PAPER 1

Constrained modeling of spin-labeled major coat protein mutants from M13 bacteriophage in a phospholipid bilayer

DENYS BASHTOVYY,[1] DEREK MARSH,[2] MARCUS A. HEMMINGA,[3] AND TIBOR PÁLI[1]

[1]Institute of Biophysics, Biological Research Centre, 6701 Szeged, Hungary
[2]Abteilung Spektroskopie, Max-Planck-Institut für Biophysikalische Chemie, 37070 Göttingen, Germany
[3]Department of Biomolecular Sciences, Wageningen University and Research Centre, 6700 ET Wageningen, The Netherlands

(RECEIVED October 18, 2000; FINAL REVISION February 5, 2001; ACCEPTED February 14, 2001)

Abstract

The family of three-dimensional molecular structures of the major coat protein from the M13 bacteriophage, which was determined in detergent micelles by NMR methods, has been analyzed by constrained geometry optimization in a phospholipid environment. A single-layer solvation shell of dioleoyl phosphatidylcholine lipids was built around the protein, after replacing single residues by cysteines with a covalently attached maleimide spin label. Both the residues substituted and the phospholipid were chosen for comparison with site-directed spin labeling EPR measurements of distance and local mobility made previously on membranous assemblies of the M13 coat protein purified from viable mutants. The main criteria for identifying promising candidate structures, out of the 300 single-residue mutant models generated for the membranous state, were 1) lack of steric conflicts with the phospholipid bilayer, 2) good match of the positions of spin-labeled residues along the membrane normal with EPR measurements, and 3) a good match between the sequence profiles of local rotational freedom and a structural restriction parameter for the spin-labeled residues obtained from the model. A single subclass of structure has been identified that best satisfies these criteria simultaneously. The model presented here is useful for the interpretation of future experimental data on membranous M13 coat protein systems. It is also a good starting point for full-scale molecular dynamics simulations and for the design of further site-specific spectroscopic experiments.

Keywords: Viral coat protein; M13 bacteriophage; site-directed spin-labeling; molecular modeling; electron paramagnetic resonance; membrane protein; lipid–protein interaction

Supplemental material: See www.proteinscience.org.

M13 is a small filamentous bacteriophage that is specific to *Escherichia coli*. It consists of ~2800 copies of the coat protein surrounding the genetic material, a circular single-stranded DNA (Marvin and Hohn 1969; Rasched and Oberer 1986). Most of the virus coat (98%) is made up of the gene 8 product that forms a 1.5–2.0 nm thick, flexible cylinder around the viral DNA: the major coat protein. The newly synthesized coat proteins are inserted in the cytoplasmic membrane of the host, where they are stored before being used in the assembly process. The major coat protein is a small multifunctional structural protein, composed of 50 amino acids (Van Wezenbeek et al. 1980). The primary sequence of the major coat protein is specifically tuned for diverse structural roles because it is capable of protein–DNA, protein–protein, and protein–lipid interactions (Hem-

Reprint requests to: Tibor Páli, Institute of Biophysics, Biological Research Centre, P.O. Box 521, 6701 Szeged, Hungary; e-mail: tpali@szbk.u-szeged.hu; fax: 36 62 433133.

Abbreviations: DOPC, dioleoylphosphatidylcholine; EPR, electron paramagnetic resonance; NMR, nuclear magnetic resonance; DodPC, dodecylphosphocholine; SDS, sodium dodecyl sulphate; 5-MSL, 5-maleimidoproxyl.

Article and publication are at www.proteinscience.org/cgi/doi/10.1110/ps.43801.

minga et al. 1993). To accomplish this, the major coat protein has three domains: a positively charged C-terminal domain, which primarily is responsible for the interaction with DNA; a hydrophobic central part of the protein, which stabilizes protein–protein interactions in the virus particle, as well as maintains a stable thermodynamic association with the host membrane during replication; and an amphipathic N-terminal domain, which is thought to be important in docking the protein at the virus assembly site (Marvin et al. 1994).

Knowledge of the three-dimensional structure of the assembly proteins in different stages of the M13 phage life cycle would be essential to understand the molecular details of viral action. Of particular interest is how the phage injects its genetic material into the host cell and then how the newly assembled phage leaves the cell without causing lysis of the host. Determination of the structure of the membranous form of the M13 major coat protein is therefore important. The classical structural methods, namely X-ray and NMR techniques, are subject to well-known limitations when applied to proteins in the membrane-bound state. In the absence of suitable crystals, possibilities to proceed are 1) to solubilize the membrane protein in detergent micelles for NMR studies, or 2) to use less direct spectroscopic structural techniques on the protein in its native membranous form.

A family of structural alternatives has been determined for the M13 major coat protein in both DodPC and SDS detergents (25 in each detergent), based on NMR studies (Papavoine et al. 1998). Although there are experimental indications that the structure of the protein may differ somewhat between detergent micelles and phospholipid bilayers (Stopar et al. 1996), these structures serve as a suitable starting point for modeling the M13 major coat protein in a phospholipid bilayer membrane. In this paper, we have tested these micellar structures against experimental constraints obtained from the protein embedded in a phospholipid bilayer to identify those most compatible with a lipid membrane environment. To achieve this, the structures of the M13 major coat protein were modified with single-residue replacements by a spin-labeled cysteine, and a first lipid-bilayer shell of DOPC was constructed around the protein. These modifications were followed by constrained molecular mechanics geometry optimization, whereby the protein backbone was left unchanged. The experimental constraints were distance data on the membrane topology of spin-labeled mutants (A25C, V31C, T36C, G38C, T46C, and A49C) of the coat protein reconstituted in a phospholipid bilayer, as well as EPR parameters characterizing the rotational freedom (namely outer hyperfine splitting) of the spin label attached at these residues (Stopar et al. 1997). It is found that the sequence profiles of these parameters are sufficiently selective to discriminate between various conformations of the membrane-bound protein. A unique subclass of structure is identified that best fits the available experimental data from membranes.

Results and Discussion

Starting structures

The backbone atoms (N, C, and C^{α}) of the M13 major coat protein structures in SDS micelles, determined by NMR and optimization techniques (Papavoine et al. 1998), are shown schematically in Figure 1. These structures were subject to further experimental constraints in the membrane-bound state, which were obtained from viable cysteine mutants in spin-label EPR experiments (Stopar et al. 1997). To do this, we built a spin-labeled cysteine, made single-residue replacements in the M13 major coat protein structures with cysteine-maleimide, optimized the modified structures, and then derived structural parameters from the models for comparison with the experimental EPR data. Some of the structures in Figure 1 that are possible in small, highly curved micelles can be eliminated immediately as suitable candidates for the structure in planar membranes, because the

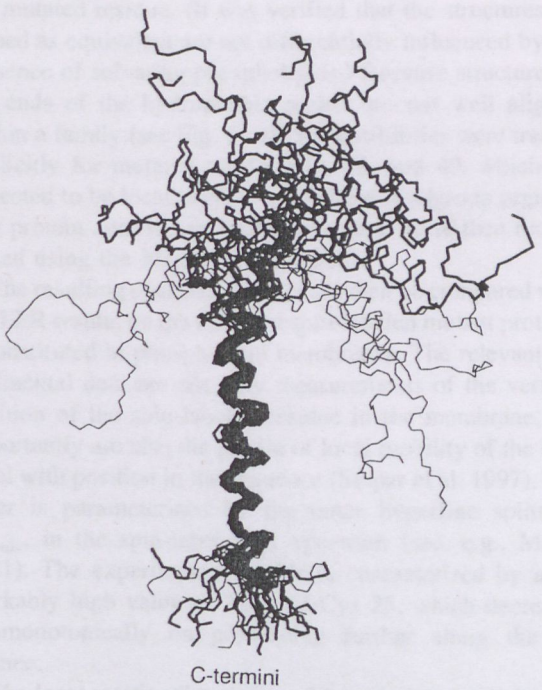

Fig. 1. Backbone atoms (N, C, and C^{α}) for the 25 low-energy structures (with the backbone aligned between residues 25 and 45) of the M13 major coat protein in SDS micelles. Structures were obtained from the Brookhaven protein data bank (entry code 2CPS) as published by Papavoine et al. (1998). Thin lines indicate U-shaped configurations that are incompatible with a planar membrane environment. The complete amino acid sequence is AEGDDPAKAAFNSLQASATEYIGYAWAMVVVIV GATIGIKLFKKFTSKAS.

charged N-terminal section would bend back into the hydrophobic interior of the membrane. These are the U-shaped structures that are indicated by light lines in Figure 1 and constitute not more than 7–8 members in each set (see Table 1). In addition, solid-state NMR studies on aligned phospholipid bilayers have indicated that the N-terminal helix (specifically the section containing Leu 14) of the closely related fd bacteriophage coat protein is oriented nearly perpendicular to the transmembrane helix (McDonnell et al. 1993). This further justifies elimination of the U-shaped structures.

Spin label structure

The 5-MSL spin label was built and minimized as the two stereoisomers arising from the chiral carbon in the proxyl ring. After connecting 5-MSL to the sulphydryl group of the cysteine residue, four stereoisomers were obtained. The eight structures obtained by rotating the proxyl ring around the N–C bond by 180° for each stereoisomer were optimized using MMFF94: One structure was chosen (Fig. 2) on the basis of lowest energy. Only two other of the alternative structures are expected to be comparably populated at room temperature. The remainder were predicted to lie higher in energy by 1.3–3.6 kcal/mol. In the chosen structure, the long axis of the maleimide spin label is oriented approximately perpendicular to the helix axis, in agreement with previous suggestions (Wolkers et al. 1997).

Table 1. *Angle between transmembrane and N-terminal helices in NMR-derived structures of M13 major coat protein in DodPC (PDB: 2CPB) and SDS (PDB: 2CPS) micelles determined by Papavoine et al. (1998)*[a]

Structure no.	DodPC	SDS
1	146°	100°
2	117°	129°
3	99°	139°
4	69°	30°
5	123°	111°
6	70°	90°
7	117°	122°
8	59°	17°
9	64°	63°[b]
10	80°	69°[b]
11	35°	108°
12	52°	146°
13	150°	113°
14	121°	111°
15	132°	61°[b]
16	153°	69°
17	100°	85°
18	50°	147°
19	18°	154°
20	96°	81°
21	65°[b]	56°
22	59°	51°
23	150°	93°
24	92°	125°
25	44°	91°

[a] Structures with an angle between helix axes of <60° were discarded in the model building.
[b] These structures were also discarded because the non-helical part of the N-terminal forms a U-shape.

Spin-labeled protein without lipid

The spin-labeled cysteine (Fig. 2) was substituted in the original NMR structures of Figure 1, and in the corresponding ones from DodPC, at positions corresponding to the single cysteine mutations (A25C, V31C, T36C, G38C, T46C, and A49C) used in the spin-label EPR studies (Stopar et al. 1997). The residue replacements were performed in MOLMOL and are indicated, on a single peptide backbone, in Figure 3. Excluding U-shaped conformations, this generates a total of ~35 possible structures for each mutant. For mutation positions definitely within the hydrophobic region (residues 31, 36, and 38), the number of independent structures was reduced to minimally 26 by grouping together equivalent structures. These are representative, on the basis of rms differences, of the local structures ±2 residues from the mutated residue. (It was verified that the structures assigned as equivalent are not differentially influenced by the presence of solvating phospholipids.) Because structures at the ends of the hydrophobic region are not well aligned within a family (see Fig. 1), all 35 possibilities were treated explicitly for mutated residues 25, 46, and 49, which are expected to be located close to the lipid headgroup regions. The protein-attached spin label structures were then reoptimized using the MMFF94 force field.

The resulting trial structures could then be compared with the EPR results on the different spin-labeled mutant proteins reconstituted in phospholipid membranes. The relevant experimental data are not only measurements of the vertical position of the spin-labeled residue in the membrane, but importantly are also the profile of local mobility of the spin label with position in the sequence (Stopar et al. 1997). The latter is parameterized by the outer hyperfine splitting, $2A_{max}$, in the spin-label EPR spectrum (see, e.g., Marsh 1981). The experimental profile is characterized by a remarkably high value of $2A_{max}$ at Cys 25, which decreases nonmonotonically on proceeding further along the sequence.

The local rotational mobility of the spin label will depend on the atomic packing density in the region of attachment of the spin label. We have chosen to characterize this packing density by a parameter (f) that is defined by

$$f = \sqrt{\sum_i \frac{m_i}{d_i^2}} \qquad (1)$$

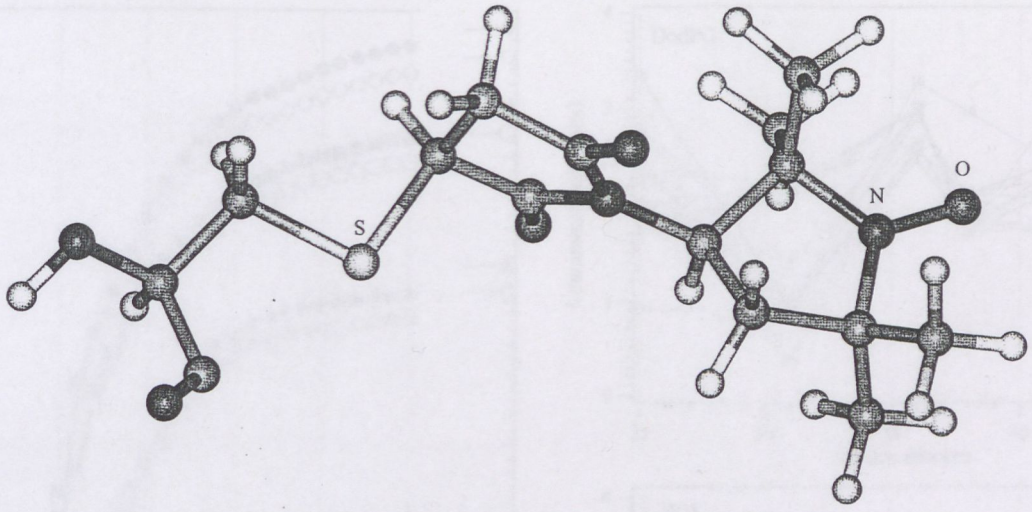

Fig. 2. Optimized structure of spin-labeled cysteine. The 5-maleimidoproxyl spin label that is covalently attached at the sulphydryl group (S) is indicated together with the nitroxide group (N–O) of the proxyl ring. The structure was built and optimized in Spartan with the MMFF94 force field, and is displayed using Insight II.

where m_i is the mass and d_i is the distance (in Å) of the ith atom measured from the reference atom. The reference position was the nitrogen in the proxyl ring of 5-MSL. All atoms of the protein, including hydrogens, were used in the summation. Only the atoms of the spin label were excluded from calculation of the f-parameter. The f-parameter in Equation 1 differs slightly from a previous one that was used successfully in a similar site-directed spin-labeling study on cytochrome c in solution (Turyna et al. 1998). (Note that the latter reference contains a printing error.) The extensions made here are to include mass weighting and to retain hydrogen atoms in the summation. Note that it is only relative values of A_{max} (i.e., the shape of the profile) with which we seek to correlate the f-parameter. For this, a qualitative correlation should suffice. However, the final fit (see Fig. 8, below) does imply an approximately linear relation, most probably because the spectra analyzed are confined to the same motional regime of spin-label dynamics (see, e.g., Marsh and Horváth 1989).

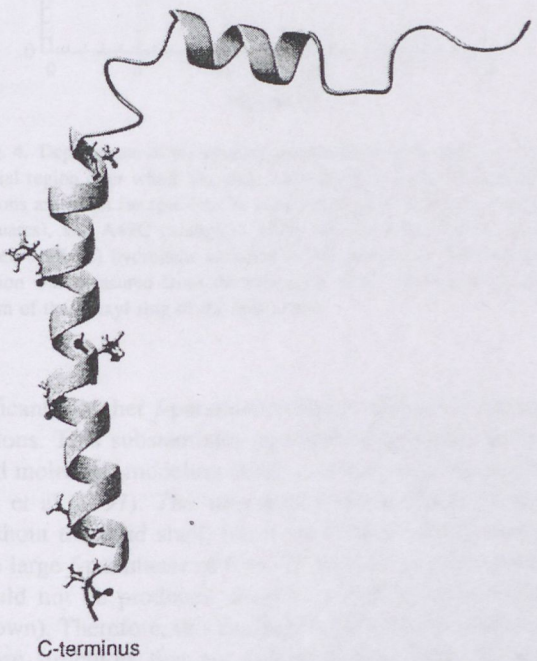

Fig. 3. Ribbon representation of the M13 major coat protein structure no. 20 of entry 2CPS from the Brookhaven Protein Data Bank (Papavoine et al. 1998). Side chains of the original residues that were changed singly to cysteine (*top* to *bottom*: A25C, V31C, T36C, G38C, T46C, and A49C), both by mutagenesis (Stopar et al. 1997) and by modeling in the present study, are shown in ball-and-stick representation. The figure was created using MOLMOL (Koradi et al. 1996).

The dependence of the f-parameter on the range over which the summation is made is shown in Figure 4. Data are given for three different mutants, and the effect of including or excluding hydrogen atoms is also shown. A 9-Å range was finally chosen for the summation, after testing various summations ranging from 4 to 50 Å, because this was found to give near-optimal discrimination between the different mutants (see Fig. 4). The 9-Å radial region is large enough to include all atoms that can affect the spin label directly and is small enough to exclude atoms that have no direct influence on the spin label.

The profiles of the calculated f-parameters are shown in Figure 5 for the two reduced sets of structures (i.e., from the families of structures in DodPC and SDS). Relative to the experimental EPR profile, the most significant feature of Figure 5 is that only structures in which the Cys 25 residue is inside the interhelix hinge region between the N-terminal and transmembrane segments (cf. Fig. 3) give rise to a sig-

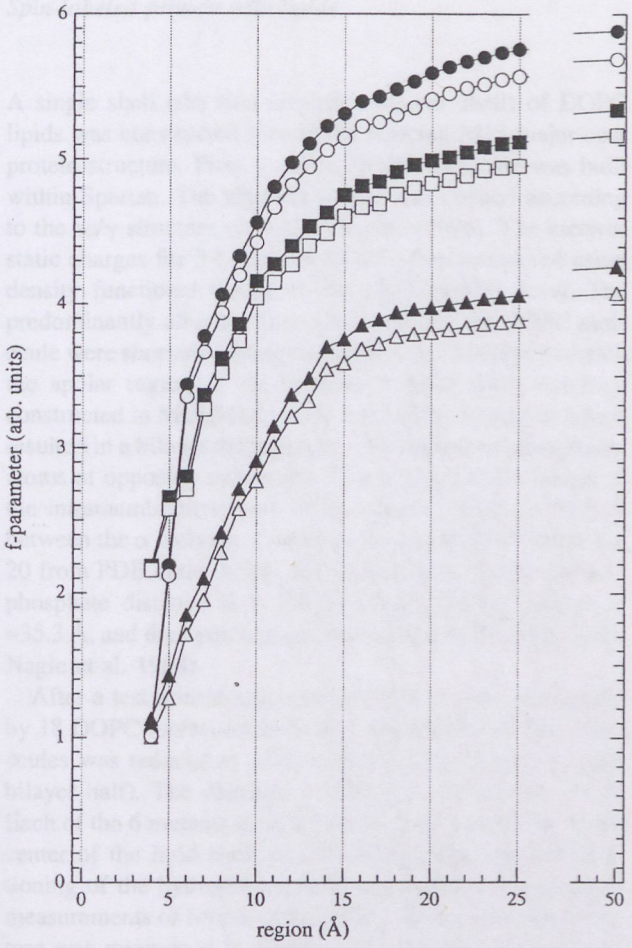

Fig. 4. Dependence of the packing parameter (*f*, defined in Eq. 1) on the radial region over which the atom summation is made. Results of calculations are given for spin-labeled cysteine mutants: A25C (circles), V31C (squares), and A49C (triangles), either with (solid symbols) or without (open symbols) hydrogens included in the evaluation. The radius of the region was measured from the reference atom, which was the nitrogen atom of the proxyl ring of the spin label.

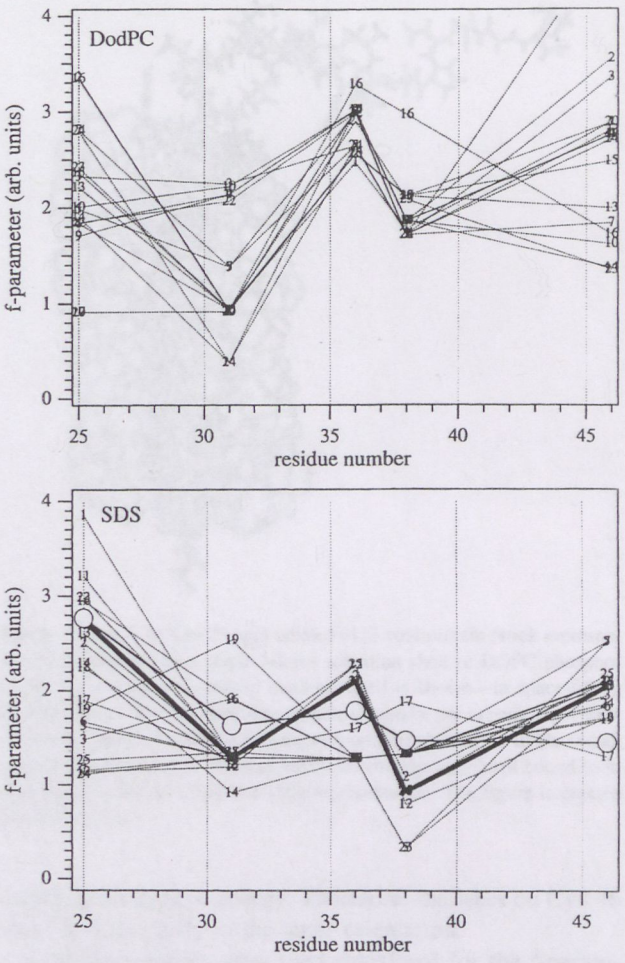

Fig. 5. Sequence profiles for the *f*-parameter of M13 coat protein structures with single cysteine–maleimide replacements, in the absence of a lipid bilayer. Each *x*-coordinate represents a single mutant of the original families of structures determined in DodPC (*top*) or in SDS (*bottom*), in which the cysteine–maleimide side chain was locally optimized after amino acid replacement. Structure numbers indicated in the figure are those in the PDB files 2CPB and 2CPS for NMR-based structures in DodPC and SDS micelles, respectively, determined by Papavoine et al. (1998). (*Bottom*) The heavy solid line indicates the selected structure no. 20; open circles are the experimental EPR outer hyperfine splittings scaled to best match this profile.

nificantly higher *f*-parameter than do the other mutant positions. This substantiates previous suggestions from EPR and molecular modeling studies on the hinge region (Wolkers et al. 1997). The structures used in Figure 5 are still without the lipid shell, but it was found subsequently that the large *f*-parameter of Cys 25, relative to other positions, could not be produced simply by adding lipids (data not shown). Therefore, this finding limits suitable candidates to those structures that are L-shaped (see Table 1) and for which Cys 25 is inside the hinge region. Surprisingly, no such structure was found in the DodPC family (i.e., PDB: 2CPB; Fig. 5, top). Out of the small subset (3–4) of very similar structures found to fulfill this condition in the SDS family (i.e., PDB: 2CPS), structure no. 20 was selected as representative for further model building because the *f*-parameter profile (see Fig. 5, bottom) most resembled the experimental mobility profile. In this subset of structures, the large *f*-parameter (and outer hyperfine splitting) for residue 25 is caused by the cysteine–maleimide being squeezed between Trp 26 and Ile 22 side chains. Additionally, putative hydrogen bonds between Trp 26 and maleimide oxygens possibly contribute to immobilization of the spin label. Structure no. 20 from the SDS family was a favorable choice also because the C-terminal half of the hinge region (residues 17–24) of the protein is better defined in SDS micelles than in DodPC micelles (Papavoine et al. 1998).

Spin-labeled protein with lipids

A single shell (the first solvation bilayer shell) of DOPC lipids was constructed around the selected M13 major coat protein structure. First, a single DOPC molecule was built within Spartan. The glycerol region was created according to the *sc*/γ structure given by Pascher (1996). The electrostatic charges for 5-MSL and DOPC were calculated using density functional theory at the pBP86/DN** level. The predominantly all-*trans* fatty acid chains of the DOPC molecule were shortened using the tether tool in Sculpt to match the apolar region of the protein. A lipid shell was then constructed in MOLMOL using this DOPC structure, which resulted in a bilayer thickness of ~33 Å between phosphorus atoms of opposing molecules. This is close to the length of the intramembranous part of the protein, which is 31.8 Å between the α carbons of residues 25 and 46 of structure no. 20 from PDB entry 2CPS. For comparison, the phosphate–phosphate distance in a fully hydrated DOPC bilayer is ≈35.3 Å, and the hydrocarbon thickness is 27.2 Å (Tristram-Nagle et al. 1998).

After a test minimization of the M13 protein surrounded by 18 DOPC molecules in Sculpt, the number of lipid molecules was reduced to 12 (one shell of six lipids for each bilayer half). The diameter of this lipid shell was ~19 Å. Each of the 6 mutants of structure no. 20 was inserted in the center of the lipid shell in MOLMOL, with vertical positioning of the hydrophobic helix according to the distance measurements of Stopar et al. (1997). This composite structure was minimized in Sculpt with only the protein backbone frozen. Additional forces (called springs in Sculpt) were applied to every atom of the lipid molecules, in a direction towards the axis of the protein hydrophobic helix. The MM3 force field was used and Van der Waals interactions were modeled with a modified Lennard-Jones potential between atoms within 6 Å of each other (Surles et al. 1994). Electrostatic interactions were treated with a Coulomb model, using a distance-dependent dielectric constant between atoms within 10 Å of each other (Surles et al. 1994). These structures were optimized until the fractional change in energy was <0.01. The resulting optimized model is shown for the spin-labeled A25C mutant in Figure 6.

Next, *f*-parameters were calculated for the minimized protein–lipid structures. In general, the lipid shell increased the *f*-parameter for residues 25, 31, 36, and 38 rather uniformly, whereas it had a discriminative effect at residues 46 and 49. For all mutants, the spin label was directed pointing away from the helix in the optimized structures (see Table 2). Exceptionally, this orientation produced an inconsistently high *f*-parameter for the maleimide connected to Cys 46, which is situated close to the lipid headgroups. However, for this particular residue, an orientation pointing outside the lipid shell (along the protein helix) produced a negligible change in energy. Therefore, the label on Cys 46 was fixed manually in the latter orientation.

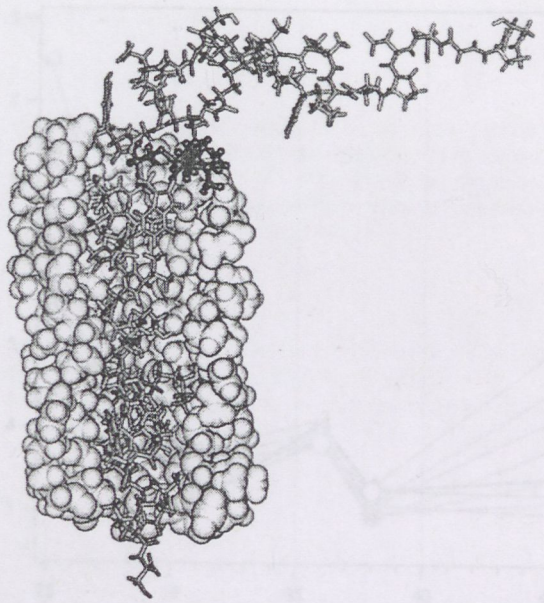

Fig. 6. Structure of Cys 25-spin labeled M13 coat protein (stick representation) surrounded by a single bilayer solvation shell of DOPC phospholipids. For clarity, only part of the lipid shell is shown—in space-filling representation. The structure was chosen to satisfy the experimental EPR constraints and was obtained by geometry optimization of the lipids and all protein side chains. Cys 25, with the 5-maleimidoproxyl label bound to it, is enhanced with dark ball-and-stick representation. The figure is created using Insight II.

Sequence profiles were then calculated for the *f*-parameter, not only with the protein in the position used for the structural optimization, but also with the protein displaced vertically in 1-Å increments, without further optimization. The resulting *f*-parameter profiles were then fitted to that of the EPR outer hyperfine splittings by using an adjustable linear scaling factor, plus offset, for the latter. The depen-

Table 2. *Orientation of the spin-label N-O bond to the axis of the transmembrane helix, for structure number 20 of the M13 coat protein (PDB: 2CPS) optimized with single cysteine-maleimide substitutions at different positions, in the absence and presence of the lipid shell*

Spin-labeled residue no.	Without lipid	With lipid[a]
25	67°	51°
31	104°	111°
36	107°	95°
38	100°	97°
46	77°	40°[b]
49	70°	70°

[a] Final structure optimized with +1 Å vertical shift of the protein.
[b] Adjusted manually, producing an isoenergetic structure (see text).

dence of the fitting errors on the vertical displacement of the protein is given in Figure 7. The minimum fitting error was not found for zero vertical displacement of the protein, but for a displacement by +2 Å. Therefore, optimization of the entire protein–lipid structure was repeated with different vertical displacements of the protein. A consistent minimum fitting error was obtained for the structure optimized with a vertical displacement of +1 Å. This represents a one-residue shift, or less, relative to the original EPR estimate. The resulting sequence profile of the *f*-parameter is compared with that of the EPR outer hyperfine splittings in Figure 8. It is seen that the packing parameters deduced from the final optimized structure reproduce the major features of the experimentally determined mobility profile rather well. Including lipids in the model considerably improves the agreement with experiment (cf. Fig. 5); the rms matching error is reduced from 0.86 to 0.32, in the presence of lipids.

Final model

In the final optimized model, residues Tyr 24 and Phe 45 are located in the (somewhat diffuse) regions of the lipid phosphates on either side of the bilayer membrane. Formation of

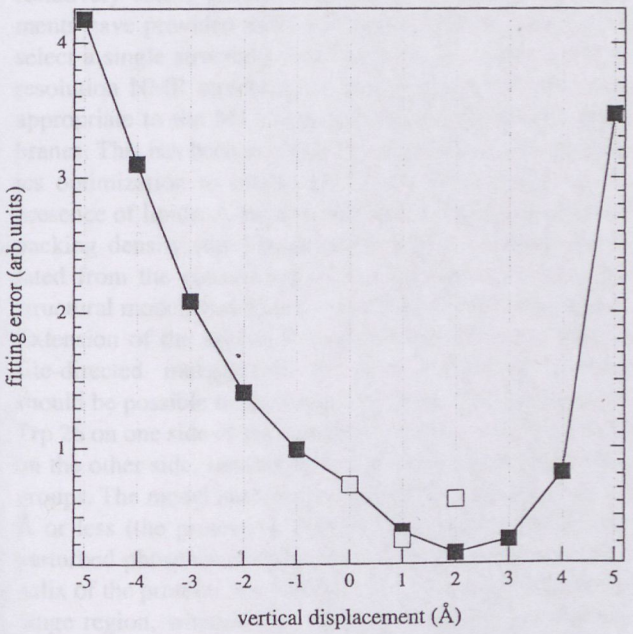

Fig. 7. Matching error between the sequence profiles of the ^{14}N outer hyperfine splitting (Stopar et al. 1997) and *f*-parameter, as a function of vertical displacement of the M13 major coat protein structure, relative to that shown in Fig. 6, in a DOPC bilayer. The minimum sum of squares of the differences was obtained by linear scaling of the outer splitting profile (with offset) to that of the *f*-parameter. Solid symbols indicate that the protein structure was not reoptimized at the new vertical position; open symbols denote reoptimization of the structure at the shifted position. Positive shift means that the N-terminal helix is moved closer to the membrane.

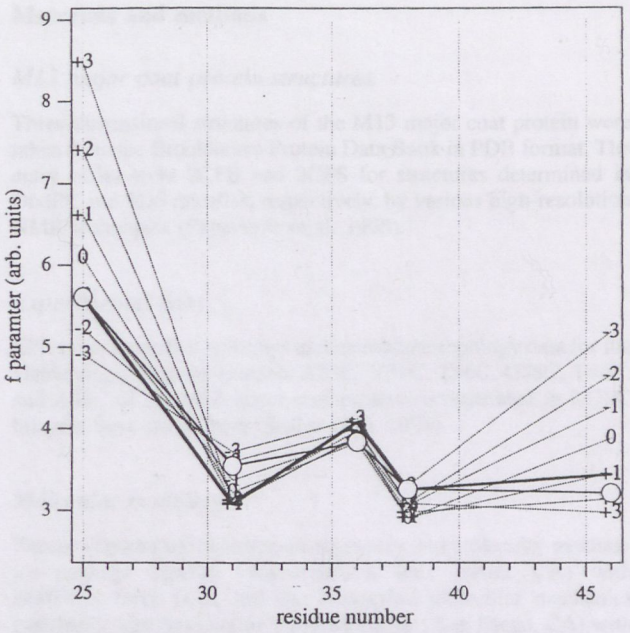

Fig. 8. Sequence profiles of the *f*-parameter from cysteine–maleimide mutants of the M13 major coat protein at different vertical shifts (broken lines), for the structure shown in Fig. 6 that was optimized with zero shift. The heavy solid line gives the sequence profile for the final structure that is reoptimized with +1 Å vertical shift. Circles are the experimental outer hyperfine splitting profile that is scaled linearly (with offset) for best match to the *f*-parameter profile of the final (+1 Å-shifted) structure.

H-bonds between the ε-amino groups of Lys 43 and Lys 44 and the carbonyl groups of the lipid fatty acid chains is suggested by the model. To some extent this resembles the snorkel effect (Monné et al. 1998). At the opposite side of the bilayer, the model places Trp 26 in a position where it can function as a membrane-anchoring residue (Schiffer et al. 1992). In the mutated structure, the ε-amino group of Lys 40 interacts with oxygens of the maleimide ring for the spin label attached to Cys 36. This interaction reduces the overall potential energy of the system by 20 kcal/mol (evaluated in Sculpt). At least in part, this may be the reason for the increased outer hyperfine splitting at Cys 36 reported in Stopar et al. (1996, 1997). 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 Cys 25 in the mutated structure. The N-terminal helix of the final structure is oriented parallel to the membrane surface (see Fig. 6), consistent with solid-state NMR results on the closely related fd coat protein in oriented membranes (McDonnell et al. 1993). An interesting feature of the amphipathic N-terminal surface helix (residues 8–16) is its azimuthal orientation relative to the membrane surface. This differs somewhat between the NMR structures determined in DodPC and SDS micelles. As already noted, the EPR data from phosphatidylcholine bilayer membranes, although

not referring specifically to the N-terminal structure, is consistent only with the SDS-family. The N-terminal helix is oriented with the face containing the charged and polar residues Lys 8 and Gln 15, and aromatic residue Phe 11, directed towards the membrane. Correspondingly, the opposite face containing alanine residues 9, 10, and 16, and Ser 13, is directed towards the aqueous phase (cf. Papavoine et al. 1994). Such an orientation may not be the optimal one, as one would expect the strongly hydrophobic Leu 14 to be oriented toward the membrane rather than equatorially as it is in the model. This orientation of the N-terminal helix is likely to be specific for the strongly negative surface potential of SDS micelles and may be relieved by a slight twist of the helix with respect to the membrane in zwitterionic lipids. The final model remains hypothetical both in the sense that the starting structures are for a micelle rather than a bilayer environment and more particularly that only one coarse constraint has been used to distinguish between the N-terminal starting structures.

Conclusions

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 coat protein in lipid bilayer membranes. This has been achieved by using molecular mechanics optimization to model the spin-labeled protein in the presence of lipids. 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. 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, Trp 26 on one side of the membrane, and Lys 43 and Lys 44 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 Cys 25 is buried inside the hinge region, whereas Cys 46 points towards the aqueous phase, in agreement with their strong and weak motional restriction, respectively. Further, Cys 36 seems to be restricted by involvement in hydrogen bonding with Lys 40 in addition to steric hindrance from Val 33 and Ile 39. The model should prove useful for the interpretation of future experimental data on membrane–M13 major coat protein systems. It is also a good starting point for full-scale molecular dynamics simulations and for the design of further site-specific spectroscopic experiments.

Materials and methods

M13 major coat protein structures

Three-dimensional structures of the M13 major coat protein were taken from the Brookhaven Protein Data Bank in PDB format. The entry codes were 2CPB and 2CPS for structures determined in DodPC and SDS micelles, respectively, by various high-resolution NMR techniques (Papavoine et al. 1998).

Experimental data

EPR outer hyperfine splittings and membrane topology data for the viable single cysteine mutants A25C, V31C, T36C, G38C, T46C, and A49C of the M13 major coat protein reconstituted in DOPC bilayers were taken from (Stopar et al. 1997).

Molecular modeling

Recent versions of the quantum chemistry and molecular mechanics package Spartan (Wavefunction Inc., Irvine, CA) with MMFF94 force field, and the interactive molecular mechanics package Sculpt (Interactive Simulations Inc., San Diego, CA) with MM3 force field, 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 (Koradi et al. 1996) was used for producing single-residue substitutions, construction of the lipid shell, and preparing the system for optimization by molecular mechanics. Insight II (Molecular Simulations Inc., San Diego, CA) was used for visualization and presentation of structures. All modeling work was performed on a Silicon Graphics (Mountain View, CA) Origin 2000 server and O2 workstations.

Electronic supplemental material

Atomic coordinates of the final optimized model, with the 5-MSL spin label bound on Cys 25, together with a single shell of DOPC lipids, is provided as exported from Sculpt (file name is m13lipid.xyz). The structure is similar to that presented in Figure 6, but was reoptimized after the protein was shifted by +1 Å relative to the lipid shell. This provided the lowest matching error (cf. Figs. 7,8). The data format follows that of the Protein Data Bank.

Acknowledgments

This work was supported by the Volkswagen-Stiftung (Germany), by the Hungarian National Science Foundation (OTKA T029458), and by the János Bolyai Foundation (Hungary). One of the SGI workstations for the modeling calculations was donated via the Max-Planck Gesellschaft (Germany).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Hemminga, M.A., Sanders, J.C., Wolfs, C.J.A.M., and Spruijt, R.B. 1993. Lipid-protein interactions involved in bacteriophage M13 infection. In *Protein-lipid interactions, new comprehensive biochemistry 25* (ed. A. Watts), pp. 191–212. Elsevier, Amsterdam.

Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14:** 51–55.

Marsh, D. 1981. Electron spin resonance: Spin labels. In *Membrane spectroscopy. Molecular biology, biochemistry and biophysics*, Vol. 31 (ed. E. Grell), pp. 51–142. Springer-Verlag, New York.

Marsh, D. and Horváth, L.I. 1989. Spin-label studies of the structure and dynamics of lipids and proteins in membranes. In *Applications in biology and biochemistry advanced EPR* (ed. A.J. Hoff), pp. 707–752. Elsevier, Amsterdam.

Marvin, D.A. and Hohn, B. 1969. Filamentous bacterial viruses. *Bacteriol. Rev.* **33:** 172–209.

Marvin, D.A., Hale, R.D., Nave, C., and Citterich, M.H. 1994. Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages Ff (fd, f1, M13), If1 and Ike. *J. Mol. Biol.* **235:** 260–286.

McDonnell, P.A., Shon, K., Kim, Y., and Opella, S.J. 1993. fd coat protein structure in membrane environments. *J. Mol. Biol.* **233:** 447–463.

Monné, M., Nilsson, I., Johansson, M., Elmhed, N., and von Heijne, G. 1998. Positively and negatively charged residues have different effects on the position in the membrane of a model transmembrane helix. *J. Mol. Biol.* **284:** 1177–1183.

Papavoine, C.H.M., Konings, R.N.H., Hilbers, C.W., and van de Ven, F.J.M. 1994. Location of the M13 coat protein in sodium dodecyl sulphate micelles as determined by NMR. *Biochemistry* **33:** 12990–12997.

Papavoine, C.H.M., Christiaans, B.E.C., Folmer, R.H.A., Konings, R.N.H., and Hilbers, C.W. 1998. Solution structure of the M13 major coat protein in detergent micelles: A basis for a model of phage assembly involving specific residues. *J. Mol. Biol.* **282:** 401–419.

Pascher, I. 1996. The different conformations of the glycerol region of crystalline acylglycerols. *Curr. Opin. Struct. Biol.* **6:** 439–448.

Rasched, I. and Oberer, E. 1986. Ff coliphages: Structural and functional relationships. *Microbiol. Rev.* **50:** 401–427.

Schiffer, M., Chang, C.H., and Stevens, F.J. 1992. The functions of tryptophan residues in membrane proteins. *Protein Eng.* **5:** 213–214.

Stopar, D., Spruijt, R.B., Wolfs, C.J.A.M., and Hemminga, M.A. 1996. Local dynamics of the M13 major coat protein in different membrane-mimicking systems. *Biochemistry* **35:** 15467–15473.

Stopar, D., Jansen, K.A.J., Páli, T., Marsh, D., and Hemminga, M.A. 1997. Membrane location of spin-labeled M13 major coat protein mutants determined by paramagnetic relaxation agents. *Biochemistry* **36:** 8261–8268.

Surles, M.C., Richardson, J.S., Richardson, D.C., and Brooks, Jr., F.P. 1994. Sculpting proteins interactively: Continual energy minimisation embedded in a graphical modelling system. *Protein Sci.* **3:** 198–210.

Tristram-Nagle, S., Petrache, H.I., and Nagle, J.F. 1998. Structure and interactions of fully hydrated dioleoylphosphatidylcholine bilayers. *Biophys. J.* **75:** 917–925.

Turyna, B., Osyczka, A., Kostrzewa, A., Blicharski, W., Enghild, J.J., and Froncisz, W. 1998. Preparation and electron paramagnetic resonance characterisation of spin labeled monoderivatives of horse cytochrome *c*. *Biochim. Biophys. Acta* **1386:** 50–58.

Van Wezenbeek, P.M.G.F., Hulsebos, T.J.M., and Schoenmakers, J.G.G. 1980. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene* **11:** 129–148.

Wolkers, W.F., Spruijt, R.B., Kaan, A., Konings, R.N.H., and Hemminga, M.A. 1997. Conventional and saturation-transfer EPR of spin-labeled mutant bacteriophage M13 coat protein in phospholipid bilayers. *Biochim. Biophys. Acta* **1327:** 5–16.

PAPER 2

Protoplasma (2003) 221: 31–40
DOI 10.1007/s00709-002-0065-0

PROTOPLASMA
Printed in Austria

Structure prediction for the di-heme cytochrome b_{561} protein family

Denys Bashtovyy[1], **Alajos Bérczi**[1], **Han Asard**[2], and **Tibor Páli**[1,*]

[1] Institute of Biophysics, Biological Research Centre, Szeged
[2] Beadle Center for Genetic Research, University of Nebraska-Lincoln, Lincoln, Nebraska

Received May 12, 2002; accepted September 20, 2002; published online May 21, 2003

Summary. Atomic models possessing the common structural features identified for the cytochrome b_{561} (cyt b_{561}) protein family are presented. A detailed and extensive sequence analysis was performed in order to identify and characterize protein sequences in this family of transmembrane electron transport proteins. According to transmembrane helix predictions, all sequences contain 6 transmembrane helices of which 2–6 are located closely in the same regions of the 26 sequences in the alignment. A mammalian (*Homo sapiens*) and a plant (*Arabidopsis thaliana*) sequence were selected to build 3-dimensional structures at atomic detail using molecular modeling tools. The main structural constraints included the 2 pairs of heme-ligating His residues that are fully conserved in the family and the lipid-facing sides of the helices, which were also very well conserved. The current paper proposes 3-dimensional structures which to our knowledge are the first ones for any protein in the cyt b_{561} family. The highly conserved His residues anchoring the two hemes on the cytoplasmic side and noncytoplasmic side of the membrane are in all proteins located in the transmembrane helices 2, 4 and 3, 5, respectively. Several highly conserved amino acids with aromatic side chain are identified between the two heme ligation sites. These residues may constitute a putative transmembrane electron transport pathway. The present study demonstrates that the structural features in the cyt b_{561} family are well conserved at both the sequence and the protein level. The central 4-helix core represents a transmembrane electron transfer architecture that is highly conserved in eukaryotic species.

Keywords: Cytochrome b_{561}; Structure prediction; Heme protein; Protein family; Ascorbic acid; Electron transport.

Abbreviations: Asc ascorbate; cyt cytochrome; MDA monodehydroascorbate; TM transmembrane; 3-D 3-dimensional.

Introduction

The high-redox-potential *b*-type cytochrome (cyt b_{561}) of chromaffin granule membranes of the mammalian adrenal medulla can be fully reduced by ascorbate (Asc). The wavelength of its characteristic alpha-band absorbance maximum in the reduced-minus-oxidized absorbance spectra is close to 561 nm. The protein is capable of transporting electrons through the chromaffin granule membrane (Njus et al. 1983, Srivastava et al. 1984, Kelley and Njus 1986, Kent and Fleming 1987). In the past decades, evidence accumulated for the presence of a similar Asc-reducible cyt b_{561} in plant plasma membranes (for a recent review, see Asard et al. 2001). This protein is also able to transfer electrons across the membrane (Asard et al. 1992) in a way that may be similar to that of the chromaffin granule membrane. Genes coding for proteins with significant homology to the mammalian cyt b_{561} have recently been identified in a large number of plant species (Asard et al. 2000). The mammalian and predicted plant cyt b_{561} proteins are highly hydrophobic and transport electrons from the cytoplasmic side of the membrane in which they are embedded to the extracellular space or into intracellular vesicles. The physiological function of the plant plasma membrane cyt b_{561} is yet to be elucidated. The first evidence for the existence of the cyt b_{561} protein family in plants and animals was presented on the basis of a sequence analysis of 9 related sequences from different eukaryotic species (Asard et al. 2000). It is generally believed that these redox proteins play an important role in a wide variety of physiological processes, including iron uptake, cell defense, nitrate reduction, and signal transduction. Recently a new member of this protein family has been located in the duodenal cells of the small intestine and demonstrated to play a role in the reduction of iron prior to its uptake (McKie et al. 2001). It was suggested that a cyt b_{561} in these cells pos-

* Correspondence and reprints: Institute of Biophysics, Biological Research Centre, P.O. Box 521, 6701 Szeged, Hungary.
E-mail: tpali@nucleus.szbk.u-szeged.hu

sesses ferric reductase activity. A similar function was proposed for cyt b_{561}-like domains in larger proteins which play a potential role in neurodegenerative disorders (Ponting 2001). This activity apparently contrasts with the activity of the chromaffin granule and plant plasma membrane cytochromes b_{561}, which are likely to function as monodehydroascorbate (MDA) reductases. These proteins have been demonstrated to receive an electron from cytoplasmic Asc and transfer it across the membrane to MDA (Kelley and Njus 1986, Harnadek et al. 1992). Details of this process, in particular the transmembrane electron transfer mechanism, are not yet resolved, but almost certainly the two heme centers are involved (Tsubaki et al. 1997, Kobayashi et al. 1998, Trost et al. 2000). These proteins therefore represent an important and unique family of transmembrane electron transport proteins because of their putative and/or yet to be identified physiological functions in a variety of eukaryotic cells. In addition, the 1 eq reaction between Asc and the proteins and the long distance between the two hemes, which almost spans the membrane interior, make them a potentially very interesting model for redox reactions between metalloproteins and organic substrates and also for transmembrane electron transfer.

Since cyt b_{561} proteins have not yet been crystallized, atomic-detail structural data about these proteins are lacking. On the other hand, essential structural features, including conserved heme-binding residues, transmembrane (TM) helices, and potential substrate binding sites have been identified (Okuyama et al. 1998, Asard et al. 2001, Takeuchi et al. 2001). Moreover, success in the purification of the plant plasma membrane cyt b_{561} (Trost et al. 2000; Bérczi et al. 2001, 2003) may render biophysical studies on plant proteins possible in the near future. Therefore, a working model of atomic detail representing the main structural features of the cyt b_{561} protein family may be highly useful. The objectives of the present work were to identify new structural similarities in the cyt b_{561} family and to build 3-dimensional (3-D) atomic models for representative plant and mammalian cyt b_{561} proteins. In the absence of direct structural data, this was done in three steps: (1) exploring the structural features of the family via sequence alignment and predicting the TM helices, (2) establishing the intra- and extracellular topology and the lipid-facing propensities of the TM helices, and (3) building 3-D atomic models that satisfy all known and newly explored structural constraints. Although we used most recent modeling tools and we built structures, our strategy turned out to be essentially similar to that followed by Hägerhäll and Hederstedt (1996). The structural features of the predicted models are discussed in relation to the structural data and the putative electron transport mechanism.

Results

Features of the cytochrome b_{561} family at the primary structural level

Recent studies involving sequence comparisons have suggested a close relationship between cyt b_{561} proteins of the plant and animal kingdom and demonstrated the conservation of a number of structural features (Asard et al. 2000, 2001; McKie et al. 2001; Ponting 2001; Asada et al. 2002). In order to identify additional conserved properties and structural features of sequences related to the well-known plant and human cyt b_{561} proteins, more members of the family had to be included in a sequence comparison. Related sequences were identified from 26 different tissues and organisms via PSI-BLAST database searches with default settings (Altschul et al. 1997) using Artb561-1, Artb561-4 and Hosb561-1 as queries (sequence names are used as defined in Asard et al. 2001). Sequences were selected which (1) had previously been reported to belong to this family (Asard et al. 2001), (2) showed conserved functionally relevant key residues in similar locations, primarily the heme-ligating histidines and the SLHSW motif (a putative MDA binding site; Tsubaki et al. 1997, Asard et al. 2001) and (3) were at least 200 residues long. This latter constraint was needed to exclude incomplete gene fragments (Asard et al. 2001). The sequences were aligned using MULTICLUSTAL (Yuan et al. 1999). Figure 1 represents the most complete and detailed sequence alignment of the cyt b_{561} family to date. All sequences contain 6 regions that are rich in highly conserved amino acid residues or residue properties. These conserved regions are separated by regions with no or very little conservation and most of the gaps can be found in these nonconserved regions too. Clearly, the sequences display largest variability in these intermediate and the terminal nonconserved regions. Notable is that the nonconserved regions are too short to form membrane-spanning alpha-helices (TM helices), especially if the gaps are also taken into account. The conserved regions, on the other hand, overlap very well

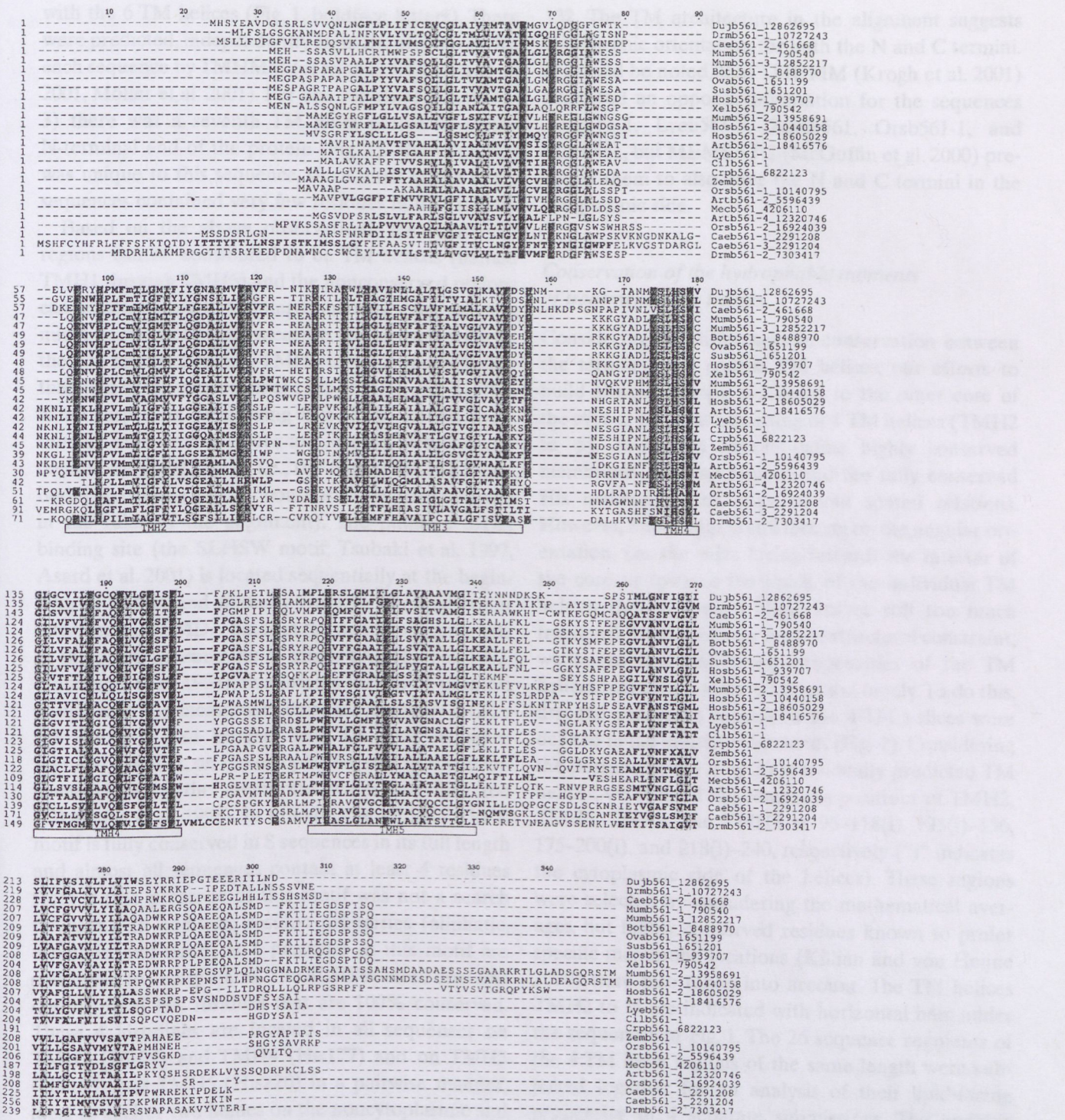

Fig. 1. MULTICLUSTAL alignment of 26 cyt b_{561} sequences. The names and numbers before and after the underscore are names according to the proposed naming convention (Asard et al. 2001) and GenBank identifiers (Benson et al. 2002), respectively. Highlighting is performed with TeXshade (Beitz 2000) above 85% threshold conservation of residue properties and according to the chemical mode. Concerning font styles and gray shades, on a relative scale between white and black of 0 to 100%, residue properties (Karlin 1985) are represented as follows: aromatic (F, W, Y), 50%; basic (H, K, R), 42%; hydroxyl (S, T), 34%; aliphatic (A, G, I, L, V), 26%; acidic (D, E), 18%; amide (N, Q), 10%; imino (P), font in italics (no shading); sulfur (C, M), lower case font (no shading). Amino acids in the TM segments, predicted by TMHMM (Krogh et al. 2001) independently for each sequence, are shown in boldface. The ruler above the alignment is numbered according to the consensus sequence. The horizontal bars indicate the consensus TM helices TMH2 to -5 that are most conserved

with the 6 TM helices (Fig. 1, boldface letters). These were predicted, independently from the alignment, for each sequence by TMHMM (version 2.0) (Krogh et al. 2001, Möller et al. 2001). For one sequence (Caeb561-3) there was a seventh TM helix predicted at the N-terminal end of the protein (Fig. 1). However, this was unique to this sequence, since most of the other sequences contained very few residues in that region.

Based on the above observations, the conserved regions can be considered to be TM helices (named TMH1 through TMH6) and the nonconserved regions can be considered to be interconnecting loops and terminal domains. By any measure, the conservation of the TM helices 2–5 (i.e., TMH2 to -5) is much higher than that of the two terminal ones (TMH1 and -6). This is not surprising considering that all the known functionally important residues are located in the region defined by TMH2 to -5, i.e., the region 95–240 (numbering according to the "consensus" sequence shown in the ruler of the alignment). The putative MDA binding site (the SLHSW motif; Tsubaki et al. 1997, Asard et al. 2001) is located sequentially at the beginning of TMH4 and spatially close to the lipid–water interface. Since MDA binds at the noncytoplasmic side of the plasma membrane in plants (Asard et al. 1992, Horemans et al. 1994), this orients the whole putative structure in the membrane. This automatically locates the putative Asc binding site, i.e., the sequence segment 114–122, on the cytoplasmic side of the membrane close to the membrane–water interface too (Okuyama et al. 1998). In this site, the ALLVYRVFR motif is fully conserved in 8 sequences in its full length and almost all sequences contain at least 4 residues from it (the presence of this motif was not a search criterion). In addition, residue properties (aliphatic, aromatic, and basic) at 4 positions in this motif are conserved to nearly 100%. The SLHSW motif contains one of the 4 His residues that are 100% conserved. These 4 histidines are located in all sequences on TMH2 (His99) and TMH4 (His177) and on TMH3 (His135) and TMH5 (His220) in a pairwise manner, ideal to anchor two hemes on the noncytoplasmic and cytoplasmic side of membrane, respectively, as proposed earlier (e.g., Okuyama et al. 1998, Tsubaki et al. 2000). A number of highly conserved aromatic amino acid residues are identified in the TM helices between the heme-ligating His residues. TMH4 appears to be richest in conserved aromatic residues. Notable is the presence of highly conserved basic residues located between TMH2 and TMH3 at positions 119, 128, and 132. The TM architecture in the alignment suggests cytoplasmic orientation for both the N and C termini. It should be noted, that TMHMM (Krogh et al. 2001) predicted an opposite orientation for the sequences Artb561-1, Lyeb561-1, Zemb561, Orsb561-1, and Mecb561, but MEMSAT 2 (McGuffin et al. 2000) predicted them to also have the N and C termini in the cytoplasmic side.

Conservation of the hydrophobic moments of transmembrane helices

Considering the differences in conservation between the terminal and central TM helices, our efforts to build structures were restricted to the inner core of the cyt b_{561} proteins consisting of 4 TM helices (TMH2 to -5). These TM helices define highly conserved membrane-spanning regions and the fully conserved His residues define their overall spatial relations. However, knowledge is still lacking on the angular orientation, i.e., the sides facing towards the interior of the core or towards the lipids, of the individual TM helices (heme ligation alone leaves still too much freedom). Therefore, as a further structural constraint, we tested the lipid-facing propensities of the TM helices and their conservation in the family. To do this, first the consensus regions of the 4 TM helices were defined in the multiple alignment (Fig. 1). Considering the position of the 4 by 26 individually predicted TM helices, we defined the consensus positions of TMH2, TMH3, TMH4, and TMH5 as 95–118(i), 133(i)–156, 175–200(i), and 218(i)–240, respectively ("i" indicates the cytoplasmic side of the helices). These regions were selected by considering the mathematical averages, but highly conserved residues known to prefer specific membrane locations (Killian and von Heijne 2000) were also taken into account. The TM helices TMH2 to -5 are indicated with horizontal bars under the sequences in Fig. 1. The 26 sequence segments of the 4 TM helix regions of the same length were subjected together for an analysis of their lipid-facing propensity in 4 separate submissions. The analysis was performed using the knowledge-based kPROT method, which scores residues with free-energy-like values (Pilpel et al. 1999). These kPROT values are related to the likelihood of a given residue to be oriented towards the lipids in TM alpha-helices. The algorithm treats TM helices as ideal alpha-helices. The vectorial sum of the kPROT values over the helical wheel of a single TM helix defines the side of the TM

helix that is most likely oriented towards the lipids. This analysis was performed for all of the consensus TM helix regions, i.e., 4 by 26 sequence segments. The radial thin solid lines on Fig. 2 show these vectorial sums for the respective TM helices of all species. Also shown are the mean vectors, which were obtained as averages over the 26 vectors, defining the (conserved) side of each TM helix most likely facing the lipids (thick lines with arrows). The scattering around the mean vector varies for each helix, is generally small, and is remarkably better for TMH2 and TMH4. The larger scattering in TMH3 and TMH5 in part origi-

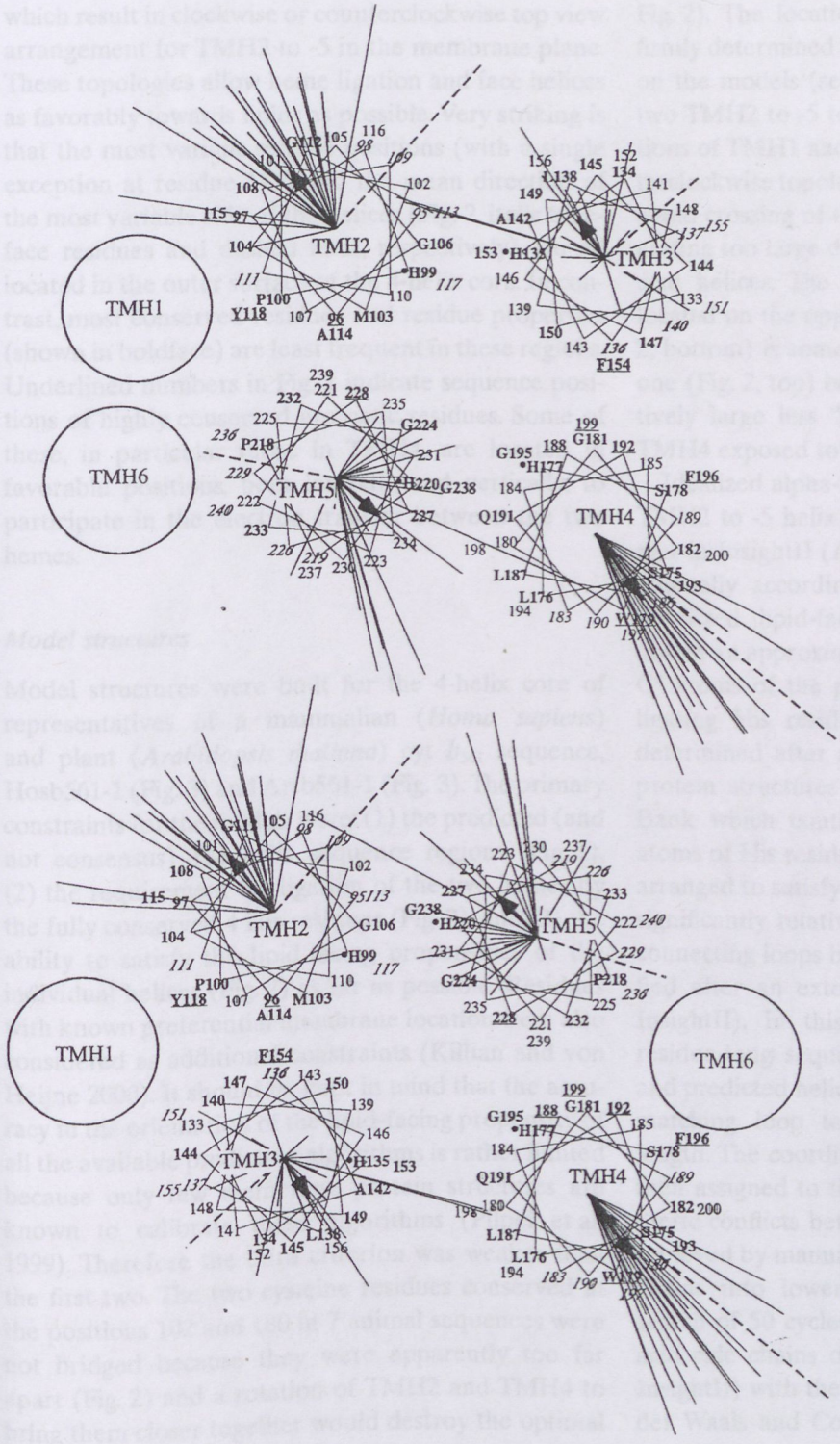

Fig. 2. Lipid-facing propensity of the individual and consensus TM helices. kPROT vectors (Pilpel et al. 1999) for the individual proteins are shown with radial thin lines (not all being visible) in the helical wheels of the consensus transmembrane helices TMH2 to -5 defined by the sequence regions 95–118(i), 133(i)–156, 175–200(i), and 218(i)–240, respectively, according to the consensus sequence numbering (gaps were removed for the prediction). The two most likely helix topologies from which structures were built are shown as viewed from the noncytoplasmic side of the membrane (top view). Accordingly, for the N-terminal cytoplasmic (TMH3, TMH5) and noncytoplasmic (TMH2, TMH4) helices, sequence numbering goes counterclockwise and clockwise, respectively. The thick solid line with an arrow represents the mean vector that is the average over all the 26 corresponding TM helices. The dashed lines indicate the direction in which variation of the lipid-facing propensity of the residues over the different sequences is largest. All lines are proportional to the numerical values they represent. Consensus residue numbers are indicated along the wheel together with residue names in single-letter code at locations where residues are highly conserved (cf. Fig. 1). The most variant residue positions are indicated in italic typeface. Identifiers in bold typeface indicate highly conserved residues and residue properties, with and without residue names, respectively. The putative heme-ligating His residues are indicated with bullets. Underlined numbers indicate sequence positions of highly conserved aromatic residues

nates from imposing consensus helices on the TM helix positions, which are less well conserved than in TMH2 and TMH4. The generally good conservation of the lipid-facing propensity in the cyt b_{561} family not only supports our multiple alignment, but it also argues that other helix geometries are very unlikely. The imposed constraints allow two alternative helix topologies, which result in clockwise or counterclockwise top view arrangement for TMH2 to -5 in the membrane plane. These topologies allow heme ligation and face helices as favorably towards lipids as possible. Very striking is that the most variant residue positions (with a single exception at residue 117) and the mean direction of the most variable side of the helices (Fig. 2, italic typeface residues and dashed lines, respectively) are all located in the outer surface of the 4-helix core. In contrast, most conserved residues and residue properties (shown in boldface) are least frequent in these regions. Underlined numbers in Fig. 2 indicate sequence positions of highly conserved aromatic residues. Some of these, in particular those in TMH4, are located in favorable positions, both laterally and vertically, to participate in the electron transfer between the two hemes.

Model structures

Model structures were built for the 4-helix core of representatives of a mammalian (*Homo sapiens*) and plant (*Arabidopsis thaliana*) cyt b_{561} sequence, Hosb561-1 (Fig. 3) and Artb561-1 (Fig. 3). The primary constraints for the models were: (1) the predicted (and not consensus) TM helix sequence regions (Fig. 1), (2) the requirement for ligation of the two hemes by the fully conserved 4 His residues (Fig. 2), and (3) the ability to satisfy the lipid-facing propensities of the individual helices (Fig. 2) as far as possible. Residues with known preferential membrane location were also considered as additional constraints (Killian and von Heijne 2000). It should be kept in mind that the accuracy in the orientation of the lipid-facing propensity of all the available prediction algorithms is rather limited because only few membrane protein structures are known to calibrate these algorithms (Pilpel et al. 1999). Therefore the third criterion was weaker than the first two. The two cysteine residues conserved at the positions 102 and 180 in 7 animal sequences were not bridged because they were apparently too far apart (Fig. 2) and a rotation of TMH2 and TMH4 to bring them closer together would destroy the optimal helix orientations. In addition, Kent and Fleming (1990) found that all cysteines were in the free sulfhydryl form in chromaffin granule cyt b_{561}. Both topologies still leave some "freedom" in orienting the helix axes relative to the membrane normal and along their long axis (Fig. 2). These uncertainties can be estimated to be below ca. 25° and 30°, respectively (see Fig. 2). The locations of TMH1 and TMH6 are not firmly determined by the primary constraints imposed on the models (see above) and are different for the two TMH2 to -5 topologies (Fig. 2). The unique locations of TMH1 and TMH6 in the clockwise and counterclockwise topologies result from the requirement to avoid crossing of the helix-connecting loops and generating too large distances between sequentially adjacent helices. The topology with TMH1 and TMH6 located on the opposite sides of the core helices (Fig. 2, bottom) is somewhat more likely than the circular one (Fig. 2, top) because the latter one leaves a relatively large less "lipophilic" surface on TMH3 and TMH4 exposed to lipids.

Idealized alpha-helices were built for the predicted TMH2 to -5 helix sequence segments using Biopolymer in InsightII (Accelrys 2000). They were arranged manually according to the two topologies and the predicted lipid-facing propensities (Fig. 2) and to obtain an approximate distance of 1.2 nm between the C^{α} atoms of the pairs of the fully conserved heme-ligating His residues. This His-to-His distance was determined after an inspection of coordinates of 34 protein structures in the Brookhaven Protein Data Bank which contained b-hemes ligated by the N^{ε} atoms of His residue pairs. The helices could be easily arranged to satisfy all the criteria without tilting them significantly relative to the membrane normal. Interconnecting loops between the TM helices were identified after an extensive search using Homology (in InsightII). In this process preflex and postflex 5-residue-long sequence matches between the known and predicted helices were scored to identify the best-matching loop template structures of the correct length. The coordinates of these loop templates were then assigned to the interconnecting loop sequences. Steric conflicts between amino acid side chains were removed by manual adjustment. In order to relax side chains into lower energy states, a conformational search of 50 cycles was performed for all the amino acid side chains of the models using Homology (in InsightII) with the default 0.8 nm cutoff for both Van der Waals and Coulomb interactions. The structures

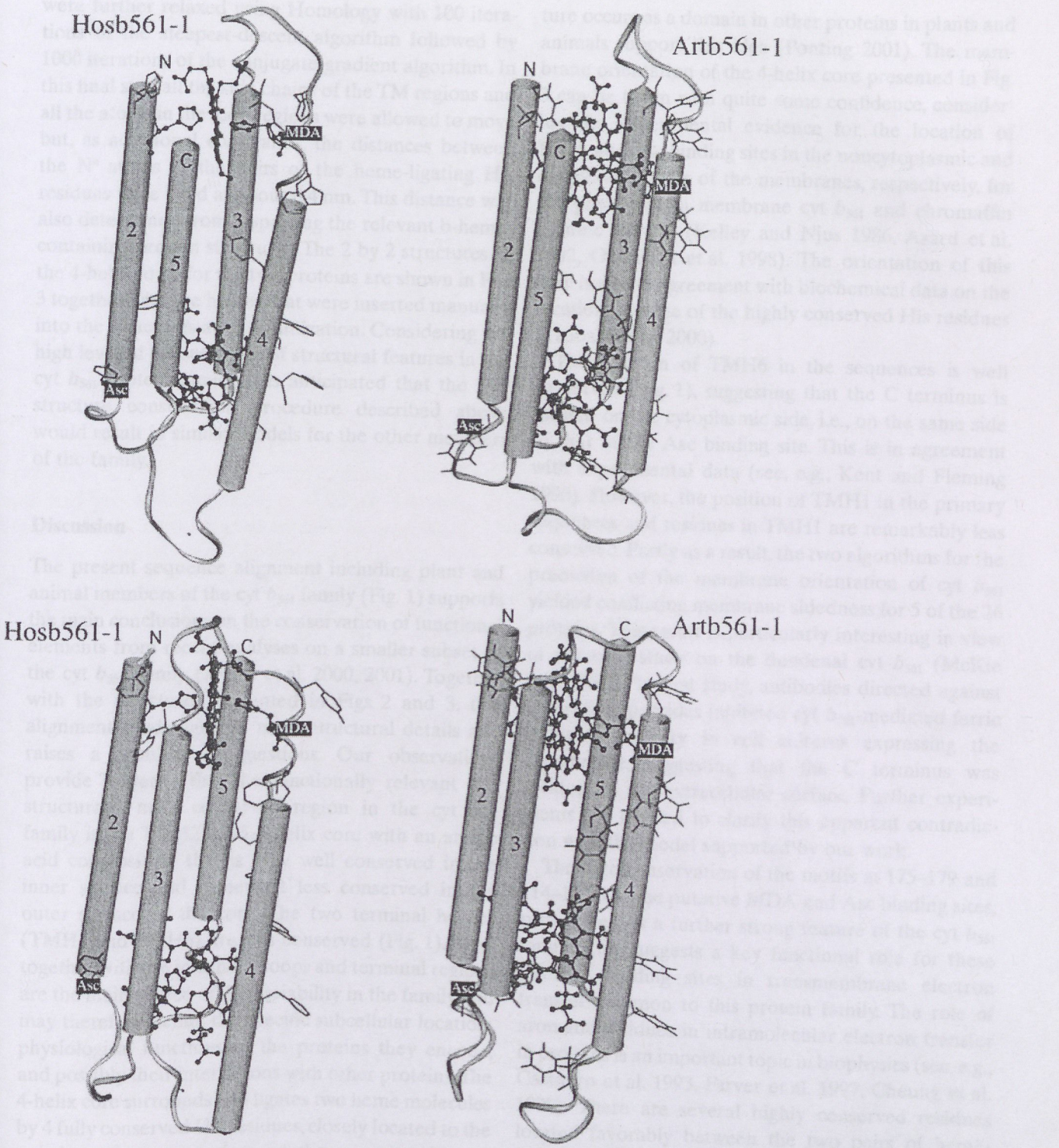

Fig. 3. Predicted 3-D structures for the *Homo sapiens* and *Arabidopsis thaliana* cyt b_{561} sequences, Hosb561-1 and Artb561-1, respectively. Two structures are shown for both sequences, namely, with clockwise and counterclockwise helix topology (top and bottom, respectively, as defined in Fig. 2). Only the most conserved helices (TMH2 to -5) are shown, together with interconnecting loops. The ribbon cartoon was created with InsightII (Accelrys 2000) and represents secondary structural forms (i.e., ideal alpha-helix and random coil). Highly conserved aromatic amino acid residues are indicated in light gray thin bar presentation. The two pairs of fully conserved His residues (His139 and His157) and the hemes ligated by them are shown in dark gray ball-and-stick presentation. The N- and C-termini of helices 2 and 5 are indicated with N and C, respectively. The TM helix numbers are indicated on the cylinders. The cytoplasmic side is below each structure. The putative MDA and Asc binding sites are indicated with the corresponding text boxes. The helix axes are closely perpendicular to the membrane plane

were further relaxed using Homology with 100 iterations of the steepest-descent algorithm followed by 1000 iterations of the conjugate-gradient algorithm. In this final step all the side chains of the TM regions and all the atoms in the loop regions were allowed to move but, as additional constraints, the distances between the N^{ε} atoms of the pairs of the heme-ligating His residues were fixed at about 0.4 nm. This distance was also determined from inspecting the relevant b-heme-containing protein structures. The 2 by 2 structures of the 4-helix cores for the two proteins are shown in Fig. 3 together with the hemes that were inserted manually into the structures after optimization. Considering the high level of conservation of structural features in the cyt b_{561} protein family, it is anticipated that the 3-D structure construction procedure described above would result in similar models for the other members of the family.

Discussion

The present sequence alignment including plant and animal members of the cyt b_{561} family (Fig. 1) supports the main conclusions on the conservation of functional elements from recent analyses on a smaller subset of the cyt b_{561} family (Asard et al. 2000, 2001). Together with the structures presented in Figs. 2 and 3, 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 cyt 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 (Fig. 1). 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 4-helix core surrounds and ligates two heme molecules by 4 fully conserved His residues, closely located to the membrane–water interface on the opposite sides of the membrane. Since the putative, well or highly conserved Asc and MDA binding sites, and other highly conserved residues with yet unknown function, are located in this region, this 4-helix core may represent a conserved transmembrane electron transfer machinery. Recent findings demonstrating that this core structure occurs as a domain in other proteins in plants and animals support this idea (Ponting 2001). The membrane orientation of the 4-helix core presented in Fig. 3 can be taken with quite some confidence, considering the experimental evidence for the location of MDA and Asc binding sites in the noncytoplasmic and cytoplasmic sides of the membranes, respectively, for the plant plasma membrane cyt b_{561} and chromaffin granule cyt b_{561} (Kelley and Njus 1986, Asard et al. 1992, Okuyama et al. 1998). The orientation of this core is also in agreement with biochemical data on the location of some of the highly conserved His residues (Tsubaki et al. 2000).

The position of TMH6 in the sequences is well conserved (Fig. 1), suggesting that the C terminus is located on the cytoplasmic side, i.e., on the same side as that of the Asc binding site. This is in agreement with experimental data (see, e.g., Kent and Fleming 1990). However, the position of TMH1 in the primary sequences and residues in TMH1 are remarkably less conserved. Partly as a result, the two algorithms for the prediction of the membrane orientation of cyt b_{561} yielded conflicting membrane sidedness for 5 of the 26 proteins. This result is particularly interesting in view of a recent study on the duodenal cyt b_{561} (McKie et al. 2001). In that study, antibodies directed against C-terminal peptides inhibited cyt b_{561}-mediated ferric reductase activity in cell cultures expressing the cytochrome, suggesting that the C terminus was located on the extracellular surface. Further experiments are needed to clarify this apparent contradiction with the model supported by our work.

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 cyt b_{561} family. This suggests a key functional role for these putative binding sites in transmembrane electron transfer common to this protein family. The role of aromatic residues in intramolecular electron transfer in proteins is an important topic in biophysics (see, e.g., Casimiro et al. 1993, Farver et al. 1997, Cheung et al. 1999). There are several highly conserved residues located favorably between the two pairs of heme-ligating His residues (Figs. 2 and 3). 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. It will be an important subject for future studies to test whether some or all

of these conserved aromatic residues are essential for transmembrane electron transfer in cyt b_{561} proteins. It is interesting that the location of the fully conserved His residues in TMH3 and TMH5 are rather close to the lipid-facing sides in these "mean" topologies (see bullets and solid vectors in Fig. 2). This raises the question whether structural details of heme ligation in the cytoplasmic side are possibly less conserved in the family, and/or heme ligation may be even less stable, than in the noncytoplasmic side. It should be noted that His139 and His157 are also well but not fully conserved. It has been argued that His139 does not participate in heme ligation because it is not present in some of the animal proteins (Asard et al. 2001, Asada et al. 2002). Chemical modification of His residues also supports this interpretation (Tsubaki et al. 2000). His residues are believed to play an important direct role in proton movement coupled to the electron transfer reaction (Njus et al. 2001, Kipp et al. 2001). Whether His139 and His157 manifest some roles in heme ligation and/or in the electron transfer mechanism in different members of this protein family remains to be explored (note that some of the residue replacements appear simultaneously at these positions in some sequences; see Fig. 1). The high structural similarity between the plant and animal cyt 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. Of the di-heme proteins with known structure, the membranous subunit c of fumarate reductase from *Wolinella succinogenes*, available at 0.22 nm resolution (Lancester et al. 1999; PDB entry 1QLA), is most relevant to the present study, although its sequence is very different from that of the cyt b_{561} family (hence not shown in the alignment). Similarly to our models, two pairs of His residues on 4 TM helices coordinate the two hemes and the heme planes have similar orientations. The orientation of the 4-helix bundle relative to the direction of the electron flow is also the same. However, the TM helices are much longer than those in the cyt b_{561} family and, consequently, are significantly tilted and kinked. Though there is quite some freedom in the helix orientations, a similar degree of TM tilts and kinks in our models would result in a large hydrophobic mismatch with the membrane. The shorter heme-to-heme distance of about 1.5 nm in subunit c of fumarate reductase is in part due to the tilts and kinks in the TM helices in addition to the fact that the heme-ligating residues are located much closer to the center of the TM helices than in our models. 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 latter one is particularly interesting because of the large distance between the heme centers, 2.2–2.9 nm in the present models, supported by an early estimate of 2.1–3.2 nm for a cyt b_{561} (Esposti et al. 1989). The present models will be further refined as new structural data emerge in the future.

Acknowledgments

This work was supported by grants from the Flemish (BIL 99/48) and Hungarian (OMFB B-4/99) governments, the Volkswagen-Stiftung, and the Hungarian National Science Fund (OTKA T-029458 and T-034488). Work by H.A. was a contribution of the University of Nebraska, Agricultural Research Division, Lincoln, Nebraska. Journal Series nr. 13700.

References

Accelrys (2000) InsightII Modeling Environment, Release 2000 (June 2000) and Biopolymer and Homology (March 2000). Accelrys Inc., San Diego, Calif

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402

Asada A, Kusakawa T, Orii H, Agata K, Watanabe K, Tsubaki M (2002) Planarian cytochrome b561: conservation of a six transmembrane structure and localization along the central and peripheral nervous system. J Biochem 131: 175–182

Asard H, Horemans N, Caubergs RJ (1992) Transmembrane electron transport in ascorbate-loaded plasma membrane vesicles from higher plants involves a b-type cytochrome. FEBS Lett 306: 143–146

- Terol-Alcayde J, Preger V, Del Favero J, Verelst W, Sparla F, Perez-Alonso M, Trost P (2000) *Arabidopsis thaliana* sequence analysis confirms the presence of cyt b561 in plants: evidence for a novel protein family. Plant Physiol Biochem 38: 905–912
- Kapila J, Verelst W, Bérczi A (2001) Higher-plant plasma membrane cytochrome b_{561}: a protein in search of a function. Protoplasma 217: 77–93

Beitz, E (2000) TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. Bioinformatics 16: 135–139

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL (2002) GenBank. Nucleic Acids Res 30: 17–20

Bérczi A, Lüthje S, Asard H (2001) *b*-Type cytochromes in plasma membranes of *Phaseolus vulgaris* hypocotyls, *Arabidopsis thaliana* leaves, and *Zea mays* roots. Protoplasma 217: 50–55

- Caubergs RJ, Asard H (2003) Partial purification and characterization of an ascorbate-reducible *b*-type cytochrome from the plasma membrane of *Arabidopsis thaliana* leaves. Protoplasma 221: 47–56

Casimiro DR, Richards JH, Winkler JR, Gray HB (1993) Electron transfer in ruthenium-modified cytochromes c: sigma-tunneling

pathways through aromatic residues. J Phys Chem 97: 13073–13077

Cheung MS, Daizadeh I, Stuchebrukov AA, Heelis PF (1999) Pathways of electron transfer in *Escherichia coli* DNA photolase: Trp306 to FADH. Biophys J 76: 1241–1249

Esposti MD, Kamensky YA, Arutjunjan AM, Konstantinov AA (1989) A model for the molecular organization of cytochrome β-561 in chromaffin granule membranes. FEBS Lett 254: 74–78

Farver O, Skov LK, Young S, Bonander N, Karlsson BG, Vanngard T, Pecht I (1997) Aromatic residues may enhance intramolecular electron transfer in azurin. J Am Chem Soc 119: 5453–5454

Hägerhäll C, Hederstedt L (1996) A structural model for the membrane-integral domain succinate:quinone oxidoreductases. FEBS Lett 389: 25–31

Harnadek GJ, Ries EA, Tse DG, Fitz JS, Njus D (1992) Electron transfer in chromaffin-vesicle ghosts containing peroxidase. Biochim Biophys Acta 1135: 280–286

Horemans N, Asard H, Caubergs R (1994) The role of ascorbate free radical as an electron acceptor to cytochrome b-mediated trans-plasma membrane electron transport in higher plants. Plant Physiol 104: 1455–1458

Jones DT, Taylor WR, Thornton JM (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry 33: 3038–3049

Karlin S, Ghandour G (1985) Multiple-alphabet amino acid sequence comparisons of the immunoglobulin kappa-chain constant domain. Proc Natl Acad Sci USA 82: 8597–8601

Kelley PM, Njus D (1986) Cytochrome b561 spectral changes associated with electron transfer in chromaffin-vesicle ghosts. J Biol Chem 261: 6429–6432

Kent UM, Fleming PJ (1987) Purified cytochrome b561 catalyzes transmembrane electron transfer for dopamine beta-hydroxylase and peptidyl glycine alpha-amidating monooxygenase activities in reconstituted systems. J Biol Chem 262: 8174–8178

– – (1990) Cytochrome b561 is fatty acylated and oriented in the chromaffin granule membrane with its carboxyl terminus cytoplasmically exposed. J Biol Chem 265: 16422–16427

Killian JA, von Heijne G (2000) How proteins adapt to a membrane-water interface. Trends Biochem Sci 25: 429–434

Kipp BH, Kelley PM, Njus D (2001) Evidence for an essential histidine residue in the ascorbate-binding site of cytochrome b561. Biochemistry 40: 3931–3937

Kobayashi K, Tsubaki M, Tagawa S (1998) Distinct roles of two heme centers for transmembrane electron transfer in cytochrome b561 from bovine adrenal chromaffin vesicles as revealed by pulse radiolysis. J Biol Chem 273: 16038–16042

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305: 567–580

Lancester CRD, Kröger A, Auer M, Michel H (1999) Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. Nature 402: 377–385

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404–405

McKie AT, Barrow D, Latunde-Dada GO, Rolfs A, Sager G, Mudaly E, Mudaly M, Richardson C, Barlow D, Bomford A, Peters TJ, Raja KB, Shirali S, Hediger MA, Farzaneh F, Simpson RJ (2001) An iron-regulated ferric reductase associated with the absorption of dietary iron. Science 291: 1755–1759

Möller S, Croning MDR, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17: 646–653

Njus D, Knoth J, Cook C, Kelley PM (1983) Electron transfer across the chromaffin granule membrane. J Biol Chem 258: 27–30

– Wigle M, Kelley PM, Kipp BH, Schlegel HB (2001) Mechanism of ascorbic acid oxidation by cytochrome b561. Biochemistry 40: 11905–11911

Okuyama E, Yamamoto R, Ichikawa Y, Tsubaki M (1998) Structural basis for the electron transfer across the chromaffin vesicle catalyzed by cytochrome b561: analyses of DNA nucleotide sequences and visible absorption spectra. Biochim Biophys Acta 1383: 269–278

Pilpel Y, Ben-Tal N, Lancet D (1999) kPROT: a knowledge-based scale for the propensity of residue orientation in transmembrane segments. Application to membrane protein structure prediction. J Mol Biol 294: 921–935

Ponting CP (2001) Domain homologues of dopamine beta-hydroxylase and ferric reductase: roles for iron metabolism in neurodegenerative disorders? Human Mol Genet 10: 1853–1858

Srivastava M (1986) *Xenopus* cytochrome b561: molecular confirmation of a general five transmembrane structure and developmental regulation at the gastrula stage. DNA Cell Biol 15: 1075–1080

– Duong LT, Fleming PJ (1984) Cytochrome b561 catalyzes transmembrane electron transfer. J Biol Chem 259: 8072–8075

Takeuchi F, Kobayashi K, Tagawa S, Tsubaki M (2001) Ascorbate inhibits the carbethoxylation of two histidyl and one tyrosyl residues indispensable for the transmembrane electron transfer reaction of cytochrome b561. Biochemistry 40: 4067–4076

Trost P, Bérczi A, Sparla F, Sponza G, Marzadori B, Asard H, Pupillo P (2000) Purification of cytochrome b-561 from bean hypocotyls plasma membrane: evidence for the presence of two heme centers. Biochim Biophys Acta 1468: 1–5

Tsubaki M, Nakayama M, Okuyama E, Ichikawa Y, Hori H (1997) Existence of two heme b centers in cytochrome b561 from bovine adrenal chromaffin vesicles as revealed by a new purification procedure and EPR spectroscopy. J Biol Chem 272: 23206–23210

– Kobayashi K, Ichise T, Takeuchi F, Tagawa S (2000) Diethylpyrocarbonate abolishes fast electron accepting ability of cytochrome b561 from ascorbate but does not influence electron donation to monodehydroascorbate radical: identification of the modification sites by mass spectrometric analysis. Biochemistry 39: 3276–3284

Yuan J, Amend A, Borkowski J, DelMarco R, Bailey W, Liu Y, Xie G, Blevins R (1999) MULTICLUSTAL: a systematic method for surveying Clustal W alignment parameters. Bioinformatics 15: 862–863