UNIVERSITY OF SZEGED

ALBERT SZENT-GYÖRGYI MEDICAL CENTER

DEPARTMENT OF BIOCHEMISTRY

MOLECULAR MODELING OF SELECTED TRANSMEMBRANE PROTEINS

Ph.D. Thesis

by

DENYS BASHTOVYY

Institute of Biophysics

Biological Research Center

Hungarian Academy of Sciences

Szeged, 2004

Introduction

Transmembrane Proteins

Transmembrane proteins share a common property: part of their structure is embedded in a lipid bilayer. Therefore, being located at an interface, it is almost inevitable that they mediate communication between both sides of the membrane; receptors, pores and channels are all signal transducers. In their lipid-embedded domain, only two types of secondary structure have been observed – β-strands and α-helices. β-strands are found in outer membranes of Gram-negative bacteria, mitochondria and chloroplasts, forming rigid pores known as β-barrels. Single and bundled transmembrane α-helices have a broader range of functionalities and complexities. In some instances, extensive extramembrane domains complement the TM ones, and some membrane proteins consist of huge multisubunit complexes (e.g. cytochrome c oxidase, which contains up to 13 subunits in mammals).

Integral membrane proteins represent an important class of proteins which are employed in a wide range of cellular roles. The fact that these proteins are found in a lipid environment means that atomic resolution experimental structures for these proteins are few. To date only about 50 unique high-resolution integral membrane protein structures have been solved, whereas several dozens of thousands structures for globular proteins are known.

In the presented work we have focused on the structures of two distinct membrane proteins, the major coat protein of the bacteriophage M13 and the representative(s) of the cytochrome b_{561} redox protein family. Our group dedicated substantial effort to experimental studies of these proteins in a membrane-bound form. However, for none of these proteins was an atomic structure of the native, membrane-bound form available. Our task was, therefore, to build and analyze molecular models of these proteins on the basis of the relevant experimentally and *in silico* obtained information, in order to stimulate and design new experiments and also to verify available data.

Aims

In the presented work various molecular modeling techniques (quantum chemical calculations, molecular mechanics optimizations, conformational searches, homology modeling; sequence database searches and alignments, transmembrane and lipid-facing propensities prediction etc) were applied to model and study selected transmembrane proteins,

namely the Major Coat Protein of the M13 Bacteriophage M13 and 2 representatives (plant and mammalian) from the cytochrome b_{561} family. The aims of these studies were as follows:

Major Coat Protein of the Bacteriophage M13

- To test the bundle of distinct M13 MCP structures, refined in detergent micelles, against experimental constraints obtained from the protein embedded in a phospholipid bilayer in order to identify those structures that are most compatible with a lipid membrane environment;
- To test and refine earlier proposed indicator of the local packing density (the *f*-parameter), which is readily calculated from the coordinates of the optimized protein–lipid structural model;
- To investigate structural reasons for the increased outer hyperfine splitting ($2A_{max}$) values, observed in earlier EPR experiments from the spin-labeled residues 25 and 36;

Cytochrome b_{561} family

- To perform comprehensive sequence analysis on the representatives of the family, predict transmembrane and lipid-facing propensities of the TM helices;
- To identify new structural similarities in the cytochrome b_{561} family and to build 3-dimensional atomic models for representative plant and mammalian cytochrome b_{561} proteins.

Detailed molecular models should aid the understanding of the available experimental data on these proteins and in the design of new experiments.

Materials and Methods

M13 MCP structures

Three-dimensional structures of the M13 MCP were taken from the Brookhaven Protein Data Bank in PDB format. The PDB IDs were 2CPB and 2CPS for structures determined in dodecylphosphocholine (DodPC) and sodium dodecyl sulfate (SDS) micelles, respectively, by various high-resolution NMR techniques.

Experimental data

EPR outer hyperfine splittings (given in parenthesis after the each mutant name) and membrane topology data for the viable single cysteine mutants A25C, V31C, T36C, G38C and T46C of the M13 major coat protein reconstituted in 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) bilayers were taken from (Stopar et al., 1997a).

Molecular modeling of M13 MCP

Quantum chemistry and molecular mechanics package Spartan v.5.0.4 with MMFF94 force field, and the interactive molecular mechanics package Sculpt v.2.1, were used for building and optimization of structures. Specifically, Spartan was used to generate the spin-labeled cysteine residue (validated by semi-empirical quantum chemical methods) and the phospholipid structure, and for reoptimization of the protein structure after single-residue replacement by spin-labeled cysteine. Additionally, Spartan was used for single-point energy calculations to obtain atomic charges. Adjustment of the phospholipid chain configuration and constrained molecular mechanics optimization of the protein–lipid assemblies were performed in Sculpt. MOLMOL was used for producing single-residue substitutions, construction of the lipid shell, and preparing the system for optimization by molecular mechanics. Insight II was used for visualization and presentation of structures. All modeling work was performed on a Silicon Graphics Origin 2000 server and O2 workstations.

Molecular modeling of cytochromes b_{561}

Related sequences were identified from 26 different tissues and organisms via PSI-BLAST database searches with default settings using Artb561-1, Artb561-4 and Hosb561-1 as queries. The sequences were aligned using MULTICLUSTAL. Lipid facing propensities of the predicted TM regions have been analyzed by means of kPROT. 3-dimensional structures were built using Biopolymer and Homology modules in InsightII.

Results and Discussion

Major Coat Protein of the Bacteriophage M13

In the final optimized model, residues Y24 and F45 are located in the somewhat diffuse regions of the lipid phosphates on either side of the bilayer membrane (cf. A25 and T46 in the original experimental paper). Our prediction of the shift in approximately

1 residue of the TM part of the M13 MCP, relative to the membrane normal towards the N-terminus, has been supported later experimentally.

Formation of H-bonds between the ε-amino groups of K43 and K44 and the carbonyl oxygens of the lipid fatty acid chains is suggested by the model. At the opposite side of the bilayer, the model places W26 in a position where it can function as a membrane-anchoring residue. In the mutated structure, the ε-amino group of K40 interacts with oxygens of the maleimide ring for the spin label attached to C36. This interaction reduces the overall potential energy of the system by 20 kcal/mol (evaluated in Sculpt). As we see it now, this may be the only reason for the increased outer hyperfine splitting at C36 reported earlier.

Putative H-bonds between the N–H group of the indole ring and the carbonyl oxygens of the fatty acid chains in adjacent lipid molecules possibly contribute to immobilization of the spin label on C25 in the mutated structure.

The N-terminal helix of the final structure is oriented parallel to the membrane surface, in accordance with later, independent experimental findings.

Cytochromes b_{561}

The created sequence alignment including plant and animal members of the cytochrome b_{561} family supports the main conclusions on the conservation of functional elements from recent analyses on a smaller subset of the cytochrome b_{561} family. Together with the 2- and 3-dimensional structures built, this alignment sheds light on more structural details and raises a number of questions. Our observations provide evidence that the functionally relevant and structurally most conserved region in the cytochrome b_{561} family is the TMH2 to -5 4-helix core with an amino acid composition that is very well conserved in the inner surface and somewhat less conserved in the outer surface of the core. The two terminal helices (TMH1 and TMH6) are less conserved. They together with the interhelix loops and terminal regions are the main source of the variability in the family and may therefore define the specific subcellular location, physiological functions of the proteins they encode, and possibly their interactions with other proteins.

The high conservation of the motifs at 175–179 and 114–122, i.e., the putative MDA and Asc binding sites, respectively, is a further strong feature of the cytochrome b_{561} family. This suggests a key functional role for these putative binding sites in transmembrane electron transfer common to this protein family. There are several highly conserved residues located favorably between the two pairs of heme ligating histidine residues. Of these, the aromatic residues could indeed constitute the putative transmembrane electron transport pathway. In

addition, there are a few additional (nonconserved) aromatic residues in many sequences that could also contribute to such a pathway.

The high structural similarity between the plant and animal cytochrome b_{561} proteins, both at the sequence and protein structural level, suggests that the conserved machinery of transmembrane electron transfer mediated by these proteins serves diverse, yet to be explored physiological processes in eukaryotic cells.

Conclusions

M13 Major Coat Protein of the Bacteriophage M13 Model

- Relatively coarse-grained site-directed spin-label measurements have provided sufficient experimental constraints to select a single structural subclass from the family of high resolution NMR structures in micelles as being that most appropriate to the M13 MCP in lipid bilayer membranes.
- General configuration and topology of the selected structural subclass agrees with new, independent findings.
- A relatively simple indicator of the local packing density (the *f*-parameter), which is readily calculated from the coordinates of the optimized protein–lipid structural model, was found to be adequate for this purpose and has been further tested and developed.
- Extension of the approach to sparse experimental data on site-directed mutagenesis of other membrane proteins should be possible in the future.
- In our protein–lipid model, Y24 on one side of the membrane, and K43 and K44 on the other side, interact preferentially with the lipid head groups. The model indicates a hydrophobic mismatch of 3.5 Å or less (the protein is slightly shorter) between the unperturbed phospholipid bilayer and the intramembranous α-helix of the protein.
- Spin-labeled C25 is buried inside the hinge region, whereas C46 points towards the aqueous phase, in agreement with their strong and weak motional restriction, respectively.
- Shift in approximately 1 residue of the TM helix along the membrane normal towards the N-terminus is predicted: Y24 – F45 from previous experimental A25 – T46. The result has also been justified with independent, new data.
- Spin-label at C36 is restricted solely by involvement of its maleimide carbonyl oxygens in hydrogen bonding with K40.

- The model proves useful for the interpretation of future experimental data on membrane–M13 MCP systems. It became good starting point for full-scale molecular dynamics simulations and for the design of further site-specific spectroscopic experiments.

Cytochrome b_{561} Model

- The most detailed and extensive sequence alignment and analysis to date for the representatives of the cytochrome b_{561} family was performed.
- Transmembrane regions and lipophilic properties of the sequences have been obtained.
- 2 possible topological models of 2-D TMH arrangement were proposed and discussed.
- 3-D atomic models of the 4-TMH core of the mammalian and plant sequences were built for both 2-D topologies, representing transmembrane electron transfer machinery.
- The present 3-D structures provide useful working models for designing combined point mutation and biophysical experiments targeting heme ligation and putative electron transport pathways.
- The present models will be further refined as new structural data emerge in the future.

Acknowledgements

I am indebted to my mentors, Dr. László Dux and Dr. Tibor Páli for their continuous scientific guidance, encouragement and support throughout my Ph.D studies. I express my gratitude to Dr. Pál Ormos, director of the Institute of Biophysics at Biological Research Center, for allowing me to work at this research facility. I am thankful for the help I received from Dr. Alajos Bérczi and Dr. Balázs Szalontai. I am grateful to Dr. Derek Marsh from the Max-Planck Institute for Biophysical Chemistry, for the opportunity to visit his research group. I would like to thank Dr. Peter Knox for his advices and support. I am grateful to Dr. Erzsébet Szatmári and the rest of my family for their trust in me. Finally, I would like to thank all the members of the Institute of Biophysics, Usenet and Debian communities for the friendly and supportive environment.

Publications

Papers involved in the thesis

I. Bashtovyy, D., Marsh, D., Hemminga, M.A. and Páli, T. 2001. Constrained modeling of spin-labeled major coat protein mutants from M13 bacteriophage in a phospholipid bilayer. *Protein Sci*. 10 (5): 979-987

II. Bashtovyy, D., Bérczi, A., Asard, H. and Páli, T. 2003. Structure prediction for the di-heme cytochrome b_{561} protein family. *Protoplasma* 221 (1-2): 31-40

Paper associated with but not involved in the thesis

Krasilnikov, P.M., Bashtovyy, D., Knox, P.P. and Paschenko, V.Z. Hydrogen bonds network in *Rb. Sphaeroides* RC serves as a regulatory factor of the temperature dependence of the photooxidized bacteriochlorophyll and primary quinone acceptor rate constant. (Accepted to be published in *Biofizika*).

Abstracts related to the thesis

Bashtovyy, D. and Páli, T.: Molecular modeling studies on the spin-labeled major coat protein of the M13 bacteriophage in a phospholipid bilayer, Straub Days Conference, Szeged, Hungary, 7-9 December, 1999

Bashtovyy, D., Hemminga, M., Marsh, D. and Páli, T.: Molecular mechanics studies of a viral coat protein in a phospholipid bilayer, 3rd European Biophysics Congress, Munich, Germany, 9-13 September, 2000

Bashtovyy, D., Bérczi, A., Asard, H. and Páli, T.: Common structure predicted for the two-heme cytochromes b_{561} supports biological function, 6th International Conference on Plasma Membrane Redox Systems and Their Role in Biological Stress and Disease, Ravenna, Italy, 23-26 March, 2002

Bashtovyy, D., Bérczi, A., Asard, H., Páli, T.: Model structure for the di-heme cytochrome b_{561} membrane protein family, International Workshop on Lipid-Peptide/Lipid-Protein Interactions, Gomadingen, Germany, 25-27 March, 2002

Bashtovyy, D., Bérczi, A., Asard, H., Páli, T.: Structure prediction for the di-heme cytochrome b_{561} protein family, HU-FL Mini-Symposium on Abiotic Stress and Signalization in Plants, Szeged, Hungary, 20-21 December, 2002