PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Extension of a protein docking algorithm to membranes and applications to amyloid precursor protein dimerization

Shruthi Viswanath,[1,2] Laura Dominguez,[3] Leigh S. Foster,[3] John E. Straub,[3] and Ron Elber[2,4]*

[1] Department of Computer Science, University of Texas at Austin, Austin, Texas 78712

[2] Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712

[3] Department of Chemistry, Boston University, Boston, Massachusetts 02215

[4] Department of Chemistry, University of Texas at Austin, Austin, Texas 78712

ABSTRACT

Novel adjustments are introduced to the docking algorithm, DOCK/PIERR, for the purpose of predicting structures of transmembrane protein complexes. Incorporating knowledge about the membrane environment is shown to significantly improve docking accuracy. The extended version of DOCK/PIERR is shown to perform comparably to other leading docking packages. This membrane version of DOCK/PIERR is applied to the prediction of coiled-coil homodimer structures of the transmembrane region of the C-terminal peptide of amyloid precursor protein (C99). Results from MD simulation of the C99 homodimer in POPC bilayer and docking are compared. Docking results are found to capture key aspects of the homodimer ensemble, including the existence of three topologically distinct conformers. Furthermore, the extended version of DOCK/PIERR is successful in capturing the effects of solvation in membrane and micelle. Specifically, DOCK/PIERR reproduces essential differences in the homodimer ensembles simulated in POPC bilayer and DPC micelle, where configurational entropy and surface curvature effects bias the handedness and topology of the homodimer ensemble.

Proteins 2015; 83:2170–2185.

Key words: protein–protein docking; transmembrane complexes; membrane potential; amyloid precursor protein; all-atom MD simulation; micelle and bilayer environments; explicit and implicit solvent.

INTRODUCTION

Predicting protein–protein interactions is a major goal in computational structural biology. The total number of proteins or genes in a genome is quite small (tens of thousands), which is puzzling considering the observed diversity of life. The prospects of abundant protein complexes of variable number of individual chains open many new possibilities for the natural design of diverse molecular machines. Protein interactions add another layer of useful complexity to the diversity of biological systems.

Predicting protein–protein interactions is conducted on multiple levels. We may start from the observation that the function or expression of one protein impacts the function or expression of the other and therefore they interact. The "interaction," however, can be indirect and realized through the action of other molecular mediators or promoters. Alternatively, the interaction can be made through a direct physical contact. In the present manuscript we focus on the latter. Given that two proteins are physically proximate and in contact, the determination of their three-dimensional structure is of considerable interest.

A number of theories, algorithms, and programs are available to predict the structures of protein complexes given the structures of the individual monomers.[1–4] The structures of the monomers need not be very precise and homology models have been used effectively in the past

Grant sponsor: National Science Foundation; Grant numbers: CHE-1114676, CHE-1362524, CCF-0833162; Grant sponsor: National Institutes of Health; Grant number: RO1 GM076688; Grant sponsor: Welch Foundation; Grant number: F-1783; Grant sponsors: Schlumberger Foundation "Faculty for the Future Program"; CONACYT.
*Correspondence to: Ron Elber, 201 East 24th St. STOP C0200, Austin, TX 78712-1229. E-mail: ron@ices.utexas.edu
Received 16 June 2015; Revised 13 September 2015; Accepted 17 September 2015
Published online 24 September 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24934

to create accurate structures of complexes (so called unbound docking[5,6]). These approaches usually focus on complexes found in aqueous solution for which more experimental data is available to learn parameter sets of the scoring functions and to test the theoretical predictions. However, predicting the structures of protein complexes in other environments, such as membrane-embedded proteins, is also of significant interest. Membrane proteins are critical for transport of material across cell boundaries and for transmitting signals into and out of cells. Furthermore, certain diseases and aggregation phenomena are associated with peptide interactions in membranes and are of general medical importance.[7]

Probing membrane protein interactions using rigid docking

One useful computational tool for deducing the structure of membrane protein complexes is rigid docking. Rigid docking can exhaustively sample the set of all possible rigid conformations of the complex on a lattice. This sampling is more comprehensive than the sampling obtained from equilibrium MD simulations. Further, docking, if established to be accurate, can be an efficient means of sampling higher order conformations of the peptide (oligomers), providing atomic detail regarding the structure of aggregates, as a quicker computational alternative to MD simulations.[8] Finally, the potentials used in a docking algorithm are based on contacts observed in protein interfaces and incorporate a information that differs in essential ways from information used to parameterize force fields employed in MD simulation.

Docking algorithms like Cluspro, Haddock, and ZDOCK have been used previously to study the structure of membrane proteins.[9–13] Indeed, in a recent study[14] a comparison was made between different algorithms for predicting membrane protein complexes. The algorithms were designed primarily for aqueous solutions. However, with only small adjustments they were used to predict the structures of membrane protein complexes. While the benchmarks are clearly useful, and point to successful prediction protocols, there remain a number of open and intriguing problems. (1) What explains the success of these algorithms when applied to structure prediction in such different environments? (2) Are protein–protein interfaces similar in membrane and aqueous solution, making the prediction less dependent on the environment? (3) What is the role of the membrane and membrane-water interface in determining the structure of protein complexes?

Incorporating membrane–protein interaction in DOCK/PIERR

Clearly, the environment in which the protein is embedded must have an impact on its structure and function. Earlier, simple one-body terms were proposed to describe the transfer energies of different molecules from aqueous solutions to membrane.[15] Can we add such a term to docking algorithms while retaining the rest of the scoring functions in the form found to be most successful for proteins in aqueous solution? With a simple correction at hand, one computational model can be used to predict the structures of protein complexes in both aqueous and membrane environments. Such an adjustment is also likely to provide better understanding of interactions that are sensitive to environment and those that are not.

In Ref. 13, the authors use the membrane orientation of models from ZDOCK as a filter for docking membrane complexes. In this study, in addition to orienting docking models in the membrane, we add a new term to the docking scoring functions, representing the membrane transfer energy. Further, in the adjustment of the docking algorithm to membranes we use our own docking method, DOCK/PIERR, which gives us easy access to the code and deeper understanding of benefits and limitations of the algorithm. The new membrane–protein interaction energy is a simple one-body term for the environmental changes the protein residues feel upon transfer from aqueous solution to membrane. The energy was designed by Tieleman *et al.*,[15] based on results from atomically detailed simulations. We add this term to the docking potentials developed in our group PISA[16] and PIE[17,18] and assess its use in predicting the structures of membrane protein complexes. A straightforward benchmark on a set of membrane proteins suggests that the adjustment significantly enhances the prediction accuracy.

Application to transmembrane protein homodimer structure prediction

We apply this docking technology to the problem of the prediction of coiled-coil dimer structures of transmembrane (TM) helical proteins, a problem that is of broad biophysical importance. The formation of homo- and hetero-dimers of TM helical proteins is known to be critical to the processing of membrane proteins as well as cellular signaling. Of particular interest is the transmembrane (TM) fragment of APP-C99 (C99), the 99 amino acid C-terminal fragment of the Amyloid Precursor Protein (APP).

C99 is processively cleaved by γ-secretase to produce the amyloid β (Aβ) peptide associated with Alzheimer's disease (AD).[19–21] Cleavage of C99 results in Aβ in a number of isoforms, ranging from 38 to 43 amino acids in length. The dominant isoform is $A\beta_{40}$, while the more amyloidogenic $A\beta_{42}$ is formed in a 10:1 $A\beta_{40}$: $A\beta_{42}$ ratio.[22,23] It is known that a variety of factors, including familial mutations of C99,[24,25] stability of the TM helix,[26] extent of homodimerization,[27–29] membrane lipid composition,[30] and cholesterol levels,[31] can

influence the Aβ product distribution and impact the onset of AD. As such, knowledge of the details of the cleavage process is critical to research on AD.[32,33]

An essential aspect of that problem is the prediction of the structure of C99 monomer and homodimer in bilayer environments.[34,35] In this work, we have focused on the study of the fragment $C99_{23-55}$, which contains the single TM region of C99 and is known to form homodimers in lipid bilayers and micelles. Early proposals for the structure of homodimers formed by association of the TM region of C99 consisted of right-handed coiled-coils stabilized by interpeptide interactions facilitated by the GxxxG motif[21,25,26,36–39] that provides a good surface for helix packing.[40] In contrast, a recently reported NMR structure of $C99_{15-55}$ (+8 amino acids) is a left-handed coiled-coil structure stabilized by a heptad-repeat motif involving G38 and A42.[41] The most recent NMR structure[42] of $C99_{23-55}$ homodimer reports a right-handed coiled-coil stabilized by interpeptide contacts in the C-terminal region, in agreement with earlier experimental findings and computational predictions. Given these differing proposals, additional studies, both experimental and computational, are clearly required to develop a complete and consistent understanding of the C99 homodimer structure.

In this study, we illustrate that rigid docking and reranking of sampled configurations of amyloid precursor peptides are consistent with atomically detailed simulations with implicit solvation. This finding concurs with,[43] where the authors perform rigid docking and implicit solvent simulation for dimers of Glycophorin-A and its mutants. Docking is shown to reproduce all the key conformations in the MD ensembles. Differences between structures of dimers obtained by different computational methods are discussed. Finally, structural differences between dimers characterized in micelle and bilayer environments are considered.

METHODS

In this section, we first describe DOCK/PIERR, our rigid docking algorithm, and the membrane score added to DOCK/PIERR to mimic the membrane environment. Second, we consider the dataset of unbound membrane protein complexes, used for establishing performance of various docking algorithms, along with a brief note about the docking algorithms whose performances were compared. Finally, we explain the simulation methods used for obtaining the 23–55 dimers of the amyloid precursor protein (APP) and the approach used to dock the APP monomers obtained from simulation.

DOCK/PIERR rigid docking algorithm

DOCK/PIERR[16,17,44] is an algorithm that predicts, in atomic resolution, the structure of the complex formed by two proteins, given their individual tertiary structures. The first phase involves rigid docking and coarse scoring. First, an exhaustive set of rigid transformations of one protein with respect to the other is sampled and scored, using Fast Fourier Transforms (FFT) on a grid with ~1 Å spacing. The scoring used is a combination of a van der Waals term for shape complementarity and a residue potential PIE[17,18] that is based on interface residue-residue contacts. The parameter set for the scoring function was optimized using a rigorous theory and extensive learning set (640 protein complexes).[18] For prediction purposes, the top scoring $2^{19} = 524388$ conformations are stored, and subject to subsequent filtering for interface clashes and clustering. Specifically, the set of conformations is clustered in rigid body space, followed by filtering out of models with >45 atomic clashes. The remaining models are re-clustered based on interface RMSD. Clustering removes models that are too similar to other models and allows for more diverse ranking.

The second phase of the calculation emphasizes reranking with atomic models.[16] Only the top scoring one thousand models selected by clustering as discussed above are considered. For each model, the side chains at the interface are remodeled, and a brief energy minimization is performed. The minimized structures are then rescored using an atomic potential designed for protein contact interfaces, PISA, combined with the residue potential, PIE. Note that the second phase is not meant as a refinement step. The adjustments to the structure are exceptionally small (RMSD ~ 0.1 Å) and are made to improve the atomic score, not the overall docking pose. Finally, the algorithm returns the top ten models of the complex as predictions.

In recent community-wide assessments of docking algorithms[45] (CAPRI), the DOCK/PIERR algorithm was judged to be one of the top four current automated protein–protein docking methods. The novelty of this docking algorithm lies in the potentials PIE and PISA used for scoring residue and atomic contacts at protein interfaces. The parameters of these potentials are derived by examining hundreds of thousands of correctly and incorrectly docked poses, using large-scale machine learning methods like structural SVMs.[17,46]

Membrane potential for reranking docking decoys

The docking algorithm described above only examines the interface contacts of the models and does not incorporate information about the environment surrounding the complex. The potentials, PISA and PIE, used for scoring interface contacts are derived empirically from datasets of experimental and decoy structures of primarily solvated protein complexes (the training set includes only 7 membrane proteins of a total of 640 and their contribution to the statistics is small).

We anticipate that learning on membrane proteins will result in different potentials. Nevertheless, it is tempting to keep the designed potentials "as are" and look for an additional term to score the effects of the membrane. This will make the docking method more modular, transferable, and general. In this study, adding such an energy term, that includes residue-specific information about membrane solvation, is shown to enhance prediction accuracy in membrane complexes. This term was added to the last reranking step of docking and not to the coarse FFT step. The complexity of the FFT code suggests restricting the use of the additional term, at least to begin with, to the final step of reranking,

Calculating membrane energy

Rather than design a membrane environment potential from scratch, we adopted a function that was developed by other investigators. Previous results from MD simulations by Tieleman *et al.* consider transfer free energy from aqueous solution to the center-of-membrane for each amino acid residue.[15] Their detailed and comprehensive simulations provided us with singe body adjustments that measure the costs (and rewards) of transferring each amino acid between the two environments. The underlying physical assumption is that the one body term captures the environment effect and that the impact of the membrane on the two body interactions is significantly smaller and can be neglected. The drawback of our choice is that the atomically detailed simulations and our machine learning procedure are not necessarily compatible and some double counting of the same effect may occur. On the other hand the combination of our potential with Tieleman's energy includes only one free parameter, making it relatively simple to verify the impact and the significance of the combination. We observe a large enhancement in prediction capacity, which suggests that the environment potential indeed captures a useful signal.

The membrane energy was calculated from these transfer energies using the following steps. First, each docking model was inserted into the membrane, by placing its center of mass at the center of the membrane, and by orienting the eigenvector corresponding to the smallest Eigenvalue of the tensor of inertia of the model along the membrane normal. This orientation is appropriate for elongated transmembrane proteins such as helical proteins, which are our prime targets in this study. For wide proteins a different orientation procedure will have to be used, since the eigenvector with the smallest Eigenvalue is not necessarily in the direction normal to the membrane. Second, for each docking model, the relative solvent accessibility of every residue was calculated using the program DSSP.[47] Finally, the membrane energy was calculated as follows: each residue whose side-chain center of mass was within a specified membrane width

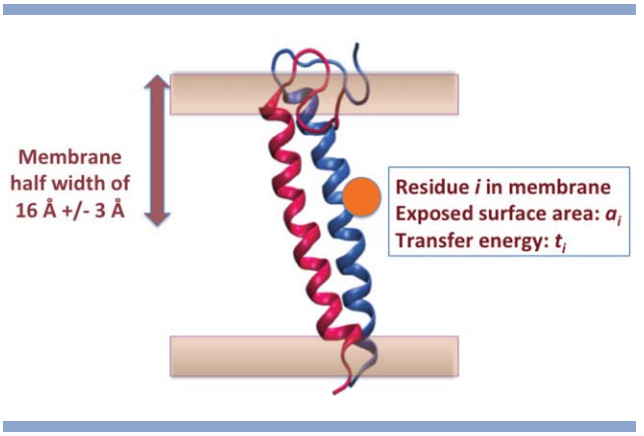

Figure 1

Example of a model oriented in the membrane, and a particular residue, *i*, inside the membrane that contributes $a_i t_i$ to the membrane energy, where a_i is the residue relative solvent exposed surface area and t_i is the residue membrane transfer energy. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

contributed to the membrane energy. The contribution from such a residue, *i*, was equal to the membrane transfer energy for that residue, t_i, weighted by its relative solvent (lipid) accessibility, a_i. As shown in Eq. (1), the membrane transfer energy, or MTE, for a model, is the sum of the transfer energy contributions from all residues *i*, within the membrane width.

$$MTE = \sum_i a_i t_i \quad (1)$$

We note that Tieleman *et al.* also provided water-to-hydrophilic membrane interface transfer energies, apart from water-to-center of membrane transfer energies. The addition of these extra parameters did not contribute to improved accuracy in ranking and hence they are not included in our docking algorithm for membrane complexes.

Membrane widths

The membrane half-width along the *Z*-axis is, of course, important for our calculations since it determines the degree of exposure of different amino acid side chains to the membrane environment or to aqueous solution. However, membrane widths are not strictly fixed and can vary among different membrane proteins.[48] For experimentally determined structures the width is known; however, for model complexes and variable composition of lipids it is not. Servers like TMDET[48] and databases like the PDBTM database[49] store precalculated widths for membrane proteins whose experimental structure has been determined. But these are difficult to use when ranking hundreds of thousands of models, with different effective membrane widths, and when studying complexes for which the experimental

Table I
Targets and Individual Components That Formed the Dataset of 30 Transmembrane Proteins

		Receptor chain		Ligand chain			
Target	Original PDB	Chain: start residue	Chain: end residue	Chain: start residue	Chain: end residue	Receptor homolog chain	Ligand homolog chain
1A91	1A91	A:1	A:42	A:43	A:79	MD	MD
1BL8	1BL8	A:23	A:119	B:23	B:119	1K4D_C	1K4D_C
1C17	1C17	A:1	A:79	B:1	B:79	MD	MD
1C3W0	1C3W	A:75	A:231	A:5	A:74	2I1X_A	1Q5J_A
1C3W1	1C3W	A:102	A:231	A:5	A:101	2I1X_A	1CWQ_A
1C3W2	1C3W	A:5	A:129	A:130	A:231	1Q5I_A	3VHZ_A
1EHK	1EHK	B:3	B:168	C:2	C:34	3S33_B	1EHK_C
1H2S	1H2S	A:1	A:225	B:23	B:82	1GU8_A	2F95_B
1H680	1H68	A:94	A:219	A:2	A:93	4GYC_A	2F93_A
1H681	1H68	A:2	A:119	A:120	A:219	2F93_A	2F93_A
1H682	1H68	A:2	A:150	A:151	A:219	2F93_A	2F93_A
1JVM	1JVM	B:24	B:123	C:24	C:120	1K4D_C	1K4D_C
1LGH	1LGH	A:1	A:56	D:1	D:56	MD	MD
1M0K0	1M0K	A:73	A:231	A:5	A:72	1C3W_A	1M0K_A
1M0K1	1M0K	A:106	A:231	A:5	A:105	1C3W_A	1Q5I_A
1M0K2	1M0K	A:5	A:128	A:129	A:231	1CWQ_A	1C3W_A
1M56	1M56	C:2	C:266	D:10	D:51	1QLE_C	1M56_D
2BHW	2BHW	A:10	A:232	B:10	B:232	MD	MD
2BRD0	2BRD	A:66	A:228	A:7	A:65	1PXR_A	2BRD_A
2BRD1	2BRD	A:103	A:228	A:7	A:102	3VHZ_A	1Q5I_A
2BRD2	2BRD	A:7	A:129	A:130	A:228	1CWQ_A	3VHZ_A
2IRV	2IRV	B:93	B:271	A:92	A:273	2O7L_A	2IC8_A
2KSE	2KSE	A:1	A:40	A:150	A:186	MD	MD
2NRF	2NRF	A:91	A:272	B:91	B:272	2IC8_A	2IC8_A
2VT4	2VT4	A:40	A:358	B:39	B:359	2Y00_A	2Y00_A
2WIE	2WIE	A:2	A:82	B:2	B:82	3V3C_A	3V3C_A
3B45	3B45	A:169	A:270	A:91	A:168	3B44_A	2IC8_A
3B4R	3B4R	A:3	A:220	B:3	B:218	MD	MD
3DWW	3DWW	A:11	A:152	C:11	C:152	MD	MD
3KCU	3KCU	A:29	A:280	B:29	B:280	3Q7K_A	3Q7K_A

Also listed are the homologous chains for those proteins whose unbound structures were obtained by homology modeling. Chains denoted by "MD" are those for which homologs were not found or for which the homology modeled structures were not sufficiently close (TM-score was lower than 0.85) to the bound structure. Molecular Dynamics was used to obtain the unbound structures for these cases. Further, for four of the complexes (1EHK, 1M0K0, 1M56, and 2BRD0) the ligand chain was retained in bound form and bound-unbound docking was performed. Details are provided in the following subsection.

data is limited. To pick a width that is consistent and optimal within our model, we use the following procedure: for each docking model, membrane transfer energies were calculated for a range of half-widths: 16 Å ± 3 Å, in steps of 0.5 Å that is, for 13.0, 13.5, 14.0, 14.5. . .16.0, 16.5, 17. . .19 Å, respectively. For each width, only protein residues whose centers of mass are within the membrane boundaries are scored according to Eq. (1) and contribute to the membrane energy for that width. The lowest (best) membrane transfer energy over the range of widths was taken as the score for the docking model. Figure 1 shows an example of a model oriented in the membrane, and a particular residue, *i*, inside the membrane that contributes to the membrane energy.

Bilayer versus micelle membrane environments

For docking membrane proteins characterized in a micelle environment, instead of the regular linear membrane model, a spherical membrane model with radius of 16 Å ± 3 Å is used to calculate the membrane energy. The code for membrane energy was implemented using the Biopython module.

Combining membrane energy with docking scores

DOCK/PIERR uses C3, a combination of atomic and residue potentials, to rerank the top one thousand minimized docking models.[16] The membrane energy (MTE) was combined with C3 in a parameter-free fashion by using the product of C3 with MTE. Such a product of docking scores/energies has been previously shown to work well to improve the accuracy of the reranking step.[16] The product energy in this study was formulated as $k * C3 * MTE$ where k=1.6 if both C3 and MTE have positive values and $k = -1.0$ otherwise. This ensures that the product energy is negative when both energies are negative (favorable) and positive otherwise. We henceforth refer to the product energy as C3*MTE. The code

for MTE was implemented using Biopython.[50] C3 energy as well as MTE are available online at http://clsb.ices.utexas.edu/web/dock_details.html.

Other docking algorithms

The performance of DOCK/PIERR was compared to Cluspro,[51] GRAMM-X[4], and ZDOCK+ZRANK.[3,52] We have compared our algorithm to these approaches in the past for the case of protein complexes in aqueous solution and it therefore makes sense to extend our comparison to membrane proteins. Previous comparative docking studies have shown that these algorithms were among the best performing algorithms for membrane protein docking.[14] Results were obtained from the servers in the cases of Cluspro and GRAMM-X. For ZDOCK+ZRANK, the ZDOCK 3.0.2 package was downloaded and docking jobs were run locally. The top 2000 models from ZDOCK were rescored using the ZRANK scoring function.

Creation of unbound membrane protein complexes dataset

A data set of 30 transmembrane protein complexes was extracted from MPStruc,[53] a database of membrane proteins from the White laboratory. Representative structures were chosen from each of the classes to ensure functional and structural diversity. The membrane span of the selected proteins was checked using the PDBTM database,[49] a database of transmembrane proteins in the PDB. Proteins selected from the MPStruc database, that had no entry in the PDBTM database, were discarded. Proteins classified as membrane proteins often do not span the entire length of the membrane and can interact with just one small region of it, for example, peripheral membrane proteins or cell-surface proteins. The PDBTM database was therefore used to determine the extent to which each protein was embedded in the membrane. Integral membrane proteins, where the majority of the structure to be docked lay in the transmembrane region, were specifically chosen.

We obtained 18 complexes for docking two separate protein chains. To increase the number of experimental complexes in this study, we also considered single-chain multi-pass trans-membrane proteins (e.g., GPCRs) that we broke into two complementary fragments, at an extramembranous loop region, and reassembled. In this way we obtained 12 more complexes. For each chosen GPCR, multiple independent splits were made, and each split produced two chains to be docked. Each independent split was taken as a separate target for unbound docking. Table I shows that we obtained 12 targets from the GPCRs 1C3W, 1H68, 1M0K, and 2BRD, 3 per GPCR, in this manner. Finally, we also discarded transmembrane chains where the binding between the chains was intricate, that is, one of the chains twisted around the other. This is because these cases are not suitable for rigid docking as one protein undergoes a large conformational change to bind with the other. For each protein complex chosen, Table I shows how we obtained the individual components to dock. The PDB IDs in the column "Original PDB" show the PDB entries that the targets are derived from.

Modeling unbound chains by homology and creating distorted structures by molecular dynamics

First, for each receptor (one of the proteins in the complex) and ligand (the complementary protein in the complex) sequence in the set of 30 transmembrane complexes, a search for homologs in the PDB was performed using PSI-BLAST.[54] For complexes for which homologs (*E*-value lower than 0.001) were found for receptor and/or ligand chains, Modeller[55,56] was used to create a structure of the unbound receptor and ligand using the homolog as template. The TM score[57] of the bound to unbound structure was measured for each homology-modeled receptor and ligand chains. Unbound (modeled) conformations that were too different (i.e., TM score lower than 0.85) from the bound (PDB) conformation were discarded.

In all, we successfully produced homologous unbound conformations for both chains in 19 of 30 complexes. Apart from these 19, 4 complexes (1EHK, 1M0K0, 1M56, 2BRD0) had one unbound chain (receptor or ligand) with TM score lower than 0.85 to the bound structure, and the other chain with a TM score higher than 0.85 to the bound structure. For these four complexes, the unbound structures with TM scores lower than 0.85 were replaced with the bound (PDB) conformation and bound–unbound docking was performed. Four other complexes (1A91, 1C17, 2BHW, 3DWW) had both receptor and ligand unbound conformations quite different (TM score lower than 0.85) from the bound conformations. And for three complexes (2KSE, 1LGH, 3B4R), homologs were not found in the first step of PSI-BLAST. Hence the latter seven complexes were treated separately and molecular dynamics was used to obtain the unbound conformations in these seven cases as described below.

For the receptor or ligand proteins for which homology modeling was unsuccessful, unbound conformations were obtained from short Molecular Dynamics MD runs on the original PDB receptor and ligand structures. The receptor and ligand were separately minimized in vacuum for 100 steps using *mini_pwl*, an energy minimization routine in the MD package MOIL,[58] in order to remove high-energy contacts and clashes in the structures before the dynamics run. The minimized structures (receptor and ligand separately) were subject to a very

short simulation of 0.1 ps at 300 K (1000 steps with a time step of 0.0001 ps). The conformations obtained after the dynamics run were used as the unbound structures. These perturbed conformations had an average all-atom RMSD of 0.717 Å to the original PDB structures, and a range of RMSDs between 0.618 and 0.859 Å. These RMSD values are smaller than typical homology models. However, MD under the above conditions tends to significantly distort the protein structures. Therefore, we did not push the simulations to longer times.

Simulation methodology for studies on APP dimers

For explicit solvent simulation, we employed all-atom models of POPC bilayer and DPC micelle environments. Initial structures for the all-atom models were taken from the results of CG simulations performed using the MARTINI force field.[50,59,60] Initially, two $C99_{23-55}$ peptides were placed 25 Å apart in a POPC bilayer. CG molecular dynamics simulation was performed for 1.5 μs for 60 independent systems using GROMACS.[61] In all cases, the CG peptides were observed to associate and form homodimers. This resulted in a diverse set of homodimer structures, each of which could be characterized as being in the Gly-in, Gly-out, or Gly-side conformational state. The CG structures were subsequently transformed to all-atom models using Pulchra[62] and embedded in the equivalent all-atom membrane using CHARMM-GUI.[63–66] The structures were minimized and pre-equilibrated at 310 K and 1 atm while restraining the protein backbone, followed by a 100 ns of molecular dynamics in the absence of restraints under a NPT semi-isotropic ensemble using CHARMM36 all-atom lipid force field (with CMAP) at 310 K and TIP3P water model.[67–69]

All-atom simulations in the DPC micelle or POPC bilayer consisted of 100 ns of MD performed on each all-atom system (following minimization and a short NVT and NPT equilibration with the protein backbone fixed). The non-bonded interactions were truncated using shift functions (between 0.9 and 1.2 nm for Lennard–Jones interactions and between 0 and 1.2 nm for electrostatics). Long-range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method[70] with a Fourier grid spacing of 0.12 nm. The pressure was set to 1 bar using a semi-isotropic coupling scheme with lateral and perpendicular pressures treated separately with coupling time 0.1 ps using the Parrinello–Rahman barostat methodology. The temperature of the system was set to 303 K and regulated using the Nosé–Hoover weak coupling algorithm.[71] The linear constraint solver (Lincs) method[72] was used to constrain all bond lengths, with a 2 fs integration step. All-atom simulations in DPC micelle were carried out under the same conditions using an isotropic coupling scheme to control the pressure. The simulations were carried out using GROMACS (v4.5.1).[61]

For the implicit solvent simulations, the replica-exchange molecular dynamics (REMD) method was used to improve sampling of the structural ensemble. Simulations were carried out using the MMTSB package[73] and CHARMM.[74] The system used an all-atom representation of the protein, and implicit representations of the lipid and water environments, which were represented by regions of varying dielectrics, with the membrane being defined as a continuous slab with a low dielectric value, in the XY-plane. The initial structures used were two $C99_{23-55}$ peptides modeled as straight helices, independently and randomly oriented in the implicit membrane. During the course of the simulation, the peptides were observed to form homodimers in the Gly-in, Gly-out, or Gly-side conformational state. The PARAM22 force field with the CMAP correction[75] was used, including corrections specific for the GBSW model[76] with updated radii. The smoothing length used at dielectric boundaries was 0.6 Å, with 24 radial integration points, and no cutoff. The surface tension coefficient was set to 0.04 kcal (mol $Å^{-2})^{-1}$. The membrane width was 40 Å with a 5 Å "head group" switching region at each end, leaving a 30 Å width as the membrane interior. The switching function for the head group region varied from the interior dielectric constant value of 1 to the solvent region dielectric of 80.

Approach for docking APP structures from simulation

A set of 50 dimers of the 23–55 segment monomer of APP-C99 corresponding to the lowest energy (based on the MD molecular mechanics energy) simulation structures obtained from implicit solvent MD simulations were used for docking.

Both bound and unbound docking was performed on each set of simulation structures. In bound docking, the monomers that is, individual helices of each simulated dimer were separated and docked. Ten top scoring models from docking were predicted for each simulation complex. For unbound docking, two simulation dimers were chosen at random (say A and B), and one helix from dimer A (say A's receptor) was docked to the other helix in dimer B (B's ligand). The docking predictions for this pair were compared to the dimer A. About 50 nonrepeating A–B receptor–ligand pairs were docked. Because the monomer conformations themselves can be quite different (>1 Å RMSD) from each other in simulations, the selection of complex B each time was constrained to those complexes where B's ligand was within 1 Å RMSD from the ligand in complex A.

Additionally, as a final post-processing step for docking APP structures and comparing the rigid docking procedure to simulations of peptide dimerization in membrane, anti-parallel dimer poses were filtered out

Table II
Docking Performance of DOCK/PIERR With C3 and C3*MTE Potentials, Gramm-X, Cluspro, and ZDOCK+ZRANK on the Dataset of 30 Unbound Membrane Protein Complexes

Docking algorithm	Top 10 number of hits within 4 Å iRMSD/number of complexes with at least one hit
DOCK/PIERR rerank with C3	2/2
DOCK/PIERR rerank with C3*MTE	14/11
ZDOCK+ZRANK	10/9
Cluspro	17/14
Gramm-X	20/17

from the final set of docking models, by making use of the additional information that the dimers found in the MD simulation are never anti-parallel. The last observation may reflect a kinetic barrier and not necessarily thermodynamic preference. However, for comparison purposes, the above filtering was found useful.

While evaluating docking methods on the APP dimers, a cutoff of 1.5 Å interface RMSD was used as the definition of "hit" or near-native structure (as the monomer helices are short and only 33 residues long). This is in contrast to the usual cutoff, which is 4 Å for an acceptable model and 2.5 Å for a high-quality model in protein-protein docking assessments such as CAPRI.[45,77] Typical RMSD increases with system size and hence a smaller value for the peptide system.

RESULTS AND DISCUSSION

In this section, we first discuss results for the prediction of membrane protein complexes. Second, we discuss the results from docking of APP dimers derived from implicit solvent simulations. Third, we discuss differences between structures obtained from alternative computational methods. Fourth, we touch upon differences in structures obtained from micelle and bilayer membrane environments.

Structure prediction of membrane protein interactions

Membrane protein interfaces can be predicted by solvated protein docking algorithms

Interfaces of membrane and water-soluble protein complexes are quite similar[14] and can be predicted with reasonable accuracy by current state-of-the-art protein-protein docking algorithms. This implies that protein-docking algorithms can be used as an additional and reliable source of information for structural studies of membrane proteins. We note that protein docking algorithms use potentials that have been trained on datasets that are primarily composed of soluble proteins; for example, Cluspro and Gramm-X use the training set in Ref. 78 which consists of 621 protein complexes out of which only 6 are membrane proteins, DOCK/PIERR is trained on a dataset of 640 complexes with a similar percentage of membrane proteins, and ZDOCK's interface contact potentials are trained on a dataset[79] of 89 complexes with one membrane protein.

In spite of being trained on interfaces of soluble proteins, these docking algorithms succeed in predicting a near-native structure in the top ten models with reasonable accuracy on membrane proteins. Table II shows the performance of four different docking algorithms on the dataset of 30 unbound transmembrane protein complexes. The measure of performance that we use here is the interface RMSD. Interface RMSD[80,81] is a widely used measure of accuracy for docking predictions, and is the RMSD measured along the interface residues of the experimental complex. The first number in column 2 of Table II shows the number of hits (near-native structures i.e., docking models that are within 4 Å interface RMSD from the experimental structure) in the top ten models cumulative across all 30 complexes (targets). The second number in column 2 of Table II shows the number of complexes for which at least one such hit was found in the top ten

Table III
The Numbers of Models With Interface RMSD <4.0 Å in the Top 10 Predictions of DOCK/PIERR With C3*MTE Potential, Gramm-X, Cluspro, and ZDOCK+ZRANK

Target	DOCK/PIERR with membrane score	ZDOCK+ ZRANK	CLUSPRO	GRAMM-X
1A91	1	1	2	1
1BL8	0	0	0	0
1C17	0	0	1	1
1C3W0	1	0	1	1
1C3W1	1	0	1	1
1C3W2	0	1	1	1
1EHK	0	0	0	0
1H2S	**3**	**0**	**0**	**0**
1H680	0	0	1	1
1H681	0	0	0	0
1H682	0	0	2	2
1JVM	**0**	**1**	**0**	**0**
1LGH	0	0	0	0
1M0K0	1	0	1	1
1M0K1	1	1	1	1
1M0K2	0	1	1	2
1M56	0	0	0	0
2BHW	0	0	0	0
2BRD0	1	0	0	1
2BRD1	0	0	1	1
2BRD2	2	0	2	1
2IRV	1	1	0	0
2KSE	0	1	0	2
2NRF	0	0	0	0
2VT4	0	0	0	0
2WIE	1	2	1	1
3B45	1	0	1	1
3B4R	**0**	**0**	**0**	**1**
3DWW	**0**	**1**	**0**	**0**
3KCU	0	0	0	0

Table IV
Bound and Unbound Docking Results for 50 Simulation Structures of the 23–55 segment of APP-C99 From Implicit Solvent

Docking type	Top 10 number of hits within 1.5 Å iRMSD to MD structure/number of complexes with at least one hit matching MD structure
Bound	43/42
Unbound	26/26

The first number in the second column is the number of hits recovered from docking across all 50 complexes: a hit is a model from docking that is within 1.5 Å interface RMSD to the corresponding simulation structure. The second number is the number of complexes for which at least one hit was found in the top ten models.

models. Depending on the algorithm, accuracy varies between 30 and 56.6% for unbound docking. Gramm-X performs the best in this study and is able to obtain a near-native structure in the top ten about 56.67% of the time in unbound docking. This is in agreement with an earlier study[14] that showed Gramm-X to have the best performance in docking membrane proteins.

Table III shows the performance of docking algorithms in terms of number of top 10 hits, split by target. DOCK/PIERR with the membrane score is able to dock complex 1H2S, which the other docking algorithms are not able to solve. Similarly, ZDOCK+ZRANK is able to solve uniquely 1JVM and 3DWW. Gramm-X is the only docking algorithm able to solve 3B4R.

Membrane energy contributes to improved recognition

As shown in Table II, the inclusion of the membrane energy significantly improves the recognition of the combination of atomic and residue potentials, C3. DOCK/PIERR is able to obtain a near-native structure in the top ten in 36.7% of complexes. We note that other docking programs can potentially benefit in accuracy from reranking using the additional membrane potential as well. Care must be used, however, when examining different algorithms and scoring functions to avoid double counting of similar energy terms.

Docking and implicit solvent MD simulations agree on structures of APP dimers

In this section, we explore the structure of the dimer formed by the 23–55 segment of the APP-C99 protein using docking and implicit solvent MD simulations. Table IV shows the performance of DOCK/PIERR for bound and unbound docking of 50 implicit solvent dimers from simulation. The docking performance was evaluated based on the number of models matching the corresponding MD structure within 1.5 Å interface RMSD. The first number in column 2 of Table IV reports the number of models in the top ten that matched the corresponding MD complex, across all 50 complexes. The second number in column 2 of Table IV is the number of complexes out of 50, for which at least one model in the top ten matched the corresponding simulation structure. Docking and MD simulation show a good agreement with 42 out of 50 dimers from bound docking matching the corresponding MD structure, and 26 out of 50 dimers from unbound docking matching the MD structure. The accuracy of unbound docking is lower than that of bound docking, which is to be expected, as the interfaces of monomers from unbound docking do not match precisely.

Figure 2 shows the probability distribution of interface RMSDs for the top 10 docking models from bound and unbound docking of the 50 simulation dimers. In other words, this is a distribution across a set of 500 bound and 500 unbound docking models. Note that since we filter out anti-parallel orientations, the interface RMSD distribution stops at 10 Å (x axis). There is a prominent tail near 1 Å, especially for bound docking indicating a significant number of near-native structures in the set of top 10 models. Another measure of confidence in docking predictions is the z scores, which is defined for a model, m, as $\frac{E_m - \mu}{\sigma}$, where E_m is the energy of the model, and μ and σc represent the average energy and standard deviation of the energy distribution respectively. The average z scores of the C3*MTE energy across the 5 best docking models (best in terms of interface RMSD) was -4.2646 among the 500 bound docking models and -3.5062 among the 500 unbound docking models. More negative z-scores indicate that the potential can distinguish near-native structures more accurately.

Further, APP dimers can be described by an order parameter based on the distance between the Gly29 in the two helices. If the distance is within 5 Å, the dimer is said to be in Gly-in conformation, if the distance is between 5 and 10 Å, the dimer is in Gly-side conformation, and if the distance is above 10 Å, the dimer is in Gly-out conformation. Gly-in structures are stabilized by

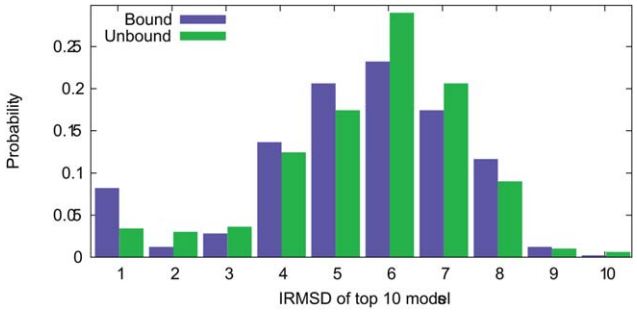

Figure 2
Probability density of the interface RMSD of top 10 docking models for 50 bound and unbound simulation dimers of APP-C99. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table V
Bound Docking Results on 40 Gly-side and 10 Gly-in Simulation Structures From Implicit Solvent

Simulation dimer type [Number of simulation dimers]	Top 10 Number of hits within 1.5 Å iRMSD to MD structure/number of complexes with at least one hit matching MD structure
Gly-side [40]	34/33
Gly-in [10]	9/9

The first number in the second column is the number of docking models within 1.5 Å interface RMSD from the corresponding simulation structure, across all complexes of the given dimer type. The second number is the number of complexes for which at least one hit was found in the top ten models for that dimer type.

interhelical contacts facilitated by the "flat face" created by the GxxxG sequence motifs. Gly-out structures are stabilized by a tetrad repeat motif that facilitates interhelical "knob-in-hole" interactions.

Based on this characterization, out of the 50 lowest energy dimers derived from implicit solvent simulation, 40 were of Gly-side type and 10 were of Gly-in type. There were no Gly-out structures in the 300 K MD ensemble. Table V shows the performance of bound docking in recovering the order parameters measured in the MD simulations. The agreement between docking and simulation dimers is high (9/10) for Gly-in type structures and good (33/40) for Gly-side structures.

Figure 3 provides a comparison of the distribution of $C99_{23–55}$ homodimer structures derived from all-atom simulation, in explicit POPC membrane or DPC micelle and in GBSW implicit membrane, with the DOCK/PIERR structures predicted by bound docking. The structures are projected on a plane defined by two order parameters.

The ensembles of C9923–55 homodimer in POPC bilayer and DPC micelle are shown in Figure 3 in terms of the order parameters ϕ_{4G} and d_{GG}. ϕ_{4G} is a dihedral angle formed by G29A-G37A-G37B-G29B, where A and B label the two C9923–55 monomers. d_{GG} is the interhelical distance between G33A-G33B. The ϕ_{4G} order parameter is positive for left-handed structures and negative for right-handed structures. The value of the d_{GG} parameter determines whether a structure is Gly-in, Gly-out or Gly-side, with Gly-in ($d_{GG} < 5$ Å), Gly-side ($d_{GG} < 5$ and 10 Å), and Gly-out ($d_{GG} > 10$ Å). Structures stabilized by interpeptide interactions facilitated by the GxxxG repeat region are Gly-in structures characterized by small values of the dGG parameter.

On the basis of the comparison of simulated ensembles and docking predictions, we can draw a number of conclusions. (1) The DOCK/PIERR predictions capture the three characteristic homodimer structural motifs, Gly-in, Gly-side, and Gly-out. (2) The predictions of DOCK/PIERR capture a number of essential trends in the environmental modulation of the $C99_{23–55}$ homodimer ensemble in bilayers and micelles, which is discussed below detail.[3] In addition, the DOCK/PIERR predictions suggest that left-handed structures are predominantly of the Gly-out topology as observed in all-atom simulations[36] and experiment.[41,42]

Figure 4 shows accurate docking predictions among the top ten models, superposed with the simulation structure from which they were assembled. The Gly-side model was within an interface RMSD of 0.5 Å from the simulation structure, while the Gly-in model was within 0.6 Å from the simulation structure. The figure shows that the backbones essentially overlap while the side-chains show minor differences.

Structural differences between the results of alternative computational methods to predict complexes of amyloid peptides

As Table V shows, DOCK/PIERR docking is reasonably accurate for Gly-in complexes generated by implicit solvent simulations using parameters for the membrane

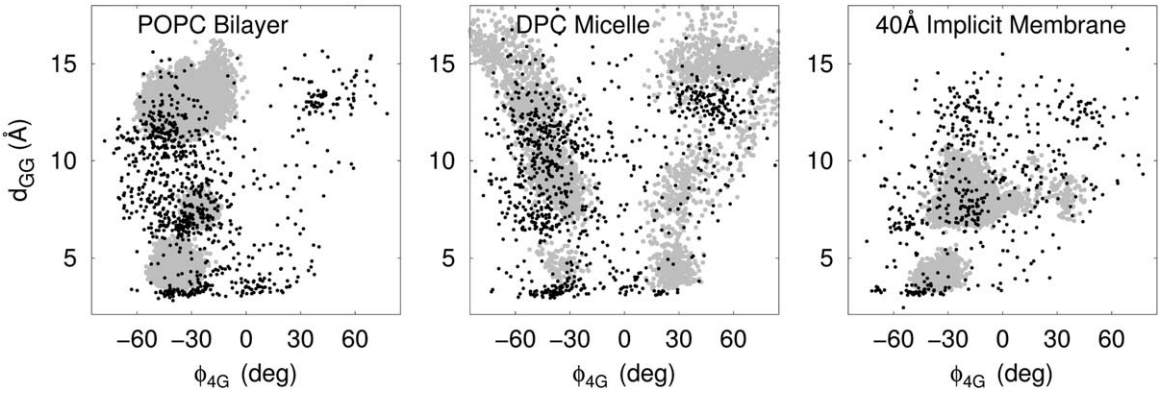

Figure 3
Distribution of homodimer structures of $C99_{23–55}$ in POPC bilayer, DPC micelle and Implicit Membrane derived from all-atom simulations (gray) and compared with the DOCK/PIERR predictions (black) projected onto the order parameters d_{GG} and ϕ_{4G}.

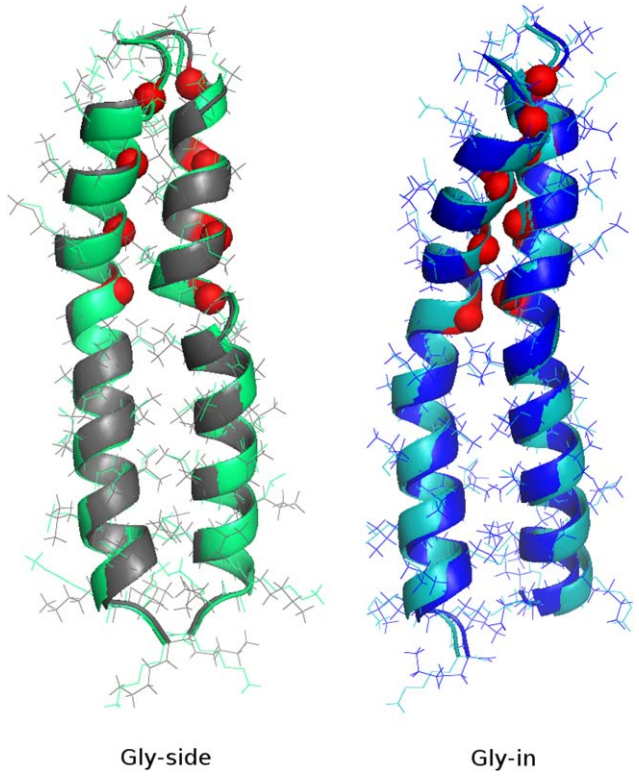

Figure 4

Left: A docking model (green) in the top 10 predictions, at an interface RMSD of 0.5 Å from the corresponding simulation structure (gray) of Gly-side type. Right: A docking prediction (cyan) in the top 10, at an interface RMSD of 0.6 Å from a Gly-in simulation structure (blue). The carbon alpha atoms of the glycine residues in the GxxxG motifs in the simulation structures are shown as red spheres.

width chosen to approximate a POPC bilayer. However, when applied to dock 30 Gly-in complexes from explicit solvent POPC bilayer, it was observed that DOCK/PIERR fails to produce a single hit in the top ten models for any of the 30 complexes. These differences in docking results hint at structural differences in the dimers from implicit and explicit solvation. The differences were investigated using the docking energy PIE, the residue contact potential used in DOCK/PIERR.

Figure 5 is a distribution of the PIE energy for the implicit and explicit solvent simulation dimers that were docked. It shows that the PIE energy is much lower for the dimers derived from implicit solvent simulations. This suggests that the number of inter-helical residue–residue contacts is higher for the implicit solvent dimers, leading to more favorable PIE energies for the latter. The contact based potentials in DOCK/PIERR favor the higher number of contacts in implicit solvent models. For this reason, docking models show better agreement with implicit solvent simulation derived dimers than with dimers derived from explicit solvent simulations.

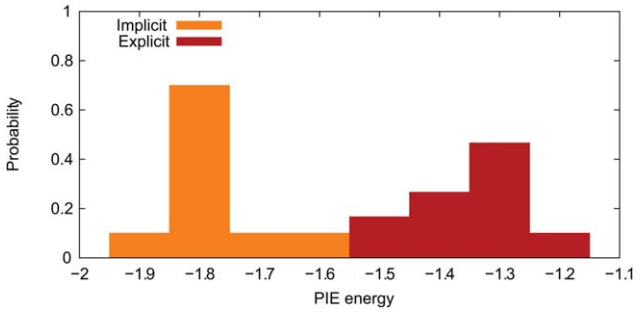

Figure 5

Probability distribution of PIE energy for 10 Gly-in implicit solvent simulation dimers and 30 Gly-in explicit solvent simulation dimers in POPC membrane that were bound docked. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Differing compactness of the helical dimers provides an explanation for the different numbers of contacts in implicit and explicit solvent. This is seen in Figure 6, which is a distribution of the smallest eigenvalue of the tensor moment of inertia for each simulation structure. The smallest eigenvalue corresponds to the long axis and can be a measure of how close the helices are to each other. The figure suggests that the implicit solvent simulation derived dimer helices are closer than the explicit solvent simulation derived dimers. In implicit solvent, the hydrophobic residues in the dimers form more contacts with each other, whereas in explicit solvent the hydrophobic residues form more contacts with the membrane. This leads to more compact dimers in implicit solvent. In explicit solvation models, protein–protein contacts can be more easily replaced by protein–water or protein–membrane contacts, while in implicit solvation model, the protein contacts are not explicitly replaced.

Figure 7 illustrates the 10 Gly-in implicit solvent models and 30 Gly-in explicit solvent models that were docked.

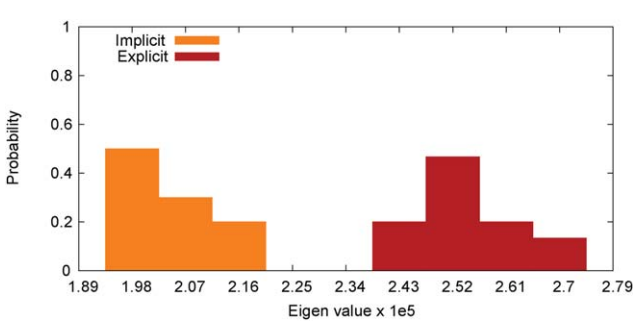

Figure 6

Distribution of the smallest Eigenvalue of the tensor moment of inertia for 10 Gly-in implicit solvent simulation dimers and 30 Gly-in explicit solvent simulation dimers in POPC membrane that were bound docked. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

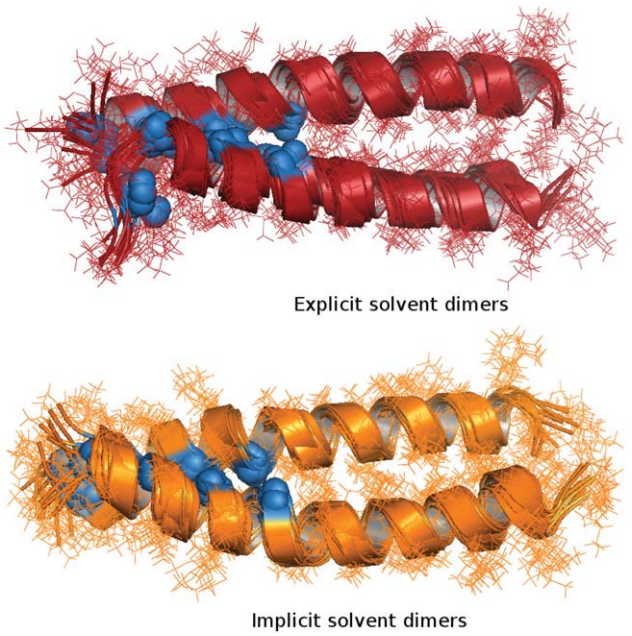

Figure 7

Top: 10 explicit solvent simulation dimers superimposed. Bottom: 10 implicit solvent simulation dimers superimposed. The dimers chosen were the top scoring simulation dimers from the MD ensemble according to C3*MTE. The carbon alpha atoms of the glycine residues in the GXXXG motifs of the dimers are shown as blue spheres.

The implicit solvent models lead to helices closer to each other at the C-terminal (right hand side) end, whereas in explicit solvent models, the helices are farther separated.

Differences between structures from DPC micelle and POPC bilayer

DOCK/PIERR is able to bound dock 17/30 simulation dimers from POPC bilayer (i.e., a model within 1.5-Å interface RMSD was found in the top ten models for 17 of 30 dimers) in Gly-out conformation. However, the same experiment repeated on the Gly-out dimers in DPC micelle results in no hits in the top ten for any of the 30 dimers derived from micelle simulations. Again the differences between the two docking accuracies hint at structural differences between dimers in the different bilayer and micelle environments.

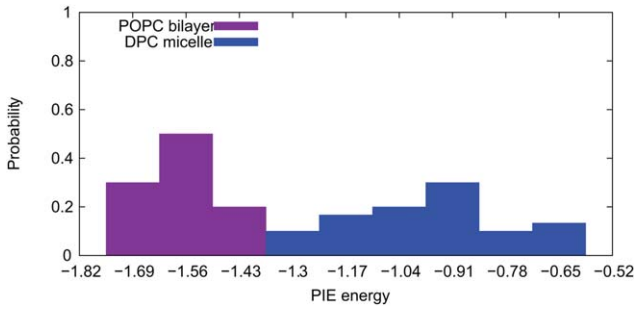

Figure 8

Probability distribution of PIE energy for 30 Gly-out explicit solvent simulation dimers in POPC bilayer and 30 Gly-out explicit solvent simulation dimers in DPC micelle that were bound docked. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

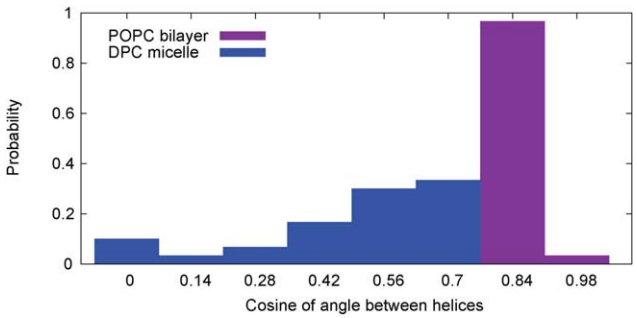

Figure 9

Distribution of the cosine of angle between helices for 30 Gly-out explicit solvent simulation dimers in POPC bilayer and 30 Gly-out explicit solvent simulation dimers in DPC micelle that were bound docked. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

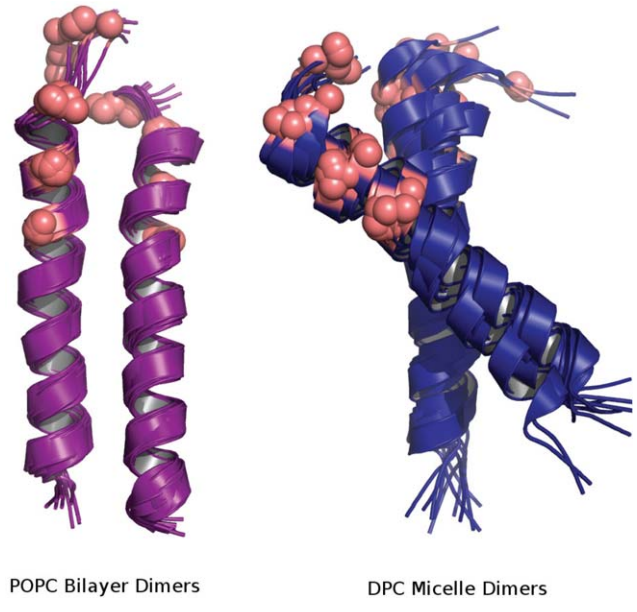

Figure 10

Left: Ten dimers derived from explicit solvent simulations in POPC membrane. Right: Ten dimers derived from explicit solvent simulations in DPC micelle. The ten dimers in each case were the top scoring dimers from the MD ensemble, as scored by C3*MTE. The carbon alpha atoms of the glycine residues in the GxxxG motifs of the dimers are shown as salmon colored spheres.

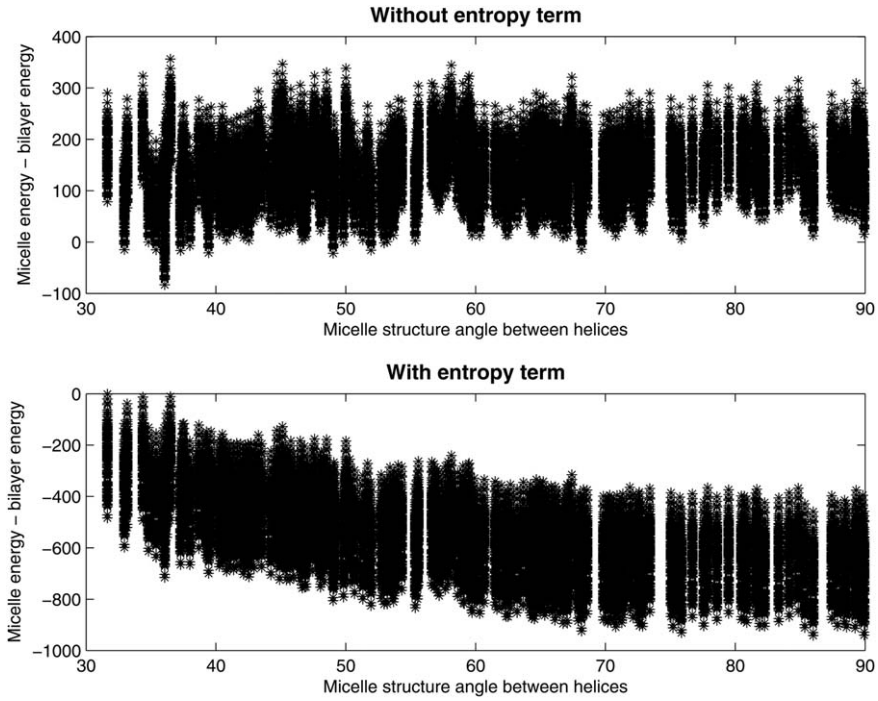

Figure 11

200 DPC micelle and 200 POPC bilayer structures of the Gly-out type were selected from the simulation ensemble. For each DPC micelle structure, the difference between its energy and the energy of each POPC bilayer structure was plotted as a function of the helical angle in the micelle structure. The top plot shows the distribution of energies without the entropy term and the bottom part shows the distribution of generalized energies after addition of the entropy term.

These differences were explored using PIE. Figure 8 shows distinct differences in the PIE energy for bilayer and micelle simulation models. The PIE energy is significantly more favorable for the bilayer models, due to a higher number of inter-helical contacts in the dimers.

Figure 9 is a plot of the absolute value of the cosine of the angle between the long axes of the helices in the dimer. There are clear differences in the distributions of angles between the helices due to the differing environments in the bilayer and micelle. Dimers in bilayers have cosine values closer to 1, indicating that the helices are more parallel. In contrast, helices in micelles have a wider range of angles and favor non-parallel orientations, which are more "X"-like.

This is also illustrated in Figure 10, which shows the 30 bilayer models with parallel helices and 30 micelle models with "X"-shaped helical angles. It is likely that helices in micelle environments tend to adopt an "X"-shaped orientation due to the influence of membrane curvature as well as an entropic effect. The "X"-shaped orientation is consistent with a greater number of configurations and larger structural fluctuations, leading to great configurational entropy of the protein in the micelle than is found in the parallel Gly-in configurations that are predominate in the membrane environment. The entropic driving force and membrane curvature may be used to explain the observed environmental dependence of dimer structures. Moreover, the observed differences point to the limited applicability of micelle environments as accurate mimics for membrane bilayers in membrane protein structure determination.

Entropic reasons for the differences between dimer structures in micelle and bilayer

Here we investigate the role of entropy as a possible reason for differences in the dimer structures observed in bilayer and micelle environments. The total energy C3*MTE is compared for a set of 200 simulation structures of Gly-out type in DPC micelle, and 200 simulation structures of Gly-out type in POPC bilayer. The top plot in Figure 11 shows that the micelle structures have a higher energy than the bilayer structures, and hence structures in the bilayer are preferred. However, the situation is reversed when we include the effect of entropy in a generalized energy. We modeled the entropy using the form $-A\log(\sin(x))$, where A is a constant and x is the small angle between helices in a dimer. The larger the angle x, the higher is the entropic contribution to energy since the helices have more rotational freedom. Upon adding this entropic term to the total energy, $E_{\text{total}} = C3 * MTE - A\log(\sin(x))$, and recalculating energy of all structures, we show in the bottom

plot of Figure 11 that for values of the constant *A* which are 610 or greater, the entropic factor makes the micelle structures, which are predominantly "X"-shaped, more favorable in generalized energy than the bilayer structures, which are parallel dimers. This analysis demonstrates that variations in entropy could explain why the "X" shaped dimers are favored in micelles compared to the parallel helices that are predominant in bilayer environments.

CONCLUSIONS

In this manuscript, we present the first comparative study of protein docking algorithms for docking unbound membrane proteins. It is also the largest comparison study including bound and unbound membrane complexes, including homo and heterodimers. We show that including information about the membrane environment as an additional one-body residue-based energy term improves the prediction capacity of our docking algorithm, DOCK/PIERR, significantly. We use this extended DOCK/PIERR method to study the dimerization of the transmembrane fragment of C99, the 99 amino acid C-terminal fragment, C99, of the Amyloid Precursor Protein.

We draw a number of conclusions that characterize the performance of the extended DOCK/PIERR method in a detailed application to the prediction of C99 homodimers in membrane and micelle environments. (1) The results from docking match well with results from implicit solvent simulation and are capable of capturing the diversity of topologically distinct coiled-coil structural states. (2) Structural ensembles derived from explicit solvent simulations differ from those derived using an implicit solvent model: explicit solvent structures have more protein–membrane contacts and implicit solvent structures have more protein–protein contacts. This difference suggests that implicit solvent models and our docking procedure are not able to reproduce the contacts formed by discrete solvent molecules. (3) Homodimer structural ensembles derived from simulations in bilayer and micelle environments display significant differences: bilayer-derived dimers have parallel helices while micelle-derived dimers display an "X"-shape, with helices oriented at a distinct crossing angle. This preference for "X"-shape can be explained on the basis of entropy (rotational freedom of the dimers in micelle) and membrane curvature.

There are several experimentally derived homodimer structures formed from fragments of APP-C99 peptide. Solid-state NMR studies of APP-$C99_{22\text{-}64}$ lead to the first experimentally derived homodimer structure,[26] in which the TM helices form a right-handed coiled-coil stabilized by contacts facilitated by the GxxxG motif and consistent with the Gly-in structures proposed in this study. Subsequently, solution NMR was used to derive homodimer structures for APP-$C99_{15\text{-}55}$ in DPC micelle[41] and APP-$C99_{28\text{-}55}$ in DPC micelle.[42] In the former study, a left-handed coiled-coil structure was proposed. In the later study, a right-handed coiled coil structure was proposed that is in good agreement with simulated structures of the APP-$C99_{15\text{-}55}$ homodimer in a DPC micelle environment.[82] However, those simulated and experimental structures for the homodimer in a DPC micelle, which are predominantly Gly-side and Gly-out structures, differ substantially from the simulated structures for APP-$C99_{15\text{-}55}$ homodimer in POPC bilayer, which is predominantly composed of Gly-in structures.[82] The results of this study of APP-$C99_{23\text{–}55}$ homodimer, and simulation studies of the APP-$C99_{15\text{-}55}$ homodimer,[36,82] suggest that the TM homodimer is best characterized by Gly-in structures in a POPC bilayer and Gly-side or Gly-out structures in a DPC micelle. As such, our simulation results, which predict a strong influence of environment on the APP-C99 homodimer structure, are consistent with the results of existing experimental studies.

Predicting the structure of higher order amyloid aggregates and developing additional potentials trained on membrane protein interfaces represent some of the promising avenues for future work in the area of membrane complex prediction.

Availability: The membrane potential and docking scores are available as downloads from http://clsb.ices.utexas.edu/web/dock_details.html.

REFERENCES

1. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res 2004;32:W96–W99.
2. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125:1731–1737.
3. Chen R, Li L, Weng ZP. ZDOCK: an initial-stage protein-docking algorithm. Proteins Struct Funct Genet 2003;52:80–87.
4. Tovchigrechko A, Vakser IA. Development and testing of an automated approach to protein docking. Proteins 2005;60:296–301.
5. Hwang H, Pierce B, Mintseris J, Janin J, Weng ZP. Protein–protein docking benchmark version 3.0. Proteins 2008;73:705–709.
6. Hwang H, Vreven T, Janin J, Weng ZP. Protein–protein docking benchmark version 4.0. Proteins 2010;78:3111–3114.
7. Sanders CR, Nagy JK. Misfolding of membrane proteins in health and disease: the lady or the tiger? Curr Opin Struc Biol 2000;10:438–442.
8. Selent J, Kaczor AA. Oligomerization of G protein-coupled receptors: computational methods. Curr Med Chem 2011;18:4588–4605.
9. Comeau SR, Camacho CJ. Predicting oligomeric assemblies: N-mers a primer. J Struct Biol 2005;150:233–244.
10. Cosconati S, Marinelli L, Lavecchia A, Novellino E. Characterizing the 1,4-dihydropyridines binding interactions in the L-type Ca2+ channel: model construction and docking calculations. J Med Chem 2007;50:1504–1513.
11. Ponomarev SY, Audie J. Computational prediction and analysis of the DR6-NAPP interaction. Proteins 2011;79:1376–1395.

12. Simon AC, Simpson PJ, Goldstone RM, Krysztofinska EM, Murray JW, High S, Isaacson RL. Structure of the Sgt2/Get5 complex provides insights into GET-mediated targeting of tail-anchored membrane proteins. Proc Natl Acad Sci USA 2013;110:1327–1332.
13. Casciari D, Seeber M, Fanelli F. Quaternary structure predictions of transmembrane proteins starting from the monomer: a docking-based approach. BMC Bioinform 2006;7:340.
14. Kaczor AA, Selent J, Sanz F, Pastor M. Modeling complexes of transmembrane proteins: systematic analysis of ProteinProtein docking tools. Mol Inform 2013;32:717–733.
15. MacCallum JL, Bennett WFD, Tieleman DP. Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. J Gen Physiol 2007;129:371–377.
16. Viswanath S, Ravikant DVS, Elber R. Improving ranking of models for protein complexes with side chain modeling and atomic potentials. Proteins 2013;81:592–606.
17. Ravikant DVS, Elber R. Energy design for protein–protein interactions. J Chem Phys 2011;135:065102.
18. Ravikant DVS, Elber R. PIE-efficient filters and coarse grained potentials for unbound protein–protein docking. Proteins 2010;78: 400–419.
19. Qi-Takahara Y, Morishima-Kawashima M, Tanimura Y, Dolios G, Hirotani N, Horikoshi Y, Kametani F, Maeda M, Saido TC, Wang R, Ihara Y. Longer forms of amyloid beta protein: implications for the mechanism of intramembrane cleavage by gamma-secretase. J Neurosci 2005;25:436–445.
20. Zhao GJ, Tan JX, Mao GZ, Cui MZ, Xu XM. The same gamma-secretase accounts for the multiple intramembrane cleavages of APP. J Neurochem 2007;100:1234–1246.
21. Kienlen-Campard P, Tasiaux B, Van Hees J, Li M, Huysseune S, Sato T, Fei JZ, Aimoto S, Courtoy PJ, Smith SO, Constantinescu SN, Octave JN. Amyloidogenic processing but not amyloid precursor protein (APP) intracellular C-terminal domain production requires a precisely oriented APP dimer assembled by transmembrane GXXXG motifs. J Biol Chem 2008;283:7733–7744.
22. Masters CL, Simms G, Weinman NA, Multhaup G, Mcdonald BL, Beyreuther K. Amyloid plaque core protein in alzheimer-disease and down syndrome. Proc Natl Acad Sci USA 1985;82:4245–4249.
23. Iwatsubo T, Odaka A, Suzuki N, Mizusawa H, Nukina N, Ihara Y. Visualization of a-beta-42(43) and a-beta-40 in senile plaques with end-specific a-beta monoclonals—evidence that an initially deposited species is a-beta-42(43). Neuron 1994;13:45–53.
24. Munter LM, Botev A, Richter L, Hildebrand PW, Althoff V, Weise C, Kaden D, Multhaup G. Aberrant amyloid precursor protein (APP) processing in hereditary forms of alzheimer disease caused by APP familial alzheimer disease mutations can be rescued by mutations in the APP GxxxG motif. J Biol Chem 2010;285:21636–21643.
25. Munter LM, Voigt P, Harmeier A, Kaden D, Gottschalk KE, Weise C, Pipkorn R, Schaefer M, Langosch D, Multhaup G. GxxxG motifs within the amyloid precursor protein transmembrane sequence are critical for the etiology of A beta 42. EMBO J 2007;26:1702–1712.
26. Sato T, Tang TC, Reubins G, Fei JZ, Fujimoto T, Kienlen-Campard P, Constantinescu SN, Octave JN, Aimoto S, Smith SO. A helix-to-coil transition at the epsilon-cut site in the transmembrane dimer of the amyloid precursor protein is required for proteolysis. Proc Natl Acad Sci USA 2009;106:1421–1426.
27. Scheuermann S, Hambsch B, Hesse L, Stumm J, Schmidt C, Beher D, Bayer TA, Beyreuther K, Multhaup G. Homodimerization of amyloid precursor protein and its implication in the amyloidogenic pathway of Alzheimer's disease. J Biol Chem 2001;276:33923–33929.
28. Gorman PM, Kim S, Guo M, Melnyk RA, McLaurin J, Fraser PE, Bowie JU, Chakrabartty A. Dimerization of the transmembrane domain of amyloid precursor proteins and familial Alzheimer's disease mutants. BMC Neurosci 2008;9–17.
29. Eggert S, Midthune B, Cottrell B, Koo EH. Induced dimerization of the amyloid precursor protein leads to decreased amyloid-beta protein production. J Biol Chem 2009;284:28943–28952.
30. Goo JH, Park WJ. Elucidation of the interactions between C99, presenilin, and nicastrin by the split-ubiquitin assay. DNA Cell Biol 2004;23:59–65.
31. Lu JX, Yau WM, Tycko R. Evidence from solid-state NMR for non-helical conformations in the transmembrane domain of the amyloid precursor protein. Biophys J 2011;100:711–719.
32. Das C, Berezovska O, Diehl TS, Genet C, Buldyrev I, Tsai JY, Hyman BT, Wolfe MS. Designed helical peptides inhibit an intramembrane protease. J Am Chem Soc 2003;125:11794–11795.
33. Wolfe MS, Guenette SY. App at a glance. J Cell Sci 2007;120:3157–3161.
34. Barrett PJ, Song YL, Van Horn WD, Hustedt EJ, Schafer JM, Hadziselimovic A, Beel AJ, Sanders CR. The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. Science 2012;336:1168–1171.
35. Dominguez L, Meredith SC, Straub JE, Thirumalai D. Transmembrane fragment structures of amyloid precursor protein depend on membrane surface curvature. J Am Chem Soc 2014;136:854–857.
36. Miyashita N, Straub JE, Thirumalai D, Sugita Y. Transmembrane structures of amyloid precursor protein dimer predicted by replica-exchange molecular dynamics simulations. J Am Chem Soc 2009; 131:3438–3439.
37. Pester O, Barrett PJ, Hornburg D, Hornburg P, Probstle R, Widmaier S, Kutzner C, Durrbaum M, Kapurniotu A, Sanders CR, Scharnagl C, Langosch D. The backbone dynamics of the amyloid precursor protein transmembrane helix provides a rationale for the sequential cleavage mechanism of gamma-secretase. J Am Chem Soc 2013;135:1317–1329.
38. Wang H, Barreyro L, Provasi D, Djemil I, Torres-Arancivia C, Filizola M, Ubarretxena-Belandia I. Molecular determinants and thermodynamics of the amyloid precursor protein transmembrane domain implicated in Alzheimer's disease. J Mol Biol 2011;408:879–895.
39. Song YL, Hustedt EJ, Brandon S, Sanders CR. Competition between homodimerization and cholesterol binding to the C99 domain of the amyloid precursor protein. Biochemistry 2013;52: 5051–5064.
40. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. Science 1997;276:131–133.
41. Nadezhdin KD, Bocharova OV, Bocharov EV, Arseniev AS. Dimeric structure of transmembrane domain of amyloid precursor protein in micellar environment. FEBS Lett 2012;586:1687–1692.
42. Chen W, Gamache E, Rosenman DJ, Xie J, Lopez MM, Li YM, Wang CY. Familial Alzheimer's mutations within APPTM increase A beta 42 production by enhancing accessibility of epsilon-cleavage site. Nat Commun 2014;5:3037.
43. Dell'Orco D, De Benedetti PG, Fanelli F. In silico screening of mutational effects on transmembrane helix dimerization: insights from rigid-body docking and molecular dynamics simulations. J Phys Chem B 2007;111:9114–9124.
44. Viswanath S, Ravikant DV, Elber R. DOCK/PIERR: web server for structure prediction of protein–protein complexes. Methods Mol Biol 2014;1137:199–207.
45. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. Proteins 2013;81:2082–2095.
46. Joachims T, Galor T, Elber R. Learning to align sequences: a maximal Margin Approach. In: Leimkuhler B, editor. New Algorithms for Macromolecular simulation. Berlin: Springer Verlag; 2005;57–69.
47. Kabsch W, Sander C. Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
48. Tusnady GE, Dosztanyi Z, Simon I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics 2005;21:1276–1277.
49. Tusnady GE, Dosztanyi Z, Simon I. PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res 2005;33:D275–D278.

50. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 2007;111:7812–7824.
51. Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li KY, Zheng JF, Vakili P, Paschalidis IC, Vajda S. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SOU, and stability analysis in CAPRI rounds 13-19. Proteins 2010; 78:3124–3130.
52. Pierce B, Weng ZP. ZRANK: reranking protein docking predictions with an optimized energy function. Proteins 2007;67:1078–1086.
53. Jayasinghe S, Hristova K, White SH. MPtopo: a database of membrane protein topology. Protein Sci 2001;10:455–458.
54. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25: 3389–3402.
55. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
56. Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, Stuart AC, Marti-Renom MA, Madhusudhan MS, Yerkovich B, Sali A. Tools for comparative protein structure modeling and analysis. Nucleic Acids Res 2003;31:3375–3380.
57. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 2005;33:2302–2309.
58. Elber R, Roitberg A, Simmerling C, Goldstein R, Li HY, Verkhivker G, Keasar C, Zhang J, Ulitsky A. Moil—a program for simulations of macromolecules. Comput Phys Commun 1995;91:159–189.
59. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J. The MARTINI coarse grained forcefield: extension to proteins. J Chem Theory Comput 2008;4:819–834.
60. Marrink SJ, de Vries AH, Tieleman DP. Lipids on the move: simulations of membrane pores, domains, stalks and curves. Biochim Biophys Acta 2009;1788:149–168.
61. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. Gromacs: fast, flexible, and free. J Comput Chem 2005;26:1701–1718.
62. Rotkiewicz P, Skolnick J. Fast procedure for reconstruction of full-atom protein models from reduced representations. J Comput Chem 2008;29:1460–1465.
63. Jo S, Kim T, Im W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. PloS One 2007;2:e880.
64. Jo S, Kim T, Iyer VG, Im W. Software news and updates—CHARNIM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 2008;29:1859–1865.
65. Cheng X, Jo S, Lee HS, Klauda JB, Im W. CHARMM-GUI micelle builder for pure/mixed micelle and protein/micelle complex systems. J Chem Inform Model 2013;53:2171–2180.
66. Wu EL, Cheng X, Jo S, Rui H, Song KC, Davila-Contreras EM, Qi Y, Lee J, Monje-Galvan V, Venable RM, Klauda JB, Im W. CHARMM-GUI membrane builder toward realistic biological membrane simulations. J Comput Chem 2014;35:1997–2004.
67. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
68. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I, MacKerell AD, Pastor RW. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. J Phys Chem B 2010;114:7830–7843.
69. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem 2013;34:2135–2145.
70. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103: 8577–8593.
71. Hoover WG. Canonical dynamics—equilibrium phase-space distributions. Phys Rev A 1985;31:1695–1697.
72. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997; 18:1463–1472.
73. Feig M, Karanicolas J, Brooks CL. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model 2004;22:377–395.
74. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm—a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
75. Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 2004;25:1400–1415.
76. Chen JH, Im WP, Brooks CL. Balancing solvation and intramolecular interactions: toward a consistent generalized born force field. J Am Chem Soc 2006;128:3728–3736.
77. Lensink M, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins 2010;78:3073–3084.
78. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at protein–protein interfaces. Proteins Struct Funct Genet 2001;43:89–102.
79. Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol 1997;267:707–726.
80. Mendez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins 2003;52:51–67.
81. Janin J. Assessing predictions of protein-protein interaction: the CAPRI experiment. Protein Sci 2005;14:278–283.
82. Dominguez L, Foster L, Meredith SC, Straub JE, Thirumalai D. Structural heterogeneity in transmembrane amyloid precursor protein homodimer is a consequence of environmental selection. J Am Chem Soc 2014;136:9619–9626.