein contains two α-helical segments, residues 7-16 and 25-45 (see Figure 19). NMR spectroscopy in dehydrated lipid bilayer [113] 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, where the helix tilt changes to 16°, which was also observed previously by solid-state NMR [33].

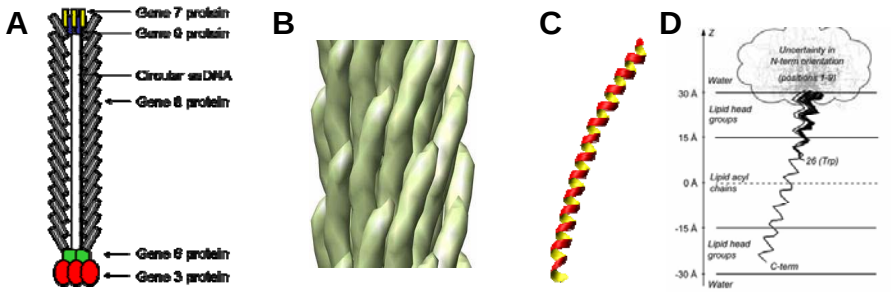

Figure 19: *The membrane coat protein of Bacteriophage M13*. **A** Schematic illustration of a bacteriophage M13 filament [170]. **B**. Arrangement of the coat protein subunits in M13 [123]. **C**. Curved α-helix model of the M13 coat protein found by X-ray crystallography [122] deposited in the Protein Data Bank as PDB file 1IFJ. **D**. FRET data-based 100 best-fit structures of AEDANS-labelled M13 coat protein in DOPC/DOPG vesicles [132].

The results obtained with SDSL-EPR spectroscopy for the protein in different lipid bilayers (14:1PC-22:1PC) [124,125,171,172] were generally in accordance with two α-helical segments model, while the modelling based on the recent fluorescence data obtained for the protein in DOPC/DOPG vesicles suggested a model of two α-helical domains with unstructured region (residues 1-9) and general tilt (residues 12-46) by 18° relative to membrane normal. The state of art regarding the model for M13 protein structure can be found in the recent publication [198].

3.4.2 Intrinsically disordered C-terminal domain of the measles virus nucleoprotein

Intrinsically disordered proteins (IDPs) are functional proteins that do not fold into well-defined, unique three-dimensional structures under physiological conditions [40,42,50,186,192]. Although there are IDPs that carry out their function while remaining permanently disordered, many of them undergo induced folding, i.e. a disorder-to-order transition upon binding to their physiological partners.

Measles virus (MV) belongs to the Paramyxoviridae family within the Mononegavirales order. This order includes several human pathogens with a strong socio-economical impact and comprises both well characterized viruses (as for instance mumps, parainfluenza, rabies and Ebola viruses) and emerging viruses, such as the Nipah and Hendra viruses that are responsible for encephalitis with a high (>50%) case-fatality rate. With approximately 800,000 deaths worldwide, measles ranks 8th as the cause of worldwide mortality and represents the main cause of childhood mortality in developing countries. Despite extensive vaccination campaigns, the disease has not been eradicated yet. Furthermore, outbreaks occur even within vaccinated populations. To date, no effective antiviral treatment exists [20].

MV's non-segmented, negative-sense, single-stranded RNA genome is encapsidated by the viral nucleoprotein (N) within a helical nucleocapsid. This latter is the substrate used by the viral polymerase complex during transcription and replication. The viral polymerase complex consists of the large protein (L) and of the phosphoprotein (P) that is an essential polymerase co-factor as it tethers the L protein onto the nucleocapsid template [128].

The MV nucleoprotein consists of two regions: a structured N-terminal moiety, NCORE (aa 1-400), which contains all the regions necessary for self-assembly and RNA-binding, and a C-terminal domain, N_{TAIL} (aa 401-525) that is intrinsically unstructured [105] and exposed at the surface of the viral nucleocapsid [62,82]. Due to its intrinsic flexible nature N_{TAIL} interacts with various partners, including host cell proteins and the P protein. The P protein is an essential subunit of the viral polymerase complex. It

has a modular organization and its C-terminal X domain (XD, aa 459-507) is responsible for binding to N_{TAIL}. Within a conserved region of N_{TAIL} (aa 489-506, Box2), an α-helical molecular recognition element (α-MoRE, aa 489-499) is involved in binding to P and in induced folding [21,81].

The current picture of the N_{TAIL}-XD complex, that is schematically presented in Figure 20, is based on the study of the crystal structure of a chimeric construct consisting of XD and of the 486-504 region of N_{TAIL} [91] as well as SDSL EPR experimental study [128] and on the small angle X-ray scattering study providing a low-resolution structural model of the complex between XD and the entire N_{TAIL} domain [22]. The later two show that the N_{TAIL} region upstream Box2 (residues 401-488) remains disordered in the complex and does not establish contacts with XD. On the other side, they also indicate that Box2 and probably also Box3 of the C-terminal region of N_{TAIL} are involved in binding to XD. Despite all the studies above and additional characterizations gained by CD spectroscopy, surface plasmon resonance and NMR spectroscopy [19,20,22,128] the structural characterization of the C-terminal domain of the measles virus (MV) nucleoprotein in its native water environment at high temperature is still not resolved.

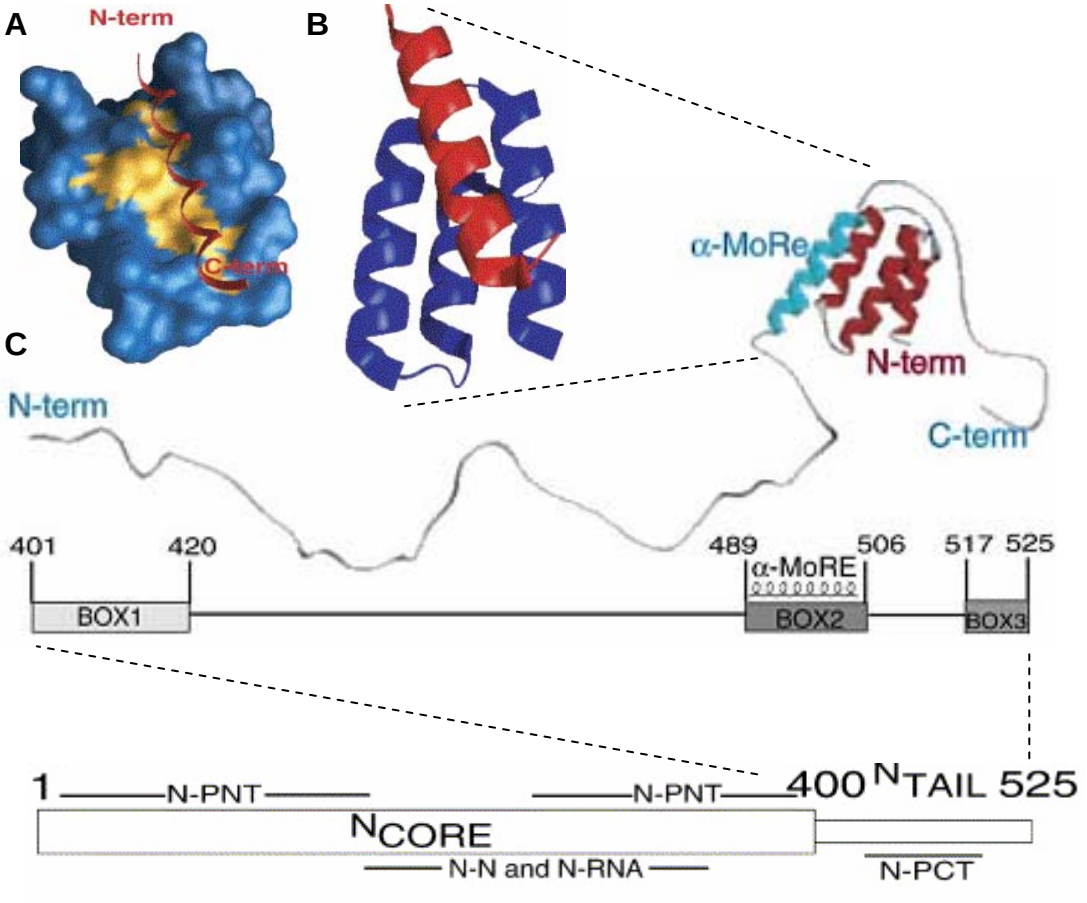

Figure 20: *Model of the N_{TAIL}-XD complex.* The predicted (A) structure was then experimentally observed (B) in the chimeric construct (encompassing residues 486-504, pdb code 1T60) as derived by small-angle X-ray scattering studies, highlighting the involvement of the α-MoRE and of the C-terminus of N_{TAIL} (Box3) in the interaction with XD [22]. The three regions of homology conserved in *Morbillivirus* members [39], namely Box1, Box2 and Box3, are highlighted (C).

3.5 Optimization of protein structure

The goal of our work is to use the series of free rotational spaces that are experimentally obtained for a membrane protein spin-labelled along its primary sequence, as constraints in optimizing its three-dimensional structure and membrane-embedment. 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 (or relative to the protein complex core), so that the calculated local restrictions fit the characteristics extracted from the experimental EPR data.

The optimization module is based on a stochastic algorithm of the Metropolis Monte Carlo family [92] with several elements of the evolutionary algorithm (mutation operator, replacement operator and elite) [44,52] (see the scheme of a single optimization run in Figure 22). However, unlike conventional evolutionary algorithm, each optimization run tunes just a single structure. One run counts for 200 generations. At each generation the current structure of the protein is modified by mutation (modification of backbone dihedral angles), crossover (introduction of the secondary structure motifs of successful structures achieved in previous generations and stored in elite) and shaking (orientation shaking, long axis rotation and position shifting) operators. For each new structure simulated restrictions are compared with the experimental SDSL-EPR data-based restriction profile. The quality of the fit at each generation is evaluated by the goodness of fit as follows:

$$\chi^2 = \frac{1}{N}\sum_i \left(\frac{\Omega_{exp,i} - \Omega_{sim,i}}{\Delta\Omega_{exp,i}} \right)^2, \tag{16}$$

where N is the number of spin-labelled 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.

Table 2: *Parameters of the algorithm for protein structure optimization*.

Parameter	Value	Description
NRuns	1000 runs	Number of optimization runs
Ngen	200 gen	Number of runs generations
sigVariationRange		Number of flexible molecules
Clashes		
MaxClashes	1000	Maximum number of allowed subsequent clashes in a single generation
Shaking dihedrals		
sigmaDih	5˚	Minimal shaking amplitude of backbone dihedral angles
sigmaDihLimit	20˚	Maximal shaking amplitude of backbone dihedral angles
DihSigArray	5-20˚	Array of backbone dihedral angles shaking amplitudes
Corossover		
CrossoverThreshold	0.05-0.3	Elitist crossover probability
Shaking chain		
sigmaCoord	2Å	Protein chain position shaking amplitude
sigmaBase	0.01rad	Protein chain direction shaking amplitude
sigmaRot	10˚	Long axis rotation amplitude
Membrane		
sigmaMemThickness	0.3 Å	Bilayer steric thickness shaking amplitude
sigmaTransMemN	3 res	Transmembrane region shaking amplitude at N-terminal end
sigmaTransMemC	1 res	Transmembrane region shaking amplitude at C-terminal end
Metropolis		
kT_threshold	-	Metropolis temperature and threshold criteria
Best replace		
NUnsucc	10-30 gen	Number of subsequent unsuccessful generations
Fine Tuning		
NFineTuneBest	3-15 gen	Number of generations for the fine-tuning after the best structure is found
NFineTuneCur	3-10 gen	Number of generations for the fine-tuning of the current structure

The goodness of fit χ^2 guides the optimization routine determining and influence acceptance of the new structures for the next generation of structural evolution. Each time a new best-fit structure found, the algorithm turns on the fine-tuning mode, which tries to do local optimization of the best-fit structure to even more improve the fit.

The settings of the optimization algorithm common for any structural optimization task are listed in Table 2. Many of the algorithm operators (backbone dihedral angles shaking amplitudes, elitist crossover, metropolis criterion, replacement by the best and fine tuning) depend on generation number.

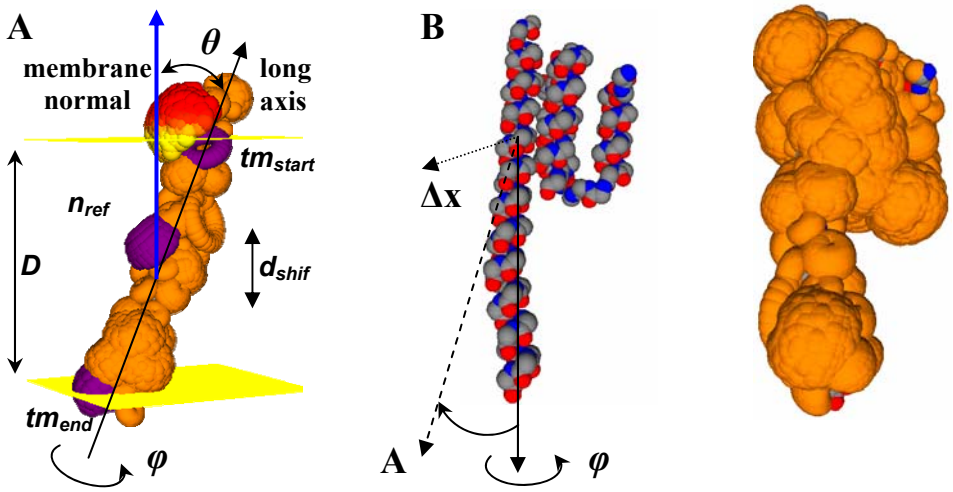

Figure 21: *Parameters for the protein structure optimization*. **A**. Relative position and orientation of membrane-embedded M13 coat protein [84] parameterised according to Table 3. The protein is shown with the conformational spaces of the amino acid side chains and spin label. The starting, tm_{start} 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**. N_{TAIL} protein is parameterized relative to the partner protein XD according to Table 4 and presented by backbone atoms (left) and conformational spaces (right).

After multiple runs of optimization, many final structures will have the similar goodness of fit, given by different Ω profiles, providing a family of structures. 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 [9,26,137].

The main structural parameters that are being optimized are the pairs of the dihedral angles $\{\varphi_i, \psi_i\}$. In addition depending on the specific protein optimization task there are parameters describing the positions and orientation of protein chains in the membrane (in case of membrane proteins) or in relationship with one another (in protein complexes). When optimizing the protein structure in a membrane, the parameters that describe the protein-lipid arrangement (see Figure 21A) are tuned simultaneously with the dihedral angles of the protein. In addition, in case of membrane proteins, the steric thickness of the bilayer [130] has to match the protein tilt angle and transmembrane length. Furthermore, 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. In case of optimization of a protein chain in a protein complex, the parameters that describe position and orientation of the chain in relation to the complex (core) are: relative chain displacement, relative orientation, long axis rotation (mainly for helical chains) (see Figure 21B).

3.5.1 How single optimization run works

The common optimization algorithm scheme is presented in Figure 22. In the beginning of the optimization run an initial protein structure is generated. In addition, optimization algorithm requires the following: a) initialization of the Ramachandran plot, which is used in mutation operator (changing the backbone dihedral angles) and which helps to reduce the search space; b) setting of the mutant positions; c) initialization of the experimental data (experimentally obtained local restrictions in terms of rotational space, Ω and the rotational diffusion, *D*); d) modelling of the lipid bilayer, in case of membrane proteins; e) setting the initial values for the optimization parameters presented in Table 2.

The first generation starts with calculation of the restrictions. At each next generation, before restrictions calculation is done (orange box in Figure 22), the current structure of a protein system is being modified by *internal* (modification of the secondary structure of a protein via mutation and crossover operator applied to the

backbone dihedral angles of a protein, see red box in Figure 22) and *external* (modification of the position and orientation of the protein towards the membrane or other protein, see the yellow box in Figure 22) operators. First of all, the optimization parameters (Table 2) and the backbone dihedral angles amplitudes are updated (unless the parameters dependence on generation is turned off). Second, the mutation operator (fine-tuning or larger structural shaking depending on algorithm stage) is applied to modify the secondary structure of the protein. Third, the elitist crossover is applied (for each structure with calculated restrictions the subsequences with the corresponding goodness of fit are stored in the Elite database and used later for elitist crossover).

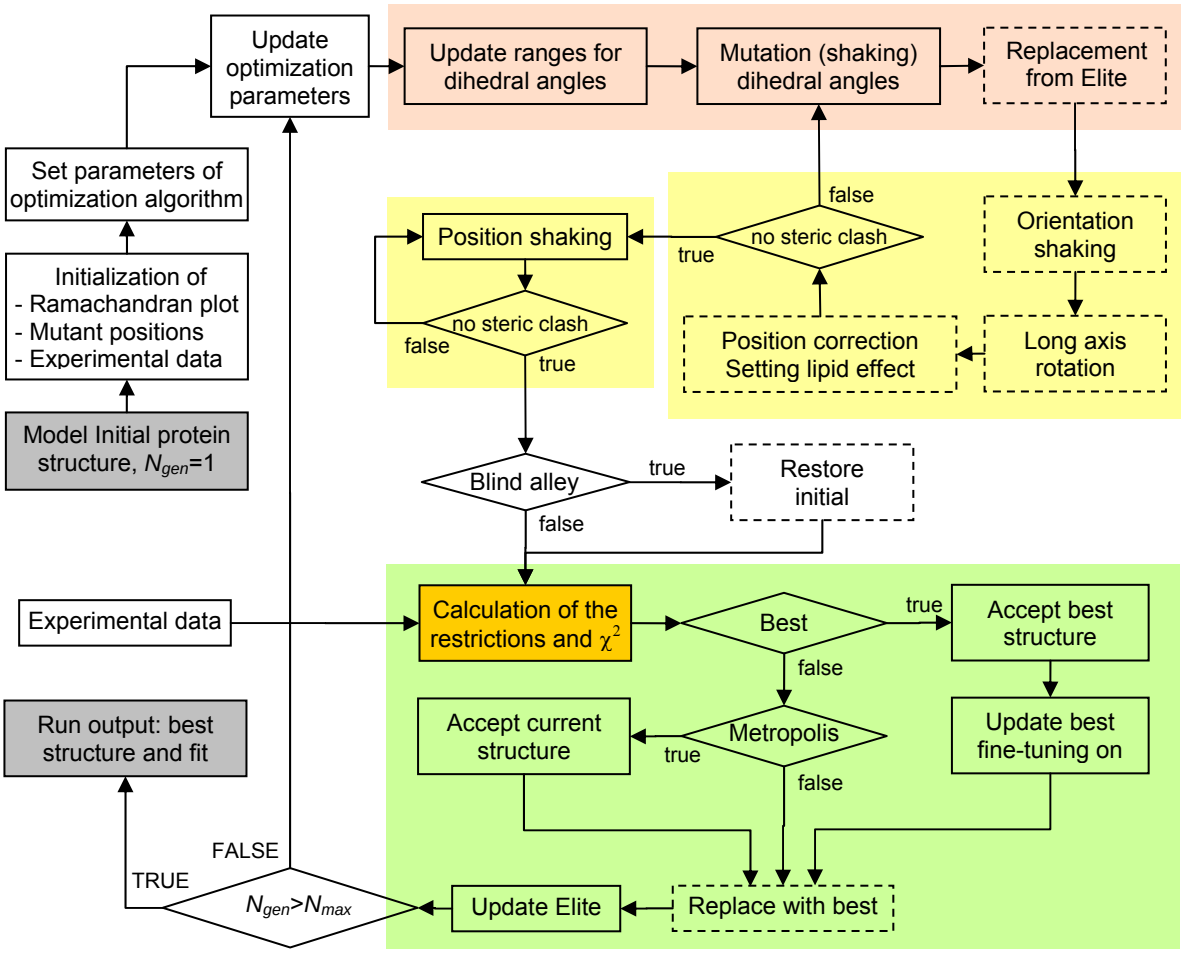

Figure 22: *The scheme of the structure optimization algorithm*. Dashed boxes represent the optional operators.

External operators modify orientation of the protein towards the membrane (in case of membrane proteins) or towards the other protein if this is the case of protein complex. This also includes the rotation of the protein around its long axis (relevant for helical chains), setting of the lipid effect or any external restrictive effect from the environment, and modification of the position of the protein towards the membrane or other protein. At several stages, before any next operator is applied, the protein system structure is checked for steric clashes. The latter relies on the original van der Waals atom radii data listed in Table 11 in Appendix B. Algorithm continues only if there no internal overlap, otherwise it returns to a previous operator and makes another try. It should be noted, that there is a maximum number of clashed structures allowed in one generation (MaxClashes in Table 2). If this number is achieved (box Blind alley in Figure 22) the current structure is replaced with the initial structure in order to avoid algorithm getting into the dead-end.

After the new structure is generated, the local restrictions at mutant positions are calculated, and the obtained restriction profile Ω_{sim} is compared with the experimental Ω_{exp}. If the goodness of fit χ^2 is better than the best found so far, the current structure becomes best and the fine-tuning mutation mode is turned on. Thus, the current structure becomes the parent for the next generation. If the new structure is not the best, it may still become the parent if the Metropolis criterion is satisfied. If the new structure is rejected, the parent stays the same as in previous generation. However, there is also some small probability, depending on the number of unsuccessful generations, that the current structure is replaced with the best.

The algorithm repeats this main loop with updating the optimization parameters and dihedral angles amplitudes as well as with generating a new structure until the maximal number of generations is reached.

3.5.2 Detection of the topology of M13 coat protein

To validate the structural optimization approach, we compared the simulation and experimental data of the conformational space of 3-maleimido proxyl spin label of 27 mutants of the M13 coat protein reconstituted in phospholipid bilayers [171,172]. A stochastic optimization algorithm was used to tune the secondary structure of the M13 protein and the relative position and orientation of the protein in lipid bilayer, so that the calculated local restrictions would correspond to the characteristics extracted from the experimental EPR data. Finally, a population of best fit structures was compared to a fluorescence-based protein model [93,131,132] of the membrane-embedded M13 protein [84].

Sample preparation and EPR spectroscopy. Various site-specific single cysteine mutants of the M13 major coat protein were prepared, purified, and labelled with 3-maleimido proxyl spin label as described previously [124,161]. Labelled mutants were reconstituted into lipid bilayers at L/P 100 as reported earlier [124,162]. For the purpose of EPR measurements, the proteoliposomes were then concentrated using lyophilization and subsequent rehydration and were collected by high-speed centrifugation [171].

Samples of reconstituted spin-labelled mutants in different lipid bilayers were filled up to 5 mm in 50-ml glass capillaries that were accommodated within standard 4-mm diameter quartz tubes. EPR spectra were recorded at room temperature on a Bruker ESP 300E EPR spectrometer (Bruker Instruments, Billerica, MA) equipped with a 108TMH/9103 microwave cavity. The EPR settings were 6.38 mW microwave power, 0.1 mT modulation amplitude, 40 ms time constant, 80 s scan time, 10 mT scan width, and 338.9 mT centre field. Up to 20 spectra were collected to improve the signal/noise ratio [172].

Data analysis by motional patterns condensation. The experimental spectra were fitted by a model of asymmetric motional restriction of a spin label, which is based on fast rotational motion approximation, with 4 spectral components as it is described in Section 3.1.3.1. Motional patterns, or a distribution of motional patterns, were resolved from the EPR spectra with the implementation of multi-run HEO (Section 3.1.3.2) and GHOST condensation algorithm (Section 3.1.3.3) that filters and groups the solutions found in the spectrum optimization runs [49,85,171-173,178].

Optimization of protein structure. By using structural optimization algorithm the parameters of the protein-lipid model (see Table 3) were optimized by improving the fit of simulated local restrictions data along the protein sequence to the available experimental free rotational space data obtained from EPR spectra of 27 mutants of the M13 coat protein reconstituted in 14:1 PC phospholipid bilayers.

Table 3: *Optimization parameters for the membrane-spanning transmembrane M13 protein system.*

Parameter	Unit	Description
$\{\varphi_i, \psi_i\}$	°	2×49 pairs of dihedral angles (first and last angles, φ_1 and ψ_{50}, are not defined)
tm_{start}	-	Starting position of the transmembrane region of the protein
tm_{end}	-	End position of the transmembrane region of the protein
D	Å	Steric thickness of the membrane (see Figure 16I)
θ	°	Tilt angle of the protein with respect to the membrane normal
d_{shift}	Å	Shift of the protein in the bilayer along the membrane normal (used for the fine-tuning of the transmembrane position of the protein)
φ	°	Rotational angle (rotation of the protein around the long axis)

Initially the secondary structure of the protein was set to an α-helical conformation ($\varphi = -57°$ and $\psi = -47°$). 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 dynamics [172]. 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 [93,131,132]. The multi-run 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 $\{\varphi_i, \psi_i\}$, by tuning the parameters of lipid bilayer (steric thickness D), and by optimizing the relative position and orientation of the protein in the lipids (tm_{start}, tm_{end}, θ, φ, d_{shift}.). For each new structure the corresponding local structural restrictions were calculated. Thus altogether about 200,000 different global structural conformations were checked.

3.5.3 Detection of conformational changes in N_{TAIL}

Herein the structure characterization approach is applied to translate multiple SDSL EPR data into structural characterization of the intrinsically disordered C-terminal domain of the measles virus (MV) nucleoprotein (N_{TAIL}, aa 401-525) [105] alone or in a complex with and the C-terminal X domain (XD, aa 459-507) of the phosphoprotein, shown to be responsible for binding with N_{TAIL} and for its α-helical folding [81].

Sample preparation and EPR spectroscopy. The details about expression, purification and preparation of N_{TAIL} as well as the details about the spin labelling could be found in [12]. EPR spectra were acquired using liquid nitrogen cryostat on an Elexsys E500 Bruker spectrometer equipped with an ELEXSYS Super High Sensitivity resonator operating at 9.3 GHz. The microwave power was 20mW and the magnetic field modulation frequency and amplitude were 100 kHz and 0.15 mT, respectively. Before EPR experiment 20 µL sample was warmed up in water bath at 296 K to equilibrate for 10 minutes, quickly transferred into quartz capillary and measured at 296 K. Then the sample was cooled down in 1 minute, equilibrated for 2 minutes and measured at temperatures in the low range (279 K, 281 K and 283 K). The temperature changing time was about 1 minute and the temperature equilibration time was 2 minutes (T stability < 0.3 K, T accuracy <0.2 K). Finally, the temperature was raised to the high range (308 K, 310 K and 312 K), with the 25 K jump taking approximately 10 minutes (including the equilibration time). The measurement time (in terms of the number of acquisitions) was modified to get satisfactory signal-to-noise ratio of the spectra.

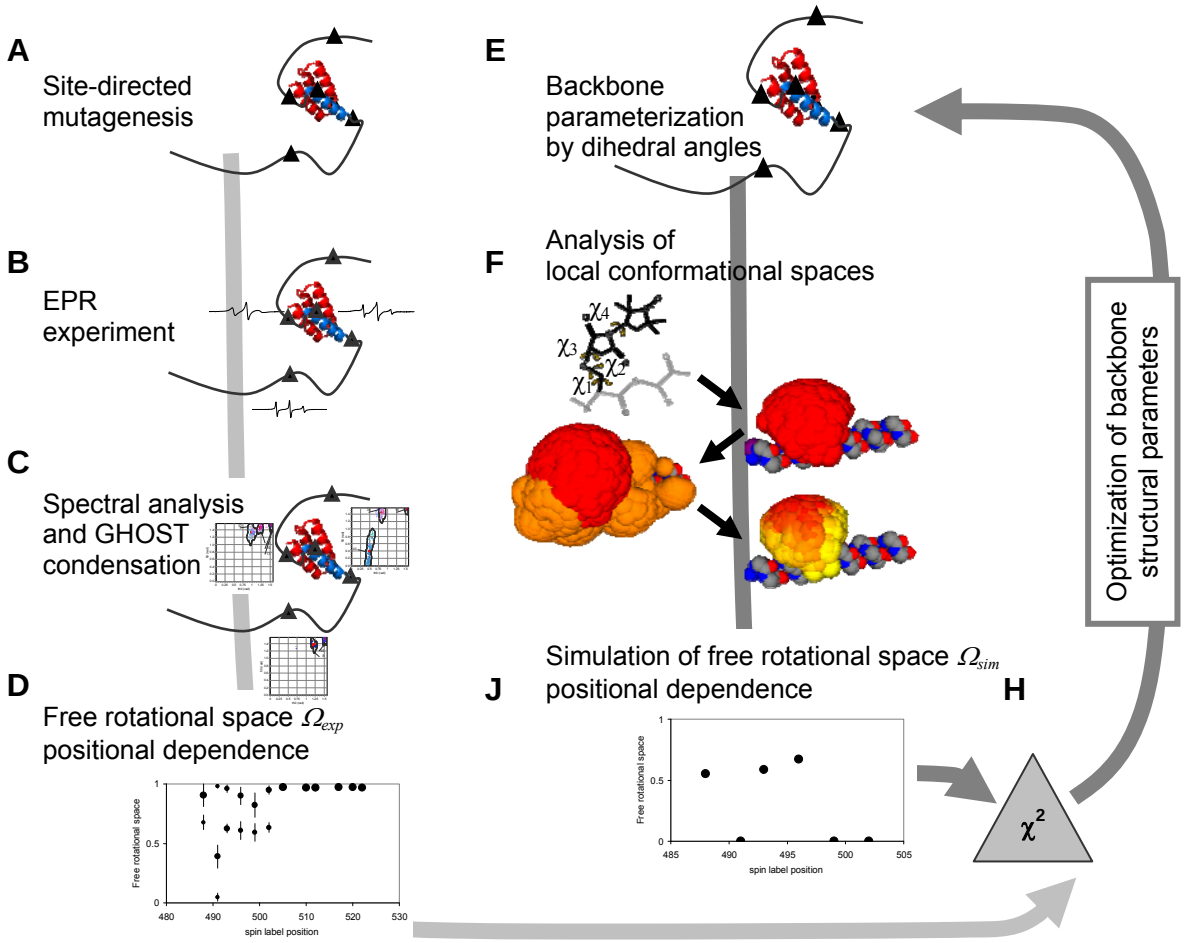

Figure 23: *The overview of the approach of structural characterization of N_{TAIL}-XD complex*. The approach is based, from one hand, on site-directed spin labelling (A), EPR spectroscopy (B), multi-component spectral analysis and GHOST condensation (C), and from the other hand, on protein structure modelling (E), conformational space modelling and restrictions calculation (F), fitting (H) of simulated restriction (J) to experimentally obtained data (D), and structure optimization (I).

Data analysis by motional patterns condensation. The experimental spectra were fitted within the fast restricted rotational motion approximation with 2, 3 or 4 spectral components (depending on the signal-to-noise value: below 100, between 100 and 150, and above 150, respectively) as it is described in Section 3.1.3 (similarly as it was done to detect motional pattern for the M13 coat protein). 20 runs of dHEO

optimization [49,85,173,178] was performed for each of the 336 spectra and condensed with GHOST condensation procedure [85,173,178] as schematically depicted in Figure 23C. All the condensed solution in terms of motional patterns was plotted against temperature for each sample and environment to detect the significant motional patterns. Those without monotonous local temperature dependence have been deleted (like isolated patterns without temperature dependence, patterns that does not have similar patterns at nearby temperature, etc). In addition, all the patterns with the weight below 10 % were also ignored as they do not contribute to the main results significantly.

Finally, the local temperature sets, 279 K - 283 K as well as 308 K - 312 K, were used to get average condensed solution pattern set for low temperature (281 K) and high temperature (310 K). The rotational restrictions were then used in terms of free rotational space Ω_{exp} and normalized rotational diffusion D_{exp} positional dependences as in [172] and schematically presented in Figure 23D.

Optimization. A stochastic optimization algorithm was used to tune the secondary structure of N_{TAIL} and the relative position of the N_{TAIL} towards XD, so that the calculated local restrictions would correspond to the characteristics extracted from the experimental EPR data. Initial structure of N_{TAIL}-XD complex was based on a chimera structure [91], available at Protein Data Bank (PDB) as 1T6O.pdb.

Table 4: *Optimization parameters of N_{TAIL} protein structure in complex with XD.*

Parameters	Unit	Description
$\{\varphi_i, \psi_i\}$	°	2x40 pairs of dihedral angles (first and last angles, φ_1 and ψ_{41}, are not defined)
Δx	Å	Displacement vector of the N_{TAIL} relative to XD
A	°	Orientation tensor of the N_{TAIL} chain relative to XD
φ	°	Rotational angle (rotation of the N_{TAIL} around its long axis relative to XD)

The parameters that are optimized are listed in Table 4. The optimization module was based on a stochastic algorithm of the Metropolis Monte Carlo family with several elements of the Evolutionary Algorithm (see section 3.5 for details).

4 Results and Discussion

In this chapter, first, the results from enhancement of EPR data analysis are presented and discussed following by the testing of the conformational space modelling, and the M13 protein structure optimization. Finally, the results of application of the proposed approach to study the changes of the secondary structure of the measles virus nucleoprotein are discussed.

4.1 Speeding-up GA for SL EPR-based characterization of biosystem complexity

Spin label EPR-based characterization [85,178] is one of the basic component of the protein structure characterization approach [176] presented in this thesis. Our goal was to make the EPR spectral analysis procedure more feasible by speeding-up of the computationally expensive spectrum optimization routine. With this goal in view, our strategy was to reduce the number of optimization runs and preserving the quality of the results. The "grid" problem was found to be the cause of loosing the solution diversity, and the problem was solved with introduction of the "shaking" operator. New modification of the spectrum optimization algorithm was tested within this work and also by the following successful application of the EPR spectral analysis and GHOST condensation method [86,94,95,135,171,172].

4.1.1 Reduction of the number of multiple runs

To measure the efficiency of the HEO algorithm modifications 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 universality of the new algorithm we analyzed two types of EPR spectra: experimental ones (from membranes and membrane proteins) and synthetic (discrete and continuous).

In our attempt to enhance the optimization algorithm, we first reduce the number of HEO runs from 200 to 20 and increase in the contribution of each run (more than one best parameter set). The results for a typical experimental spectrum are shown in the Figure 24 where the GHOST diagram (Figure 24B) and runs contribution histogram (Figure 24C) are compared with the original GHOST diagram based on the 200 runs (Figure 24A). It can be easily seen that the distribution of solution is not maintained, so this can not be the right approach to reduce the computational demand of the problem. In addition, it can be also seen on the runs contribution histogram in the Figure 24C, that only a few runs (such as the first, seventh, ninth and seventeenth) contribute to the GHOST presentation, whereas the other runs (*i.e.* the third, fourth, tenth, *etc.*) have no contribution at all. This causes a loss of solution diversity, a worse distribution of χ^2 (see minimum value and distribution width in "20 runs" column of Table 5) and a wrong solution domains determination (Figure 24B). In addition one can also see a higher solution density as a consequence of the crowding in the phase-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 26. The bad GHOST picture arises from the fact that the contribution of the runs is extremely uneven (Figures 26B), originating in a solution crowding.

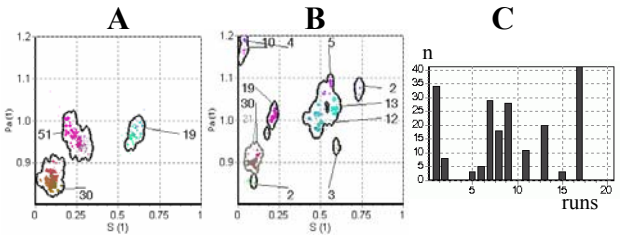

Figure 24: *Typical characterization of spin labelled membrane*. **A.** GHOST as a result of 200 runs of HEO where only one solution is extracted from a single run. **B.** GHOST as a result of 20 runs of the same HEO algorithm where on average 10 solutions are taken from each run. **C.** Runs contribution 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.

Table 5: *Optimization parameters after 200 and 20 runs for the real membrane spectrum* (for the experimental preparation see the caption to the Figure 11).

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

According to the literature, the sharing implementation could change the result [55,110]. To test the sharing approach the continuous problem was chosen (Figure 26A). The results of this test in terms of the runs contribution histogram and GHOST cross-section are shown in the Figures 26C. It can be seen that the GHOST representation better resembles the original one, and also the contribution histogram becomes more even. However, the distribution of χ^2 is worse (see the minimum value and the distribution width in "sharing" column of Table 6). 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), modified the elite set and changed the selection strategy in the GA algorithm.

4.1.2 Detection of the "grid" problem and implementation of the "shaking" operator

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. The latter is one of the most important operators in the GA algorithm and assures that "genetic material" related to good model parameters can spread and copy through the population. After a few generation decades the population forms a "grid" in the search space (Figure 25A) as a consequence of the rough action of the 3-point crossover operator. This leads to the loss of solution diversity.

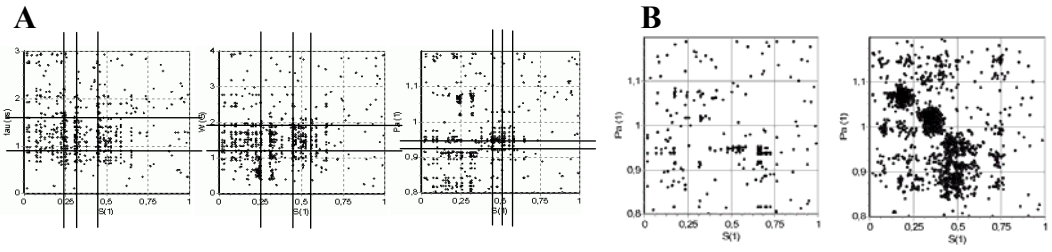

Figure 25: *Schematic presentation of the "grid" problem (A) for three cross-sections of the phase-space and its solution (B) in single run*. **A.** Due to the standard multipoint crossover, subgroups of parameters are "transferred" between generations untouched, resulting in a grid-like distribution 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 diminishing the effect of the “grid” problem.

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 in the Figure 25B for a continuous problem that represents the most extreme case of the complexity.

The results of the implemented shaking operator are shown in the Figures 26. One can see that the shaking operator considerably improves the result of a single run as the GHOST pattern from 20 runs (Figure 26D) is very similar to the original one (Figure 26A), the runs histogram is very even Figure 26D), and finally the distribution of χ^2 is very good (Table 6).

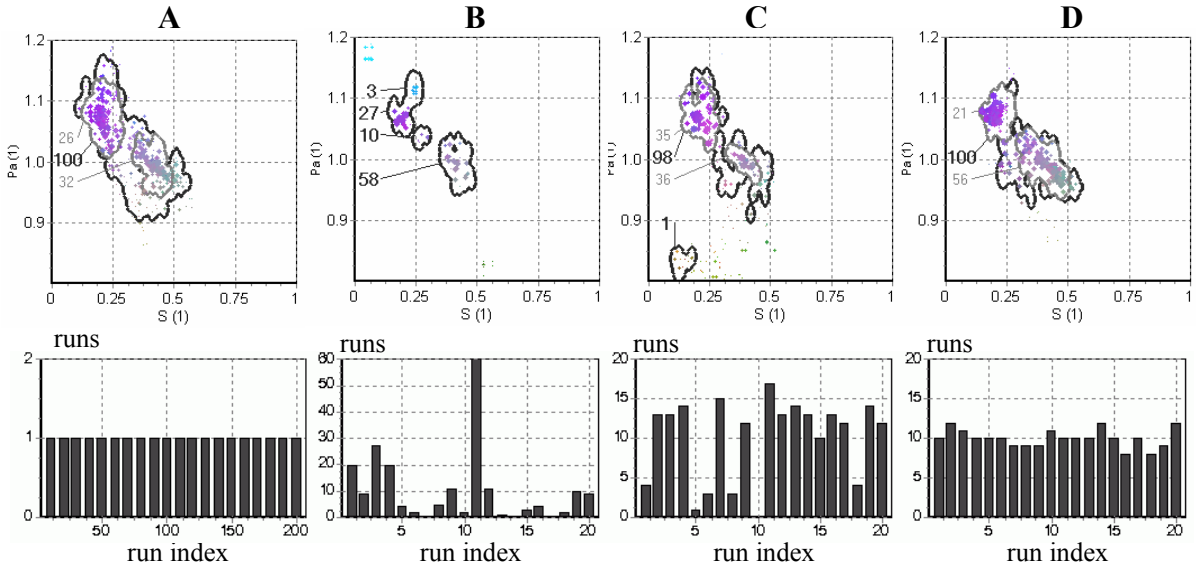

Figure 26: *Comparison of the effectiveness of different multi-run HEO-GHOST approaches on the synthetic 15-component spectrum together with runs contribution histogram.* Comparison includes also runs contribution histograms. **A.** GHOST and runs contribution as a result of 200 runs of original HEO 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 6: *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 26).

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

4.1.3 Testing of the modified algorithm

In further testing, 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 the characterization of four different examples are shown in the Figure 27, 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).

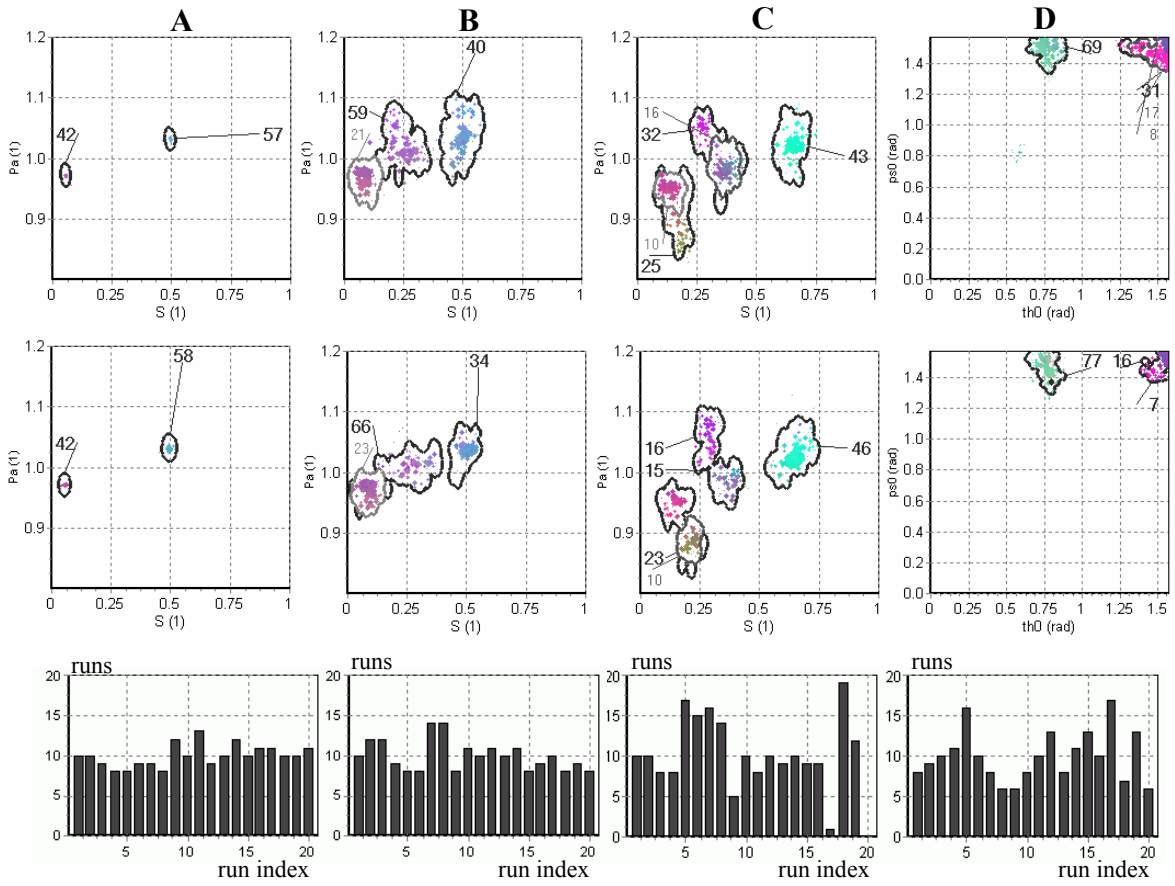

Figure 27: *Comparison of GHOST plots of original-HEO approach versus shaking-modified-HEO together with runs contribution histogram for the shaking-modified-HEO based on 20 runs.* The original-HEO approach with 200 runs (above) is compared versus modified-HEO (with shaking) based on 20 runs (bellow). **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-continues spectrum (see the caption to the Figure 26). **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-doxyl palmitate, MeFASL(10,3), and measured under the same conditions as the membranes of horse neutrophils (see the caption to the Figure 11). **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 bilayer [170].

Thus the approach enables us to reduce the number of HEO runs while preserving the quality of the final result for the synthetic 15-component spectrum (Figures 26D, and Table 6) as well as for the other testing experimental and synthetic spectra (Figures 27). Therefore by using this new algorithm, we are able to speed-up the optimization process by a factor of 5-10.

4.2 Testing the sensitivity of the spin label conformational space to the primary and secondary structure and to the lipid environment

Before local restrictions in proteins obtained via site-directed EPR spectroscopy, spectral simulation and optimization (see Sections 3.1.3 and 4.1) are compared to the ones modelled by the spin label conformational space restrictions simulations as described in Section 3.3, the sensitivity of the conformational space had to be extensively tested according to variation in primary and secondary structure of the protein, as well as according to the effect of lipids.

4.2.1 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 ($\varphi = -57°$ and $\psi = -47°$). Since the secondary structure in tests was set to be uniform among all oligopeptides, the resulting differences in Ω (Figure 28A) can be assigned to the properties of the amino acid residues, i.e., to the primary structure. As can be seen in Figure 28A, the restrictive effect of the primary structure depends on the flexibility and on the size of the amino acid side chains. 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 28A).

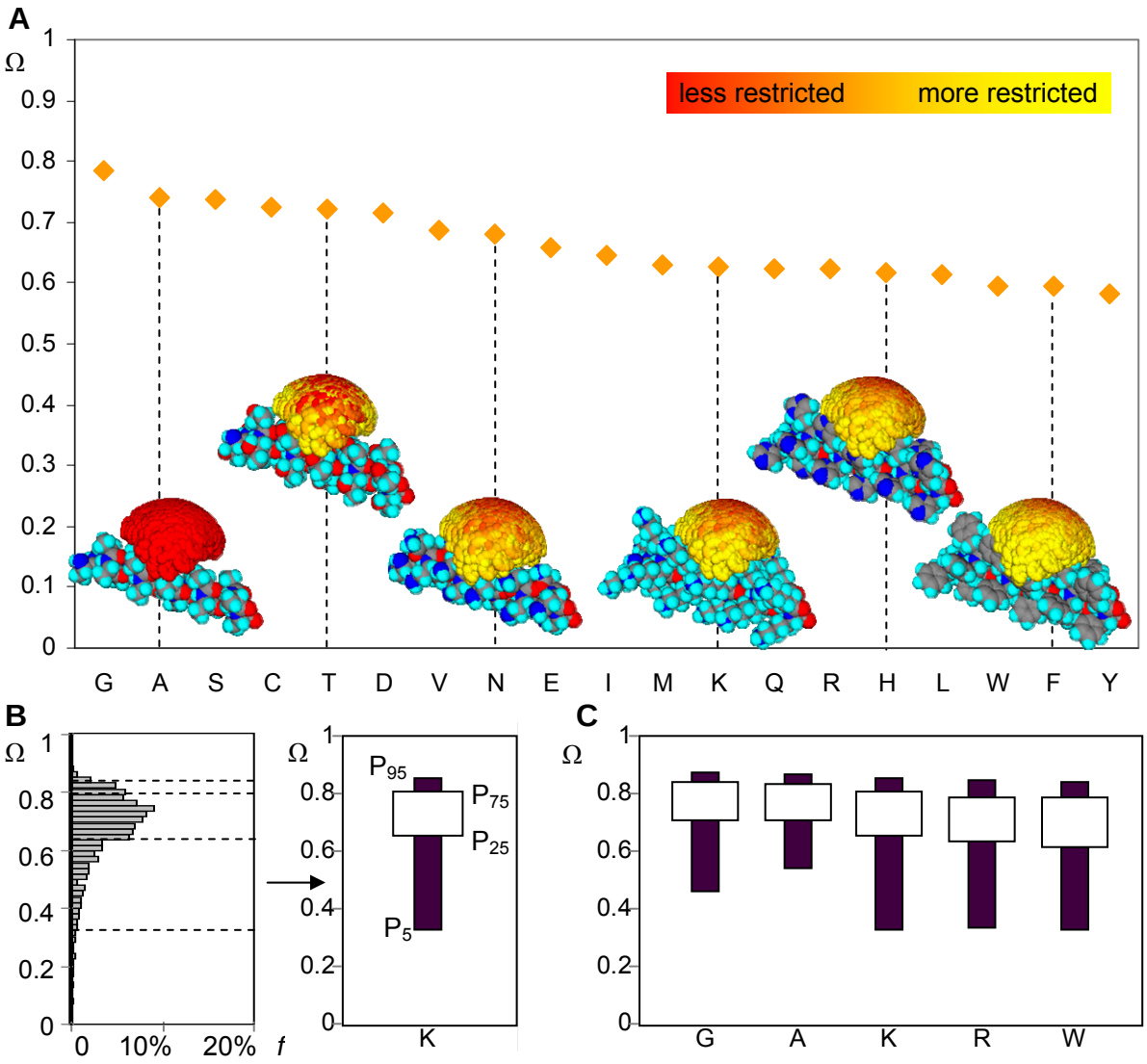

Figure 28: *Spin label conformational space sensitivity to primary structure*. **A.** Normalized free rotational space Ω (orange data points) for artificially-designed 19-residue peptide homopolymers in an α-helical conformation ($\varphi = -57°$ and $\psi = -47°$) with the 3-maleimido proxyl spin label attached to the central (10th) cysteine residue. For a few oligopeptides the conformational space of the spin label is illustrated in a molecular model with the yellow colour 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.

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 28B). This Ω distribution was simplified by plotting the inter-quartile ranges that correspond to 5^{th}, 25^{th}, 75^{th}, and 95^{th} percentiles of the Ω distribution. For a number of typical oligopeptides the inter-quartile ranges are plotted in Figure 28C. 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 28C, columns G and A). In this case, the spin label will be less sensitive to different elements of the 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 28C).

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 28C, 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 28C, columns R and W). This indicates that the spin label is more sensitive to the secondary structure rather than to the primary structure.

4.2.2 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 29A), 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.

To investigate how the secondary structure at neighbouring amino acid positions affects the conformational space of the spin label, we repeated the calculation of R_B by changing the dihedral angles further away from the *i*-th labelled site. The result is shown in Figure 29B 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 the secondary structure, the range of backbone restrictions varies from 30 to 100% (Figure 29B). Based on this finding it may not be necessary to have all positions labelled. Instead, sites could be labelled 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 $i\pm3$ and $i\pm4$). 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 [32].

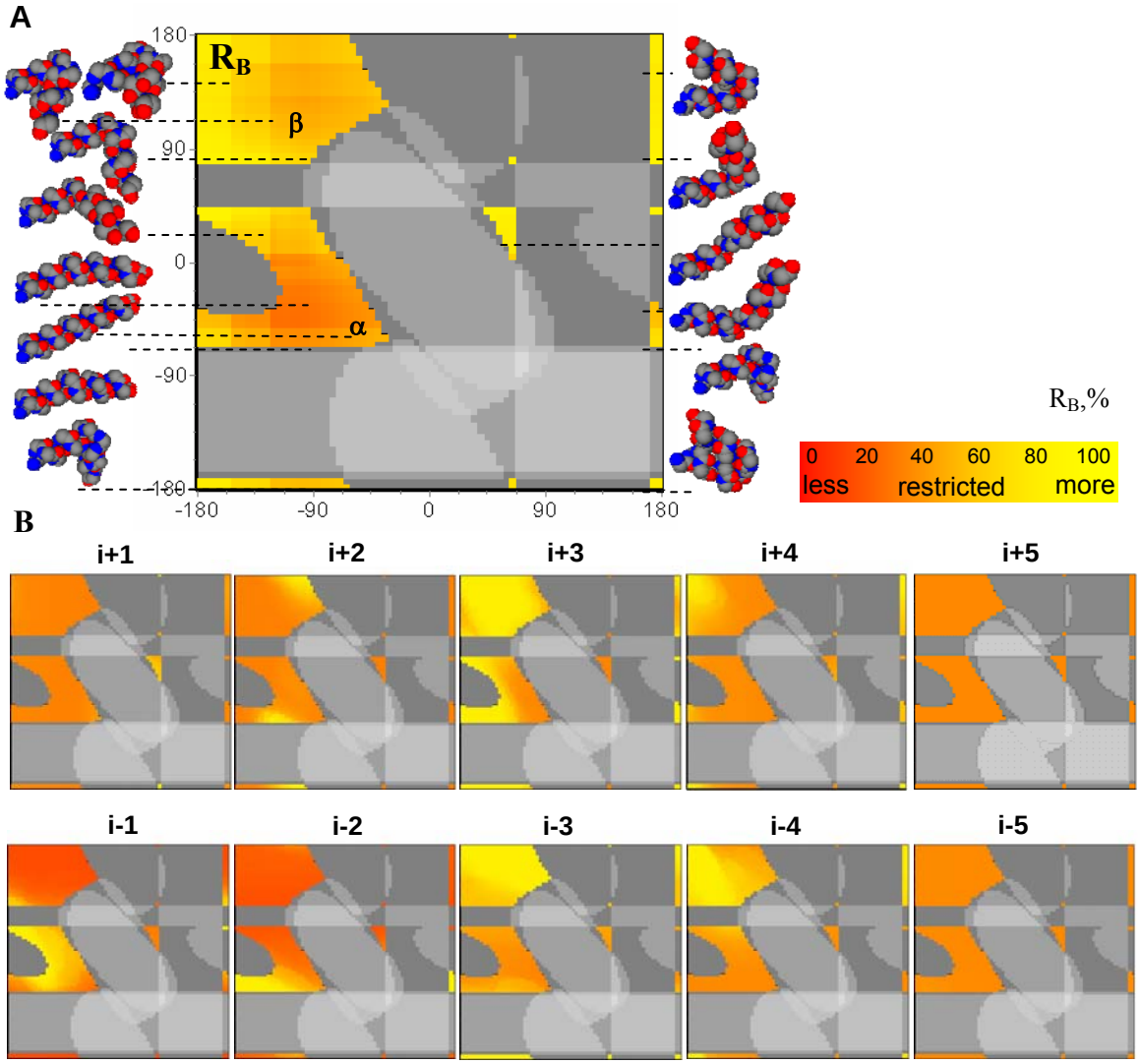

Figure 29: *Sensitivity of the conformational space of the spin label to the secondary structure*. The 3-maleimido proxyl spin label is attached to the central (10th) cysteine residue of a number of artificially-designed 19-residue oligoalanine peptide in an α-helical conformation ($\varphi = -57°$ and $\psi = -47°$). **A.** Backbone restrictions R_B (i.e., the reduction of the sum statistical weight due to overlap with the backbone) at the spin-labelled 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 coloured dot in the Ramachandran plot with the coordinates corresponding to the angles φ and ψ at the spin label position (the red-yellow colour 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-labelled site i=10 arising from adjacent amino acid positions up to five to the N and C-terminal end (i.e., $i\pm1$, $i\pm2$, $i\pm3$, $i\pm4$, $i\pm5$). The remaining secondary structure was taken as α-helix. For labelling of the axes and for the other details, see (A).

4.2.3 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 through the normalized free rotational space Ω of the membrane-embedded spin-labelled M13 protein, for which we assumed that the protein is in an α-helical conformation [93,131,132,197]. To enable comparison with experimental spin-label EPR data, we assume that the protein is reconstituted in 1,2-dierucoyl-sn-glycero-3-phosphocholine (22:1PC) phospholipid bilayers [172] 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 labelling [93,131,132]. For comparison we also examine the protein without a lipid environment.

For almost all amino acid positions in the transmembrane region of the M13 coat protein, the normalized free rotational space Ω is reduced by 20-40% due to the lipid effect (Figure 30A). Also, the trend of the values for Ω is slightly changed by the lipids. This is due to the fact that at each spin-labelled 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 alkyl chains and the spin label side chain conformations).

The effect of the orientation of the 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 30B) and for each amino acid position collected in Figure 30C. 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-EPR data combined with modelling of the membrane-embedded protein may reveal the correct orientation angle of the protein in the membrane.

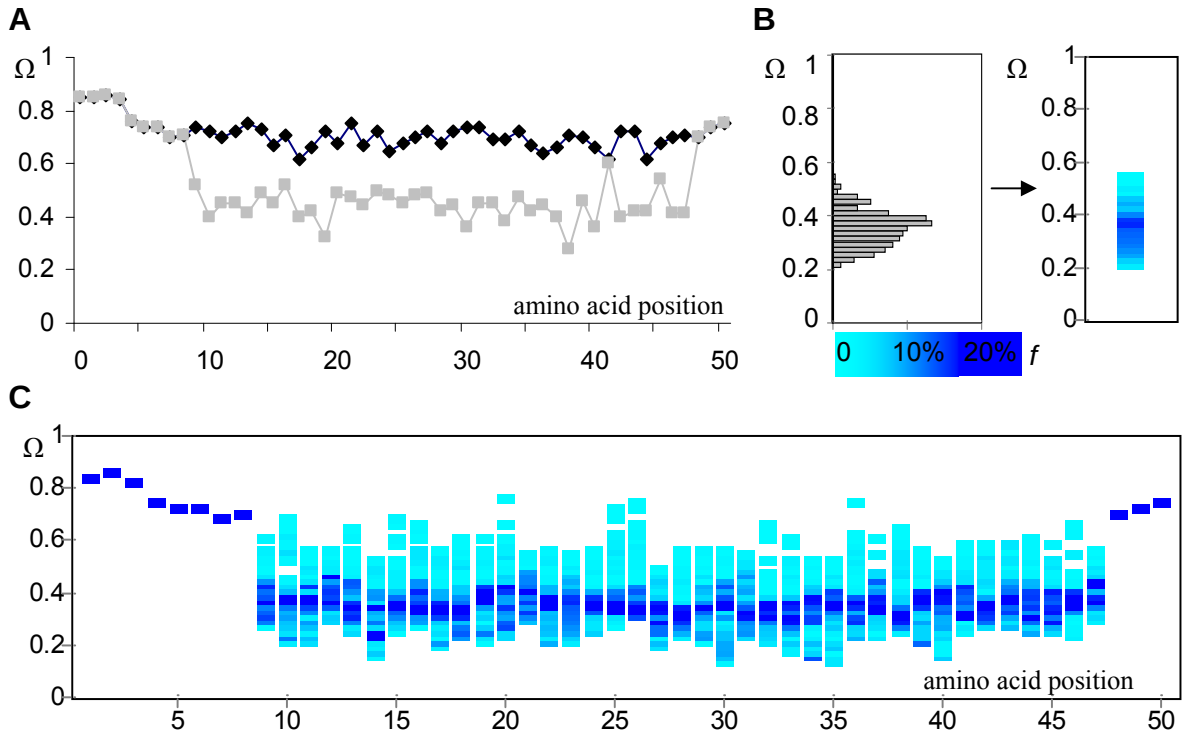

Figure 30: *Sensitivity of the conformational space of the spin label to the lipid environment.* The 3-maleimido proxyl spin label is attached to the membrane-embedded M13 protein in an α-helical conformation ($\varphi = -57°$ and $\psi = -47°$). The transmembrane region is defined between amino acid positions 9 and 47, corresponding to a tilt angle of 20°. **A.** 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 the M13 coat 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 colour coding (right). The spin label is at position 23. **C.** Amino acid position dependence of the Ω data as determined in (B).

4.2.4 Contribution of different restrictive factors to conformational space restriction

To summarize the tests of conformational space sensitivity we will discuss the contribution of different restrictive factors to conformational space restrictions.

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:

$$P_{SUM} = P_{SUM}^{initial}\left(1 - R_B - R_S - R_L\right), \tag{17}$$

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 neighbouring 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 the overlap of the side chain conformations with the backbone (Table 7). The conformational overlap with neighbouring 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 alkyl chains).

Table 7: *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 initial probability P_{SUM}
Overlap with the backbone	R_B	~ 50-60 %
Side chain neighbourhood	R_S	~ 30 %
Lipids	R_L	~ 10 %

4.2.5 Analysis of side chain rotational restrictions of membrane-embedded proteins

In order to determine the potential ranges of the rotational restrictions the protein modelling was tested against recently published experimental data of free rotational space of the membrane-bound M13 major coat protein [172] (Figure 31, 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 (see also section 4.3.1) of the 3-maleimido proxyl spin label for the protein reconstituted in phospholipid bilayers consisting of 1,2-dierucoyl-*sn*-glycero-3-phosphocholine [171,172].

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 31). For simplicity, we assumed a membrane-embedment of the protein based on a recently published model, using an α-helical protein with a tilt angle of 18º with respect to the membrane normal and with membrane crossing points at positions 9 and 47 [93,131,132]. To analyze the effect of protein conformation and membrane-embedment 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±30º and -47±30º, respectively. The Ω values related to the original α-helical protein model (φ = -57º and ψ = -47º) are indicated with white triangles in Figure 31. 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 31A). 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 31). 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-EPR experimental data and the simulated data for all spin label positions (Figure 31B). 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. This problem is addressed by introducing an optimization procedure in our calculation, which tunes the backbone dihedral angles and in fact eventually would produce an optimized ensemble of best-fitting structures.

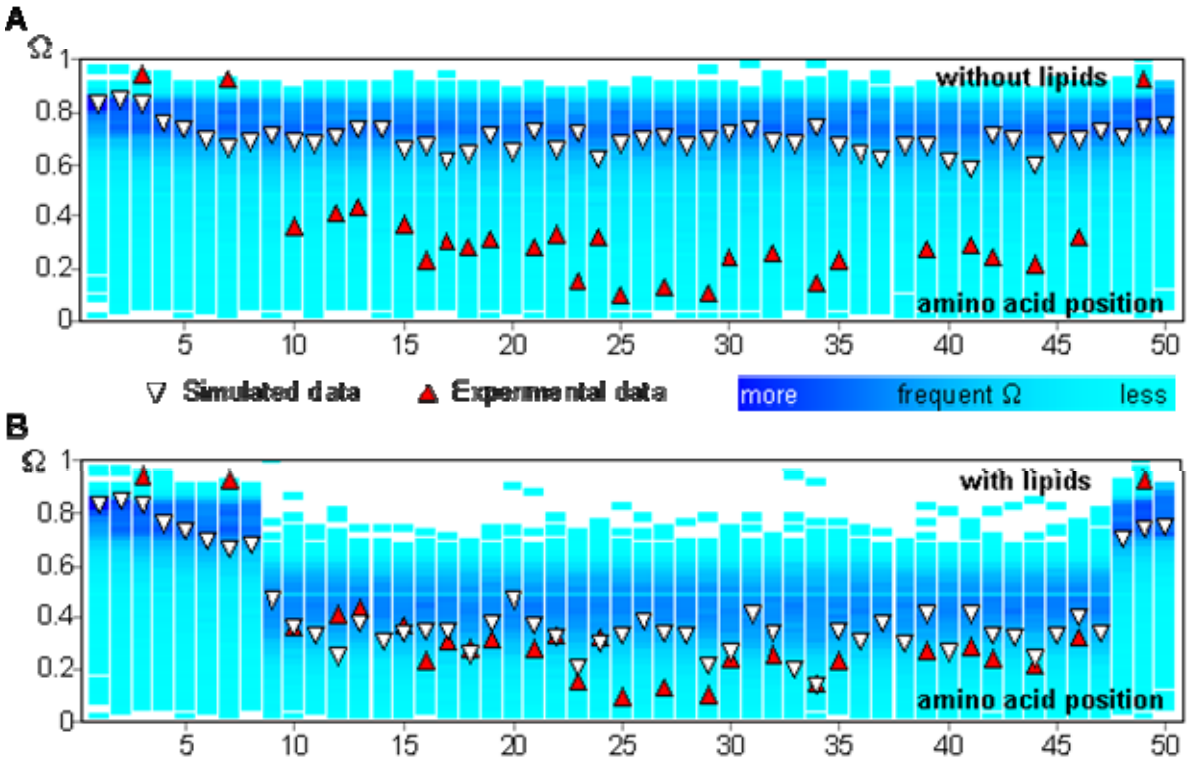

Figure 31: *Sensitivity of the free rotational space to the primary sequence and variations of the protein secondary structure and the effect of the lipids for the membrane-embedded spin-labelled M13 coat protein*. The histograms of the relative frequency of a given value of Ω (colour-coded by continuous shades of blue, such that cyan is the lowest frequency and dark blue is the highest frequency within the set of 5000 modelled 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 (φ = -57° and ψ = -47°) as defined in [93,131,132].

4.3 Optimization of the membrane-embedded M13 protein structure by fitting simulated restrictions to experimentally obtained restrictions

The secondary structure of the 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-EPR spectra of 27 mutants of the M13 coat protein reconstituted in 14:1PC bilayers [171,172]. 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 the α-helical conformation (φ = –57° and ψ = –47°). 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 dynamics [172]. 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 [93,131,132]. The multi-run 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.

4.3.1 Site directed spin labelling and EPR experiments

First, to resolve the local conformation of the protein in 14:1 PC lipid bilayers, 27 out of 50 amino acid residues were replaced for a cysteine residue, spin labelled, reconstituted into lipid bilayers, and the EPR spectra were recorded. To extract the local structural information, the EPR spectra were simulated with a multi-component model of asymmetric motional restriction [49,171,172,178] and characterized via a multi-run hybrid evolution optimization method [178]. The goodness of fit was chosen to be the reduced χ^2 function (Eq. 1) and for all 27 mutants, the quality of the simulated EPR spectra is excellent – the reduced χ^2 is between 3 and 5 at a signal/noise ratio between 250 and 400.

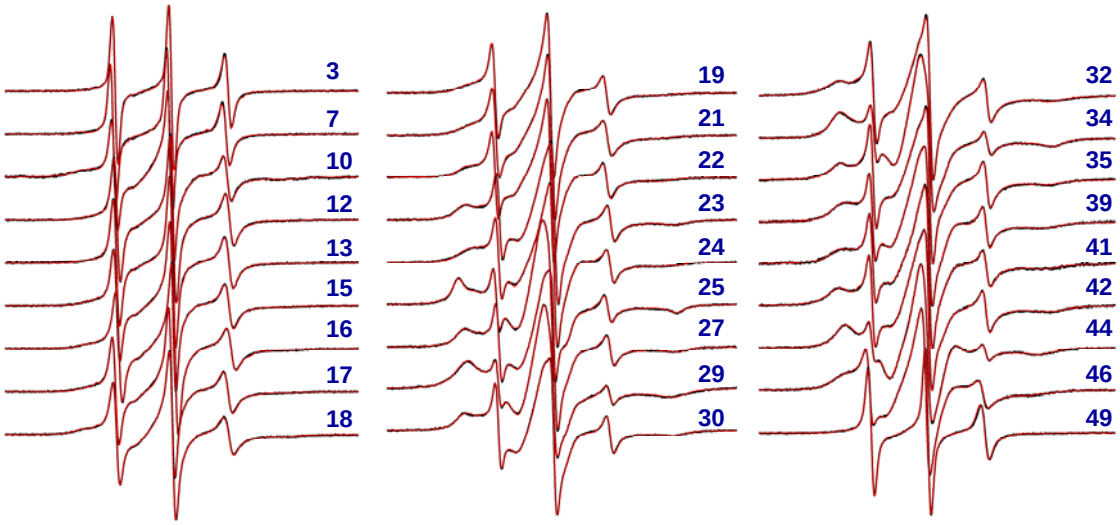

Figure 32: *Amplitude normalized EPR spectra of the spin-labelled M13 coat protein samples reconstituted in 14:1 PC lipid bilayers*. Label positions are 3, 7, 10, 12, 13, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 27, 29, 30, 32, 34, 35, 39, 41, 42, 44, 46, and 49. The total horizontal scan range is 10 mT. Spectral line heights are normalized to the same central line height (left peak). The simulated spectra are shown in red and grey is for experimental data.

The experimental and simulated EPR spectra of the 27 spin-labelled mutants in 14:1 PC lipid bilayers are shown in Figure 32, while the resolved spectroscopic parameters of multiple solutions are condensed in the GHOST plots as shown in Figure 33. GHOST presentation provides the most significant and probable groups of solutions (motional patterns) of spectral parameters.

Figure 33: *GHOST condensation plots of the spin-labelled M13 coat protein samples reconstituted in 14:1 PC lipid bilayers*. GHOST plots present the optimized multiple solutions represented in a two-dimensional distribution of the angles ϑ-φ and ϑ-dependent distribution of τ_c of the spin-labelled M13 coat protein at 27 spin label positions (the position is indicated with red number in the left bottom corner of the ϑ-φ GHOST plot). For the details about GHOST presentation see Figure 11 in Section 3.1.3.3.

For the further analysis and structural optimization, the motional patterns along the protein are collected in so-called bubble diagrams in terms of free rotational space and rotational diffusion (see Figure 34).

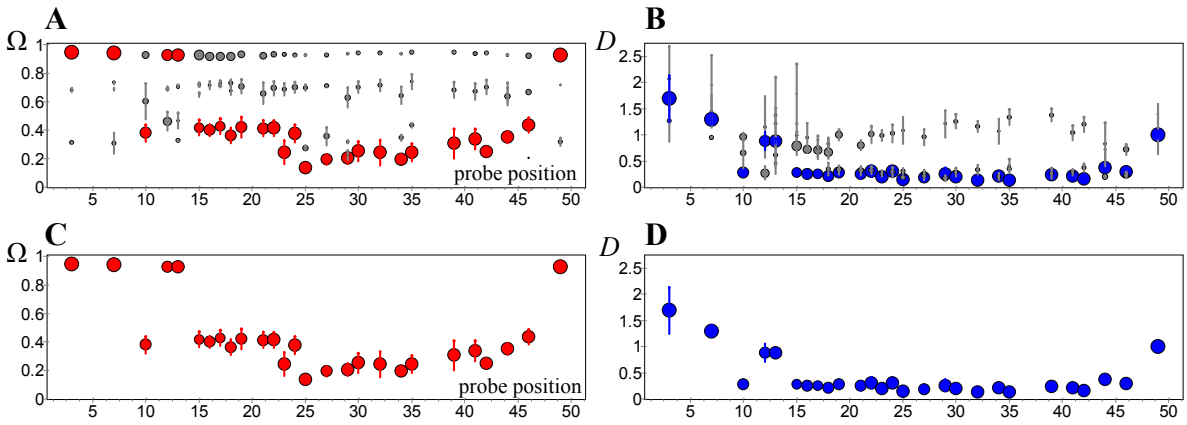

Figure 34: *Bubbles-condensed presentation of the motional patterns for the spin-labelled M13 coat protein samples reconstituted in 14:1 PC lipid bilayers.* Each bubble presents the motional patterns for 27 mutant positions along the M13 coat protein sequence in terms of free rotational space Ω (A, C) or rotational diffusion *D* (B, D) before patterns clearing (A, B) and after (C, D). In the cleaned bubble diagrams the most important motional pattern series is presented only. Each bubble represent a motional pattern, the size of bubble is proportional to the relative contribution of the particular pattern in the total spectrum; the vertical bar at each bubble represents the size of the motional pattern in the phase space (as in Figure 33). For the details about bubble diagram presentation see Figure 12 in Section 3.1.3.4.

The EPR line shapes for the different mutants in 14:1 PC along the protein primary sequence are different, ranging from mobile and isotropic (positions 3, 7, 49) to moderately immobilized (positions 15-22, and 46) and then to very anisotropic (positions 25, 27, 29, and 35). Different rotational restrictions of the spin labels as suggested by different spectral line shapes are well resolved by the GHOST plots. For example, the rotational space for position 13 is completely open as suggested by the large ϑ and φ values of the dominating motional pattern (Figure 33). On the other hand the rotational space for spin label at position 25 is very restricted as suggested by the green-coloured component. As was discussed previously [171], it is expected that the "most restricted" component represents the transmembrane state of the protein. Other components resolved by this methodology could involve non-specific labelling and other local conformations with lower probabilities [172].

4.3.2 Characterization of the membrane-embedded M13 coat protein structure

The data about the local conformational restriction in M13 protein in 14:1 PC lipid bilayer, which was obtained with EPR spectroscopy at 27 sites along the protein according to spin labelled mutants, was then used to guide structure optimization algorithm.

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 35A). 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 35B. 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° ± 4, as determined from quantitative fluorescence site-directed analysis [93]). 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 [7,130]. 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 35B. The corresponding goodness of fit of this structure is χ^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 35C) 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.

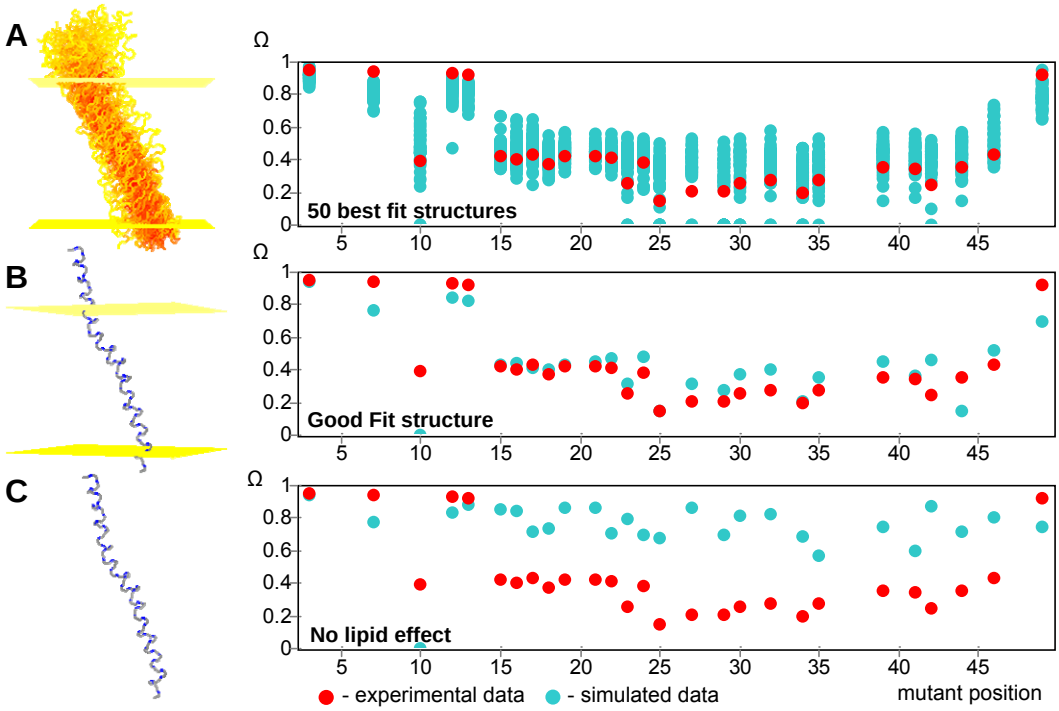

Figure 35: *Optimization of the structure and the membrane-embedment M13 coat protein.* Protein structure is optimized by fitting the calculated local restrictions (blue circles) to the restrictions profile obtained from SDSL-EPR (red circles). **A.** A population of best-fit structures (50 structures with a goodness of fit χ^2 in the range from 2.7 to 3.7) of the M13 coat protein in 14:1PC lipid bilayers. The yellow-red colour 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 the M13 coat protein calculated without taking into account the lipid effect ($\chi^2 = 33.0$).

Extensive testing of the model (see section 4.2) 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. Calculated restrictions from 5000 randomly simulated structures of the M13 coat protein reconstituted in phospholipids produced a wide range of Ω values that cover all 27 points of the experimental data [176]. 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 35A) obtained in 1000 runs of optimization are in excellent agreement with a recently proposed fluorescence-based The M13 coat protein model [93,131,132]. This nice accordance implies that our modelling 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-EPR that employ molecular dynamics (MD) simulations [10,24,38,60,97,144,157,168] 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-EPR data from all available spin-labelled 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 MD simulations. In the present case, extending the computation time would impose a severe limit to the calculations, as the optimization of the 200,000 structures resulting in Figure 35A already takes four weeks of CPU time on a 20-core computer cluster (6×Opteron Double Core 2.4 GHz and 8×Athlon Single Core 2.13 GHz).

Analysis of best-fit single structures (like the one in Figure 35B) indicates that the optimization algorithm successfully provided 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 35B), which results in the ranges of calculated Ω values for the family of best-fit structures (Figure 35A). 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.

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 mutant positions revealing a series of experimentally detected motional patterns 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 simultaneously [49]. 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.

4.4 Detection of conformational changes in N_{TAIL} by SDSL-EPR spectroscopy and conformational space modelling

The approach developed (sections 3.3 and 3.5) and numerically tested (sections 4.2 and 4.3) on small synthetic peptides and on membrane-embedded M13 coat protein was then applied to reveal the conformational changes in measles virus N_{TAIL} (see introduction in section 3.4.2) according to temperature changes and according to the presence of XD partner protein.

4.4.1 Site directed spin labelling and EPR experiments

Twelve N_{TAIL} cysteine-spin labelled mutants (S488, S491, D493, L496, Q499, A502, S505, S510, T512, V517, D520, N522) were strategically designed [12] (see Figure 20) to be the most sensitive to different possible local structural arrangements of N_{TAIL} relative to XD protein in both N_{TAIL} Box2 and N_{TAIL} Box3 parts with the reference complex structure being the chimera crystal structure.

The EPR spectra of all 12 mutants in two protein systems, N_{TAIL} alone and of N_{TAIL} with XD, each in a buffer solution with or without sucrose, were measured within the temperature scan 279 K - 281K - 283 K - 296 K - 308 K - 310 K - 312 K (see Figure 36). The sucrose was added to check the slow down of the backbone motion effect on the structural characteristics of the complex Since the proteins were expected to posses more than one local conformation at a site in addition of possible traces of non-specific labelling, measured spectra were analyzed with a multi-component model of asymmetrically restricted rotational motion of a spin label [49,171,172,178]. Excellent S/N of experimental spectra allowed such an analysis as well as optimization of the fitted spectra employing a multi-run multi-solution hybrid evolutionary algorithm [178]. The goodness of fit, the reduced χ^2 function (Eq. 1) for final spectral fits appeared always to be within less than 3 noise amplitudes relative to the experimental spectra. In some case, especially for highly disordered spectra at high temperature, e.g. for the mutant positions 499, 522 of N_{TAIL} alone and positions 512, 522 for N_{TAIL}-XD complex (both at 310 K), the lower run flatness of a value around 50% indicates over-fitting problem - the number of spectral parameters were too high for such a simple spectrum.

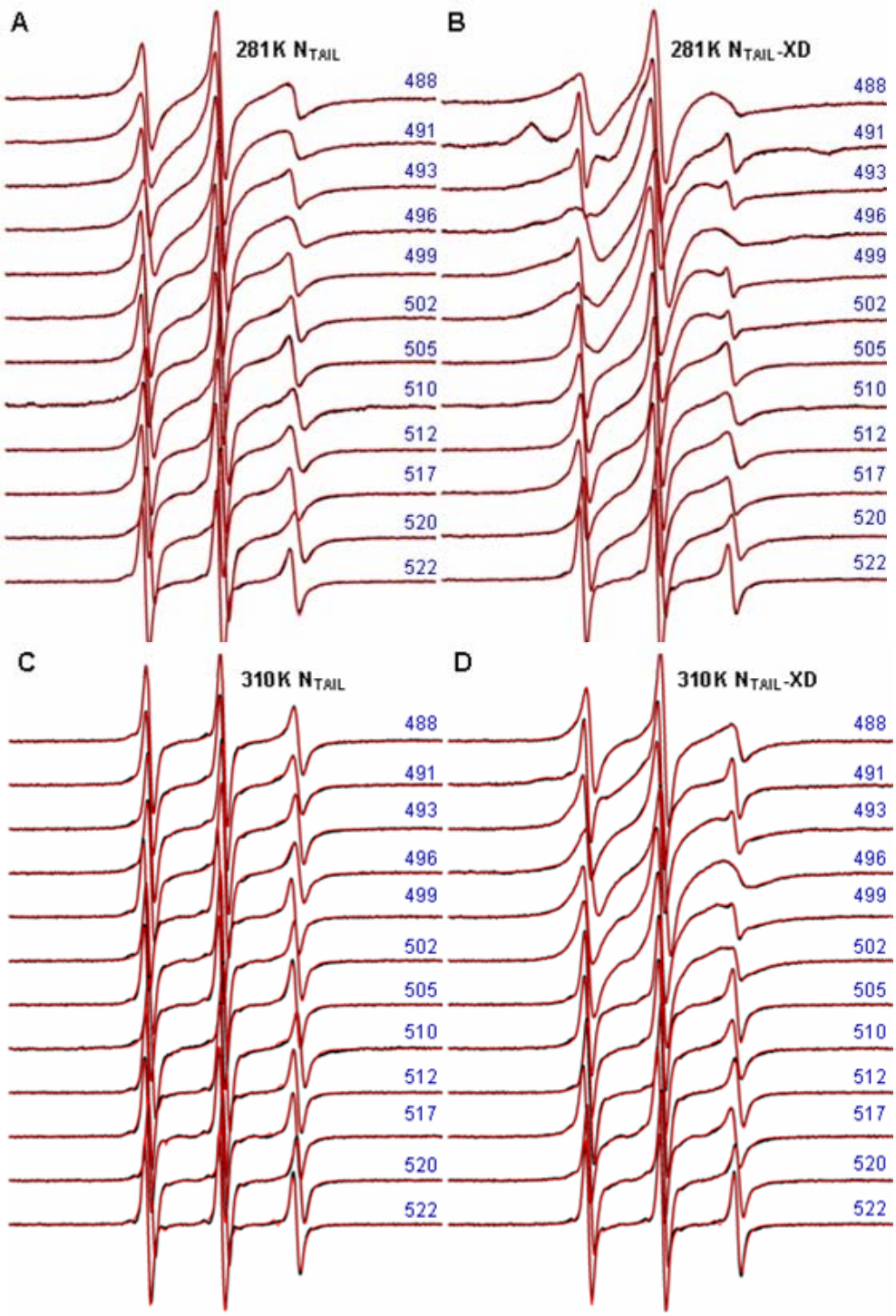

Figure 36: *Amplitude normalized EPR spectra of spin-labelled N_{TAIL} protein*. EPR measurements were done for spin-labelled N_{TAIL} samples (100 μM) in the absence (left panel) or presence of a molar excess of XD (right panel) at low (281 K) (top) and high temperature (310 K) (bottom) at 12 spin label positions (S488, S491, D493, L496, Q499, A502, S505, S510, T512, V517, D520, N522). 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 and the black behind is for experimental data.

The results from multiple spectrum optimizations were condensed with GHOST methodology and summarized in GHOST plots shown in Figure 37 – an example of the most different sites (positions S491 and V517 of Box2 and Box3 correspondingly) are presented with a ϑ–ψ distribution in Figure. In that way the most significant and probable groups of motional patterns of spectral parameters were recognized. Since each of the patterns describes local restrictions to the rotational motion of the probe it can be used to detect structural properties at each specific site in protein. In addition, the weight of a motional pattern represents the relative probability, with which corresponding restriction is detected for particular spin label mutant.

Figure 37: *GHOST condensation plots of spin-labelled N_{TAIL} protein*. GHOST plots shows the optimized multiple solutions represented in a two-dimensional distribution of the angles ϑ-φ and ϑ-dependent distribution of τ_c of spin-labelled N_{TAIL} protein samples (100 µM) in the absence or presence of a molar excess of XD at low (281 K) and high (310 K) temperatures at 12 label positions (the position is indicated with the number in the right bottom corner of the ϑ - φ GHOST plot). For the details about GHOST presentation see Figure 11 in Section 3.1.3.3.

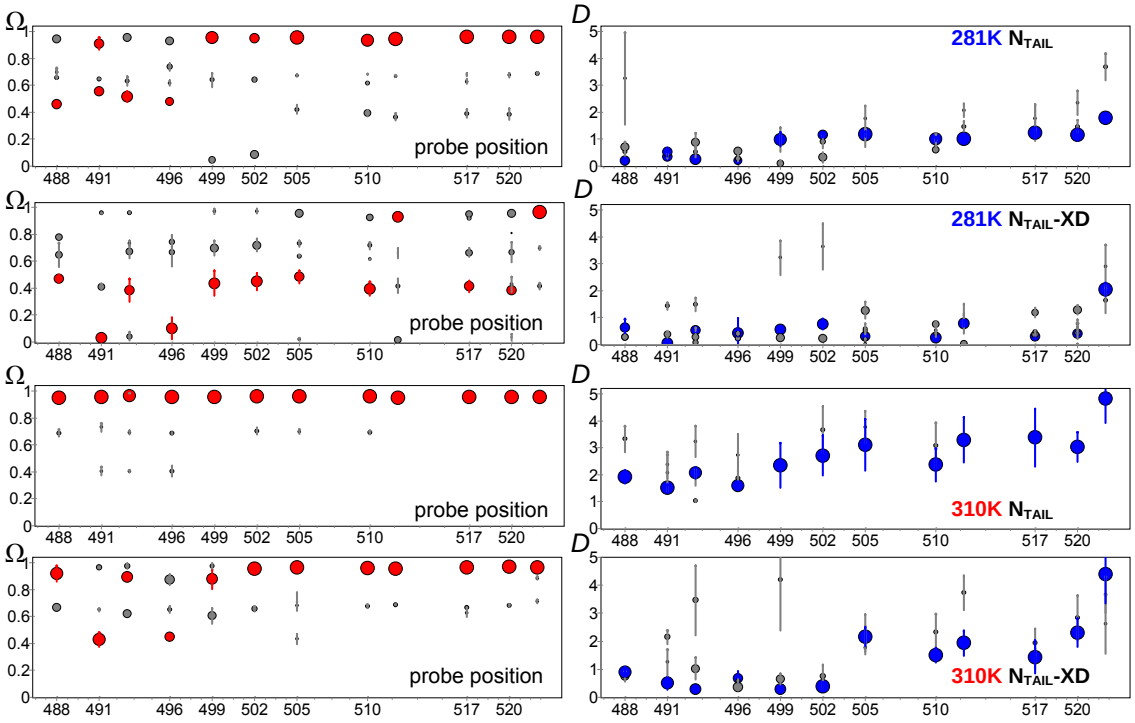

Figure 38: *Bubbles-condensed presentation of the motional patterns for N_{TAIL} protein.* Each bubble diagram presents the motional patterns for 12 mutant positions along N_{TAIL} protein sequence in terms of free rotational space Ω (left) and rotational diffusion *D* (right). With grey colour the less important motional patterns are presented. For the details about bubble diagram presentation see Figure 12 in Section 3.1.3.4.

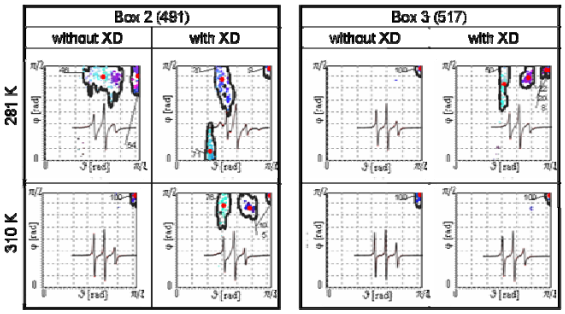

Figure 39: *GHOST condensation plots with corresponding EPR spectra of spin-labelled N_{TAIL} protein at selected mutant positions*. GHOST plots shows the optimized multiple solutions in terms of the opening and asymmetry angles ϑ and φ of spin-labelled N_{TAIL} proteins at positions 491 (Box2) and 517 (Box3) at low (281 K) and high (310 K) temperatures for N_{TAIL} in both the free and bound form. For the details about GHOST presentation see Figure 11 in Section 3.1.3.3.

For example, at position 491 (Box2) of N_{TAIL}-XD complex at 291 K (Figure 39) the three patterns, i.e. degrees of local restrictions, are resolved: very restricted motion with the probability of 71%, less restricted with a probability of 20% and fully non-restricted with only 9% contribution. Obviously, such a complexity can be explained in terms of coexisting local conformations. However, at least in case of non-restricted sites one have to be aware also of explanation in terms of too short-lived structure if non-specific labelling can be excluded.

It has to be noted that almost all the experiments with the mutants of N_{TAIL} alone or in complex with XD in a buffer without sucrose show mainly unrestricted rotational motion at all temperatures except for N_{TAIL}-XD complex at lower temperatures. Consequently these experimental results cannot be used for structure determination and are therefore not shown. However, the addition of sucrose decreases the normalized diffusion (see next section) at all positions indicating that the backbone motion slows down for a factor of 2 to 3. The latter is than enough that significant and stable restrictions appear at many positions at various temperatures (Figure 40) thus enabling us to use this information as structural constraints for protein structure optimization. It should be noted that the temperature scans help us to clear those motional patterns that are not significant as discussed in the Methodology section.

4.4.2 Scanning motional restriction along primary sequence

For all available mutants the detected motional patterns were averaged for lower temperatures (279 K - 281 K - 283 K) and for higher temperatures (308 K - 310 K - 312 K) thus enabling us to compare the local properties at two distinguished temperatures, namely "281 K" and "310 K". In further analysis two physical quantities has been precisely addressed: free rotational space Ω, a normalized product of the cone opening angle ϑ and cone anisotropy angle φ (Eq. 15), as well as normalized rotational diffusion D, a rotational mean square displacement, approximated with a quotient of a product of ϑ and φ angles, in an effective rotational correlation time τ_C:

$$D = \vartheta\phi/4\tau_C \,, \tag{18}$$

Free rotational space Ω measuring the space angle, i.e., the surface of the cone, left for local spin label wobbling is shown for all twelve spin-labelled N_{TAIL} and N_{TAIL} with XD samples at 281 and 310 K (Figure 40A and 40C). High values of Ω (close to 1) correspond to nearly unrestricted motional patterns of the spin label, whereas very low values (below 0.2) imply very high restrictions preventing almost all the side-chain conformational motions.

Rotational diffusion D indicates true mobility of the spin label normalized to the full rotational space. Unlike the conventional correlation time τ_C, which is correlated to the free rotational space [172], normalized diffusion allows comparison between the spectra of different rotational restrictions (Figure 40B and 40D).

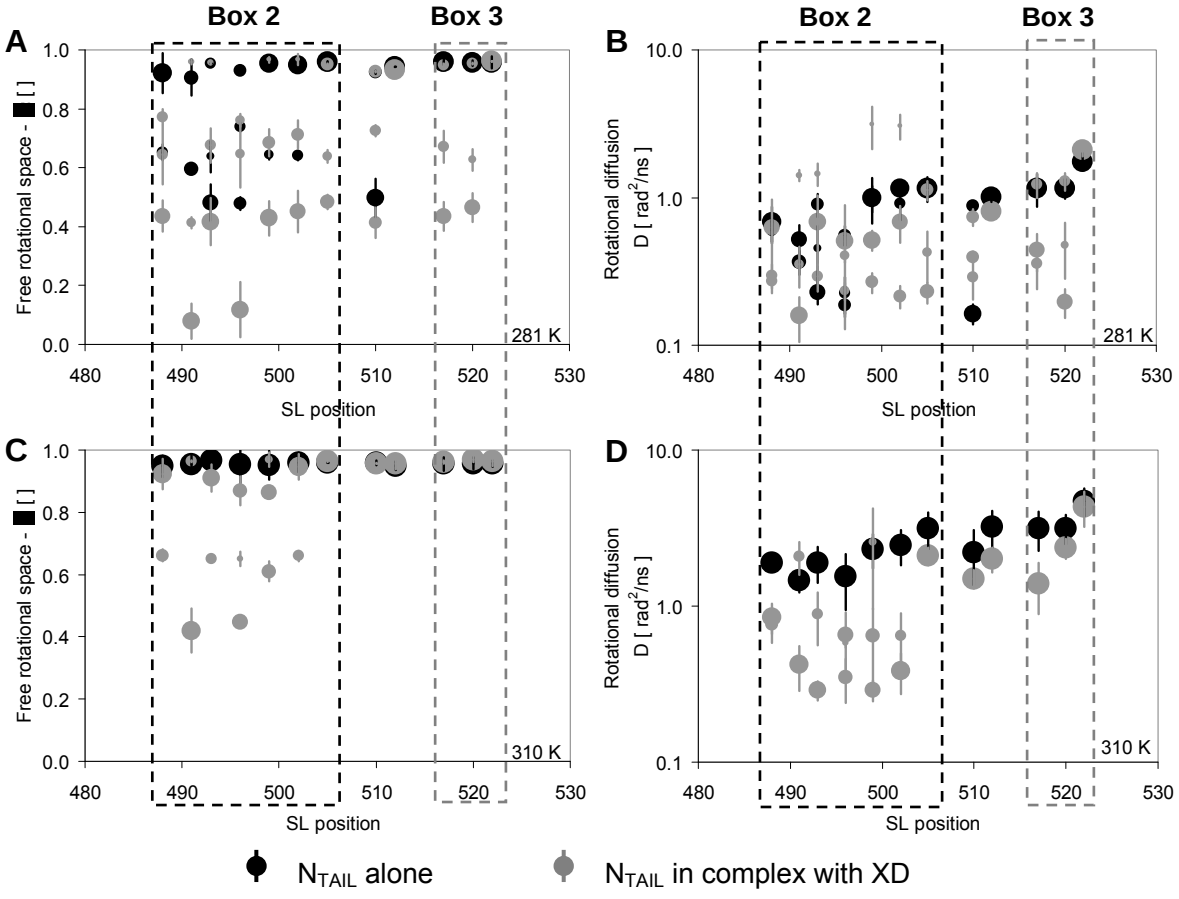

Figure 40: *Free rotational space Ω and diffusion D of N_{TAIL} in 30% sucrose.* The data represents the cleaned motional patterns of bubble diagrams in Figure 38. The data was obtained for N_{TAIL} measured alone (black empty circles) and in complex with XD (grey solid circles). The horizontal axis indicates the spin label position, the vertical axes give Ω, and D, and the error bars point to their second moments. High values of Ω (close to 1) correspond to unrestricted motional patterns of the spin label, whereas low values (between 0 and 0.2) imply very high restrictions. Mutant positions of N_{TAIL} in Box 2 and Box3 are indicated with the black and grey dashed rectangle correspondingly. **A.** Free rotational space Ω of N_{TAIL} at 281K. **B.** Rotational diffusion constant D of N_{TAIL} at 281K. **C.** Free rotational space Ω of N_{TAIL} at 310K. **D.** Rotational diffusion constant D of N_{TAIL} at 310K.

Firstly, in case of N_{TAIL} alone (without partner XD protein) there is no stable restriction at higher

temperatures (Figure 40C, open black circles), while at lower temperatures in Box 2 region some restriction appeared (Figure 40A, open black circles). Normalized diffusion of N_{TAIL} alone mutants reveals too fast motion at higher temperature (Figure 40D, open black circles), the real reason why we cannot detect any motional restrictions. On the other hand, at lower temperature the normalized diffusion is reduced, i.e. the motion slows down, for a factor of 2 in Box 3 region and for almost an order of magnitude at some positions in Box 2 (Figure 40B, open black circles) indicating significant impact of the temperature on the protein disorder in this small temperature interval.

On the other hand N_{TAIL}-XD complex stabilizes the local conformations so the restrictions can be detected already at higher temperatures (Figure 40C, grey filled circles). Also the diffusion is slower, especially in Box 2 regions where stable restrictions are detected (Figure 40D, grey filled circles). At lower temperatures there are even more restrictions detected (Figure 40A and 40B, grey filled circles). Very strong restrictions at positions 491 and 496 also indicate that the spin label is squeezed between two heavy objects, i.e. proteins' backbones of N_{TAIL} and XD.

For many positions more than one different restriction is resolved, from restricted to unrestricted. As we are interested in the most stable and well-defined global structure we will further focus on the lower envelope of these motional patterns, i.e. on the most restricted motional patterns detected at each mutant position. It should be however noted that the protein global conformations can be in principle (but not very likely) such that all possible combinations of detected restrictions at different sites can exist.

4.4.3 Conformational space modelling

Modelling of conformational spaces of N_{TAIL}-XD complex allows tracking of the restriction effect at particular amino acid side chains in different conformations of the protein complex (see Figure 41).

It is expected that the increase of the temperatures the entropic term in free energy function for protein side chains will become more and more important. Thus the inter-chain interactions that stabilize N_{TAIL}-XD complex will become weaker and the N_{TAIL} will search for other free-energy-minima conformations. We surmise that at such conditions rotationally more flexible bulky amino acids at the side of XD, which is in contact with N_{TAIL}, would force N_{TAIL} to tilt relatively to XD in order to minimize the overlap of side chains conformational spaces.

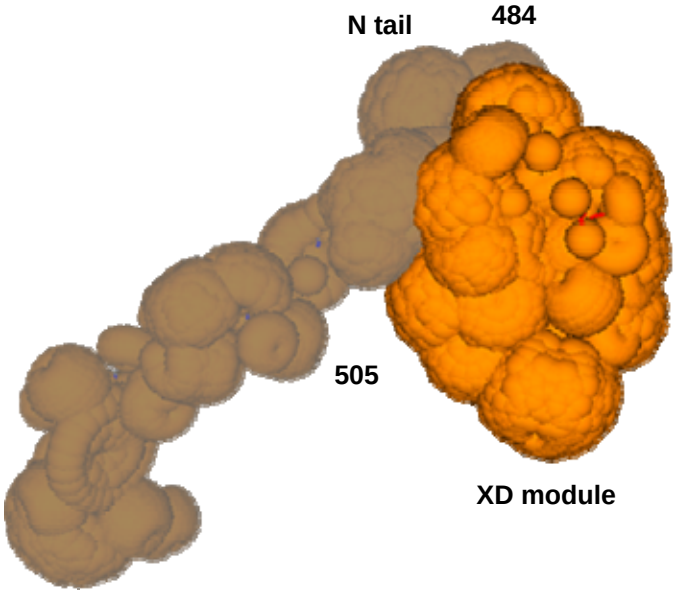

Figure 41: *Conformational space modelling of of N_{TAIL}-XD complex*.

Similar effect of minimization of the overlap between conformational spaces of the side chains is suspected also within single chains, thus, resulting in the unfolding and losing of the secondary structure (helicity).

4.4.4 Protein structure optimization

To enhance the interpretation of EPR results of both systems at low and high temperature we performed an appropriate structural modelling and searched for best-fit-structures at 281 K and 310 K as described previously in methods section. In case of N_{TAIL}-XD complex we firstly checked so-called chimera structure [91] that includes residues 486-505 of N protein with six of our mutant positions (S488, S491, D493, L496, Q499, A502). The calculated free rotational space of the spin label at these sites for the chimera structure is shown in Figure 42A. As can be seen from modelling, positions S491, Q499, A502 in chimera structure are completely restricted, forbidding any conformation of spin label side-chain. Other three positions look to be less restricted. Comparing these restrictions with the ones of the low temperature N_{TAIL}-XD data series (Figure 40A, grey filled circles), we found a reasonable agreement at positions 488 - 491 - 493, but significant difference at positions 496 - 499 - 502.

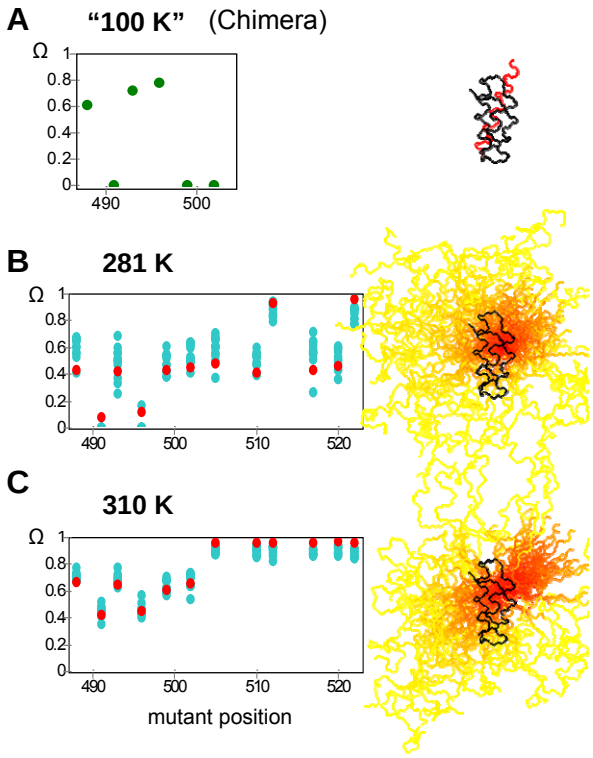

Figure 42: *Optimization of the structure of N_{TAIL} in complex with fixed XD.* Protein complex structure is optimized by fitting the calculated local restrictions (blue circles) to the restrictions profile obtained from SDSL-EPR (red circles). (A) Local restrictions calculated for the fixed chimera structure of N_{TAIL} in complex with XD. (B) and (C) present populations of 50 best-fit structures of N_{TAIL} at 281 K and 310 K correspondingly (together with 10 best fits). The fixed XD is presented with dark colour, while N_{TAIL} structures – with the yellow-red. Colour gradient here represents the structural density in the population of N_{TAIL} structures.

To search what kind of different structure can explain this mismatch we optimized the structures of N_{TAIL} in the N_{TAIL}-XD complex based on the fitting the modelled restrictions to the most probable restrictions detected by SDSL-EPR [176]. As we worked with bigger part of N_{TAIL} protein, we added appropriate 20 residues of N_{TAIL} in the 506-525 region and optimized the structures against data at 281 K and 310 K (high and low temperature). The secondary structure of the missing part of N_{TAIL} was initially set to α-helix.

Fifty best fits from 1000 optimization runs, both for N_{TAIL}-XD complex at 281 K and 310 K, are plotted in Figure 42B and 42C together with the corresponding structures. The orientation of the complex was defined by referring to the tree helical chains of XD from the chimera study (black) as in front and putting N_{TAIL} structures in background (yellow-red coloured; more red indicating more similar structure).

4.4.5 Analysis of secondary structure changes

To summarize the changes in the secondary structure of NTAIL according to the presence of XD and to the temperature, the resulting best-fit structures were analyzed in terms of helicity. For that purpose the backbone dihedral angles of the best-fit structures were collected for each amino acid position in so-called Ramachandran plots [147] and the percentage of α-helical angle pairs was compared against all other pairs (Figure 43). Clearly helical domains are resolved in Box3 region in the case of N_{TAIL} at 281 K (Figure 43A) and of N_{TAIL}-XD at 281 K (Figure 43B) and 310K (Figure 43C). For N_{TAIL} at 310 K helical domains could not be detected since there are no stable restrictions.

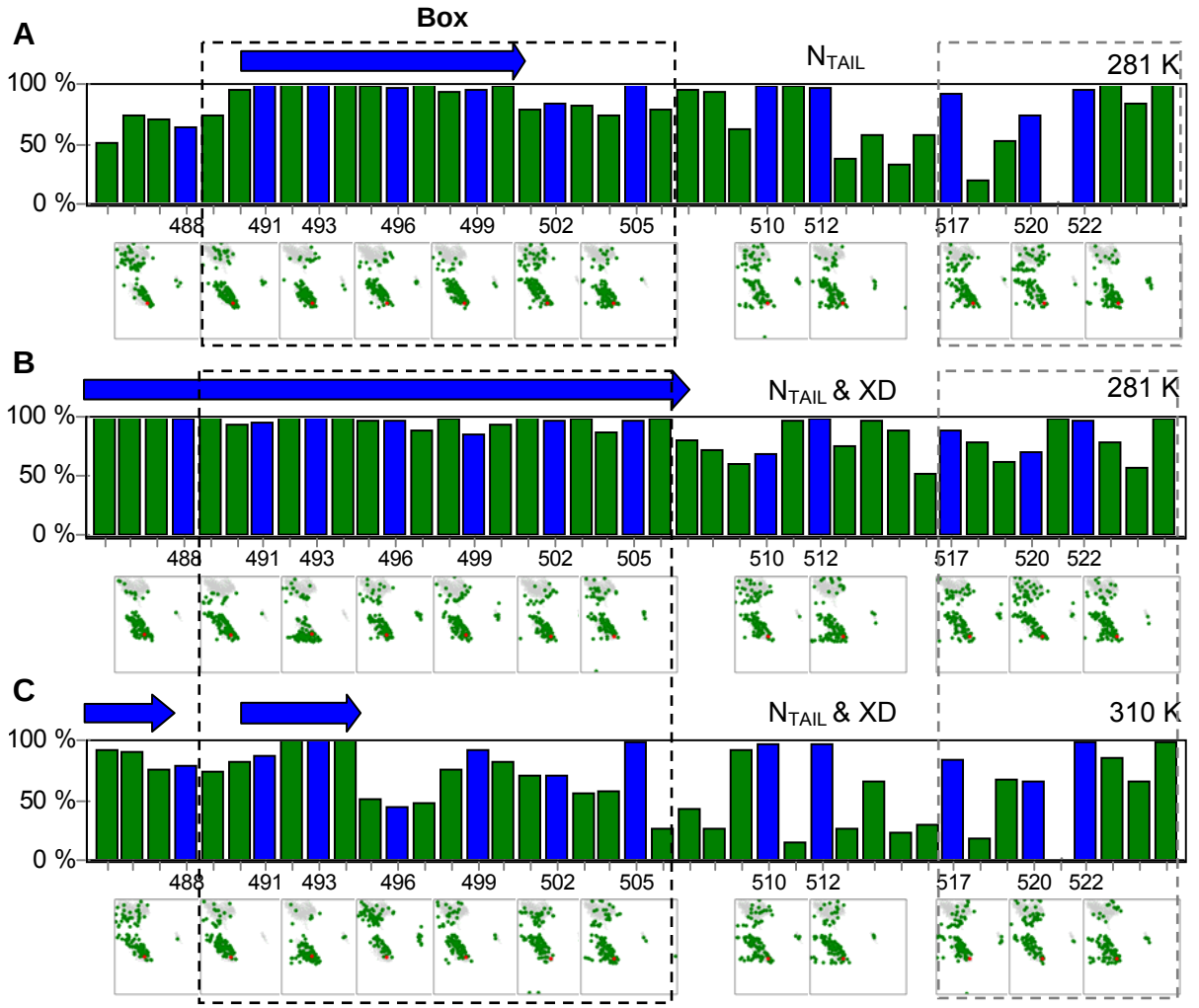

Figure 43: *Analysis of the secondary structure of N_{TAIL} at different conditions*. The height of black (mutant positions) and grey (other positions) bars represent the amount of the helical conformations among 100 best-fit structures. The Ramachandran plots below each mutant position represent the corresponding distribution of the backbone dihedral angles φ and ψ of 100 best-fit structures. Mutant positions of N_{TAIL} in Box 2 and Box3 are indicated with the black and grey dashed rectangle correspondingly. **A.** N_{TAIL} alone. **B.** N_{TAIL} in complex with XD at 281K. **C.** N_{TAIL} in complex with XD at 310K.

4.4.6 Characterization of conformational changes in N_{TAIL}

Determination of a protein dynamics and the corresponding protein function is the final goal of all structural investigations. If protein global conformations are frozen in any way, high resolution methods provide various snapshots of protein structure's rearrangement. Low temperature conditions of such experiments implies that the local protein dynamics is frozen, energy-minimized states are strongly favoured, entropic effects are diminished or ignored [193]. Any fast global structural rearrangement is highly annoying as it is incoherent with slow or even static time-window of the high-resolution methods. It is therefore straightforward why short-lived protein structures like intrinsically disordered proteins still represent not only experimental but also conceptual problem to the scientific community [42].

Two of the main features of the spin labelling EPR spectroscopy are the sensitivity to rotational motion anisotropy as well as the nanosecond time window [68]. To apply them in structure determination one have to search for appropriate dynamical events rather than static structural relations. Taking into account that side-chain and backbone dynamics are in the range of sub nanoseconds and several tens or hundreds of nanoseconds [66,83], consequently, the EPR time window is perfectly suited to distinguish those two motional phenomena. In this case the sensitivity to motional anisotropy can be used to detect restrictions to side-chain rotational conformational spaces, which can then be translated into local structural constraints. And since the backbone is much slower, i.e. static for the EPR spectroscopy window, its conformation could be characterized at EPR nanosecond time window. Noteworthy, the life-time of such a structure should be just longer than EPR nanosecond time window. If backbone experiences a structural transition from folded conformation into a random coil, the time scale of backbone suddenly speeds up and begins to overlap with the time scale of the side-chains. In such a case, EPR is also perfectly suited to detect destabilization of backbone structure. Exactly these phenomena were detected in our study of N_{TAIL} structure in N_{TAIL}-XD complex.

The ability of detection of the short-lived backbone conformations can be understood in a simple experiment by addition of sucrose. At the first glance, the experiment in a sucrose-free solution seems very boring – no stable restriction can be detected at any positions, therefore being not interesting enough to be shown. However, addition of sucrose slows down the backbone motion (for a factor 2 to 3), pushing it away from the side-chain time scale and on another side of EPR time window. Consequently, it enables detection of the local restrictions at EPR nanosecond time scale, indicating that the lifetime of the backbone conformation after addition of sucrose increases from few nanoseconds to ten or few tens of nanoseconds. Obviously, this makes SDSL EPR a unique method capable of detecting nanoseconds-lived structures.

As it can be seen in Figure 40C (black circles) the spin label positioned at any mutation site of N_{TAIL} without partner XD protein at 310 K experiences no restriction at all, while at 281 K (Figure 40A, black circles) significant restrictions appear in the Box 2 region of a length of approximately 14 amino acids. Approximating the number of degrees of freedom in this region to be roughly 900 (9 degrees of freedom per bond and approximately 100 bonds in this segment), the thermal energy difference would correspond to breaking of about 5 to 6 hydrogen bonds. The later is very similar to 7 to 8 hydrogen bonds stabilizing the short 3-turn α-helix of these 14 amino acids. This result strongly favours our hypothesis that the Box 2 helix cannot be stable for few nanoseconds at 310 K but rather at 281 K (Figure 40A and 43A), corroborating with some other experiments speculating that small part of N_{TAIL} possesses helical structure even without partner protein.

After partner protein is brought to N_{TAIL} protein, the complex structure becomes much more stable as can be seen in the restrictions of the spin probe conformational space (Figure 40A an 40C, gray circles). Beside normally restricted sites, additional almost fully restricted sites appear resembling direct overlap of the spin label at N_{TAIL} with conformational space and backbone of XD protein. Such points indicate strong interaction in the neighbourhood. In addition, such points also represent strong distant constraints in structural determination methodology.

Among strong stabilization of N_{TAIL} Box 2 region in form of α-helix by the interaction through XD protein, which can be seen in secondary structure analysis in Figure 43B, we can also identify indirect or weak stabilization in Box 3 region. The interaction is not strong enough to stabilize this region at 310 K (Figure 40C, grey circles). However, when the temperature is decreased to 281 K, the stable restrictions appear (Figure 40A, grey circles). Remember, that this region cannot stabilizes exclusively due to temperature drop (Figure 40C to Figure 40A black circles), indicating indirect stabilization of Box 3.

Despite all the above characterizations, in order to make conclusion about global conformation of the protein, a structure modelling is needed. The usage of molecular dynamics (MD) simulations would be feasible and straightforward if the structure determination would not include fitting experimental data. Since few 10-ns simulations would take few weeks on small cluster, the inverse problem solving would take enormous amount of computer time. On the other hand, to explain restrictions found by SDSL EPR we do not need full time trajectories of all the atoms in the system but only time averages of rotational motion anisotropies. Since EPR is not sensitive to atom coordinates but rather to anisotropy of spin label rotational space, we also do not need atomistic accuracy. Therefore we can approach this problem from the point of view of side-chain conformational spaces [176], which however can be fully explored by the side-chains only at appropriate temperatures. Under such conditions, entropic contribution to the free energy becomes more important and side-chains are not stacked in some of the energetically favourable conformers (rotamers). At this point the modelling approximation exactly meets the experimental conditions and the protein structure determination based on the detection and analysis of the rotational conformational space restrictions under physiological conditions can become very powerful.

By the means of the optimization method we were able to tune the dihedral angles of the N_{TAIL} protein backbone and its relative orientation to XD structure at different temperatures, i.e. whenever the restrictions were detected. By fitting the modelled restrictions of the local conformational spaces to the measured ones we derived the family of structures that equally good describe the experimental data (Figure 42B and 42C). The most striking result is the reorienting of the N_{TAIL} helical part (Box 2 region) relative to the XD helices with increasing temperature. If one starts with the chimera structure (for a bit smaller part of N_{TAIL} protein), detected with X-ray crystallography at lower temperatures, one can see that the "281 K" structure (Figure 42B), based on SDSL EPR measurements, is much less strictly positioned but in average the N_{TAIL} helix tilts away from the direction of the XD helices (Figure 42B, black structure). At "310 K" (Figure 42C) this phenomenon is even more pronounced since the tilt is more obvious and much clearly defined. By exploring the overlap of the side-chains conformational spaces both at N_{TAIL} and XD sites we found out that the tilt is actually due to minimization of the rotational spaces overlap, i.e. maximization of the entropy. The more disordered side-chains of different main chains increase the steric repulsions between chains. Therefore the family of structures modelled at high temperature represents the tendency of the flexible N_{TAIL} to minimize the side-chains contact with XD (Figure 42). It is therefore obvious that at physiological temperatures like 310 K, the structure prediction based on free energy minimization should involve also entropic part and not only energy part.

With the temperature increase from 281 K to 310 K we can see that the structures become less helical (Figure 42B and 42C), supporting our ideas derived exclusively on the spectroscopically detected motional restriction. The same results can be quantitatively confirmed by calculating the proportion of dihedral angle pairs of the best-fit structures found in α-helical region of Ramachandran space (Figure 43). One can see, that the loss of helicity when temperature is increased from 281 K (Figure 43B) to 310 K (Figure 43C) is obvious. As discussed previously, the Box 3 region structure melts completely, while Box 2 helicity melts substantially. At 310 K only a small part of Box 2 can still be found in α-helical conformations.

It should be noted, that the quality of the structure determination strongly depends on the number of spin labelled sites. Since the spin label rotational space is twice as big as an average rotational space of amino acid side-chain, its restrictions arise not only from the first amino cid neighbour (forward and backward) but also from several amino acids in the vicinity, the last being not restricted to the same backbone. It is therefore sufficient to label in average every 3^{rd} or 4^{th} site, enabling one to avoid mutation on some of the critical sites at the same time maintaining the resolution of the method. As it can be seen from our study, 12 mutants enable us to derive temperature dependent backbone conformations of 40 amino acid long short-lived N_{TAIL} protein with an accuracy of a part of nanometre. If there would not be a strong interaction site between N_{TAIL} and XD in Box 2 region which provides strong constraint in the modelling, we would certainly need additional mutants on XD site to resolve N_{TAIL} orientation relative to XD more accurately. With that one should also note that the structure of XD was not optimized but approximated with the XD low temperature structure derived within X-ray crystallography chimera experiment.

5 Conclusions

Fast proteins structural rearrangements are very likely to be observed in protein systems under physiological conditions, especially in case of intrinsically disordered proteins. Also the specific environment (e.g. membrane) can be very critical for protein functionality. Therefore the development of alternative structure determination methodology is crucial to complement the structural picture provided by the well-established high-resolution techniques. In this perspective we herein presented a novel combination of site directed spin labelling EPR spectroscopy and molecular modelling that both describe restrictions to side-chain conformational spaces, stable at EPR nanosecond time window. The restrictions are then used to navigate the optimization of a protein backbone conformation, which finally provides a family of equally-good global conformations of the protein.

In the first part of the thesis we developed and implemented a novel shaking operator in order to reduce the computational demand of the original multiple HEO approach and carried out an extensive testing on various spectra that represent a wide range of possible applications. With this new modification of the optimization algorithm we succeeded to keep the quality of the EPR-based characterization, thereby considerably reducing the computational time of the multiple EPR spectral analysis by a factor of 5-10 making the application of this advanced EPR spectra characterization to complex biosystems, such as proteins and biological membranes, more feasible. Further numerical calculations on both synthetic and experimental data [86,94,135,171,172,190] proved the efficiency of the enhanced algorithm and opened new possibilities for its application [84,176].

In the second part of this work, the method of protein structure modelling, conformational spaces simulations, and local structural restriction calculations was developed. The sensitivity of the modelled conformational space restrictions to primary and secondary structural elements as well as to lipid arrangements (in case of membrane proteins) was extensively tested and proven. The method characterizes the simulated restrictions in the same way as it is done in the analysis of experimental EPR spectra.

In the third part, we developed an optimization algorithm, which tunes a protein structure and fits simulated restrictions of the spin label free rotational space to the experimental data obtained from SDSL-EPR spectra. In case of a reference membrane protein (the membrane-embedded M13 coat protein reconstituted in 14:1PC bilayers), a multi-run optimization results in a family of favourable protein structures, which not only agree with the available SDSL-EPR data, but also are consistent with the previously published model based on site-directed fluorescence labelling [93,131,132].The proposed method was then applied to study structural characteristics of partially disordered measles virus N_{TAIL}-XD complex in physiological conditions. We discovered that increase of the temperature in the system is accompanied by the tilting of N_{TAIL} helix relative to XD helices in order to maximize N_{TAIL}-XD contact entropy as well as by the temperature-induced N_{TAIL} helix melting. By application of sucrose to the protein solution we found out that N_{TAIL} structure last for about few to few tens of nanoseconds revealing SDSL EPR as the unique experimental method able to determine the structure of such a short lived protein structure in its native environment.

Our work clearly demonstrates that the simultaneous analysis of available SDSL-EPR data and structural modelling can provide information about protein structure. Let us stress, that the structural modelling can easily include additional data from primary structure analysis (secondary structure predictions, hydropathy index [102,202] calculation) and results from other experimental techniques (X-ray crystallography high-resolution structures, structural information from NMR spectroscopy, fluorescence spectroscopy, infrared spectroscopy, and circular dichroism). In this perspective, structural modelling is thought to be a connecting link, which transfers multiple structural data into a high-resolution structure or structural characterization revealing the functional properties of intrinsically disordered proteins, membrane proteins, not being limited to other classes of proteins. The present method therefore provides a challenging starting point for the development of a powerful methodology for the protein structure characterization, an alternative approach to conventional techniques.

6 Acknowledgements

First of all, I would like to thank my 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. Together with our lab colleagues I have been learning from you how we should encourage and motivate each other as you always try to make us feel positive and optimistic about what we do. 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.

Marcus, thank you! It was a great opportunity to learn Life Sciences 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.

I would like to thank our 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 Ziherl 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. Thank you to colleagues from Marseille, with whom we worked on N_{TAIL}-XD complex: Valérie Belle, Sabrina Rouger, Stéphanie Costanzo, André Fournel, Bruno Guigliarelli, and Sonia Longhi.

Thank you to Prof. Dr. Igor Muševič, the head of our department of solid state physics at JSI for supporting and stimulating our research work! 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 that made the GHOSTMaker a more powerful and user-friendly! Thank you, Daniele, for the critical look at the text of the dissertation!

Thank you to MPS, to Prof. Dr. Aleksander Zindanšek, Sergeja, Zvonka, and to all colleagues from IPS for motivating support and momentary help during my postgraduate studies and thesis preparation!

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 Laptenok (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!

7 References

1 Alexov, E. Role of the protein side-chain fluctuations on the strength of pair-wise electrostatic interactions: comparing experimental with computed pK(a)s. *Proteins* **50** (1), 94-103 (2003).

2 Allen, J.P. *Biophysical chemistry* (Wiley-Blackwell Pub., Oxford; Hoboken, NJ, 2008).

3 Almeida, F.C. & Opella, S.J. fd coat protein structure in membrane environments: structural dynamics of the loop between the hydrophobic trans-membrane helix and the amphipathic in-plane helix. *Journal of Molecular Biology* **270** (3), 481-495 (1997).

4 Altenbach, C.; Oh, K.J.; Trabanino, R.J.; Hideg, K. & Hubbell, W.L. Estimation of inter-residue distances in spin labeled proteins at physiological temperatures: experimental strategies and practical limitations. *Biochemistry* **40** (51), 15471-15482 (2001).

5 Anderson, R.G.W. & Jacobson, K. A role for lipid shells in targeting proteins to caveolae, rafts, and other lipid domains. *Science* **296** (5574), 1821-1825 (2002).

6 Ansari, A.; Berendzen, J.; Bowne, S.F.; Frauenfelder, H.; Iben, I.E.; Sauke, T.B.; Shyamsunder, E. & Young, R.D. Protein states and proteinquakes. *Proceedings of the National Academy of Sciences of the United States of America* **82** (15), 5000-5004 (1985).

7 Balgavy, P.; Dubnickova, M.; Kucerka, N.; Kiselev, M.A.; Yaradaikin, S.P. & Uhrikova, D. Bilayer thickness and lipid interface area in unilamellar extruded 1,2-diacylphosphatidylcholine liposomes: a small-angle neutron scattering study. *Biochimica et Biophysica Acta* **1512** (1), 40-52 (2001).

8 Bashtovyy, D.; Marsh, D.; Hemminga, M.A. & Pali, T. Constrained modeling of spin-labeled major coat protein mutants from M13 bacteriophage in a phospholipid bilayer. *Protein Science* **10** (5), 979-987 (2001).

9 Bax, A. Two-dimensional NMR and protein structure. *Annual Review of Biochemistry* **58**, 223-256 (1989).

10 Beier, C. & Steinhoff, H.-J. A structure-based simulation approach for electron paramagnetic resonance spectra using molecular and stochastic dynamics simulations. Biophysical Journal **91** (7), 2647-2664 (2006).

11 Belle, V.; Fournel, A.; Woudstra, M.; Ranaldi, S.; Prieri, F.; Thome, V.; Currault, J.; Verger, R.; Guigliarelli, B. & Carriere, F. Probing the opening of the pancreatic lipase lid using site-directed spin labeling and EPR spectroscopy. *Biochemistry* **46** (8), 2205-2214 (2007).

12 Belle, V.; Rouger, S.; Costanzo, S.; Liquiere, E.; Štrancar, J.; Guigliarelli, B.; Fournel, A. & Longhi, S. Mapping alpha-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. *Proteins* **73** (4), 973-988 (2008).

13 Berliner, L.J.; Eaton, G.R. & Eaton, S.S. *Distance measurements in biological systems by EPR* (Kluwer Academic/Plenum Publishers, New York, 2000).

14 Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T. & Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology* **112** (3), 535-542 (1977).

15 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).

16 Bond, P.J.; Holyoake, J.; Ivetac, A.; Khalid, S. & Sansom, M.S.P. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *Journal of Structural Biology* **157** (3), 593-605 (2007).

17 Borbat, P.P. & Freed, J.H. Multiple-quantum ESR and distance measurements. *Chemical Physics Letters* **313** (1-2), 145-154 (1999).

18 Borbat, P.P.; McHaourab, H.S. & Freed, J.H. Protein structure determination using long-distance constraints from double-quantum coherence ESR: study of T4 lysozyme. *Journal of the American Chemical Society* **124** (19), 5304-5314 (2002).

19 Bourhis, J.-M.; Canard, B. & Longhi, S. Structural disorder within the replicative complex of measles virus: Functional implications. *Virologie* **9** (5), 367-383 (2005).

20 Bourhis, J.-M.; Canard, B. & Longhi, S. Structural disorder within the replicative complex of measles virus: Functional implications. *Virology* **344** (1), 94-110 (2006).

21 Bourhis, J.-M.; Johansson, K.; Receveur-Bréchot, V.; Oldfield, C.J.; Dunker, K.A.; Canard, B. & Longhi, S. The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Research* **99** (2), 157-167 (2004).

22 Bourhis, J.-M.; Receveur-Brechot, V.; Oglesbee, M.; Zhang, X.; Buccellato, M.; Darbon, H.; Canard, B.; Finet, S. & Longhi, S. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains

predominantly unfolded. *Protein Science* **14** (8), 1975-1992 (2005).
23 Bower, M.J.; Cohen, F.E. & Dunbrack, R.L., Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *Journal of Molecular Biology* **267** (5), 1268-1282 (1997).
24 Budil, D.E.; Sale, K.L.; Khairy, K.A. & Fajer, P.G. Calculating slow-motional electron paramagnetic resonance spectra from molecular dynamics using a diffusion operator approach. **110** (10), 3703-3713 (2006).
25 Capaldi, S.; Guariento, M.; Saccomani, G.; Fessas, D.; Perduca, M. & Monaco, H.L. A single amino acid mutation in zebrafish (Danio rerio) liver bile acid-binding protein can change the stoichiometry of ligand binding. *Journal of Biological Chemistry* **282** (42), 31008-31018 (2007).
26 Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K. & Oschkinat, H. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420** (6911), 98-102 (2002).
27 Charles M. Deber, S.-C.L. Peptides in membranes: Helicity and hydrophobicity. *Biopolymers* **37** (5), 295-318 (1995).
28 Choi, V. On updating torsion angles of molecular conformations. *Journal of Chemical Information and Modeling* **46** (1), 438-444 (2006).
29 Columbus, L. & Hubbell, W.L. A new spin on protein dynamics. *Trends in Biochemical Sciences* **27** (6), 288-295 (2002).
30 Columbus, L. & Hubbell, W.L. Mapping backbone dynamics in solution with site-directed spin labeling: GCN4-58 bZip free and bound to DNA. *Biochemistry* **43** (23), 7273-7287 (2004).
31 Columbus, L.; Kalai, T.; Jeko, J.; Hideg, K. & Hubbell, W.L. Molecular motion of spin labeled side chains in alpha-helices: analysis by variation of side chain structure. *Biochemistry* **40** (13), 3828-3846 (2001).
32 Creighton, T.E. *Proteins: Structures and molecular properties*, 2nd ed (W.H. Freeman, New York, 1993).
33 Cross, T.A. & Opella, S.J. Protein structure by solid state nuclear magnetic resonance. Residues 40 to 45 of bacteriophage fd coat protein. *Journal of Molecular Biology* **182** (3), 367-381 (1985).
34 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).
35 de Planque, M.R. & Killian, J.A. Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring. *Molecular Membrane Biology* **20** (4), 271-284 (2003).
36 de Planque, M.R.R.; Bonev, B.B.; Demmers, J.A.A.; Greathouse, D.V.; Koeppe, R.E.; Separovic, F.; Watts, A. & Killian, J.A. Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide-lipid interactions. *Biochemistry* **42** (18), 5341-5348 (2003).
37 Deb, K. Master thesis, Genetic Algorithms in Multimodal Function Optimization. University of Alabama, 1989.
38 DeSensi, S.C.; Rangel, D.P.; Beth, A.H.; Lybrand, T.P. & Hustedt, E.J. Simulation of nitroxide electron paramagnetic resonance spectra from brownian trajectories and molecular dynamics simulations. *Biophysical Journal* **94** (10), 3798-3809 (2008).
39 Diallo, A.; Barrett, T.; Barbron, M.; Meyer, G. & Lefevre, P.C. Cloning of the nucleocapsid protein gene of peste-des-petits-ruminants virus: relationship to other morbilliviruses. *Journal of General Virology* **75 (Pt 1)**, 233-237 (1994).
40 Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; Ausio, J.; Nissen, M.S.; Reeves, R.; Kang, C.; Kissinger, C.R.; Bailey, R.W.; Griswold, M.D.; Chiu, W.; Garner, E.C. & Obradovic, Z. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling* **19** (1), 26-59 (2001).
41 Dunker, A.K. & Obradovic, Z. The protein trinity--linking function and disorder. *Nature Biotechnology* **19** (9), 805-806 (2001).
42 Dyson, H.J. & Wright, P.E. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **6** (3), 197-208 (2005).
43 Edidin, M. Lipids on the frontier: a century of cell-membrane bilayers. *Nature Reviews Molecular Cell Biology* **4** (5), 414-418 (2003).
44 Eiben, A.E. & Smith, J.E. *Introduction to evolutionary computing* (Springer, New York, 2003).
45 Engh, R.A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica Section A* **47** (4), 392-400 (1991).
46 Fanucci, G.E. & Cafiso, D.S. Recent advances and applications of site-directed spin labeling. *Current Opinion in Structural Biology* **16** (5), 644-653 (2006).
47 Fasman, G.D. *Circular dichroism and the conformational analysis of biomolecules* (Plenum Press, New York, 1996).
48 Fernandes, F.; Loura, L.M.; Koehorst, R.; Spruijt, R.B.; Hemminga, M.A.; Fedorov, A. & Prieto, M. Quantification of protein-lipid selectivity using FRET: application to the M13 major coat protein. *Biophysical Journal* **87** (1), 344-352 (2004).
49 Filipič, B. & Štrancar, J. Tuning EPR spectral parameters with a genetic algorithm. *Applied Soft Computing* **1** (1), 83-90 (2001).
50 Fink, A.L. Natively unfolded proteins. *Current Opinion in Structural Biology* **15** (1), 35-41 (2005).
51 Fleishman, S.J.; Unger, V.M. & Ben-Tal, N. Transmembrane protein structures without X-rays. *Trends in*

Biochemical Sciences **31** (2), 106-113 (2006).
52 Fogel, D.B.; Bäck, T. & Michalewicz, Z. *Evolutionary computation* (Institute of Physics Publishing, Bristol; Philadelphia, 2000).
53 Gawrisch, K. The dynamics of membrane lipids. In: Yeagle, P. (ed.) *The structure of biological membranes*. 147-171 (CRC Press, Boca Raton, Fla, 2005).
54 Goldberg, D.E. *Genetic algorithms in search, optimization, and machine learning* (Addison-Wesley Pub. Co., Reading, Mass., 1989).
55 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).
56 Gordon, V.S.; Whitley, D. & Bohn, A. Dataflow parallelism in genetic algorithms. In: Männer, R. & Manderick, B. (ed.) *Parallel problem solving from nature, 2: Proceedings of the Second Conference on Parallel Problem Solving from Nature*. 533−542 (Elsevier Science, Amsterdam, The Netherlands, 1992).
57 Grigoryan, G.; Ochoa, A. & Keating, A.E. Computing van der Waals energies in the context of the rotamer approximation. *Proteins* **68** (4), 863-878 (2007).
58 Gu, J. & Hilser, V.J. Predicting the energetics of conformational fluctuations in proteins from sequence: a strategy for profiling the proteome. *Structure* **16** (11), 1627-1637 (2008).
59 Gumbart, J.; Wang, Y.; Aksimentiev, A.; Tajkhorshid, E. & Schulten, K. Molecular dynamics simulations of proteins in lipid bilayers. *Current Opinion in Structural Biology* **15** (4), 423-431 (2005).
60 Håkanssona, P.; Westlund, P.O.; Lindahl, E. & Edholm, O. A direct simulation of EPR slow-motion spectra of spin labelled phospholipids in liquid crystalline bilayers based on a molecular dynamics simulation of the lipid dynamics. *Physical Chemistry Chemical Physics* **3**, 5311-5319 (2001).
61 Hancock, J.F. Lipid rafts: contentious only from simplistic standpoints. *Nature Reviews Molecular Cell Biology* **7** (6), 456-462 (2006).
62 Heggeness, M.H.; Scheid, A. & Choppin, P.W. Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proceedings of the National Academy of Sciences of the United States of America* **77** (5), 2631-2635 (1980).
63 Hemminga, M.A. Introduction and future of site-directed spin labeling of membrane proteins. In: Hemminga, M.A. & Berliner, L. (ed.) *ESR Spectroscopy in Membrane Biophysics*. 1-16 (Springer, 2007).
64 Hemminga, M.A. & Berliner, L.J. *ESR spectroscopy in membrane biophysics* (Springer, New York, 2007).
65 Henderson, R. Realizing the potential of electron cryo-microscopy. *Quarterly Reviews of Biophysics* **37** (1), 3-13 (2004).
66 Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450** (7172), 964-972 (2007).
67 Ho, B.K.; Thomas, A. & Brasseur, R. Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the α-helix. *Protein Science* **12** (11), 2508-2522 (2003).
68 Hoff, A.J. *Advanced EPR: applications in biology and biochemistry* (Elsevier, Amsterdam New York, 1989).
69 Hovmoller, S.; Zhou, T. & Ohlson, T. Conformations of amino acids in proteins. *Acta Crystallographica Section D* **58** (5), 768-776 (2002).
70 Hubbell, W.L. & Altenbach, C. Investigation of structure and dynamics in membrane proteins using site-directed spin labeling. *Current Opinion in Structural Biology* **4** (4), 566-573 (1994).
71 Hubbell, W.L.; Cafiso, D.S. & Altenbach, C. Identifying conformational changes with site-directed spin labeling. *Nature Structural Biology* **7** (9), 735-739 (2000).
72 Hubbell, W.L.; McHaourab, H.S.; Altenbach, C. & Lietzow, M.A. Watching proteins move using site-directed spin labeling. *Structure* **4** (7), 779-783 (1996).
73 Hustedt, E.J.; Stein, R.A.; Sethaphong, L.; Brandon, S.; Zhou, Z. & Desensi, S.C. Dipolar coupling between nitroxide spin labels: the development and application of a tether-in-a-cone model. *Biophysical Journal* **90** (1), 340-356 (2006).
74 Israelachvili, J.N. Refinement of the fluid-mosaic model of membrane structure. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **469** (2), 221-225 (1977).
75 Jacobson, K.; Mouritsen, O.G. & Anderson, R.G.W. Lipid rafts: at a crossroad between cell biology and physics. *Nature Cell Biology* **9** (1), 7-14 (2007).
76 Jensen, M.Ø. & Mouritsen, O.G. Lipids do influence protein function - the hydrophobic matching hypothesis revisited. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1666** (1-2), 205-226 (2004).
77 Jeschke, G.; Bender, A.; Paulsen, H.; Zimmermann, H. & Godt, A. Sensitivity enhancement in pulse EPR distance measurements. *Journal of Magnetic Resonance* **169** (1), 1-12 (2004).
78 Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Physical Chemistry Chemical Physics* **9** (16), 1895-1910 (2007).
79 Jeschke, G.; Wegener, C.; Nietschke, M.; Jung, H. & Steinhoff, H.J. Interresidual distance determination by four-pulse double electron-electron resonance in an integral membrane protein: the Na+/proline transporter PutP of Escherichia coli. *Biophysical Journal* **86** (4), 2551-2557 (2004).
80 Johansson, A.C. & Lindahl, E. Amino-acid solvation structure in transmembrane helices from molecular dynamics simulations. *Biophysical Journal* **91** (12), 4450-4463 (2006).

81 Johansson, K.; Bourhis, J.M.; Campanacci, V.; Cambillau, C.; Canard, B. & Longhi, S. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *Journal of Biological Chemistry* **278** (45), 44567-44573 (2003).

82 Karlin, D.; Longhi, S. & Canard, B. Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association. *Virology* **302** (2), 420-432 (2002).

83 Karplus, M. & McCammon, J.A. The internal dynamics of globular proteins. *CRC Critical Reviews in Biochemistry* **9** (4), 293-349 (1981).

84 Kavalenka, A.; Hemminga, M.A. & Štrancar, J. Optimization of membrane protein structure based on SDSL-ESR constraints and conformational space modeling. Submitted (2009).

85 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 Modeling* **45** (6), 1628-1635 (2005).

86 Kavalenka, A.A.; Spruijt , R.B.; Wolfs, C.J.A.M.; Štrancar, J.; Croce, R.; Hemminga, M.A. & van Amerongen, H. Site-directed spin labeling study of the light-harvesting complex CP29. *Biophysical Journal* **96** (9), 3620-3628 (2009).

87 Kelly, S.M. & Price, N.C. The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science* **1** (4), 349-384 (2000).

88 Killian, J.A. & Nyholm, T.K.M. Peptides in lipid bilayers: the power of simple models. *Current Opinion in Structural Biology* **16** (4), 473-479 (2006).

89 Killian, J.A. & von Heijne, G. How proteins adapt to a membrane-water interface. *Trends in Biochemical Sciences* **25** (9), 429-434 (2000).

90 King, M.W. Inborn Errors in Metabolism. http://themedicalbiochemistrypage.org/inborn.html, (2009).

91 Kingston, R.L.; Hamel, D.J.; Gay, L.S.; Dahlquist, F.W. & Matthews, B.W. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proceedings of the National Academy of Sciences of the United States of America* **101** (22), 8301-8306 (2004).

92 Kirkpatrick, S.; Gelatt, C.D., Jr. & Vecchi, M.P. Optimization by simulated annealing. *Science* **220** (4598), 671-680 (1983).

93 Koehorst, R.B.M.; Spruijt, R.B.; Vergeldt, F.J. & Hemminga, M.A. Lipid bilayer topology of the transmembrane α-helix of M13 major coat protein and bilayer polarity profile by site-directed fluorescence spectroscopy. *Biophysical Journal* **87**, 1445-1455 (2004).

94 Kubale, V.; Abramovic, Z.; Pogacnik, A.; Heding, A.; Sentjurc, M. & Vrecl, M. Evidence for a role of caveolin-1 in neurokinin-1 receptor plasma-membrane localization, efficient signaling, and interaction with beta-arrestin 2. *Cell and Tissue Research* **330** (2), 231-245 (2007).

95 Kuzdzal, M. EPR studies of a drug in liposomes and cells. Submitted (2009).

96 Lacapere, J.-J.; Pebay-Peyroula, E.; Neumann, J.-M. & Etchebest, C. Determining membrane protein structures: still a challenge! *Trends in Biochemical Sciences* **32** (6), 259-270 (2007).

97 LaConte, L.E.; Voelz, V.; Nelson, W.; Enz, M. & Thomas, D.D. Molecular dynamics simulation of site-directed spin labeling: experimental validation in muscle fibers. *Biophysical Journal* **83** (4), 1854-1866 (2002).

98 Lee, A.G. Lipid-protein interactions in biological membranes: a structural perspective. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1612** (1), 1-40 (2003).

99 Lee, S.W. & Belcher, A.M. Virus-based fabrication of micro- and nanofibers using electrospinning (2004), Vol. 4, pp. 387-390.

100 Lehninger, A.L.; Nelson, D.L. & Cox, M.M. *Lehninger principles of biochemistry*, 4th ed (W.H. Freeman, New York, 2005).

101 Li, J.-P.; Balazs, M.E.; Parks, G.T. & Clarkson, P.J. A species conserving genetic algorithm for multimodal function optimization. *Evolutionary Computation* **10** (3), 207-234 (2002).

102 Li, S.-C. & Deber, C.M. A measure of helical propensity for amino acids in membrane environments. *Nature Structural and Molecular Biology* **1** (6), 368-373 (1994).

103 Lietzow, M.A. & Hubbell, W.L. Motion of spin label side chains in cellular retinol-binding protein: correlation with structure and nearest-neighbor interactions in an antiparallel β-sheet. **43** (11), 3137-3151 (2004).

104 Lindahl, E. & Sansom, M.S. Membrane proteins: molecular dynamics simulations. *Current Opinion in Structural Biology* **18** (4), 425-431 (2008).

105 Longhi, S.; Receveur-Brechot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S. & Canard, B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *Journal of Biological Chemistry* **278** (20), 18638-18648 (2003).

106 Lovell, S.C.; Davis, I.W.; Arendall, W.B., 3rd; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S. & Richardson, D.C. Structure validation by Cα geometry: φ,ψ and Cβ deviation. Proteins **50** (3), 437-450 (2003).

107 Luckey, M. *Membrane structural biology: with biochemical and biophysical foundations* (Cambridge University Press, Cambridge ; New York, 2008).

108 MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.;

Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D. & Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102** (18), 3586-3616 (1998).

109 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).

110 Mahfoud, S.W. Ph.D. thesis, *Niching Methods for Genetic Algorithms*. University of Illinois at Urbana-Champaign, 1995.

111 Makowski, L. Terminating a macromolecular helix. Structural model for the minor proteins of bacteriophage M13. *Journal of Molecular Biology* **228** (3), 885-892 (1992).

112 Mao, C.; Solis, D.J.; Reiss, B.D.; Kottmann, S.T.; Sweeney, R.Y.; Hayhurst, A.; Georgiou, G.; Iverson, B. & Belcher, A.M. Virus-based toolkit for the directed synthesis of magnetic and semiconducting nanowires. *Science* **303** (5655), 213-217 (2004).

113 Marassi, F.M. & Opella, S.J. Simultaneous assignment and structure determination of a membrane protein from NMR orientational restraints. *Protein Science* **12** (3), 403-411 (2003).

114 Marrink, S.J.; Risselada, H.J.; Yefimov, S.; Tieleman, D.P. & de Vries, A.H. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **111** (27), 7812-7824 (2007).

115 Marsh, D. Electron Spin Resonance: Spin Labels. In: Grell, E. (ed.) *Membrane Spectroscopy*. 51-142 (Springer-Verlag, Berlin New York, 1981).

116 Marsh, D. Electron spin resonance in membrane research: Protein-lipid interactions. Methods **46** (2), 83-96 (2008).

117 Marsh, D. Protein modulation of lipids, and vice-versa, in membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1778** (7-8), 1545-1575 (2008).

118 Marsh, D. & Horvath, L.I. Spin label studies of the structure and dynamics of lipids and proteins in membranes. In: Hoff, A.J. (ed.) *Advanced EPR: Applications in Biology and Biochemistry*. 707-752 (Elsevier, 1989).

119 Marsh, D. & Horvath, L.I. Structure, dynamics and composition of the lipid-protein interface. Perspectives from spin-labelling. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **1376** (3), 267-296 (1998).

120 Marsh, D. & Pali, T. The protein-lipid interface: perspectives from magnetic resonance and crystal structures. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1666** (1-2), 118-141 (2004).

121 Martin, W.N., Lienig, J., Cohoon, J.P. Island (migration) models: Evolutionary algorithms based on punctuated equilibria. In: Bäck, T., Fogel, D.B., & Michalewics, Z. (ed.) *Handbook of Evolutionary Computation*. C6.3:1-C6.3:1 (Institute of Physics Publishing, Bristol, UK, 1998).

122 Marvin, D.A. Filamentous phage structure, infection and assembly. *Current Opinion in Structural Biology* **8** (2), 150-158 (1998).

123 Marvin, D.A.; Welsh, L.C.; Symmons, M.F.; Scott, W.R. & Straus, S.K. Molecular structure of fd (f1, M13) filamentous bacteriophage refined with respect to X-ray fibre diffraction and solid-state NMR data supports specific models of phage assembly at the bacterial membrane. *Journal of Molecular Biology* **355** (2), 294-309 (2006).

124 Meijer, A.B.; Spruijt, R.B.; Wolfs, C.J. & Hemminga, M.A. Configurations of the N-terminal amphipathic domain of the membrane-bound M13 major coat protein. *Biochemistry* **40** (16), 5081-5086 (2001).

125 Meijer, A.B.; Spruijt, R.B.; Wolfs, C.J. & Hemminga, M.A. Membrane-anchoring interactions of M13 major coat protein. *Biochemistry* **40** (30), 8815-8820 (2001).

126 Mellgren, R.L. Structural biology: Enzyme knocked for a loop. *Nature* **456** (7220), 337-338 (2008).

127 Mitra, K.; Ubarretxena-Belandia, I.; Taguchi, T.; Warren, G. & Engelman, D.M. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proceedings of the National Academy of Sciences of the United States of America* **101** (12), 4083-4088 (2004).

128 Morin, B.; Bourhis, J.M.; Belle, V.; Woudstra, M.; Carriere, F.; Guigliarelli, B.; Fournel, A. & Longhi, S. Assessing induced folding of an intrinsically disordered protein by site-directed spin-labeling electron paramagnetic resonance spectroscopy. *The Journal of Physical Chemistry B* **110** (41), 20596-20608 (2006).

129 Muller, D.J. & Engel, A. Strategies to prepare and characterize native membrane proteins and protein membranes by AFM. *Current Opinion in Colloid & Interface Science* **13** (5), 338-350 (2008).

130 Nagle, J.F. & Tristram-Nagle, S. Structure of lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1469** (3), 159-195 (2000).

131 Nazarov, P.V.; Koehorst, R.B.M.; Vos, W.L.; Apanasovich, V.V. & Hemminga, M.A. FRET study of membrane proteins: simulation-based fitting for analysis of protein structure, membrane embedment and association. *Biophysical Journal* **91**, 454-466 (2006).

132 Nazarov, P.V.; Koehorst, R.B.M.; Vos, W.L.; Apanasovich, V.V. & Hemminga, M.A. FRET study of membrane proteins: determination of the tilt and orientation of the N-terminal domain of M13 major coat protein. *Biophysical Journal* **92** (4), 1296-1305 (2007).

133 Nordio, P.L. General magnetic resonance theory. In: Berliner, L.J. (ed.) *Spin labeling: theory and applications*. 5-51 (Academic Press, New York, 1976).

134 Opella, S.J.; Zeri, A.C. & Park, S.H. Structure, dynamics, and assembly of filamentous bacteriophages by

nuclear magnetic resonance spectroscopy. *Annual Review of Physical Chemistry* **59**, 635-657 (2008).
135 Pabst, G.; Hodzic, A.; Štrancar, J.; Danner, S.; Rappolt, M. & Laggner, P. Rigidification of neutral lipid bilayers in the presence of salts. *Biophysical Journal* **93** (8), 2688-2696 (2007).
136 Palsdottir, H. & Hunte, C. Lipids in membrane protein structures. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1666** (1-2), 2-18 (2004).
137 Papavoine, C.H.; Aelen, J.M.; Konings, R.N.; Hilbers, C.W. & Van de Ven, F.J. NMR studies of the major coat protein of bacteriophage M13. Structural information of gVIIIp in dodecylphosphocholine micelles. *European Journal of Biochemistry* **232** (2), 490-500 (1995).
138 Papavoine, C.H.; Christiaans, B.E.; Folmer, R.H.; Konings, R.N. & Hilbers, C.W. Solution structure of the M13 major coat protein in detergent micelles: a basis for a model of phage assembly involving specific residues. *Journal of Molecular Biology* **282** (2), 401-419 (1998).
139 Papavoine, C.H.; Konings, R.N.; Hilbers, C.W. & van de Ven, F.J. Location of M13 coat protein in sodium dodecyl sulfate micelles as determined by NMR. *Biochemistry* **33** (44), 12990-12997 (1994).
140 Pebay-Peyroula, E. *Biophysical analysis of membrane proteins: investigating structure and function* (Wiley-VCH, Weinheim, 2008).
141 Petoukhov, M.V. & Svergun, D.I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical Journal* **89** (2), 1237-1250 (2005).
142 Petrowski, A. A clearing procedure as a niching method for genetic algorithms. In: *Proceedings of the IEEE International Conference of Evolutionary Computation (ICEC '96)*. 798-803 (Nagoya, Japan, 1996).
143 Pietzsch, J. Mind the membrane. In: *A living frontier – exploring the dynamics of the cell membrane* (Palazzo Arzaga, Italy, 2004).
144 Pistolesi, S.; Ferro, E.; Santucci, A.; Basosi, R.; Trabalzini, L. & Pogni, R. Molecular motion of spin labeled side chains in the C-terminal domain of RGL2 protein: A SDSL-EPR and MD study. *Biophysical Chemistry* **123** (1), 49-57 (2006).
145 Rainey, J.K. & Goh, M.C. Statistically based reduced representation of amino acid side chains. *Journal of Chemical Information and Computer Sciences* **44** (3), 817-830 (2004).
146 Ramachandran, G.N.; Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**, 95-99 (1963).
147 Ramachandran, G.N. & Sasisekharan, V. Conformation of polypeptides and proteins. *Advances in Protein Chemistry* **23**, 283-438 (1968).
148 Robinson, B.; Thomann, H.; Beth, A.; Fayer, P. & Dalton, L.R. The phenomenon of magnetic resonance: Theoretical considerations. In: Dalton, L.R. (ed.) *EPR and Advanced EPR Studies of Biological Systems*. 11−110. (CRC Press, Boca Raton, Fla., 1985).
149 Sale, K.; Sar, C.; Sharp, K.A.; Hideg, K. & Fajer, P.G. Structural determination of spin label immobilization and orientation: A Monte Carlo minimization approach. *Journal of Magnetic Resonance* **156** (1), 104-112 (2002).
150 Sale, K.; Song, L.; Liu, Y.S.; Perozo, E. & Fajer, P. Explicit treatment of spin labels in modeling of distance constraints from dipolar EPR and DEER. *Journal of the American Chemical Society* **127** (26), 9334-9335 (2005).
151 Sanders, J.C.; Haris, P.I.; Chapman, D.; Otto, C. & Hemminga, M.A. Secondary structure of M13 coat protein in phospholipids studied by circular dichroism, Raman, and Fourier transform infrared spectroscopy. *Biochemistry* **32** (46), 12446-12454 (1993).
152 Sareni, B. & Krahenbuhl, L. Fitness sharing and niching methods revisited. *IEEE Transactions on Evolutionary Computation* **2** (3), 97-106 (1998).
153 Schindler, H. & Seelig, J. EPR spectra of spin labels in lipid bilayers. *The Journal of Chemical Physics* **59** (4), 1841-1850 (1973).
154 Schneider, D.J. & Freed, J.H. Calculating slow motional magnetic resonance spectra: A user's guide. In: Berliner, L.J. & Reuben, J. (ed.) *Biological Magnetic Resonance: Spin Labeling, Theory and Applications*. 1−76 (Plenum Press, New York, 1989).
155 Schweiger, A. & Jeschke, G. *Principles of pulse electron paramagnetic resonance* (Oxford University Press, Oxford, UK ; New York, 2001).
156 Scott, K.A.; Bond, P.J.; Ivetac, A.; Chetwynd, A.P.; Khalid, S. & Sansom, M.S.P. Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* **16** (4), 621-630 (2008).
157 Sezer, D.; Freed, J.H. & Roux, B. Simulating electron spin resonance spectra of nitroxide spin labels from molecular dynamics and stochastic trajectories. *Journal of Chemical Physics* **128** (16), 165106 (2008).
158 Shetty, R.P.; De Bakker, P.I.; DePristo, M.A. & Blundell, T.L. Advantages of fine-grained side chain conformer libraries. *Protein Engineering* **16** (12), 963-969 (2003).
159 Singer, S.J. & Nicolson, G.L. The fluid mosaic model of the structure of cell membranes. *Science* **175** (23), 720-731 (1972).
160 Spears, W. Simple subpopulation schemes. In: Sebald, A.V. & Fogel, L.J. (ed.) *Proc. Third Annual Conf. Evolutionary Programming (EP'94)*. 296-307 (World Scientific, Singapore, 1994).
161 Spruijt, R.B.; Meijer, A.B.; Wolfs, C.J. & Hemminga, M.A. Localization and rearrangement modulation of the N-terminal arm of the membrane-bound major coat protein of bacteriophage M13. *Biochimica et Biophysica Acta* **1509** (1-2), 311-323 (2000).

162 Spruijt, R.B.; Wolfs, C.J. & Hemminga, M.A. Aggregation-related conformational change of the membrane-associated coat protein of bacteriophage M13. *Biochemistry* **28** (23), 9158-9165 (1989).
163 Spruijt, R.B.; Wolfs, C.J.; Verver, J.W. & Hemminga, M.A. Accessibility and environment probing using cysteine residues introduced along the putative transmembrane domain of the major coat protein of bacteriophage M13. *Biochemistry* **35** (32), 10383-10391 (1996).
164 Standfuss, J.; Terwisscha van Scheltinga, A.C.; Lamborghini, M. & Kuhlbrandt, W. Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 A resolution. *EMBO Journal* **24** (5), 919-928 (2005).
165 Steinhoff, H.-J.; Müller, M.; Beier, C. & Pfeiffer, M. Molecular dynamics simulation and EPR spectroscopy of nitroxide side chains in bacteriorhodopsin. *Journal of Molecular Liquids* **84** (1), 17-27 (2000).
166 Steinhoff, H.-J.; Savitsky, A.; Wegener, C.; Pfeiffer, M.; Plato, M. & Mцbius, K. High-field EPR studies of the structure and conformational changes of site-directed spin labeled bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1457** (3), 253-262 (2000).
167 Steinhoff, H.J. & Hubbell, W.L. Calculation of electron paramagnetic resonance spectra from Brownian dynamics trajectories: application to nitroxide side chains in proteins. *Biophysical Journal* **71** (4), 2201-2212 (1996).
168 Stoica, I. Using molecular dynamics to simulate electronic spin resonance spectra of T4 lysozyme. *The Journal of Physical Chemistry B* **108** (5), 1771-1782 (2004).
169 Stopar, D.; Spruijt, R.B. & Hemminga, M.A. Anchoring mechanisms of membrane-associated M13 major coat protein. *Chemistry and Physics of Lipids* **141** (1-2), 83-93 (2006).
170 Stopar, D.; Spruijt, R.B.; Wolfs, C.J. & Hemminga, M.A. Protein-lipid interactions of bacteriophage M13 major coat protein. *Biochimica et Biophysica Acta* **1611** (1-2), 5-15 (2003).
171 Stopar, D.; Štrancar, J.; Spruijt, R.B. & Hemminga, M.A. Exploring the local conformational space of a membrane protein by site-directed spin labeling. *Journal of Chemical Information and Modeling* **45** (6), 1621-1627 (2005).
172 Stopar, D.; Štrancar, J.; Spruijt, R.B. & Hemminga, M.A. Motional restrictions of membrane proteins: a site-directed spin labeling study. *Biophysical Journal* **91** (9), 3341-3348 (2006).
173 Štrancar, J. Advanced ESR spectroscopy in membrane biophysics. In: Hemminga, M.A. & Berliner, L. (ed.) *ESR Spectroscopy in Membrane Biophysics*. 49-93 (Springer, 2007).
174 Štrancar, J. EPRSIM-C: A Spectral Analysis Package. In: Hemminga, M.A. & Berliner, L. (ed.) *ESR Spectroscopy in Membrane Biophysics*. 323-341 (Springer, 2007).
175 Štrancar, J. EPRSIM-C: A Spectral Analysis Package. http://www.ijs.si/ijs/dept/epr/EPRSIMC_overview.htm, (2009).
176 Štrancar, J.; Kavalenka, A.; Ziherl, 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).
177 Štrancar, J.; Koklic, T. & Arsov, Z. Soft picture of lateral heterogeneity in biomembranes. *Journal of Membrane Biology* **196** (2), 135-146 (2003).
178 Štrancar, J.; Koklic, T.; Arsov, Z.; Filipic, B.; Stopar, D. & Hemminga, M.A. Spin label EPR-based characterization of biosystem complexity. *Journal of Chemical Information and Modeling* **45** (2), 394-406 (2005).
179 Štrancar, J.; Sentjurc, M. & Schara, M. Fast and accurate characterization of biological membranes by EPR spectral simulations of nitroxides. *Journal of Magnetic Resonance* **142** (2), 254-265 (2000).
180 Streichert, F.; Stein, G.; Ulmer, H. & Zell, A. A clustering based niching method for evolutionary algorithms. In: Cantú-Paz, E. *et al.* (ed.) *Genetic and Evolutionary Computation (GECCO 2003)*. 644-645 (Springer, Berlin, 2003).
181 Svergun, D.I. & Koch, M.H.J. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* **66** (10), 1735-1782 (2003).
182 Taylor, R.D.; Jewsbury, P.J. & Essex, J.W. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *Journal of Computational Chemistry* **24** (13), 1637-1656 (2003).
183 Thompson, M.A. ArgusLab 4.0. http://www.ArgusLab.com, (2009).
184 Tombolato, F.; Ferrarini, A. & Freed, J.H. Dynamics of the nitroxide side chain in spin-labeled proteins. *The Journal of Physical Chemistry B* **110** (51), 26248-26259 (2006).
185 Tombolato, F.; Ferrarini, A. & Freed, J.H. Modeling the effects of structure and dynamics of the nitroxide side chain on the ESR spectra of spin-labeled proteins. *Journal of Physical Chemistry B* **110** (51), 26260-26271 (2006).
186 Tompa, P. Intrinsically unstructured proteins. *Trends in Biochemical Sciences* **27** (10), 527-533 (2002).
187 Torres, J.; Stevens, T.J. & Samso, M. Membrane proteins: the 'Wild West' of structural biology. *Trends in Biochemical Sciences* **28** (3), 137-144 (2003).
188 Tsvetkov, Y.D.; Milov, A.D. & Maryasov, A.G. Pulsed electron–electron double resonance (PELDOR) as EPR spectroscopy in nanometre range. *Russian Chemical Reviews* (6), 487 (2008).
189 Ulmschneider, M.B. & Ulmschneider, J.P. Folding Peptides into Lipid Bilayer Membranes. *Journal of Chemical Theory and Computation* **4** (11), 1807-1809 (2008).

190 Urbančič, I. & Štrancar, J. High-throughput spin label EPR spectra analysis reveals biased reporting. Submitted (2009).
191 Ursem, R.K. Multinational evolutionary algorithms. In: *Congress of Evolutionary Computation (CEC 1999)*. 1633-1640 (IEEE Press, Washington, DC, USA, 1999).
192 Uversky, V.N. Natively unfolded proteins: a point where biology waits for physics. *Protein Science* **11** (4), 739-756 (2002).
193 van Gunsteren, W.F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D.P.; Glattli, A.; Hunenberger, P.H.; Kastenholz, M.A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N.F. & Yu, H.B. Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie. International Ed. In English* **45** (25), 4064-4092 (2006).
194 Vasquez, M. An evaluvation of discrete and continuum search techniques for conformational analysis of side chains in proteins. *Biopolymers* **36** (1), 53-70 (1995).
195 Vasquez, M. Modeling side-chain conformation. *Current Opinion in Structural Biology* **6** (2), 217-221 (1996).
196 Venturoli, M.; Maddalena Sperotto, M.; Kranenburg, M. & Smit, B. Mesoscopic models of biological membranes. *Physics Reports* **437** (1-2), 1-54 (2006).
197 Vos, W.L.; Koehorst, R.B.M.; Spruijt, R.B. & Hemminga, M.A. Membrane-bound conformation of M13 major coat protein: A structure validation through FRET-derived constraints. *Journal of Biological Chemistry* **280**, 38522-38527 (2005).
198 Vos, W.L.; Nazarov, P.V.; Koehorst, R.B.M.; Spruijt, R.B. & Hemminga, M.A. From 'I' to 'L' and back again: the odyssey of membrane-bound M13 protein. *Trends in Biochemical Sciences*, in press (2009).
199 Watts, A. Solid-state NMR approaches for studying the interaction of peptides and proteins with membranes. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **1376** (3), 297-318 (1998).
200 Webpage. RCSB Protein Data Bank. http://www.rcsb.org/pdb/Welcome.do, (2009).
201 White, S. Membrane proteins of known 3D structure. http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html, (2009).
202 Wimley, W.C. & White, S.H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural and Molecular Biology* **3** (10), 842-848 (1996).
203 Word, J.M.; Lovell, S.C.; LaBean, T.H.; Taylor, H.C.; Zalis, M.E.; Presley, B.K.; Richardson, J.S. & Richardson, D.C. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *Journal of Molecular Biology* **285** (4), 1711-1733 (1999).
204 Xiang, Z. & Honig, B. Extending the accuracy limits of prediction for side-chain conformations. *Journal of Molecular Biology* **311** (2), 421-430 (2001).
205 Yin, X. & Germay, N. A fast genetic algorithm with sharing scheme using cluster analysis methods in multimodal function optimization. In: Albrecht, R.F., Reeves, C.R., & Steele, N.C. (ed.) *Interernational Conference on Artificial Neural Nets and Genetic Algorithms*. 450-457 (1993).
206 Zhang, M. & Kavraki, L.E. A new method for fast and accurate derivation of molecular conformations. *Journal of Chemical Information and Computer Sciences* **42** (1), 64-70 (2002).

Index of Figures

Index of Tables

Appendix A

GHOSTMaker software from EPRSIM-C: a spectral analysis package

The EPRSIM-C software package provides a tool for nitroxide-based characterization of:

- Specifically and nonspecifically labelled biological membranes
- Site-directed spin labelled proteins
- Biopolymer networks explored by specific labelling or concentration imaging
- Nanomaterials
- Other interesting complex biological systems labelled with different labelling techniques

EPRSIM-C package is capable of performing an automatic nitroxide-based characterization of a biological system by implementing simpler and still accurate simulation models, powerful optimization routines as well as data condensation techniques that enable the user to parameterize the complex system in a more understandable way [173]. For increased efficiency the different tasks are handled within different programs (see Figure 44).

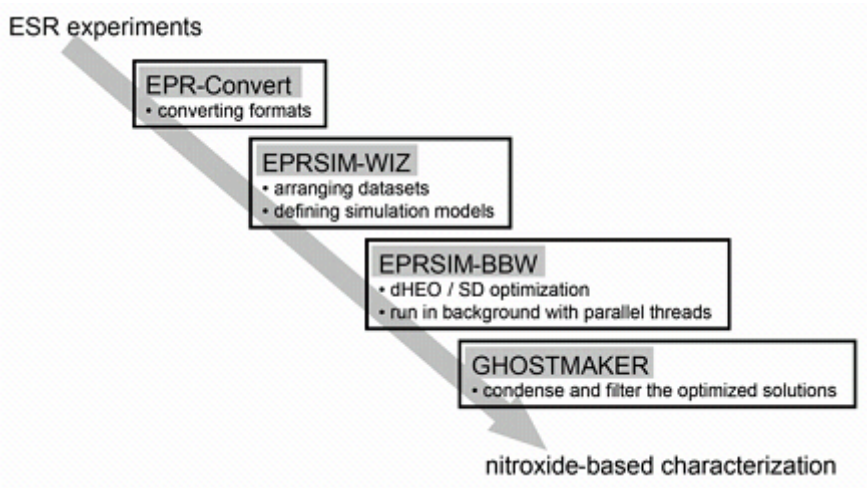

Figure 44: *EPRSIM-C software package concept.* Specific tasks of the individual program modules are indicated.

All the intensive numerical tasks – in fact, the optimizations – are done in background within EPRSIM-BBW in a way that one can submit the tasks from a network in a simple way. The simulation models are based on partial motional averaging [173]. The user should understand that the complex phase space can only be searched by a stochastic population-based optimization routine. This approach takes a great deal of computational time. However, much faster local search methods are not applicable in general, as the user can never be sure that there are not equally good solutions in other parts of the phase space.

The user interface of EPRSIM-WIZ helps to arrange and check the experimental and simulated data sets and is not meant to show the progress of the optimization routines. The main idea of having this kind of interface separated is to arrange clearly the experimental and simulation data. Various formats of experimental series can be read directly or translated into a proper format within EPR-Convert.

Inverse problem solving is separated from the data condensation routine implemented in GHOSTMAKER to increase the efficiency of the calculations. This module filters the data sets resulting from the optimization routine according to goodness of fit and solution density. In addition, it helps to condensate the solutions into groups of various complexity [174].

GHOSTMaker software is used to condense and present the results of the analysis of EPR spectra (by spectral simulations and optimization).

The purpose of GHOSTMaker is to condense the multi-run dHEO solutions into GHOSTs to be able to represent and distinguish groups of solutions. In general, one can extract at least the best solution from each run when a genetic algorithm is applied for optimization. As this kind of optimization is purely stochastic, one should compare at least the best solutions from each run to determine the uncertainty of solutions. If the proposed model complexity is large enough, the uncertainty represents the pure deviation of solutions. However, if the proposed complexity is too low and the optimization finds larger regions of the phase space to describe the spectral data, these uncertainties also represent the distribution of solutions [174].

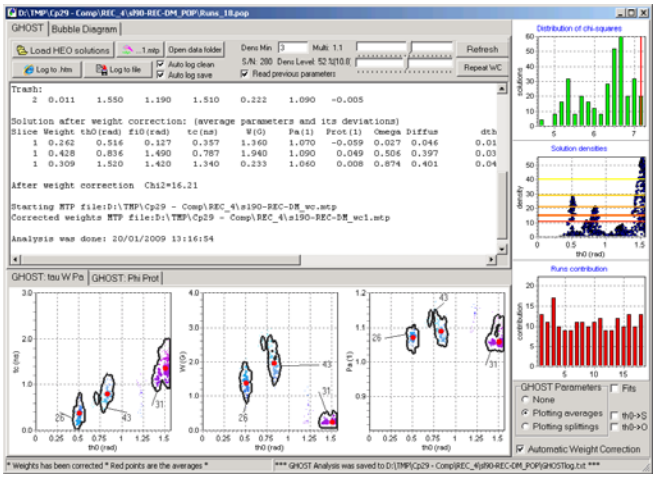

Figure 45: *GHOSTMaker software working window (single EPR data analysis)*.

The program operates in the following modes:

- Single EPR data analysis (view simulations of single spectrum)
- Multiple EPR data analysis (simulations of series of spectra)
- Several series comparison

In Single EPR data mode (GHOST page, see Figure 45) the results of single spectrum simulations can be loaded, condensed, viewed graphically or exported if needed. In Multiple EPR data mode (Bubbles page, see Figure 46) the GHOST results (from single mode) can be viewed and analyzed simultaneously with so-called bubble diagram (see section 3.1.3.4). In addition, multiple series of data could be loaded and compared in bubble diagram (Figure 46).

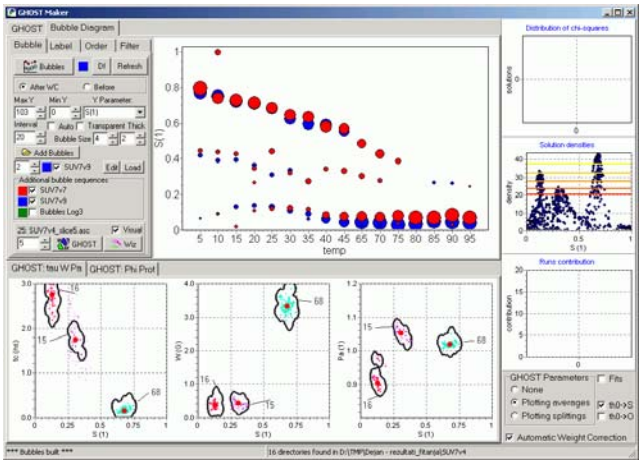

Figure 46: *GHOSTMaker software working window (multiple EPR data analysis)*.

Appendix B

Protein structure modelling

The modelling of the protein structure makes use of previously reported fixed bond lengths and bond angles [45,108,203] that are presented in Tables 8 and 9.

Tuning interatomic contact distances with Ramachandran plot calculation

The secondary structure of a protein is parameterized with pairs of backbone dihedral angles φ_i and ψ_i (Figure 47A). 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 plot [146]. To derive the φ_i - ψ_i distribution, a list of backbone atom pairs has to be checked for steric interatomic contacts [67]. 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 plots [67,69,106,146]. As a consequence the van der Waals radii of the atoms (Table 10) needed to be reduced in agreement with the literature [57,67,158,184]. Note that the reduction of the "apparent" van der Waals radii simulates the effect of atomic interactions when checking steric overlaps [23,194,195]. 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 10). The atom pairs in Table 10 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 47B) and the distribution of the angles φ and ψ based on PDB structures analysis [106] (Figure 47C). 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 reference [67,69,106,146] 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 47D). 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 11 with the reduced van der Waals atoms radii.

Table 8: *Values of the chemical bond lengths used in the modelling of the protein structure*. These values are based on previously reported constants [45,108,203].

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 9: *Values of the chemical bond angles used in the modelling of the protein structure*. These values are based on previously reported constants [45,108,203].

Bond	Bond angle (º)
– C –	109.5
– C =	120.0
– N –	120.0
O–H	104.5
S–H	104.5

Table 10: *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 [57,67,184].

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_{β}, 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 11: *Reduced van der Waals atoms radii*. This radii values were used in the modelling of the conformational space of the amino acid side chains in accordance with [23,57,158], 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

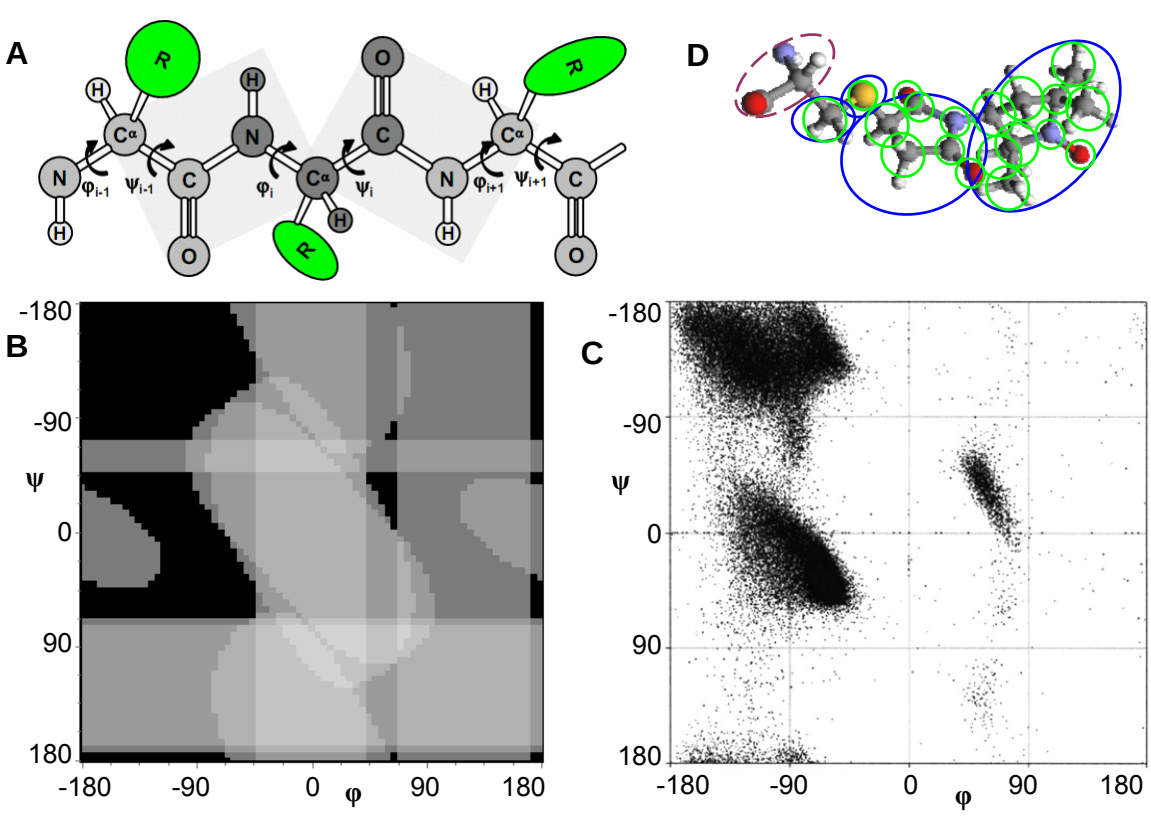

Figure 47: *Protein secondary structure parameterization*. **A.** Definition of the backbone dihedral angles. Available backbone free rotations and corresponding dihedral angles are shown on a three-residue model. Amino acid side chains are schematically presented with the green ovals marked with 'R'. **B.** Ramachandran plot: the distribution of allowed φ and ψ backbone dihedral angles (islands coloured 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 Bank [106]. **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.

Appendix C

Publications

Papers in international scientific journals:

1. 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).
2. Štrancar, J.; Kavalenka, A.; Ziherl, 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).
3. 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).

   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.

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: Krasnoproshin, 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: Krasnoproshin, V. & Aluja, J.G. (ed.) *MS'2004 - International Conference on Modelling and Simulation*. 130-134 (Minsk, Belarus, 2004).

Curriculum vitae

Aleh Kavalenka received his master degree and started his postgraduate studies at the Department of Systems Analysis at Belarusian State University (Minsk, Belarus) under supervision of Prof. Dr. Vladimir Apanasovich in 2003. In years 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 on Advanced Spectroscopy under supervision of Dr. Marcus Hemminga. From 2004 he has been 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. His scientific work is related to Membrane Proteins Structure Determination, Spectroscopic Data Analysis, Protein Structure Modelling, and Evolutionary Optimization. From November 2006 till September 2009 he was carrying out his PhD studies at the Jožef Stefan International Postgraduate School.

Speeding Up a Genetic Algorithm for EPR-Based Spin Label Characterization of Biosystem Complexity

Aleh A. Kavalenka,[†] Bogdan Filipič,[‡] Marcus A. Hemminga,[§] and Janez Štrancar*,[‖]

Department of Systems Analysis, Belarusian State University, F. Skorina Avenue 4, Minsk 220050, Belarus, Department of Intelligent Systems and Laboratory of Biophysics, Solid State Physics Department, Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia, and Laboratory of Biophysics, Wageningen University, Dreijenlaan 3, NL-6703 HA Wageningen, The Netherlands

Received April 28, 2005

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 Paramagnetic Resonance (EPR) 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 "quasicontinuous" 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 EPR 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.

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). The complexity of a biological membrane, for example, originates in its biochemical composition of a few hundred 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.

EPR spectroscopy in combination with nitroxide spin labeling (SL-EPR) has proven to be a powerful technique for the exploration of heterogeneity and motion in biological systems.[1,2] The time scale of SL-EPR 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 applied.[3] 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 developed.[1] 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 ap-

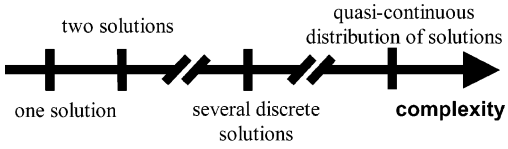

Figure 1. Biosystem complexity axis of increasing complexity from simple single-solution to quasicontinuous distribution of solutions.

* Corresponding author phone: +386 1 477 32 26; fax: +386 1 477 31 91; e-mail: Janez.strancar@ijs.si.
[†] Belarusian State University.
[‡] Department of Intelligent Systems, Jožef Stefan Institute.
[§] Wageningen University.
[‖] Laboratory of Biophysics, Solid State Physics Department, Jožef Stefan Institute.

10.1021/ci0501589 CCC: $30.25
Published on Web 08/23/2005

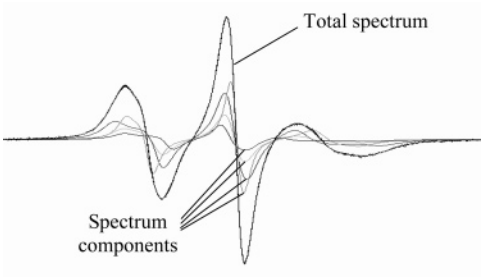

Figure 2. Superimposed four component EPR spectrum. The total EPR spectrum is a sum of four spectral components derived within a described simulation model and determined by four sets of the following spectral parameters $\{\vartheta, \varphi, \tau_c, W, p_A, \text{Prot}, d\}$.

proach named Hybrid Evolutionary Optimization (HEO) was shown to be powerful enough to study complex heterogeneous systems,[1] 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 quasicontinuous 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 h 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.

THEORY AND METHODOLOGY

EPR Spin Labeling. EPR 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 analogues 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 nonlabeled molecules. This fact enables us to use EPR to explore biological systems in vivo so that there is no need for (bio)chemical extraction of the subsystem of interest. In this way, various coexisting states of the system can be detected and characterized.

As was mentioned in the Introduction, EPR spin labeling inherits a unique sensitivity to the motional and polarity properties of the labeled molecules providing an opportunity to extract 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 EPR spectrum of such a labeled system (Figure 2).

EPR Spectrum Modeling. Generally, to describe the EPR spectra of spin labels, the stochastic Liouville equation should be used.[4−6] However, under physiological conditions the majority of the local rotational motions is fast with respect to the EPR 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 emphasize only the physical background of the spectral parameters involved in our calculations. First, 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 ϑ (that defines the maximal tilt angle) and asymmetry cone angle φ (that describes the maximal restriction of spinning) will be applied. Second, the traces of the interaction tensors **g** and **A** are linearly corrected with p_A[9] and Prot parameters to take into account the effects of polarity and proticity, respectively. Third, when calculating the convolution of the magnetic field distribution and basic line shape, in addition two line width parameters are applied: a Lorentzian-type line is defined in the motional narrowing approximation[10] 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 line shape parameters ϑ, 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-EPR which is around 30 parameters, this allows the usage of at most 4 spectral components.

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

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

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 elsewhere,[11] 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).[1,11] 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.

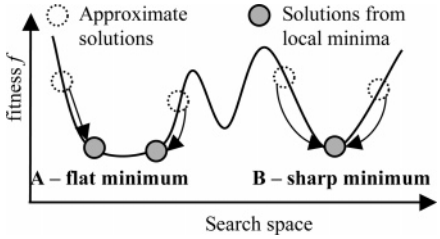

Figure 3. Schematic presentation of parameters search space and the effect of the local mutation procedure responsible for fine-tuning. Due to the noisy spectra and finite resolution of the local optimization routine starting approximations (white circles) are optimized into more accurate solutions (gray 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.

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. To do that, one should clearly understand the HEO as well as the problem search space.

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.

Population Diversity in GA. 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)[12] 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[12−14] 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;[15,16] parallel niching methods (sharing,[17] crowding,[14,18] and clearing[13]);

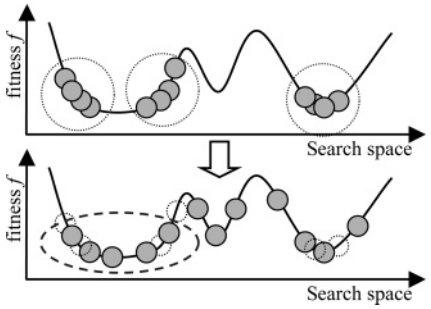

Figure 4. Schematic presentation of the fitness sharing operator function. Top: In a nonsharing routine crowding at the local minima is allowed, since there is no operator that would maintain diversity. Bottom: In a sharing-routine, fitness function is increased according to the density of solution, aiming to prevent crowding.

speciation methods[19−21] and clustering;[22,23] and multipopulation methods[24] (island models[25,26] and migration models[27]).

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.[28] 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.

Sharing. Sharing[14,17] requires that fitness is shared as a single resource among similar individuals in a population of solutions.[29] 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[30]

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

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

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

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 4).

Shaking. Shaking is a new operator that was developed to provide small Gaussian-like deviations to the spectral parameters (Figure 5) 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).

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 quasicontinuous 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

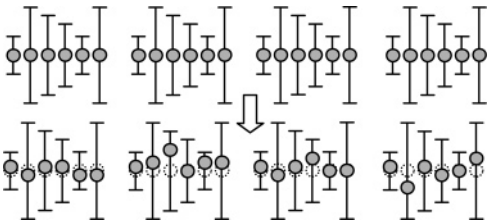

Figure 5. Schematic presentation of the Gaussian shaking operator. Shaking operator implies a Gaussian random generator that provides a small deviation to the value of each parameter. The error bars indicate the width of Gaussian probability distribution of these deviations. The standard relative uncertainties of the spectral parameters $\{\vartheta, \varphi, \tau_c, W, p_A, \text{Prot}, d\}$ 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.

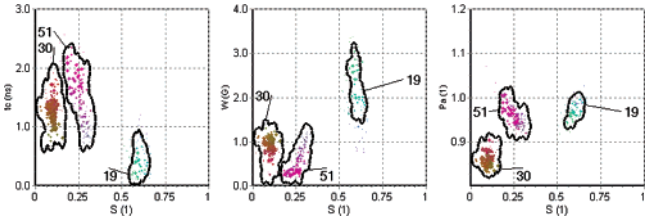

Figure 6. An example of the GHOST solution presentation of the spin labeled horse neutrophils taken from bronchoalveolar fluid (BAL) from horses suffering from the chronic obstructive pulmonary disease (COPD). Horses were sedated with medetomidine purchased from Domosedan (Turku, Finland). A 2.5-m long endoscope was introduced through the precleaned and topically anesthetized nostril and advanced until it wedged in a bronchus. Three hundred milliliters of prewarmed sterile physiological saline solution was infused through the biopsy channel into the bronchus and immediately reaspirated 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 measured at Bruker ELEXSYS E500 9.6 GHz spectrometer (field sweep of 10 mT; modulation: 0.15 mT, 100 kHz; 5 scans of 40 s with 40 ms of time constant), fitted with EPRSIM BBW software and characterized using GHOST condensation procedure.

the most important features of the system (EPR 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 before,[1] the GHOST condensed results are presented in 2D cross-sections $\{S\text{-}\tau_c, S\text{-}W, S\text{-}p_A\}$ (Figure 6). 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, and p_A in their definition intervals {0−3 ns}, {0−4 G}, and {0.8−1.2}, respectively (Figure 7). 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.

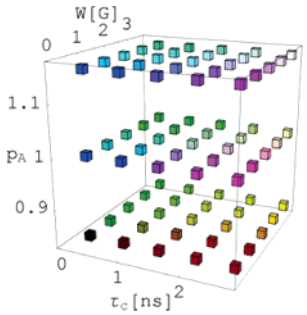

Figure 7. Color legend. The RGB (red, green, and blue) color of any particular solution point codes the relative values of parameters τ_c, W, and p_A in their definition intervals.

Table 1. Optimization Parameters after 200 and 20 Runs for the Real Membrane Spectrum[a]

criteria	200 runs	20 runs
χ^2_{min}	3.4	4.1
$\sigma(\chi^2)$	2.0	1.9
ρ_{max}	64.2	71.5

[a] For the experimental preparation see the caption to Figure 6.

RESULTS AND DISCUSSIONS

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 EPR spectra: experimental ones (from membranes and membrane proteins) and synthetic (discrete and continuous).

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 Figure 8 where the GHOST diagram (Figure 8b) and contribution histogram (Figure 8c) are compared with the original GHOST diagram based on the 200 runs (Figure 8a). 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 8b) does not resemble the original one (Figure 8a). 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 Figure 8c, 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 8b). In addition one also can see a higher solution density as a consequence of the crowding in the search space. An 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 Figure 11. 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 12b), originating in a solution crowding.

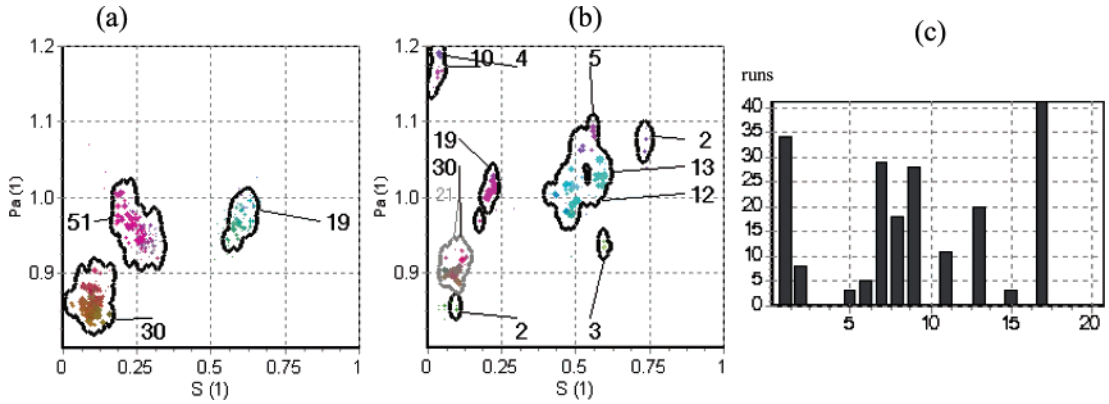

Figure 8. Typical characterization of spin labeled real membrane (see the caption to Figure 6): (a) GHOST as a result of 200 runs of HEO where only one solution is extracted from a single run; (b) GHOST as a result of 20 runs of the same HEO algorithm where on average 10 solutions are taken from each 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 a particular run) along the *y*-axis.

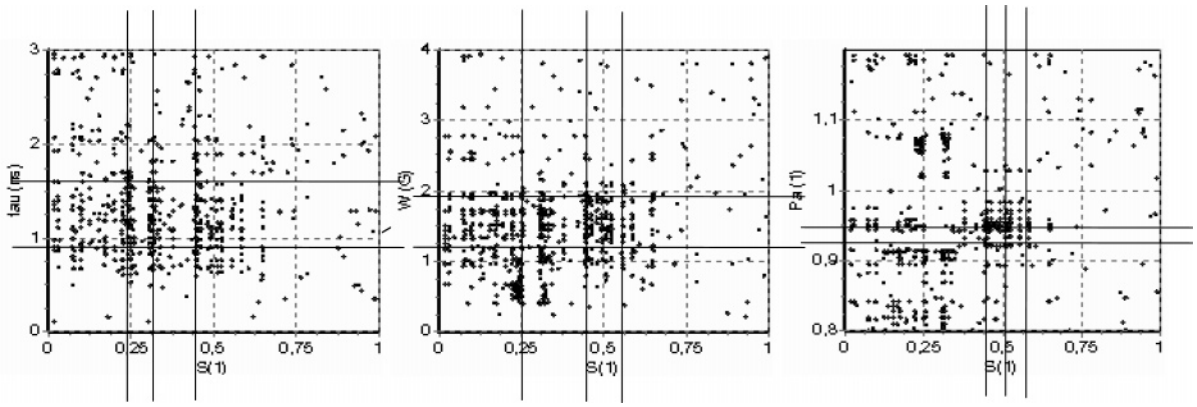

Figure 9. Schematic presentation of the "grid" problem for three cross-sections of the phase-space. Due to the standard multipoint crossover, subgroups of parameters are "transferred" between generations untouched, resulting in a gridlike distribution of the GHOST solution (single run). The lines indicate very high vertical and horizontal densities of solutions that evolve from copying of parts of parameter sets within the optimization routine.

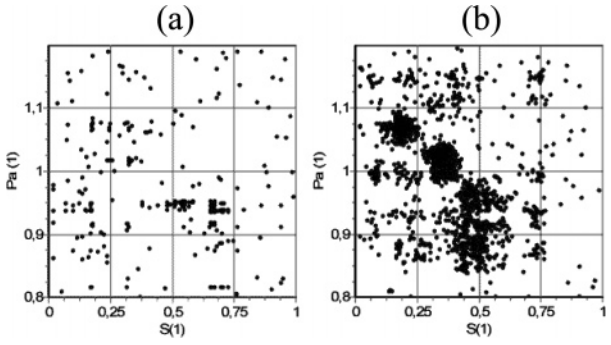

Figure 10. Single run GHOSTs (with population size 600): (a) original version with crowding problem—several solutions are crowded in many regions and (b) version with shaking that maintains diversity—solutions crowded in each point previously now spread over the flat minima region with the help of shaking operator.

According to the literature, the sharing implementation could change the result.[14,17] To test the sharing approach the continuous problem was chosen (Figure 11a and 12a). The results of this test in terms of the runs contribution histogram and GHOST cross-section are shown in Figures 11c and 12c. 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 2. Comparison of the χ^2 Distributions and Solution Densities for the Different Multirun HEO−GHOST Approaches on the Synthetic 15-Component Spectrum that Simulates Quasicontinuous Distribution of Spectral Parameters[a]

criteria	200 runs	20 runs	sharing	shaking
χ^2_{min}	1.2	1.2	1.7	1.2
$\sigma(\chi^2)$	0.9	0.4	1.3	0.9
ρ_{max0}	69.5	75.7	69	66.1

[a] See also caption to Figure 12.

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 9) as a consequence of the rough action of the 3-point crossover operator. This leads to the loss of solution diversity.

In the HEO algorithm only a local search operator is capable of restoring the diversity and eliminating the "grid", but due to the high computational cost and extremely high

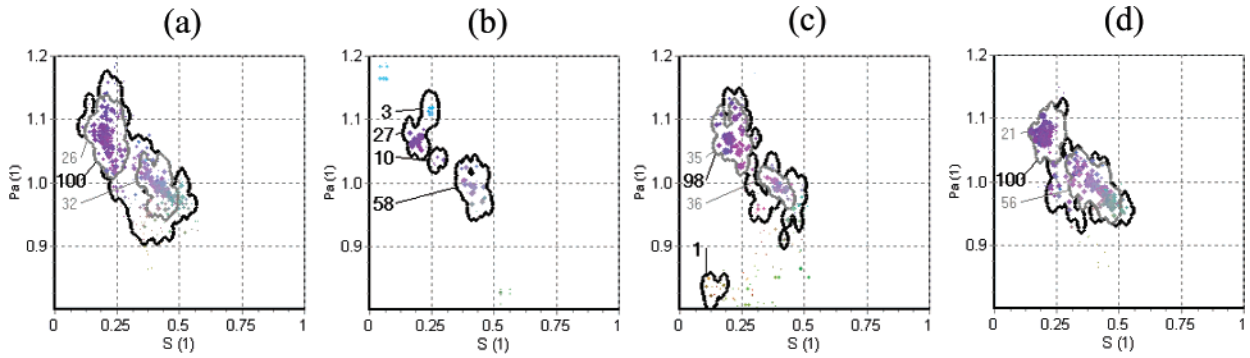

Figure 11. Comparison of the effectiveness of different multirun HEO−EnDashGHOST approaches on the synthetic 15-component spectrum that simulates a quasicontinuous distribution of spectral parameters. (a) GHOST as a result of 200 runs of original HEO routine; (b) GHOST as a result of 20 runs of the original HEO routine; (c) GHOST as a result of 20 runs of the modified HEO routine that includes sharing operator; and (d) GHOST as a result of 20 runs of the modified HEO routine that includes shaking operator as described in the text.

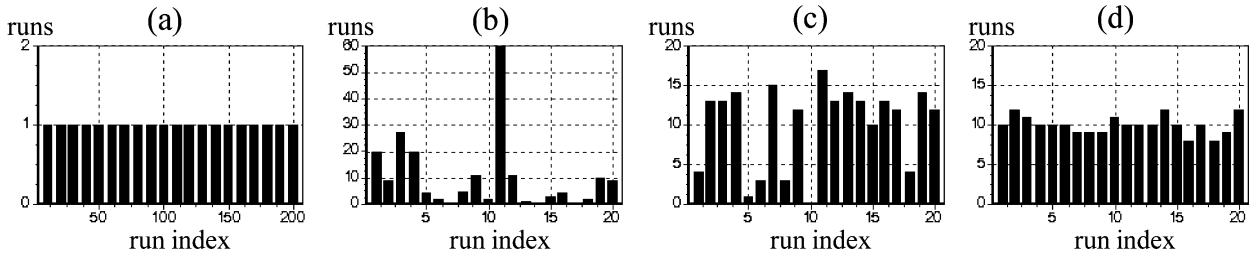

Figure 12. Comparison of runs contribution histograms of different multirun HEO −GHOST approaches on the synthetic 15-component spectrum that simulates a quasicontinuous distribution of spectral parameters. (a) GHOST as a result of 200 runs of original HEO routine; (b) GHOST as a result of 20 runs of the original HEO routine; (c) GHOST as a result of 20 runs of the modified HEO routine that includes sharing operator; and (d) GHOST as a result of 20 runs of the modified HEO routine that includes shaking operator as described in the text.

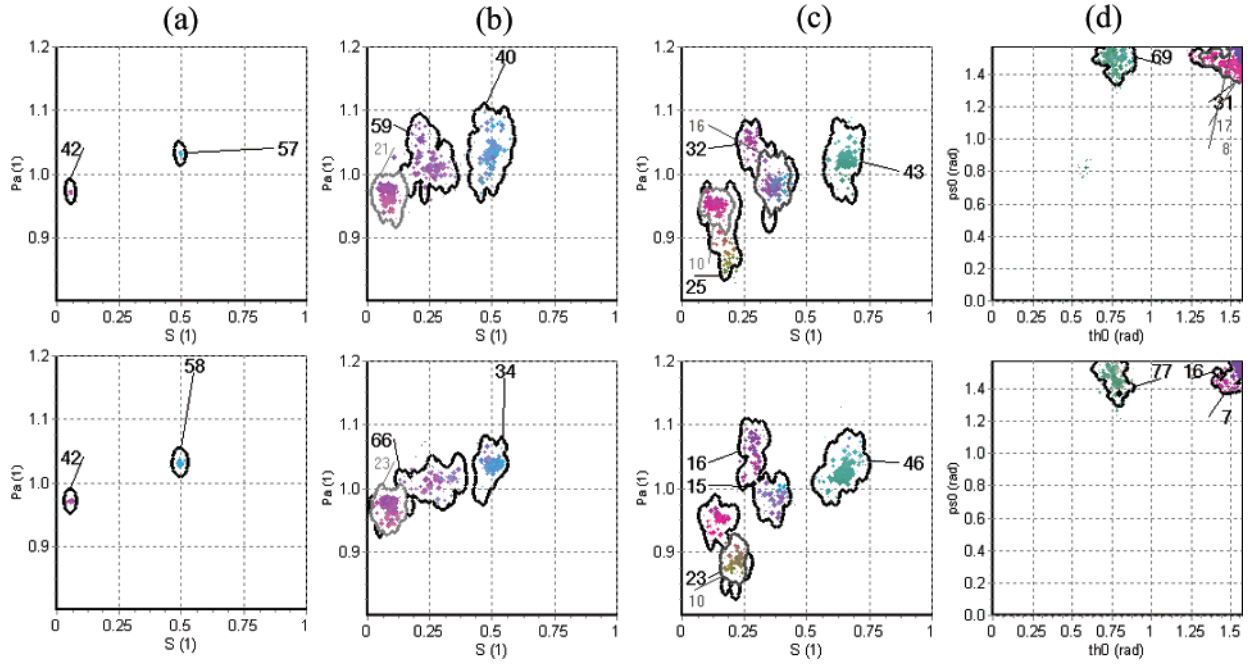

Figure 13. GHOSTs comparison of the original-HEO approach with 200 runs (above) versus modified-HEO (with shaking) based on 20 runs (bellow): (a) for synthetic discrete 2D spectrum that was constructed from two spectral components with the known parameter set and optimized as unknown one; (b) synthetic quasicontinues spectrum (see the caption to Figure 12); (c) 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 labeled with the methyl ester of 5-doxyl palmitate, MeFASL(10,3), and measured under the same conditions as the membranes of horse neutrophils (see the caption to Figure 6); (d) spectrum of the spin labeled (maleimide spin label) cystein mutant of major coat protein of bacteriophage M13 at amino acid position 46 reconstituted in dimyristoylphosphatidylcholine lipid bilayer.[31]

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 and thereby eliminates the effect of the "grid".

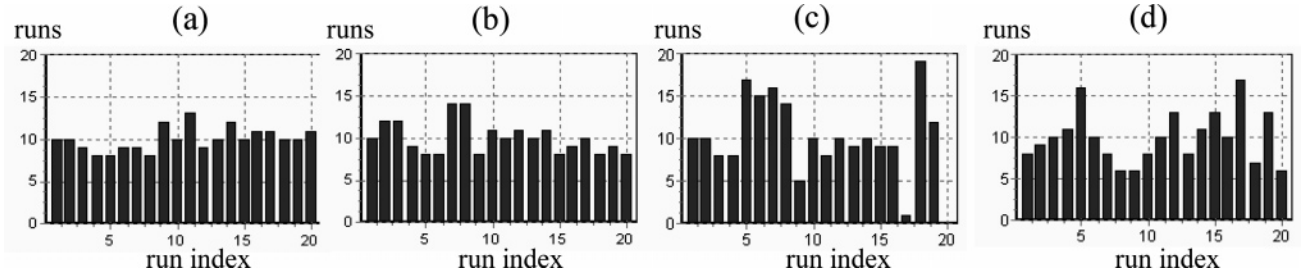

Figure 14. Runs contribution histogram for the modified HEO approach based on 20 runs with the implemented shaking operator. For the description of the samples (a)−(d) see the caption to Figure 13.

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 in Figure 10 for a continuous problem that represents the most extreme case of the complexity.

The results of the implemented shaking algorithm are shown in Figures 11 and 12. One can see that the shaking operator considerably improves the result of a single run as the GHOST pattern from 20 runs (Figure 11d) is very similar to the original one (Figure 11a), the runs contribution histogram is very even (Figure 12d), and finally the distribution of χ^2 is very good (Table 2). This approach therefore enables us to reduce the number of HEO runs while preserving the quality of the final result (Figures 11d and 12d and Table 2). Therefore by using this new algorithm, we are able to speed up the optimization process by a factor of 5−7.

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 Figure 13, 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).

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 in keeping the quality of the characterization, thereby considerably reducing the computational time of the EPR spectrum analysis by a factor of 5−7. With this successful modification the application of advanced EPR spectra analysis[1] 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 EPR spectra analysis.

ACKNOWLEDGMENT

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 AND NOTES

(1) Štrancar, J.; Koklič, T.; Arsov, Z.; Filipič, B.; Stopar, D.; Hemminga, M. A. Spin Label EPR-based Characterization of Biosystem Complexity. *J. Chem. Inf. Model.* **2005**, *45*, 394−406.

(2) Columbus, L.; Hubbell, W. L. A new spin on protein dynamics. *Trends Biochem. Sci.* **2002**, *27*, 288−295.

(3) Štrancar, J.; Koklič, T.; Arsov, Z. Soft picture of lateral heterogeneity in biomembranes. *J. Membr. Biol.* **2003**, *196*, 135−146.

(4) Budil, D. E.; Lee, S.; Saxena, S.; Freed, J. H. Nonlinear-Least-Squares Analysis of Slow-Motion EPR Spectra in One and Two Dimensions Using a Modified Levenberg−Marquardt Algorithm. *J. Magn. Reson. A* **1996**, *120*, 155−189.

(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*; Dalton, L. R., Ed.; CRC Press: Boca Raton, FL, 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*; Berliner, L. J., Reuben, J., Eds.; Plenum Press: New York, 1989; pp 1−76.

(7) Štrancar, J.; Šentjurc, M.; Schara, M. Fast and accurate characterization of biological membranes by EPR spectral simulations of nitroxides. *J. Magn. Reson.* **2000**, *142*, 254−265.

(8) Schindler, H.; Seelig, J. EPR spectra of spin labels in lipid bilayers. *J. Chem. Phys.* **1973**, *59*, 1841−1850.

(9) Marsh, D. Electron Spin Resonance: Spin Labels. In *Membrane Spectroscopy*; Grell, E., Ed.; Springer-Verlag: Berlin, 1981; pp 51−142.

(10) Nordio, P. L. General magnetic resonance theory. In *Spin Labeling, Theory and Application*; Berliner, L. J., Ed.; Academic Press: New York, 1976; pp 5−51.

(11) Filipič, B.; Štrancar, J. Tuning EPR spectral parameters with a genetic algorithm. *Appl. Soft Computing* **2001**, *1*, 83−90.

(12) Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, PA, 1989.

(13) Pétrowski, A. A. Clearing procedure as a niching method for genetic algorithms. In *3rd IEEE International Conference on Evol. Comput.*; 1996; pp 798−803.

(14) Mahfoud, S. W. Niching Methods for Genetic Algorithms, Ph.D. Thesis, University of Illinois at Urbana-Champaign, Urbana, 1995.

(15) Beasley, D.; Bull, D. R.; Martin, R. R. A Sequential Niche Technique for Multimodal Function Optimization. *Evol. Comput.* **1993**, *1*, 101−125.

(16) Glover, F. Tabu Search − Part I. *ORSA J. Comput.* **1989**, *1*, 190−206.

(17) Goldberg, D.; Richardson, J. Genetic algorithms with sharing for multimodal function optimization. In *2nd International Conference on Genetic Algorithms*; Grefenstette, Ed., 1987; pp 41−49.

(18) De Jong, K. A. An Analysis of the Behavior of a Class of Genetic Adaptive Systems, Doctoral Dissertation, University of Michigan, 1975.

(19) Li, J. P.; Balazs, M. E.; Parks, G. T.; Clarkson, P. J. A species conserving genetic algorithm for multimodal function optimization. *Evol. Comput.* **2003**, *11*, 107−109.

(20) Spears, W. Simple Subpopulation Schemes. In *3rd Annual Conference on Evol. Programming*; 1994; pp 296−307.

(21) Deb, K. Genetic Algorithms in Multimodal Function Optimization, Master's Thesis, University of Alabama, 1989.
(22) Streichert, F.; Stein, G.; Ulmer, H.; Zell, A. A Clustering Based Niching Method for Evolutionary Algorithms. In *Genetic Evol. Comput. Conf., Lect. Notes Comput. Sci.* **2003**, *2723*, 644−645.
(23) Yin, X.; Germay, 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*; 1993; pp 450−457.
(24) Ursem, R. K. Multinational Evolutionary Algorithms. In *Congress of Evol. Comput.*; 1999; Vol. 3, pp 1633−1640.
(25) Gordon, V. S.; Whitley, D.; Bohn, A. Dataflow parallelism in genetic algorithms. In *Parallel Problem Solving from Nature 2*; Manner, R., Manderick, B., Eds.; Elsevier Science: Amsterdam, The Netherlands, 1992; pp 533−542.
(26) Bessaou, M.; Pétrowski, A.; Siarry, P. Island model cooperating with speciation for multimodal optimization. In *6th International Conference on Parallel Problem Solving From Nature*; Schoenauer, M. et al., Eds.; Paris, France, 2000; pp 437−446.
(27) Martin, W. N.; Lienig J.; Cohoon, J. P. Island (migration) models: evolutionary algorithms based on punctuated equilibria. In *Handbook of Evolutionary Computation*; Back, T., Fogel, D. B., Michalewicz, Z., Eds.; Institute of Physics Publishing: Bristol, U.K., 1997; pp C6.3:1−C6.3:1.
(28) Darwen, P. J.; Xin, Y.. Every Niching Method has its Niche: Fitness Sharing and Implicit Sharing Compared. In *4th Int. Conf. Parallel Problem Solving from Nature, Lect. Notes Comput. Sci.* **1996**, *1141*, 398−407.
(29) Mahfoud, S. W. *Simple Analytical Models of Genetic Algorithms for Multimodal Function Optimization*; Technical Report IlliGAL Report No 93001; 1993.
(30) Sareni, B.; Krähenbühl, L. Fitness Sharing and Niching Methods Revisited. In *IEEE Trans. Evol. Comput.*; 1998; Vol. 2.
(31) Stopar, D.; Spruijt, R. B.; Wolfs, C. J. A. M.; Hemminga, M. A. Protein−lipid interactions of bacteriophage M13 major coat protein. *Biochim. Biophys. Acta* **2003**, *1611*, 5−15.

CI0501589

Journal of Magnetic Resonance 197 (2009) 245–248

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Communication

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

Janez Štrancar [a], Aleh Kavalenka [a], Primož Ziherl [a,b], David Stopar [c], Marcus A. Hemminga [d,*]

[a] Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia
[b] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia
[c] Biotechnical Faculty, University of Ljubljana, Večna pot 111, SI-1000 Ljubljana, Slovenia
[d] Laboratory of Biophysics, Wageningen University, Dreijenlaan 3, NL-6703 HA Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 9 April 2008
Revised 23 October 2008
Available online 24 December 2008

Keywords:
Site-directed spin-labeling (SDSL)
Electron spin resonance (ESR, EPR)
Membrane protein
Structure determination
Modeling of side chain conformational space
Computer simulations

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 [P.V. Nazarov, R.B.M. Koehorst, W.L. Vos, V.V. Apanasovich, M.A. Hemminga, FRET study of membrane proteins: determination of the tilt and orientation of the N-terminal domain of M13 major coat protein, Biophys. J. 92 (2007) 1296–1305].

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 structure [3]. 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 difficulties [4]. 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 field [5–10].

The basis of this technique is a 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 tremendously [7]. 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 previously [11–13]. Here we develop a molecular model to describe local 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.

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

* Corresponding author. Fax: +31 317 482 725.
E-mail address: marcus.hemminga@wur.nl (M.A. Hemminga).
URL: http://ntmf.mf.wau.nl/hemminga/ (M.A. Hemminga).

1090-7807/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2008.12.019

(iii) the restrictions imposed by the surrounding lipids. These effects are illustrated in Fig. 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 chains [14]. 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 (Fig. 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:

$$p_i = \left\{ \begin{array}{l} 0, \text{backbone overlap} \\ 1, \text{no backbone overlap} \end{array} \right\} \times \left[\prod_{k \in \text{neighboring amino acids}} \left(1 - \frac{N^k_{k-i\,\text{overlap}}}{N^k_{all}} \right) \right] (1 - \sin \vartheta_i), \qquad (1)$$

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 (Fig. 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 (Fig. 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. 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 Cβ to the oxygen atom of the nitroxide) and in a first approximation described by $(1 - \sin\vartheta)$ (Fig. 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.

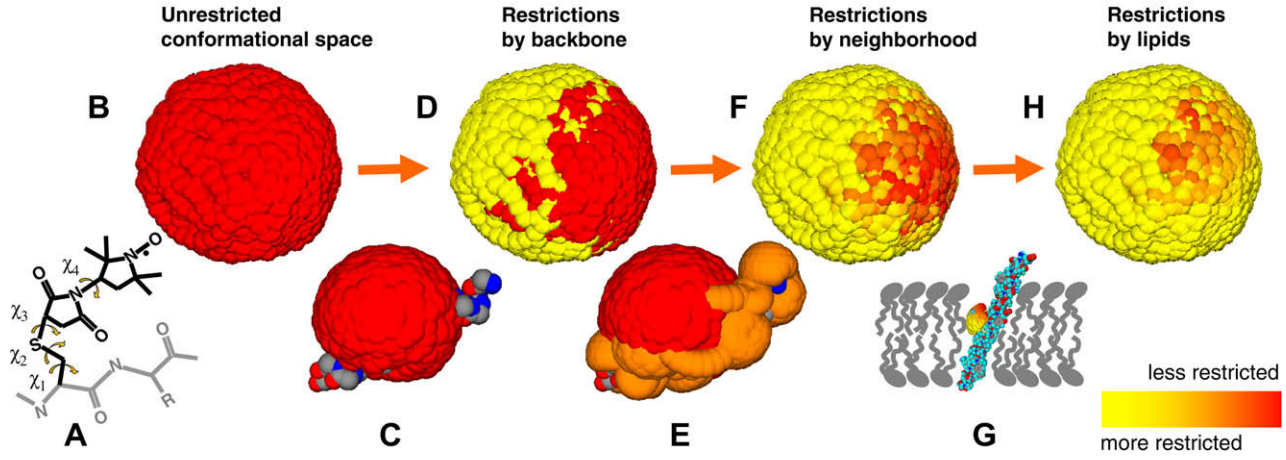

Fig. 1. Schematic illustration of the conformational space of the spin-label side chain for membrane-embedded 3-maleimido proxyl spin-labeled M13 coat protein. For simplicity, the protein is assumed to be in a perfect α-helical conformation and embedded in a bilayer of 1,2-dierucoyl-*sn*-glycero-3-phosphocholine between amino acid positions 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.

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 sites [11–13]. 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 (Fig. 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:

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

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.

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 data [12] (Fig. 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 (Fig. 2). For simplicity, we assumed a membrane-embedment 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-embedment 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: $-57 \pm 30°$ and $-47 \pm 30°$, respectively. The Ω values related to the original α-helical protein model ($\varphi = -57°$ and $\psi = -47°$) are indicated with white triangles in Fig. 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 (Fig. 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 Fig. 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 (red triangles). 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 (Fig. 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.

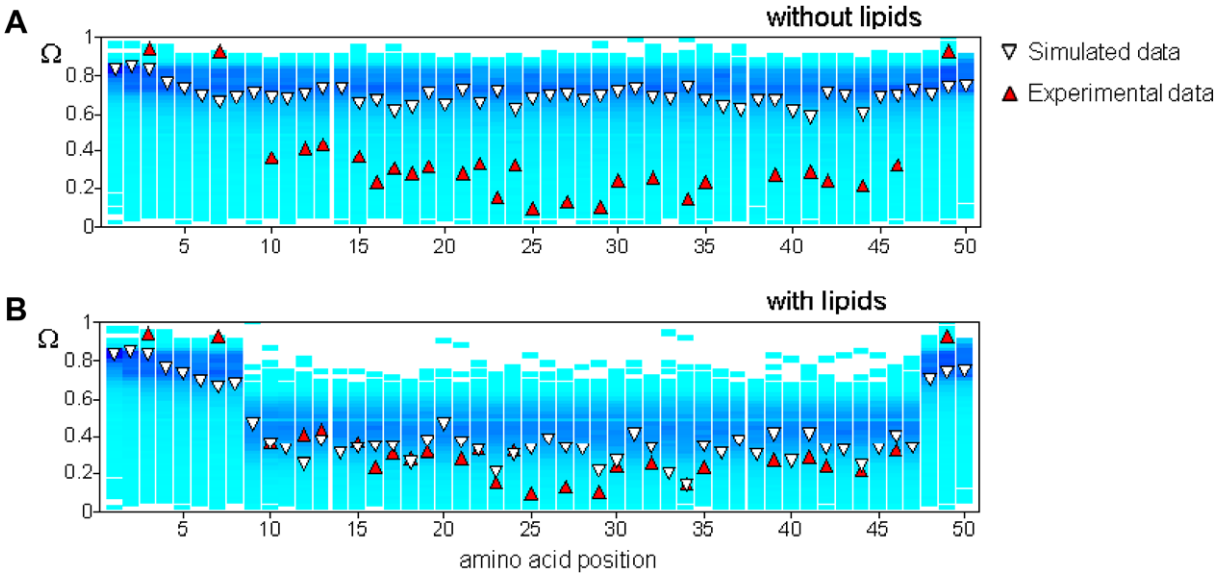

Fig. 2. Sensitivity of the free rotational space to the primary sequence, variations of the protein secondary structure and the effect of the lipids for membrane-embedded spin-labeled M13 coat protein. The histograms of the relative frequency of a given value of Ω (color-coded by continuous shades of 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 ($\varphi = -57°$ and $\psi = -47°$) as defined in [19,21,22].

4. Conclusions

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 structures [23]. 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 4 weeks using twenty 4-GFLOP processors, which makes this approach highly competitive compared to other high-resolution methodologies.

As compare 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 Ni^{2+} 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].

Acknowledgments

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] A. Arora, L.K. Tamm, Biophysical approaches to membrane protein structure determination, Curr. Opin. Struct. Biol. 11 (2001) 540–547.
[2] J.-J. Lacapère, E. Pebay-Peyroula, J.-M. Neumann, C. Etchebest, Determining membrane protein structures: still a challenge!, Trends Biochem Sci. 32 (2007) 259–270.
[3] S. White, Membrane proteins of known 3D structure, 2008, <http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html>.
[4] J. Torres, T.J. Stevens, M. Samsó, Membrane proteins: the 'Wild West' of structural biology, Trends Biochem. Sci. 28 (2003) 137–144.
[5] W.L. Hubbell, D.S. Cafiso, C. Altenbach, Identifying conformational changes with site-directed spin labeling, Nat. Struct. Biol. 7 (2000) 735–739.
[6] L. Columbus, W.L. Hubbell, A new spin on protein dynamics, Trends Biochem. Sci. 27 (2002) 288–295.
[7] M.A. Hemminga, L.J. Berliner (Eds.), ESR Spectroscopy in Membrane Biophysics, Springer, New York, USA, 2007.
[8] C. Beier, H.-J. Steinhoff, A structure-based simulation approach for electron paramagnetic resonance spectra using molecular and stochastic dynamics simulations, Biophys. J. 91 (2006) 2647–2664.
[9] D.E. Budil, K.L. Sale, K.A. Khairy, P.G. Fajer, Calculating slow-motional electron paramagnetic resonance spectra from molecular dynamics using a diffusion operator approach, J. Phys. Chem. A 110 (2006) 3703–3713.
[10] L.E.W. LaConte, V. Voelz, W. Nelson, M. Enz, D.D. Thomas, Molecular dynamics simulation of site-directed spin labeling: experimental validation in muscle fibers, Biophys. J. 83 (2002) 1854–1866.
[11] D. Stopar, J. Štrancar, R.B. Spruijt, M.A. Hemminga, Exploring the local conformational space of a membrane protein by site-directed spin labeling, J. Chem. Inf. Model. 45 (2005) 1621–1627.
[12] D. Stopar, J. Štrancar, R.B. Spruijt, M.A. Hemminga, Motional restrictions of membrane proteins: a site-directed spin labeling study, Biophys. J. 91 (2006) 3341–3348.
[13] J. Štrancar, T. Koklič, Z. Arsov, B. Filipič, D. Stopar, M.A. Hemminga, Spin label EPR-based characterization of biosystem complexity, J. Chem. Inf. Model. 45 (2005) 394–406.
[14] M. Karplus, J.A. McCammon, The internal dynamics of globular proteins, CRC Crit. Rev. Biochem. 9 (1981) 293–349.
[15] B. Ho, R. Brasseur, The Ramachandran plots of glycine and pre-proline, BMC Struct. Biol. 5 (2005) 14.
[16] B.K. Ho, A. Thomas, R. Brasseur, Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the α-helix, Protein Sci. 12 (2003) 2508–2522.
[17] J.M. Word, S.C. Lovell, T.H. LaBean, H.C. Taylor, M.E. Zalis, B.K. Presley, J.S. Richardson, D.C. Richardson, Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms, J. Mol. Biol. 285 (1999) 1711–1733.
[18] Z. Xiang, B. Honig, Extending the accuracy limits of prediction for side-chain conformations, J. Mol. Biol. 311 (2001) 421–430.
[19] R.B.M. Koehorst, R.B. Spruijt, F.J. Vergeldt, M.A. Hemminga, Lipid bilayer topology of the transmembrane α-helix of M13 major coat protein and bilayer polarity profile by site-directed fluorescence spectroscopy, Biophys. J. 87 (2004) 1445–1455.
[20] D. Stopar, R.B. Spruijt, M.A. Hemminga, Anchoring mechanisms of membrane-associated M13 major coat protein, Chem. Phys. Lipids 141 (2006) 83–93.
[21] P.V. Nazarov, R.B.M. Koehorst, W.L. Vos, V.V. Apanasovich, M.A. Hemminga, FRET study of membrane proteins: simulation-based fitting for analysis of protein structure, membrane embedment and association, Biophys. J. 91 (2006) 454–466.
[22] P.V. Nazarov, R.B.M. Koehorst, W.L. Vos, V.V. Apanasovich, M.A. Hemminga, FRET study of membrane proteins: determination of the tilt and orientation of the N-terminal domain of M13 major coat protein, Biophys. J. 92 (2007) 1296–1305.
[23] A. Bax, Two-dimensional NMR and protein structure, Annu. Rev. Biochem. 58 (1989) 223–256.
[24] B. Filipič, J. Štrancar, Evolutionary computational support for the characterization of biological systems, in: G.B. Fogel, D. Corne, (Eds.), Evolutionary Computation in Bioinformatics, Elsevier Science, San Francisco, USA 2003, pp. 279–294.
[25] A.A. Kavalenka, B. Filipič, M.A. Hemminga, J. Štrancar, Speeding up a genetic algorithm for EPR-based spin label characterization of biosystem complexity, J. Chem. Inf. Model. 45 (2005) 1628–1635.

Site-Directed Spin-Labeling Study of the Light-Harvesting Complex CP29

Aleh A. Kavalenka,[†‡] Ruud B. Spruijt,[†] Cor J. A. M. Wolfs,[†] Janez Štrancar,[‡] Roberta Croce,[§] Marcus A. Hemminga,[†*] and Herbert van Amerongen[†]

[†]Laboratory of Biophysics, Wageningen University, Dreijenlaan 3, NL-6703HA Wageningen, The Netherlands; [‡]Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia; [§]Department of Biophysical Chemistry/Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands

ABSTRACT The topology of the long N-terminal domain (~100 amino-acid residues) of the photosynthetic Lhc CP29 was studied using electron spin resonance. 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 electron spin resonance spectra were analyzed in terms of a multicomponent 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.

INTRODUCTION

Photosynthesis in green plants and algae occurs in chloroplasts. Their highly folded thylakoid membranes provide a home for the multisubunit protein complexes PSI and PSII, which 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 transfer it to the reaction center where it can be used to drive charge separation. All antenna complexes of higher plants belong to the Lhc 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 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 nonphotochemical quenching, a mechanism that protects PSII against photoinhibition (5). 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 A5 (6), leading to the harmless dissipation of excess excitation energy.

The structure of LHCII has been resolved at 2.72 Å (7) showing three transmembrane helices, two amphipathic helices on the lumenal side of the membrane, and the positions of 14 Chl and 4 xanthophyll molecules per monomeric subunit. Structural information on the minor antenna complexes CP24, CP26, and CP29 is still lacking, but sequence analysis (8) 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 (~100 amino-acid residues), which contains a phosphorylation site (11). Phosphorylation takes place, for instance, under cold stress and is accompanied by a structural change of the protein (12). 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 mechanisms (13). However, details are lacking on both the structure and the structural changes.

Submitted November 18, 2008, and accepted for publication January 28, 2009.

*Correspondence: marcus.hemminga@wur.nl

Chl, chlorophyll; CP29, chlorophyll-*a*/*b*-binding protein 29 (a minor antenna complex of photosystem II); DM, n-Dodecyl β-D-maltoside; ESR, electron spin resonance; GHOST, condensation algorithm that filters and groups the solutions found in optimization runs; LDS, lithium dodecyl sulfate; Lhc, light-harvesting complex; LHCII, light-harvesting chlorophyll-*a*/*b*-binding protein of photosystem II; MTS-SL, (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate spin label; PS, photosystem.

Editor: David D. Thomas.

0006-3495/09/05/3620/9 $2.00

doi: 10.1016/j.bpj.2009.01.038

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 ESR appears to show the highest potential to further develop the field (14).

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 spectroscopy (15). 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.

This work 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 10 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.

MATERIALS AND METHODS

Construction and isolation of overexpressed CP29 apoprotein

Lhcb4.1 cDNA of *Arabidopsis thaliana* (*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 naturally 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* (*E. 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.

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

Purified pigments were obtained from spinach. Concentrations of pigments were determined spectroscopically: Chls as described by Porra et al. (19) and carotenoids as described by Davies (20). Just before labeling, inclusion bodies containing CP29 apoprotein were freshly purified from dithiothreitol and dissociated in LDS reconstitution buffer (2% LDS, 12.5% sucrose, 20 mM Na_2HPO_4 pH 7.6). CP29 apoproteins were labeled at room temperature for 3 h with a five-times molar excess of the spin label MTS-SL (methanethiosulfonate from TRC, Toronto, Canada). Excess spin label was removed using affinity chromatography on a His-Trap column. Before 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 Na_2HPO_4, 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 Chl *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 Chls and carotenoids in all preparations.

ESR measurements

All washed and concentrated spin-labeled CP29 preparations in DM buffer (final protein concentration between 0.07 mM and 0.2 mM) were transferred to 50 μl capillaries up to 1 cm height and placed in a standard 4-mm quartz ESR tube. Spectra were measured on an X-band Bruker Elexsys E-500 ESR system (Bruker, Rheinstetten, Germany) 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 (Eurotherm, Leesburg, VA). Spectra were recorded at 10-mT scan widths with a microwave power of 5 mW at 6°C. To improve signal/noise, 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.

ESR spectral simulation, optimization, and solution condensation

The ESR spectra of spin-labeled CP29 samples were simulated with a multicomponent model as described previously (16,23). The spectral parameters $\{\vartheta, \varphi, \tau_c, W, p_A, prot\}$ of each component of the simulated spectra were simultaneously optimized with a multirun 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 label (25). The dynamics of the spin probe gives rise to a motion in a cone (27), which can be described with three parameters: a maximum opening cone angle ϑ, a cone asymmetry angle φ, and an effective correlation time τ_c. The magnetic interaction tensors **g** and **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 **g** tensor (28). It was found that the relative error for parameter *prot* was quite large. Therefore this parameter is not used in our further discussion (23). 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 $\{\vartheta, \varphi, \tau_c, W, p_A, prot\}$ and four relative contributions following a previous approach (23). Typically, 20 runs of the population-based hybrid evolutionary optimization 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, and b), protein backbone fluctuations (27). 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 levels (30). Visual analysis of the resulting GHOST plots, which present a combination of two parameters (φ and ϑ, τ_c and ϑ, p_A and ϑ), 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.

RESULTS

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.

ESR experiments

The ESR spectra of the reconstituted holoprotein complexes and of the apoproteins in detergent solution are shown in Fig. 1. In all cases the spectra have a multicomponent 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

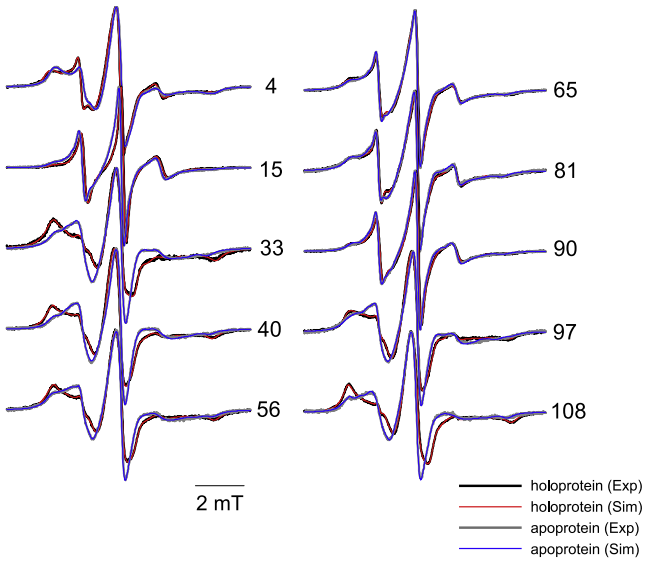

FIGURE 1 ESR spectra of MTS-SL spin-labeled CP29 protein samples at label positions 4, 15, 33, 40, 56, 65, 81, 90, 97, and 108 reconstituted in DM with (holoprotein, *black line*) and without (apoprotein, *gray line*) pigments. The total horizontal scan range is 10 mT. Spectral line heights are normalized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

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 multicomponent ESR spectra, we used a multicomponent model of asymmetric motional restriction (16,23) and optimized the fitted spectra employing a multirun multisolution hybrid evolutionary method (25). The goodness of fit was chosen to be the reduced χ^2 function:

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

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 10 spin-labeled CP29 holo- and apoprotein samples, the quality of the simulated ESR spectra is good (see Fig. 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 $\vartheta-\varphi$ GHOST shown for

positions 65 and 108 in Fig. 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 (Fig. 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) (Fig. 2). In most cases, the motional patterns in the GHOST plots (as shown in Fig. 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/noise 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 (see Fig. 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 (Fig. 1) are presented in the GHOST plots with colored triangles (Fig. 2). The size of a triangle is proportional to the contribution of the corresponding component in the 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 as (23):

$$\Omega = \frac{\vartheta\varphi}{(\pi/2)^2}. \tag{2}$$

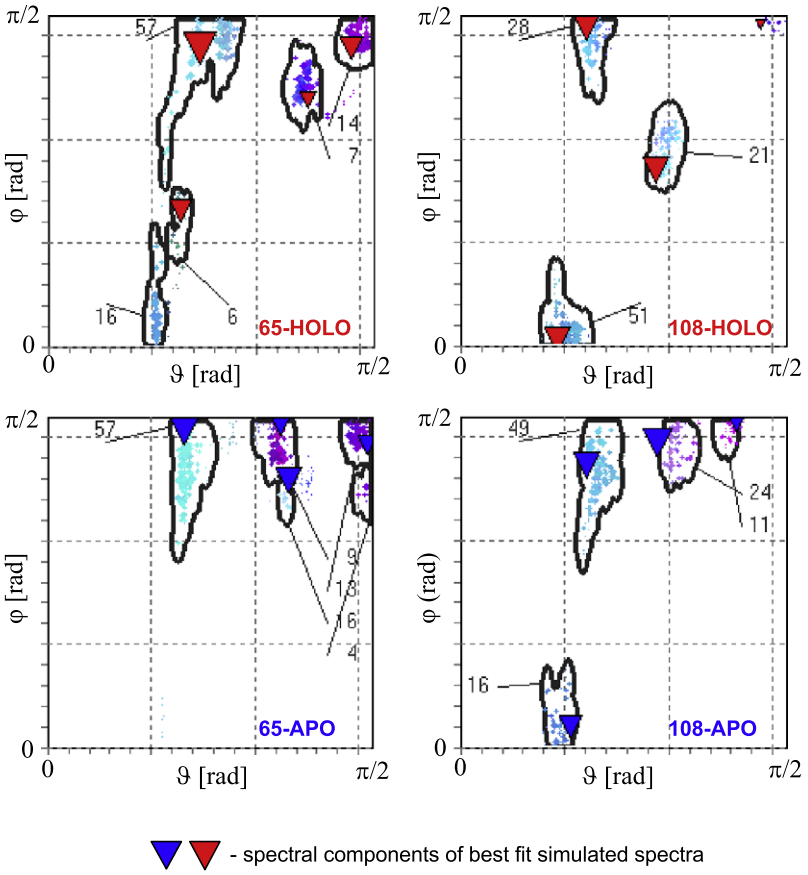

FIGURE 2 GHOST plots showing the optimized multiple solutions represented in a two-dimensional distribution of the angles ϑ and φ of MTS-SL spin-labeled CP29 protein samples at positions 65 and 108 reconstituted in DM with (*top*, holoprotein) and without (*bottom*, apoprotein) pigments. The components 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 ns}, {0–0.4 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, whereby the area of each triangle is proportional to the relative contribution of the corresponding component in the simulated spectrum.

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 10 spin-labeled CP29 holo- (Fig. 3 *A*) and apoprotein samples (Fig. 3 *B*). 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 **g** and **A** of the spin label (16,28). These parameters are presented in Fig. 3, *A* 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 Fig. 3 *C*. Fig. 3 *D* compares the weighted averages of the motional patterns of the spin-labeled holo- and apoproteins. In general, it can be seen in Fig. 3 that high values of the free rotational space Ω correspond to high values of the effective rotational correlation time τ_c.

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 is not a very strong denaturing detergent providing a relatively compact protein-detergent complex (33). If the pigments are absent, the structure of CP29 protein is looser and it may be partly unfolded (33). For LHCII in DM, the spectroscopic properties are similar to those observed in the intact thylakoid membrane (34). Because LHCII and CP29 have a strong sequence homology in the transmembrane protein body (9), 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

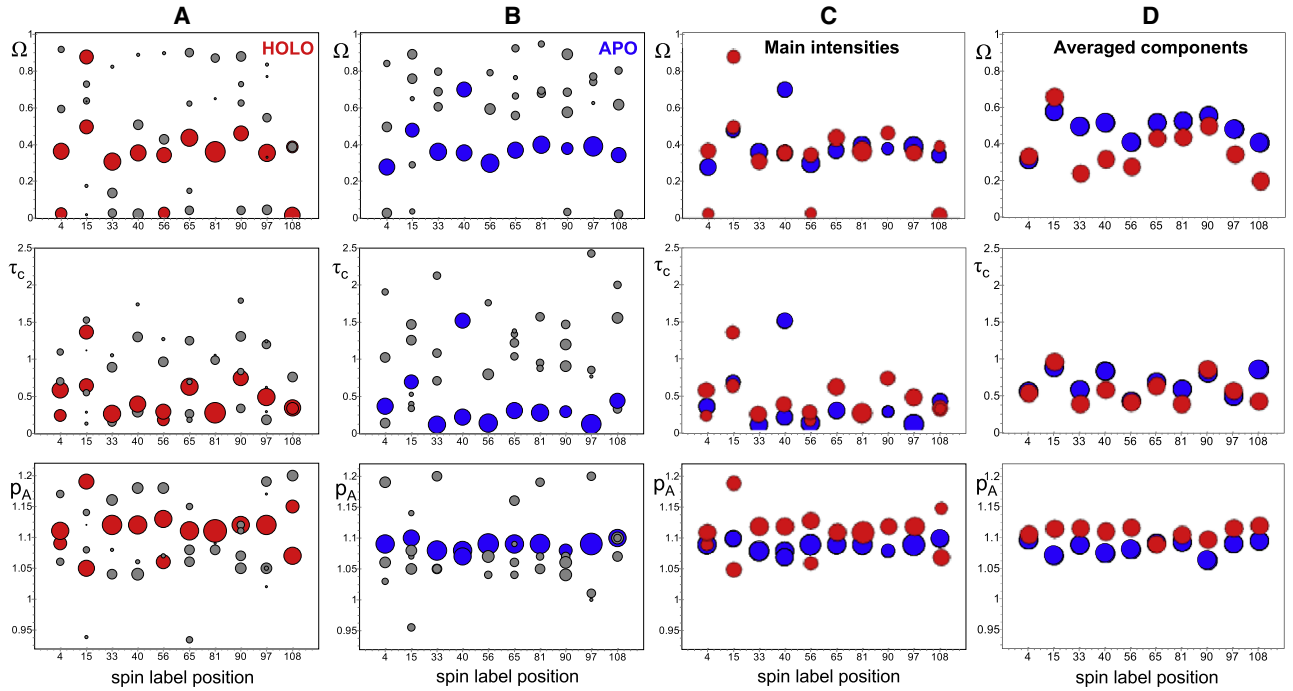

FIGURE 3 ESR data of MTS-SL spin-labeled CP29 protein samples reconstituted in DM with (*A*) (holoprotein, *red circles*) and without (*B*) (apoprotein, *blue circles*) pigments. Less-pronounced motional patterns with a contribution below 25% are represented by gray circles. The horizontal axis indicates 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.

starting point for a comparative spin-label ESR study addressing the questions given above.

From a qualitative analysis of the ESR spectra (Fig. 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 pigments (11). 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 protein (11). Also positions 4 and 15 at the N-terminal end are just slightly affected by pigment reconstitution. Position 4 displays a clear two-component characteristic 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, because 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 multicomponent ESR spectra, we carried out a spectral decomposition based on a multicomponent model of asymmetric motional restriction (16,23), followed by a multirun, multisolution hybrid evolutionary approach (25). The multicomponent model turned out to be robust enough to cover many different combinations of coexisting local motional patterns. The multisolution 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), because 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 Fig. 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 been seen in Fig. 3, *A* and *B*, the GHOST analysis results in a number of motional patterns. There are several factors that can contribute to a multicomponent 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 5%. As can be seen in Fig. 3 *A* even by discarding motional patterns with small contributions ($<$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 (Fig. 3 *C*). 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 (Fig. 3 *D*). 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 Fig. 3, *A* 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 Fig. 3 *A*, in which the spin-label motion is highly restricted. These motional patterns are almost gone in Fig. 3 *B*. In turn, in Fig. 3 *B* 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. Because 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 Fig. 3 *C*, 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 Fig. 3, *A* 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 (Fig. 3 *D*). These data represent the general trend, but details about the various components are lost. In Fig. 3 *D*, apart from information about the average free rotational space Ω, 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 Fig. 3 *D*, 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 Fig. 3 *D* (a high value for p_A reflects an increased local polarity (25)), 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 Fig. 3 *D*, closely follows the qualitative interpretation of the ESR spectra in Fig. 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 Fig. 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 4 (15). 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.

Summarizing model

CP29 has a strong sequential homology with LHCII, the major difference being an N-terminal insert from amino-acid residue 56 to 98 (9). 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 spinach (7) as a starting point for constructing a model for CP29 (Fig. 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 (Fig. 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 (Fig. 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 protein (7) (Fig. 4). The next labeled position toward the N-terminal end, position 15, shows a high value for Ω suggesting rather unrestricted spin-label motion. This indicates that this

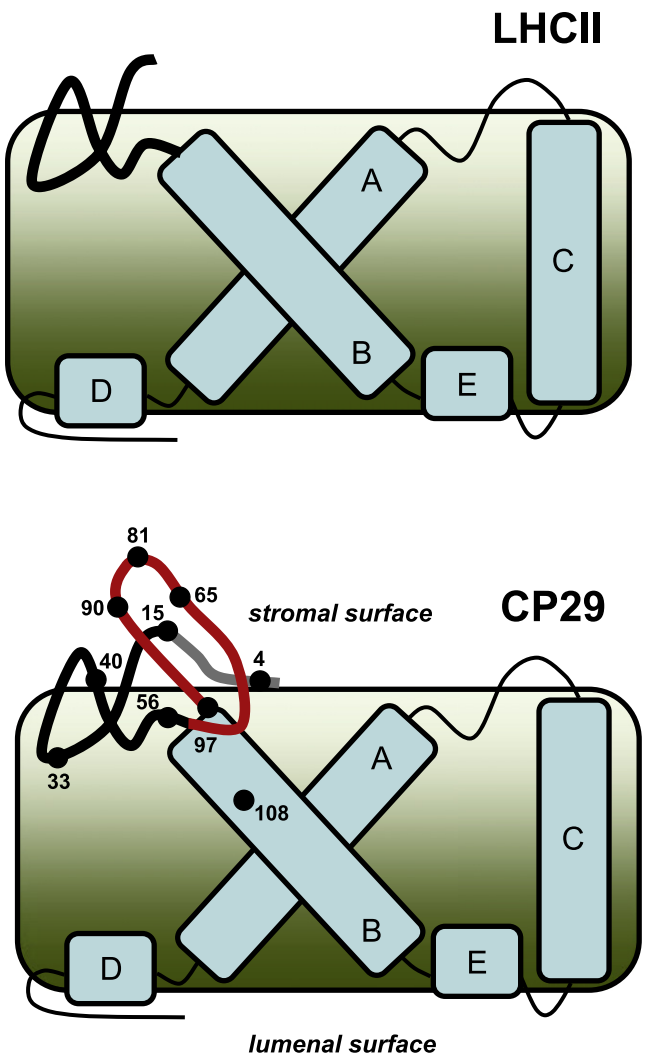

FIGURE 4 Schematic structural model of CP29 based on the crystal structure of LHCII from spinach (7) (Protein Data Bank ID: 1RWT). The main helical structures (*A–E*) of the transmembrane protein body are shown in light blue. The extra N-terminal insert of CP29 (as compared to LHCII) is shown as a red loop extruding from the main transmembrane protein body. The N-terminus from amino-acid residue 1–14 is indicated in gray, as this part of the structure is not resolved in the crystal structure of LHCII. The numbers refer to the labeled positions (*black dots*).

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 (Figs. 1 and 3 *C*). It is slightly affected by the absence of pigments; however, its value for Ω (Fig. 3 *D*) 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 ~80% of the cases the N-terminus is folded back on the hydrophobic core (15). 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.

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 Chl *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.

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., and C. F. Yocum. 2006. Structure and function of photosystems I and II. *Annu. Rev. Plant Biol.* 57:521–565.
2. Jansson, S. 1999. A guide to the Lhc genes and their relatives in Arabidopsis. *Trends Plant Sci.* 4:236–240.
3. Caffarri, S., R. Croce, L. Cattivelli, and R. Bassi. 2004. A look within LHCII: differential analysis of the Lhcb1–3 complexes building the major trimeric antenna complex of higher-plant photosynthesis. *Biochemistry.* 43:9467–9476.
4. Dekker, J. P., and E. J. Boekema. 2005. Supramolecular organization of thylakoid membrane proteins in green plants. *Biochim. Biophys. Acta.* 1706:12–39.
5. Avenson, T. J., T. K. Ahn, D. Zigmantas, K. K. Niyogi, Z. Li, et al. 2008. Zeaxanthin radical cation formation in minor light-harvesting complexes of higher plant antenna. *J. Biol. Chem.* 283:3550–3558.
6. Ahn, T. K., T. J. Avenson, M. Ballottari, Y.-C. Cheng, K. K. Niyogi, et al. 2008. Architecture of a charge-transfer state regulating light harvesting in a plant antenna protein. *Science.* 320:794–797.
7. Liu, Z., H. Yan, K. Wang, T. Kuang, J. Zhang, et al. 2004. Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature.* 428:287–292.
8. Green, B. R., and D. G. Durnford. 1996. The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47:685–714.
9. Bassi, R., R. Croce, D. Cugini, and D. Sandonà. 1999. Mutational analysis of a higher plant antenna protein provides identification of chromophores bound into multiple sites. *Proc. Natl. Acad. Sci. USA.* 96:10056–10061.
10. Sandonà, D., R. Croce, A. Pagano, M. Crimi, and R. Bassi. 1998. Higher plants light harvesting proteins. Structure and function as revealed by mutation analysis of either protein or chromophore moieties. *Biochim. Biophys. Acta.* 1365:207–214.
11. Testi, M. G., R. Croce, P. P.-D. Laureto, and R. Bassi. 1996. A CK2 site is reversibly phosphorylated in the photosystem II subunit CP29. *FEBS Lett.* 399:245–250.

12. Croce, R., J. Breton, and R. Bassi. 1996. Conformational changes induced by phosphorylation in the CP29 subunit of Photosystem II. *Biochemistry*. 35:11142–11148.

13. Mauro, S., P. Dainese, R. Lannoye, and R. Bassi. 1997. Cold-resistant and cold-sensitive maize lines differ in the phosphorylation of the photosystem II subunit, CP29. *Plant Physiol*. 115:171–180.

14. Hemminga M. A., and L. J. Berliner, editors. 2007. ESR spectroscopy in membrane biophysics. Springer, New York.

15. Van Oort, B., S. Murali, E. Wientjes, R. B. M. Koehorst, R. B. Spruijt, et al. 2009. Ultrafast resonance energy transfer from a site-specifically attached fluorescent chromophore reveals the folding of the N-terminal domain of CP29. *Chem. Phys*. 357:113–119.

16. Stopar, D., J. Štrancar, R. B. Spruijt, and M. A. Hemminga. 2005. Exploring the local conformational space of a membrane protein by site-directed spin labeling. *J. Chem. Inf. Model*. 45:1621–1627.

17. Nagai, K., and H. C. Thøgersen. 1987. Synthesis and sequence-specific proteolysis of hybrid proteins produced in Escherichia coli. *Methods Enzymol*. 153:461–481.

18. Paulsen, H., U. Rümler, and W. Rüdiger. 1990. Reconstitution of pigment-containing complexes from light-harvesting chlorophyll a/b-binding protein overexpressed inEscherichia coli. *Planta*. 181:204–211.

19. Porra, R. J., W. A. Thompson, and P. E. Kriedemann. 1989. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta*. 975:384–394.

20. Davies, B. 1965. Analysis of carotenoid pigments. *In* Chemistry and Biochemistry of Plant Pigments. T. W. Goodwin, editor. Academy Press, New York. 489–532.

21. Giuffra, E., D. Cugini, R. Croce, and R. Bassi. 1996. Reconstitution and pigment-binding properties of recombinant CP29. *Eur. J. Biochem*. 238:112–120.

22. Croce, R., M. G. Muller, S. Caffarri, R. Bassi, and A. R. Holzwarth. 2003. Energy transfer pathways in the minor antenna complex CP29 of photosystem II: a femtosecond study of carotenoid to chlorophyll transfer on mutant and WT complexes. *Biophys. J*. 84:2517–2532.

23. Stopar, D., J. Štrancar, R. B. Spruijt, and M. A. Hemminga. 2006. Motional restrictions of membrane proteins: A site-directed spin labeling study. *Biophys. J*. 91:3341–3348.

24. Filipič, B., and J. Štrancar. 2001. Tuning EPR spectral parameters with a genetic algorithm. *Appl. Soft Comput*. 1:83–90.

25. Štrancar, J., T. Koklič, Z. Arsov, B. Filipič, D. Stopar, et al. 2005. Spin label EPR-based characterization of biosystem complexity. *J. Chem. Inf. Model*. 45:394–406.

26. Kavalenka, A. A., B. Filipič, M. A. Hemminga, and J. Štrancar. 2005. Speeding up a genetic algorithm for EPR-based spin label characterization of biosystem complexity. *J. Chem. Inf. Model*. 45:1628–1635.

27. Columbus, L., and W. L. Hubbell. 2004. Mapping backbone dynamics in solution with site-directed spin labeling: GCN4–58 bZip free and bound to DNA. *Biochemistry*. 43:7273–7287.

28. Steinhoff, H. J., A. Savitsky, C. Wegener, M. Pfeiffer, M. Plato, et al. 2000. High-field EPR studies of the structure and conformational changes of site directed spin labeled bacteriorhodopsin. *Biochim. Biophys. Acta*. 1457:253–262.

29. Štrancar, J., M. Šentjurc, and M. Schara. 2000. Fast and accurate characterization of biological membranes by EPR spectral simulations of nitroxides. *J. Magn. Reson*. 142:254–265.

30. Štrancar, J., T. Koklič, and Z. Arsov. 2003. Soft picture of lateral heterogeneity in biomembranes. *J. Membr. Biol*. 196:135–146.

31. Caffarri, S., F. Passarini, R. Bassi, and R. Croce. 2007. A specific binding site for neoxanthin in the monomeric antenna proteins CP26 and CP29 of Photosystem II. *FEBS Lett*. 581:4704–4710.

32. Paulsen, H., B. Finkenzeller, and N. Kühlein. 1993. Pigments induce folding of light-harvesting chlorophyll *a*/*b*-binding protein. *Eur. J. Biochem*. 215:809–816.

33. Horn, R., G. Grundmann, and H. Paulsen. 2007. Consecutive binding of chlorophylls *a* and *b* during the assembly *in vitro* of light-harvesting chlorophyll-*a*/*b* protein (LHCIIb). *J. Mol. Biol*. 366:1045–1054.

34. Hemelrijk, P. W., S. L. S. Kwa, R. van Grondelle, and J. P. Dekker. 1992. Spectroscopic properties of LHC-II, the main light-harvesting chlorophyll a/b protein complex from chloroplast membranes. *Biochim. Biophys. Acta*. 1098:159–166.

35. Gradinaru, C. C., I. H. M. van Stokkum, A. A. Pascal, R. van Grondelle, and H. van Amerongen. 2000. Identifying the pathways of energy transfer between carotenoids and chlorophylls in LHCII and CP29. A multicolor, femtosecond pump-probe study. *J. Phys. Chem. B*. 104:9330–9342.

36. Pascal, A., C. Gradinaru, U. Wacker, E. Peterman, F. Calkoen, et al. 1999. Spectroscopic characterization of the spinach Lhcb4 protein (CP29), a minor light-harvesting complex of photosystem II. *Eur. J. Biochem*. 262:817–823.

37. Mozzo, M., F. Passarini, R. Bassi, H. van Amerongen, and R. Croce. 2008. Photoprotection in higher plants: The putative quenching site is conserved in all outer light-harvesting complexes of photosystem II. *Biochim. Biophys. Acta*. 1777:1263–1267.

38. Mozzo, M., L. Dall'Osto, R. Hienerwadel, R. Bassi, and R. Croce. 2008. Photoprotection in the antenna complexes of photosystem II: role of individual xanthophylls in chlorophyll triplet quenching. *J. Biol. Chem*. 283:6184–6192.

39. Jeschke, G., A. Bender, T. Schweikardt, G. Panek, H. Decker, et al. 2005. Localization of the N-terminal domain in light-harvesting chlorophyll a/b protein by EPR measurements. *J. Biol. Chem*. 280:18623–18630.