A new method to model membrane protein structure based on silent amino acid substitutions

3 AUTHORS, INCLUDING:

Jaume Torres

Nanyang Technological University

131 PUBLICATIONS **2,344** CITATIONS

SEE PROFILE

A New Method to Model Membrane Protein Structure Based on Silent Amino Acid Substitutions

John A.G. Briggs, Jaume Torres, and Isaiah T. Arkin[*]
Cambridge Centre for Molecular Recognition. Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

ABSTRACT The importance of accurately modeling membrane proteins cannot be overstated, in lieu of the difficulties in solving their structures experimentally. Often, however, modeling procedures (e.g., global searching molecular dynamics) generate several possible candidates rather then pointing to a single model. Herein we present a new approach to select among candidate models based on the general hypothesis that silent amino acid substitutions, present in variants identified from evolutionary conservation data or mutagenesis analysis, do not affect the stability of a native structure but may destabilize the non-native structures also found. The proof of this hypothesis has been tested on the α-helical transmembrane domains of two homodimers, human glycophorin A and human CD3-ζ, a component of the T-cell receptor. For both proteins, only one structure was identified using all the variants. For glycophorin A, this structure is virtually identical to the structure determined experimentally by NMR. We present a model for the transmembrane domain of CD3-ζ that is consistent with predictions based on mutagenesis, homology modeling, and the presence of a disulfide bond. Our experiments suggest that this method allows the prediction of transmembrane domain structure based only on widely available evolutionary conservation data. Proteins 2001;44:370–375.
© 2001 Wiley-Liss, Inc.

Key words: membrane proteins; CD3-ζ; glycophorin; molecular dynamics; proteomics; molecular modeling

INTRODUCTION

One of the new strategies in structural biology to obtain valid structural models for proteins is the use of proteins that share a high degree of sequence identity, because these proteins also share a high structural similarity.[1] Thus, the use of a structure determined experimentally allows the structure prediction of homologous proteins.[2] Even when no experimentally determined structures exist for a family of homologous proteins, homology between sequences can be useful in the prediction of structural features such as secondary structure or solvent accessibility[3–7] and also the topology of membrane proteins.[8] Information derived from homology has also been used to improve threading algorithms.[9] By simultaneously optimizing a set of homologous proteins and penalizing structural differences between the simulations, Keasar et al.[10,11] have enhanced *ab initio* algorithms that search for global energy minima in conformational space.

It would be particularly useful to apply information derived from evolutionary conservation data to transmembrane domains in which high-resolution methods for structure determination have proved difficult to implement. The prediction of transmembrane domain structure is facilitated by their tendency to adopt an α-helical fold, limiting the possible conformations. In this respect, Brunger and coworkers[12,13] have developed a procedure to predict transmembrane domain structure based on the results of mutagenesis studies and global searching molecular dynamics simulations. In this method, multiple symmetric bundles of helices are constructed, each differing from the other by the rotation of the helices about their axes. These are then used as starting positions for molecular dynamics and energy minimization protocols. The output structures from these simulations are compared, yielding clusters where multiple simulations converged to the same structure, indicating the presence of a possible oligomeric form. The preferred cluster can then be selected on the basis of existing experimental data. Past studies have used mutagenesis, as in dimerizing glycophorin A[14,15] or pentamerizing phospholamban[16] due to the fact that both proteins, abnormally, remain in the oligomeric form in SDS-PAGE. Also, site-directed infrared dichroism of M2[17] and phospholamban[18] has been used.
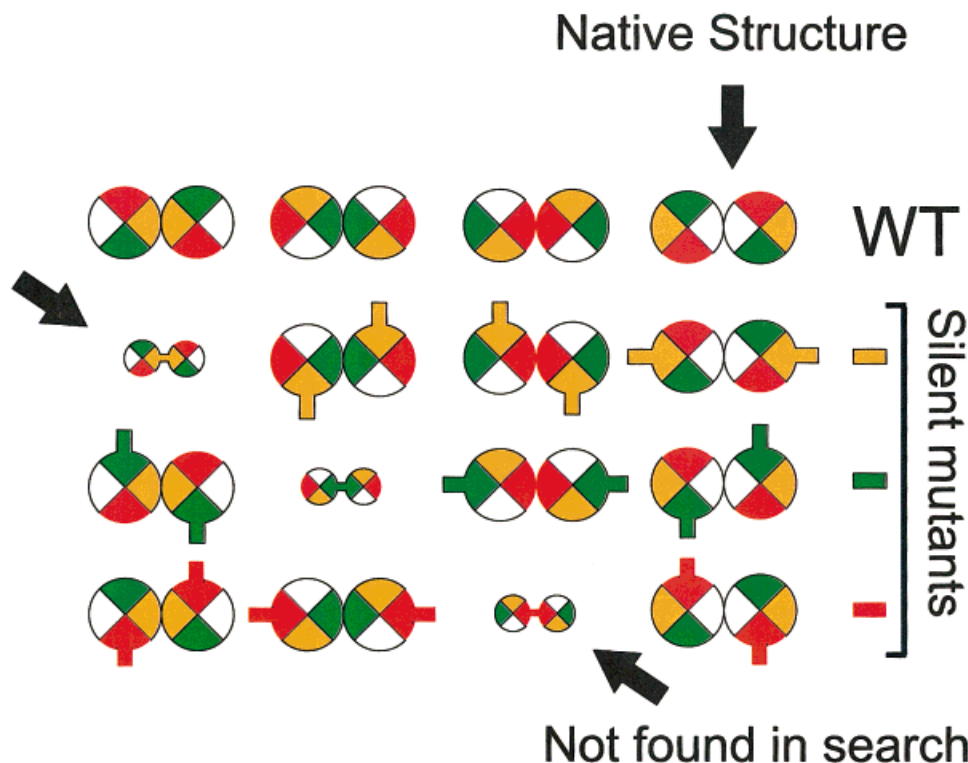
Fig. 1. Schematic representation of structures obtained from global search protocols on helical dimers. Each helix is dissected into four quadrants to illustrate the different sides of the helix. The top row represents four possible structures identified from a global search using the wild-type sequence. The lower three rows represent the structures identified in global searches using three sequences carrying silent amino acid substitutions, where the mutation is represented by a small protrusion from the circle. Structures that are destabilized by the mutation and therefore are not found by the search protocol are shrunk. Note that only one structure (last column) is identified in all the simulations and is taken as the native structure. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

When using mutagenesis, the method is limited by the fact that simple oligomerization assays of the type used on phospholamban[16] and glycophorin A[15] are possible only in a few cases. Also, although loss of oligomerization is expected to occur when the mutated residue is located on the helix-helix interface, this need not be the case. For example, a residue may be located elsewhere, but the mutation may alter the structure of the helix monomer.

We report here an improvement to this approach that has the great advantage of allowing the use of homologous sequences, available in the genome databases, in the place of mutagenesis data. We focus on silent amino acid substitutions, working on the premise that they will not affect the native structure, but may destabilize non-native possible structures generated during a global searching molecular dynamics protocol. Silent amino acid substitutions are any mutation that does not affect folding to the native state and can be identified experimentally using functional assays as well as oligomerization assays. Alternatively, evolutionary homologues can be considered as sequences carrying silent amino acid substitutions. Independent global searches are therefore conducted by using variants that either carry silent amino acid substitutions or correspond to different evolutionary homologues. If enough variants are used, and provided the native structure corresponds to an energy minimum in the molecular dynamics simulation, only one model, corresponding to the native structure, will have persisted in all searches (Fig. 1). This hypothesis has been tested on two homodimers. Glycophorin A, a protein of known structure[19] on which extensive mutagenesis analyses have been conducted[15] and CD3-ζ, a protein of unknown structure that is incorporated as the final and rate-limiting step in the assembly of the T-cell receptor (TCR) complex.[20,21] It is known that the two monomers of CD3-ζ associate via their transmembrane domains[22] that contain a putative glycine-based dimerization motif similar to that in the transmembrane helices of glycophorin A.[23] For both GPA and CD3-ζ, sequences containing silent amino acids substitutions identified in mutagenesis experiments, and homologous sequences identified from a genomic database, are used independently to arrive at the same model.

MATERIALS AND METHODS
Global Search Molecular Dynamics

All calculations were performed by using PCNS, the parallel-processing version of the Crystallography and NMR System (CNS Version 0.3),[24] with the OPLS parameter set and united atom topology,[25] explicitly describing only polar and aromatic hydrogens. A global search was conducted as described elsewhere, by using CHI (CNS searching of helical interactions), assuming a symmetrical interaction between the two helices.[13,26] An α-helical conformation was maintained during the search by apply-

ing 3.2 Å maximal distance restraints between O_i and N_{i+4}. Calculations were performed *in vacuo*.

For the simulations, the human glycophorin A transmembrane domain consisted of residues 73–91: ITLIIF-GVMAGVIGTILLI. The human CD3-ζ transmembrane domain consisted of residues 31–51: LCYLLDGIL-FIYGVILTALFL. The mutation C32G was present in all simulated CD3-ζ peptides to eliminate the disulfide bond, facilitating the global search. This silent mutation does not dramatically reduce cell surface expression of the TCR and is therefore unlikely to affect dimer structure.[23] The aspartate residue was modeled in its protonated form, as would be expected on the basis of energy considerations.[27]

Close evolutionary homologues of human glycophorin A and CD3-ζ were obtained by searching the OWL database.[28] For the purposes of testing our method, we ensured that all sequences modeled represented species with the same structure and function. All nonidentical sequences in the database that were annotated as glycophorin were used except those that contained the cysteine characteristic of glycophorin B, as were all nonidentical sequences annotated as CD3-ζ. The glycophorin A mutants selected were those that exhibited significant levels of dimerization in both the TOXCAT assays[29] and SDS-PAGE assays,[15] and that are therefore likely to adopt the same native structure as wild-type glycophorin A. The F40A, F40L, F40V, G43A, and G43S mutants of CD3-ζ were selected on the basis of their ability to form a functional TCR complex at the cell surface.[23]

Trials were conducted starting from both left and right 50° crossing angles, with the helices rotated 360° about their helical axes in 10° increments. Sampling was further increased by performing four trials from each starting configuration using different initial random velocities, making a total of 288 trials. Clusters of output structures were identified, containing 10 or more structures within 1.0 Å RMSD from any other structure within the cluster. Consequently, some clusters overlap, and output structures may be members of more than one cluster. The output structures in a cluster were averaged and subjected to a further simulated annealing protocol, as in the initial search. This average was taken as representative of the cluster.

Analysis of the Simulations

The results from the simulations were analyzed taking into account that the helix crossing angle, Ω, is the torsion angle between the helix axes[30] (Fig. 2). The rotational orientation of the helix at a particular residue is given by ϕ (see above), the angle between the vector perpendicular to the helix axis that joins it to the middle of the C═O bond, and another vector perpendicular to, and from the same position on the helix axis, pointing in the direction of the helix tilt. This angle is therefore 0 in the direction of the tilt. The helical axis is a vector from the mean of the coordinates of the five α carbons to the N-terminal side of the defined residue to the mean of the coordinates of the five α carbons to the C-terminal side of the defined residue.
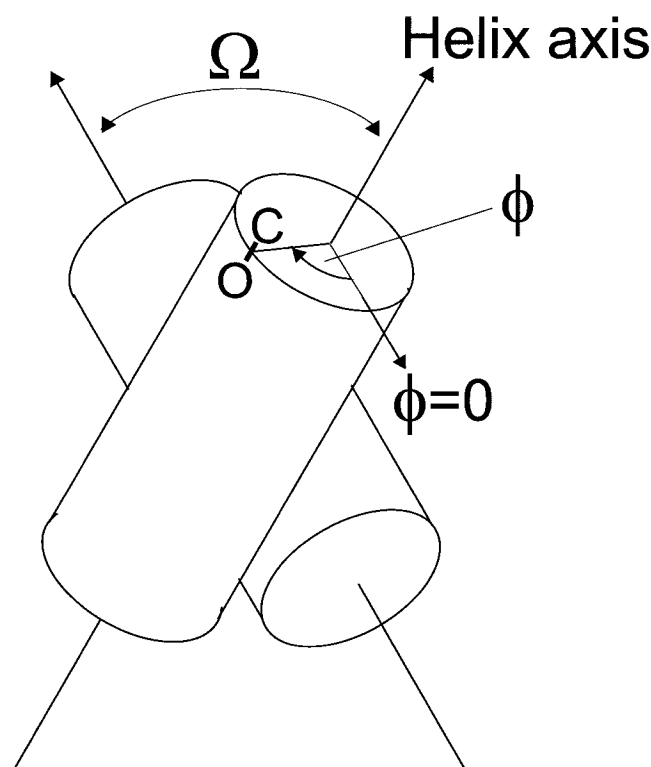

Fig. 2. The geometric parameters that were used to define the orientation of a transmembrane helix. The crossing angle Ω and the rotational orientation ϕ are illustrated.

To analyze the data, the structures identified were plotted against the ϕ angle of an arbitrarily specified residue, which was G83 in glycophorin A and I41 in CD3-ζ. Comparisons between the average structures of clusters from different variants were made by calculating the RMSD between their α carbon backbones.

RESULTS AND DISCUSSION
Glycophorin A

Glycophorin A contains the best characterized instance of noncovalent homo-oligomerizing transmembrane α-helices. The structure of the dimerizing transmembrane α-helices of glycophorin A has been determined experimentally in detergent micelles by solution NMR.[19] Furthermore, both mutagenesis[15] and evolutionary conservation[28] data are available, making it an ideal test case for our new methodology.

The results of the global search molecular dynamics protocol for different homologues of glycophorin A are shown in Figure 3, upper panel. Each cluster is represented by a circle. It is not possible to reliably determine which cluster or group of clusters is representative of the native structure based on energy alone (the energies of the average structures in the human variant only range between −32 and −52 kcal/mol). Even if the different averaged structures exhibited substantially different energies, it would still be difficult to select the native structure, both because calculations of energy depend on the accuracy and applicability of the force fields used, and because
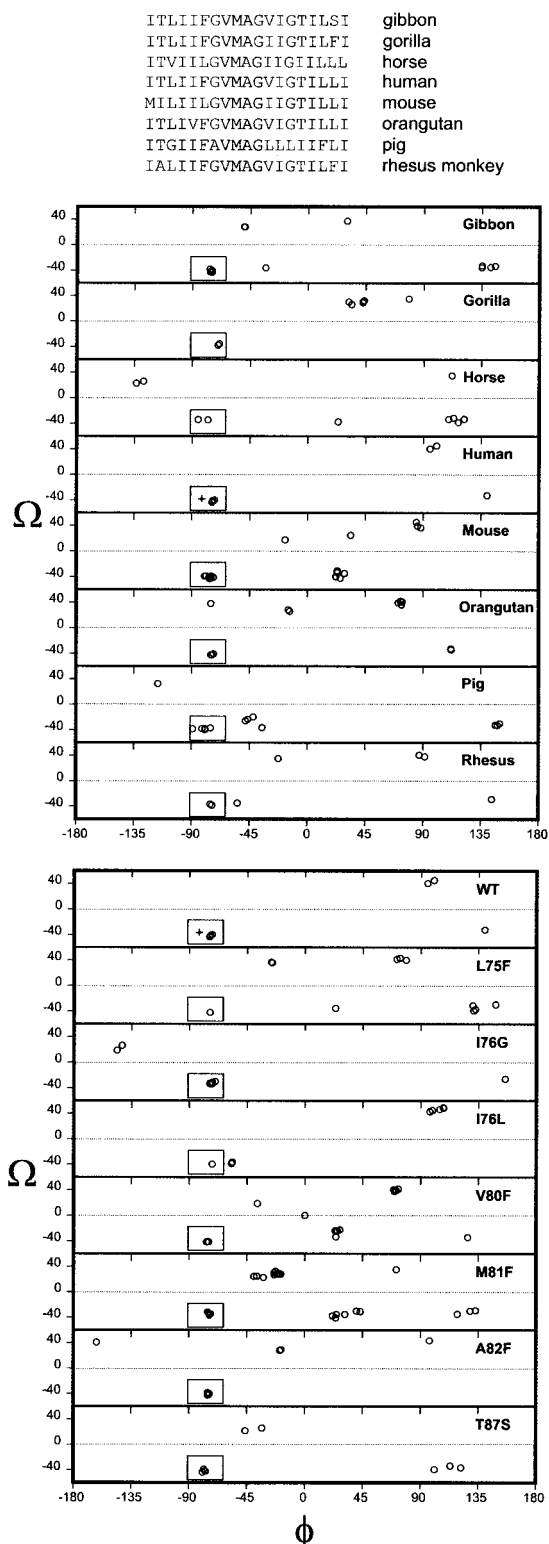
```
ITLIIFGVMAGVIGTILSI    gibbon
ITLIIFGVMAGIIGTILFI    gorilla
ITVIILGVMAGIIGIILLL    horse
ITLIIFGVMAGVIGTILLI    human
MILIILGVMAGIIGTILLI    mouse
ITLIVFGVMAGVIGTILLI    orangutan
ITGIIFAVMAGLLLIIFLI    pig
IALIIFGVMAGVIGTILFI    rhesus monkey
```


Fig. 3. Sequence alignment of similar glycophorin A homologues from the OWL database.[28] Results of the global search molecular dynamics protocol for close homologues of glycophorin A (upper panel) and wild type and seven mutants of human glycophorin A (lower panel). The clusters obtained are indicated in terms of their rotational and interhelix tilt angles. The rectangle indicates the position at which clusters are identified in all searches. The plus sign in the Human/WT panels indicates the position of the NMR structure of glycophorin A.

the simulations are undertaken *in vacuo* (due to CPU time limitations). We note that repeating the global search molecular dynamics protocol gives similar results. Comparing the results of all of these global searches reveals that at only one position, $\Omega = -40°$, $\phi = -80°$, are clusters found in searches using all of the homologues (Fig. 3, upper panel, see rectangles). The RMSD between any pair within this set of clusters (a "complete set") is <0.7 Å Cα RMSD. No other such set can be defined within 2 Å Cα RMSD.

The results of global searching molecular dynamics protocols for variants containing silent amino acid substitutions (Fig. 3, lower panel) also point to a single structure that persists in all instances. This structure is the same as that identified by using homologues. No pair of structures within the complete set differs by >0.7 Å Cα RMSD, and no other such set can be defined within 3 Å Cα RMSD. The members of the set that were obtained during the search using the human wild-type sequence were averaged and subjected to a simulated annealing protocol. The resulting structure differed from the published NMR structure of glycophorin A[19] by a Cα RMSD of <1.0 Å. Thus, both procedures, using either mutagenesis data or evolutionary conservation, point to the same structure that is identical to that obtained from NMR.

CD3-ζ

We then examined the transmembrane domain of the CD3-ζ homodimer, whose structure is not known. Searches were performed by using close homologues identified from the OWL database[28] (Fig. 4). The results (Fig. 4, top panel), as in glycophorin A, point to a single structure that persists in all of the simulations, forming a complete set at <0.9 Å Cα RMSD. No other set could be found at <1.54 Å Cα RMSD.

Searches were then conducted by using variants containing silent amino acid substitutions, identified from mutagenesis[23] (Fig. 4, lower panel). In this instance, a complete set can be defined at <0.8 Å Cα RMSD, and no other can be defined within 2 Å Cα RMSD. The complete sets obtained by using homologues and silent amino acid substitutions correspond to the same structure that we conclude is likely to represent the native structure. We note that the clusters selected were never the ones with lowest energy, reaffirming our previous statement regarding the difficulty of selecting the correct structure based on energy criteria alone.

A representative structure was produced by averaging the five clusters inside the rectangle in the simulation corresponding to the human variant (these structures were members of the complete sets produced by using both mutagenesis and evolutionary conservation) and subjecting them to a simulated annealing protocol. This structure is depicted in Figure 5 as slices through the bundle axis of the CD3-ζ dimer.

The model is consistent with previous observations. For example, the α carbons at residue 32 in the two monomers (where glycines were simulated in the place of cysteines to facilitate the global search) have orientation and separation (5.86 Å) consistent with the formation of a disulfide
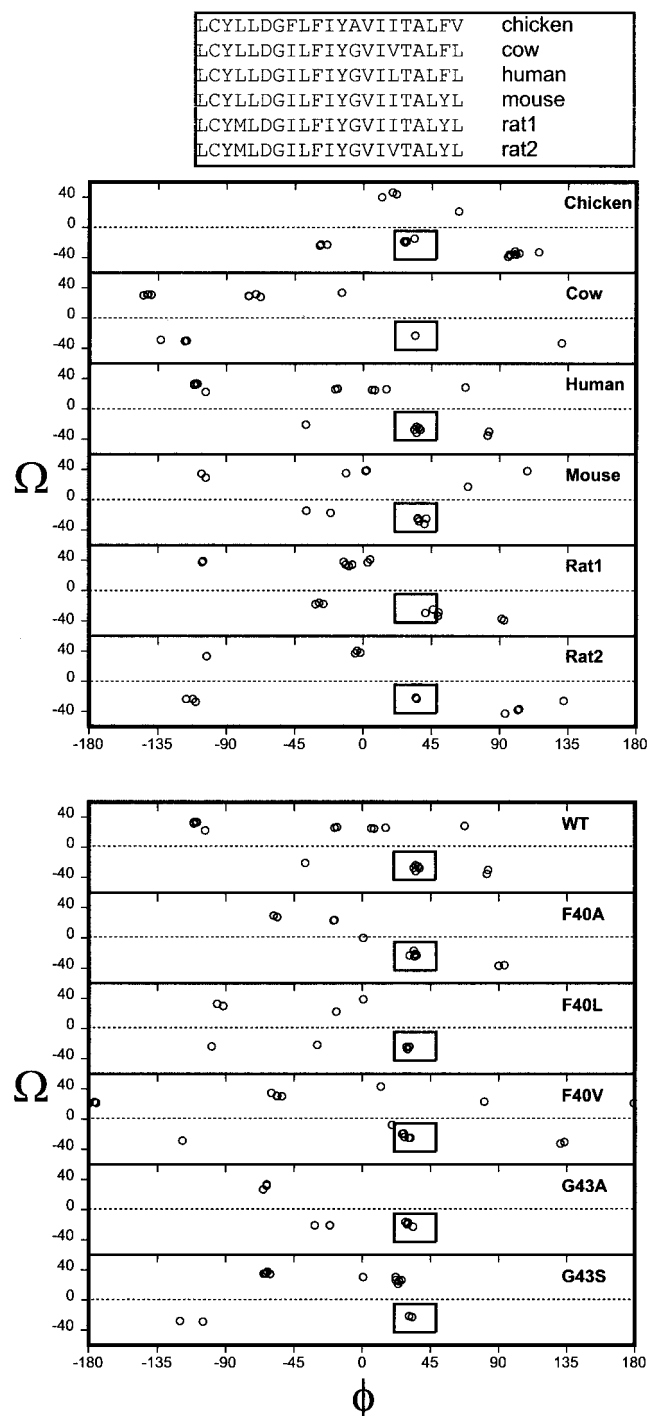
```
LCYLLDGFLFIYAVIITALFV    chicken
LCYLLDGILFIYGVIVTALFL    cow
LCYLLDGILFIYGVILTALFL    human
LCYLLDGILFIYGVIITALYL    mouse
LCYMLDGILFIYGVIITALYL    rat1
LCYMLDGILFIYGVIVTALYL    rat2
```


Fig. 4. Sequence alignment of CD3-ζ homologues from the OWL database.[28] Results of the global search molecular dynamics protocol for close homologues of CD3-ζ (upper panel) and the wild type and five mutants of CD3-ζ (lower panel). The clusters obtained are indicated in terms of their rotational and interhelix tilt angles. The rectangle indicates the position at which clusters are identified in all searches.
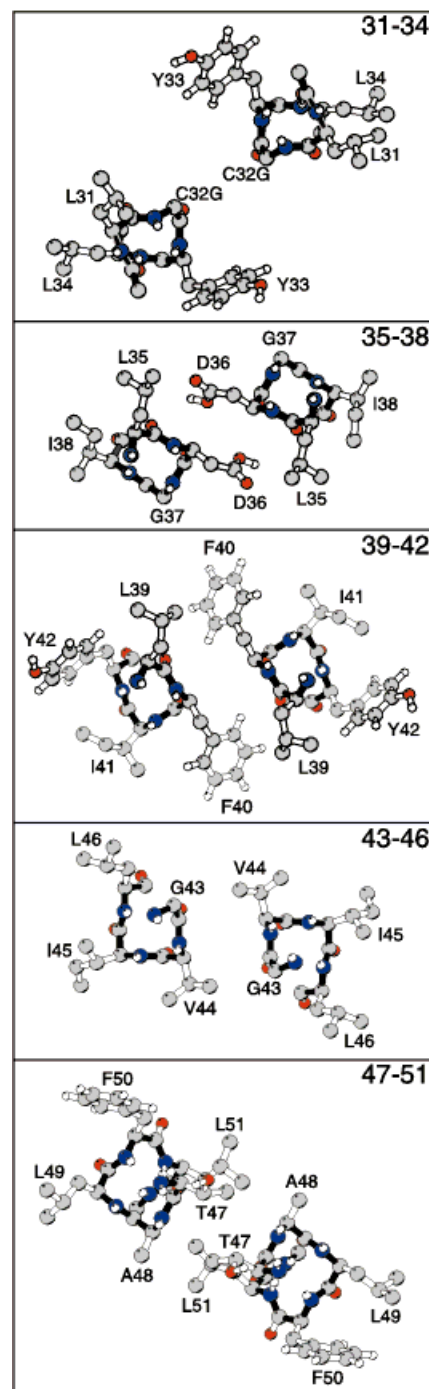
Fig. 5. Slices through the CD3-ζ homodimer structure in ball and stick representation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

bridge between these two residues. In addition, this structure is similar to that produced by modeling CD3-ζ onto the solved structure of glycophorin A,[23] with which it has primary sequence homology. Furthermore, the glycines at position 43 are located in the helix-helix interface (Fig. 5)

but rotated slightly away from the interface relative to the equivalent glycines in glycophorin A, consistent with the lower sensitivity of CD3-ζ to mutation to larger residues at this position.[22,23]

CONCLUSIONS

Using only the information from silent amino acid substitutions, we have obtained a structure for glycoph-

orin A that is virtually identical to that determined by solid state NMR. Also, a plausible model has been derived for the CD3-ζ homodimer. We point out, however, that the method is clearly limited by the ability of the global searching molecular dynamics simulation to obtain the native structure, or a model that represents it closely, as one of the clusters. Because of the limitations of simulations, such as inaccuracies in the force field or, in this case, the absence of an explicit bilayer, the native structure may not always be identified. In that case, the implementation during the simulation of structural restraints obtained from experimental techniques, such as ATR-FTIR, might be necessary. In the case of homo-oligomers, a simple helix tilt should be sufficient. For more complicated proteins, information on individual helix tilts and relative helix register, derived from electron microscopy data or EPR, may be required. The wealth of readily available evolutionary conservation data provides potential for the widespread application of this approach, not only to membrane proteins using global searching molecular dynamics but in concert with any technique that generates multiple candidate protein structures.

REFERENCES

1. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 1991;9:56–68.
2. Greer J. Comparative modeling of homologous proteins. Methods Enzymol 1991;202:239–252.
3. Maxfield F, Scheraga H. Improvements in the prediction of protein backbone topography by reduction of statistical errors. Biochemistry 1979;18:697–704.
4. Niermann T, Kirschner K. Use of homologous sequences to improve protein secondary structure prediction. Methods Enzymol 1991;202:45–59.
5. Rost B, Sander C. Progress of 1D protein structure prediction at last. Proteins 1995;23:295–300.
6. Zvelebil M, Barton G, Taylor W, Sternberg M. Prediction of protein secondary structure and active sites using the alignment of homologous sequences. J Mol Biol 1987;195:957–961.
7. Thompson M, Goldstein R. Predicting solvent accessibility: higher accuracy using Bayesian statistics and optimized residue substitution classes. Proteins 1996;25:38–47.
8. Persson B, Argos P. Topology prediction of membrane proteins. Protein Sci 1996;5:363–371.
9. Defay T, Cohen F. Multiple sequence information for threading algorithms. J Mol Biol 1996;262:314–323.
10. Keasar C, Tobi D, Elber R, Skolnick J. Coupling the folding of homologous proteins. Proc Natl Acad Sci USA 1998;95:5880–5883.
11. Keasar C, Elber R, Skolnick J. Simultaneous and coupled energy optimization of homologous proteins: a new tool for structure prediction. Folding Design 1997;2:247–258.
12. Treutlein H, Lemmon M, Engelman D, Brunger A. The glycophorin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. Biochemistry 1992;31:12726–12732.
13. Adams P, Arkin I, Engelman D, Brunger A. Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. Nat Struct Biol 1995;2:154–162.
14. Lemmon M, Flanagan J, Hunt J, Adair B, Bormann B, Dempsey C, Engelman D. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. J Biol Chem 1992;267:7683–7689.
15. Lemmon M, Flanagan J, Treutlein H, Zhang J, Engelman D. Sequence specificity in the dimerization of transmembrane alpha-helices. Biochemistry 1992;31:12719–12725.
16. Arkin I, Adams P, MacKenzie K, Lemmon M, Brunger A, Engelman D. Structural organization of the pentameric transmembrane alpha-helices of phospholamban, a cardiac ion channel. Embo J 1994;13:4757–4764.
17. Kukol A, Adams PD, Rice LM, Brunger AT, Arkin IT. Experimentally based orientational refinement of membrane protein models: a structure for the influenza A M2+H channel. J Mol Biol 1999;286:951–962.
18. Torres J, Adams P, Arkin I. Use of a new label, $^{13}C=^{18}O$, in the determination of a structural model of phospholamban in a lipid bilayer: spatial restraints resolve the ambiguity arising from interpretations of mutagenesis data. J Mol Biol 2000;300:677–685.
19. MacKenzie K, Prestegard J, Engelman D. A transmembrane helix dimer: structure and implications. Science 1997;276:131–133.
20. Manolios N. Hierarchy of T cell antigen receptor assembly. Immunol Cell Biol 1995;73:544–548.
21. Jacobs H. Pre-TCR/CD3 and TCR/CD3 complexes: decamers with differential signalling properties? Immunol Today 1997;18:565–569.
22. Rutledge T, Cosson P, Manolios N, Bonifacino J, Klausner R. Transmembrane helical interactions: zeta chain dimerization and functional association with the T cell antigen receptor. Embo J 1992;11:3245–3254.
23. Bolliger L, Johansson B. Identification and functional characterization of the zeta-chain dimerization motif for TCR surface expression. J Immunol 1999;163:3867–3876.
24. Brunger A, Adams P, Clore G, Gros W, Grosse-Kunstleve R, Jiang J, Kuszewski J, Nilges M, Pannu N, Read R, Rice L, Simonson T, Warren G. Crystallography and NMR system: a new software system for macromolecular structure determination. Acta Crystallogr D 1998;905–921.
25. Jorgensen W, Tirado-Rives J. The OPLS potential function for proteins, energy minimization for crystals of cyclic peptides and crambin. J Am Chem Soc 1988;110:1657–1666.
26. Adams P, Engelman D, Brunger A. Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. Proteins 1996;26:257–261.
27. Engelman D, Steitz T, Goldman A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem 1986;15:321–353.
28. Bleasby A, Akrigg Y, Attwood T. OWL—a non-redundant composite protein sequence database. Nucleic Acids Res 1994;22:3574–3577.
29. Russ W, Engelman D. TOXCAT: a measure of transmembrane helix association in a biological membrane. Proc Natl Acad Sci USA 1999;96:863–868.
30. Chothia C, Levitt M, Richardson D. Helix to helix packing in proteins. J Mol Biol 1981;145:215–250.