Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane)

Gabriel Studer1,2, Marco Biasini1,2 and Torsten Schwede1,2,*

1Biozentrum, University of Basel, Basel, 4056, Switzerland and 2SIB Swiss Institute of Bioinformatics, Basel, 4056, Switzerland

ABSTRACT

Motivation: Membrane proteins are an important class of biological macromolecules involved in many cellular key processes including signalling and transport. They account for one third of genes in the human genome and >50% of current drug targets. Despite their importance, experimental structural data are sparse, resulting in high expectations for computational modelling tools to help fill this gap. However, as many empirical methods have been trained on experimental structural data, which is biased towards soluble globular proteins, their accuracy for transmembrane proteins is often limited.

Results: We developed a local model quality estimation method for membrane proteins ('QMEANBrane') by combining statistical potentials trained on membrane protein structures with a per-residue weighting scheme. The increasing number of available experimental membrane protein structures allowed us to train membrane-specific statistical potentials that approach statistical saturation. We show that reliable local quality estimation of membrane protein models is possible, thereby extending local quality estimation to these biologically relevant molecules.

Availability and implementation: Source code and datasets are available on request.

Contact: torsten.schwede@unibas.ch

Supplementary Information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Protein modelling plays a key role in exploring sequence structure relationships when experimental data are missing. Modelling techniques using evolutionary information, in particular homology/comparative modelling, developed into standardized pipelines over recent years. An indispensable ingredient of such a pipeline is the accuracy estimation of a protein model, directly providing the user with information regarding the range of its possible applications (Baker and Sali, 2001; Schwede, 2013; Schwede et al., 2009). In this context, global model quality assessment tools are important for selecting the best model among a set of alternatives, whereas local model estimates assess the plausibility and likely accuracy of individual amino acids (Benkert et al., 2011; Fasnacht et al., 2007). Various techniques have been developed to address this question, with consensus methods and knowledge-based approaches showing best results in blind assessments (Kryshtafovych et al., 2014). Consensus approaches require an ensemble of models with structural variety, reflecting alternative conformations (Roche et al., 2014; Skwark and Elofsson, 2013).

In contrast, knowledge-based methods (such as statistical potentials) can be applied to single models but are in general less accurate than consensus methods and exhibit strong dependency on the structural data they have been trained on.

The unique physicochemical properties of biological membranes give rise to interactions that are energetically discouraged in soluble proteins, and vice versa (White, 2009). However, most scoring functions using knowledge-based methods (Benkert et al., 2011; Lathy et al., 1992; Ray et al., 2012; Sippl, 1993; Zhou and Zhou, 2002) have been trained on soluble proteins. Thus, they perform poorly when applied to models of membrane proteins. This specific, but highly relevant, important aspect of protein model quality assessment has received only little attention in recent years (Heim and Li, 2012; Ray et al., 2010). With the growing amount of available high resolution membrane protein structures (Garman, 2014; White, 2004) the template situation for homology modelling procedures is improving quickly and, even more important for this work, it is gradually becoming possible to adapt knowledge-based methods to this class of models.

As a result of such efforts, we present QMEANBrane, a combination of statistical potentials targeted at local quality estimation of membrane protein models in their naturally occurring oligomeric state: after identifying the transmembrane region using an implicit solvation model, specifically trained statistical potentials get applied on the different regions of a protein model (Figs 1 and 2). To overcome statistical saturation problems, a novel approach for deriving statistical potentials from sparse training data has been devised. We have benchmarked the performance of the approach on a large heterogeneous test set of models and illustrate the result on the example of alignment errors in a transmembrane model.

2 METHODS

2.1 Target function

The similarity/difference between a model and a reference structure can be expressed in the form of distances between corresponding atoms in the model and its reference structure after performing a global superposition. However, this global superposition approach fails to give accurate results in case of domain movements. To overcome such problems, e.g. in the context of the CASP (Moult et al., 2014) experiments, the structures are manually split into so-called assessment units and evaluated separately (Taylor et al., 2014). This manual procedure is time consuming and not suitable for automate large-scale evaluation, e.g. such as performed by CAMEO (Haas et al., 2013). Alternatively, similarity/difference between a model and reference structure can be expressed in the form of superposition-free measures such as the local Distance Difference Test (iDDT) score (Mariani et al., 2013) assessing the differences in interatomic
2.2 Membrane segment definition

The OPM database (Orientations of Proteins in Membranes; Lomize et al., 2006a) applies minimization of a free energy expression to predict the transmembrane part of a protein structure. In this work, we use a similar but simplified approach, still resulting in a robust approximation of the membrane segment definition. The energy expression is defined as

\[
\Delta G = \sum \sigma^\text{wat} \cdot \text{bil} f(z_i) \cdot \text{ASA}_i
\]

with \(\sigma^\text{wat} \cdot \text{bil}\) representing the transfer energy from water to decadiene for atom \(i\) per \(\text{Å}^2\) (Lomize et al., 2004), \(f(z_i)\) the hydrophobicity as a function of the distance to the membrane centre \(z_i\) and \(\text{ASA}_i\), the accessible surface area of atom \(i\) in \(\text{Å}^2\) as calculated with NACCESS (www.bioinf.man.ac.uk/naccess). Not all surface-facing atoms, as determined by NACCESS are in contact with the membrane, even if they fall in between the lipid bilayer, e.g. as is the case for hydrophilic pores. To determine the subset of surface atoms in direct contact with the lipid bilayer, the protein structure surface as calculated by MSMS (Santer et al., 1996) is placed onto a 3D grid, marking every cube in the grid containing surface vertices. The application of a flood fill algorithm (http://lodev.org/cgtutor/floodfill.html) on every layer along the z-axis then allows the generation of a subset of potentially membrane facing atoms.

The parameters describing the membrane (i.e. tilt angle relative to z-axis, rotation angle around z-axis, membrane width and distance of membrane centre to origin) first undergo a coarse grained sampling to identify the 10 best parameter sets for further refinement using a Levenberg–Marquardt minimizer. This procedure is repeated several times with different initial orientations of the structure to find the set of parameters leading to the lowest total free energy.

The bilayer consists of a hydrocarbon core flanked by interface regions with a large chemical heterogeneity (White et al., 2001). It is known that the properties of a membrane protein are strongly influenced by the interaction with the phospholipid bilayer, and a simple split into a membrane and soluble part would not faithfully reflect the variation of molecular properties along the membrane axis (Bernsel et al., 2008). To catch these variations along the membrane axis, we split the transmembrane proteins into three parts, which are treated separately: an interface part consisting of all residues with their Ca atom positions within 5Å of the membrane defining planes, a core membrane part consisting of all residues with their Ca atom positions in between the two membrane defining planes not intersecting with the interface residues and finally, a soluble protein part consisting of all remaining residues.

2.3 Model quality predictors

To assess the membrane protein models quality, we mainly rely on statistical potential terms, combined with the relative solvent accessibility of each residue as calculated by DSSP (Kabsch and Sander, 1983). The four statistical potential terms (their exact parameterizations are described in the Supplementary Material), are the following:

1. **All-atom Interaction Term**: Pairwise interactions are considered between all chemically distinguishable heavy atoms. A sequence separation threshold has been introduced to allow focusing on long-range interactions and reduce the influence of local secondary structure. Interactions originating from atoms of residues closer in sequence than this threshold are neglected.

2. **Cβ Interaction Term**: This term assesses the overall fold by only considering pairwise interactions between Cβ positions of the 20 standard amino acids. In case of glycine, a representative of the Cβ position gets constructed using the backbone as anchor. The same sequence separation as in the all-atom interaction is applied.

3. **Solvation Term**: Statistics are created by counting surrounding atoms around all chemically distinguishable heavy atoms not belonging to the assessed residue itself.

4. **Torsion Term**: The central \(\phi/\psi\) angles of three consecutive amino acids are assessed based on the identity of the involved amino acids using a grouping scheme described by Solis and Rachovsky (Solis and Rackovsky, 2006).
 Assessing the local structural quality of transmembrane protein models

...
2.7 Per amino acid weighting scheme

QMEANBrane uses a linear model fitted on the per-residue lDDT score to combine the single quality predictors. To remove amino acid-specific biases, such a linear model is trained for every standard amino acid:

\[ s_i = \sum w_j S_{ij} \]

where \( s_i \) is the combined score of residue at position \( i \), \( w_j \) the weight of quality predictor \( j \) and \( S_{ij} \) the score of quality predictor \( j \) at position \( i \).

2.8 Implementation

QMEANBrane is designed on a modular basis, implementing computationally expensive tasks in a C++ layer. All functionality is made fully accessible from the Python language and can directly be embedded into the computational structural biology framework OpenStructure (Biasini et al., 2010, 2013), allowing to assemble custom assessment pipelines to address more specific requirements.

3 RESULTS AND DISCUSSION

3.1 Membrane prediction accuracy

To evaluate the performance of our membrane finding algorithm, a comparison with the result obtained by OPM has been performed on the 200 structures used for training of the membrane-specific statistical potentials. At this point, OPM is assumed to be the gold standard, even though it is a calculation by itself. By further considering the membrane width as the main feature of accuracy, 95% of the absolute width deviations are \(<4\) Å. In terms of translational distances, this corresponds to a ‘misprediction’ of 2–3 residues for helices and about 1–2 residue for sheets (Fig. 1). Interestingly, using this approach, it is not only possible to automatically detect transmembrane regions but also to distinguish between transmembrane and soluble structures in general (Supplementary Fig. S3).

3.2 Performance on the test dataset

For a first analysis of performance on predicting local scores of membrane-associated residues in transmembrane protein models, we used the previously described model set for training the linear weights. Clusters have been built by applying kClust on the target sequences with a sequence identity threshold of 30%. The local scores for the membrane-associated residues of one cluster have then been predicted using linear models trained on all residues from models not belonging to that particular cluster (Table 1; Supplementary Fig. S6).

3.3 Independent performance evaluation on models of the GPCR Dock experiments

Not many independent compilations of membrane protein models with known target structures exist. For a performance evaluation and comparison with other widely used quality assessment tools, we rely on the models generated during the GPCR Dock experiments 2008/2010 (Kufareva et al., 2011; Michino et al., 2009) (Fig. 2). A total of 491 models for three different targets, the human dopamine receptor, the human adenosine receptor and the human chemokine receptor were available. Receiver operating characteristic (ROC) analysis with the local lDDT as target value has been performed on all membrane-associated residues as defined by OPM, showing a clear superiority of QMEANBrane over other methods such as ProQ2 (Ray et al., 2012), QMEAN (Benkert et al., 2011), ProQM (Ray et al., 2010), Prosa (Wiederstein and Sippl, 2007), Verify3D (Luthy et al., 1992) or DFire (Zhou and Zhou, 2002) (Fig. 3). Removing all GPCR/Rhodopsin structures from the training data has only a minor effect. See Supplementary Figure S4 for a more detailed performance analysis taking other measures of similarity into account. Because ProQM is the only other method specifically developed for the particular case of membrane protein model quality assessment, we also performed a direct comparison of QMEANBrane and ProQM on the dataset used to test/train ProQM in Supplementary Figure S5.

3.4 Retrospective analysis of modelling examples

To illustrate the usefulness of QMEANBrane in tackling problems as they occur in real modelling cases, two targets with known structures have been selected for a more detailed analysis using the recently released SWISS-MODEL workspace (Biasini et al., 2014). The H+ translocating pyrophosphatase from Vigna...
radiata (PDB ID: 4A01) and a dopamine transporter of Drosophila melanogaster (PDB ID: 4M48). Models based on different target-template alignments have been compared to test QMEANBrane’s capability of detecting incorrect alignments, particularly alignment shifts in transmembrane helices. (Alignments are available in the Supplementary Materials.)

The pyrophosphatase has, with the sodium translocating pyrophosphatase from Thermotoga maritima (PDB ID: 4AV3), a rather close homologue (sequence identity >40%). Nevertheless, the alignments provided by BLAST (Altschul et al., 1990) and HHBlits differ significantly. Because the BLAST alignment has a lower coverage, not including the first transmembrane helix, only the part covered by both alignments is considered. Supplementary Figure S7 shows a comparison of the QMEANBrane scores from the two models built with the different alignments. Two transmembrane helices contain an alignment shift of three residues, resulting in a clear local increase of the QMEANBrane scores of the model built with the HHBlits alignment relative to the model built with the BLAST alignment. The higher quality of the HHBlits model gets confirmed by its global lDDT of 0.63 versus 0.59 of the BLAST model.

For the dopamine transporter example, we chose an amine transporter from Aquifex aeolicus VF5, identified by HHBlits with a sequence identity of ~24%, as the primary template. Despite the good coverage, a major problem occurs in transmembrane helix 5. The initial HHBlits alignment has an insertion of three residues enforcing a helix break and an unnatural bulge within the transmembrane part. To analyse possible modifications of the initial alignment, we rely on QMEANBrane to compare the relative differences in the models with alternative alignments with the initial model (Figs 4 and 5).

Three different alternative alignments were considered: the first is to shift the helix insertions towards the C-terminus. Despite the increase of the QMEANBrane score at the location of the alignment modification, the scores in helix 5 towards the C-terminus drop significantly, suggesting no improvement of the overall model quality. As second alternative, the insertion has been shifted into the loop connecting transmembrane helices 4 and 5. Because of their proximity, a distortion of both involved helix endings was inevitable, thus unfavourable. The third alternative, shift of the insertion towards the N-terminus before helix 4, and introducing an additional deletion in the aforementioned loop increasing the local sequence identity in helix 4, consistently increases the QMEANBrane scores in helices 4 and 5, as well as the helices close in space. These findings are confirmed by the global lDDT scores of the models built based on those alignments (initial alignment: 0.54, shift into middle: 0.54, shift towards C-terminus: 0.53, shift towards N-terminus: 0.57).

4 CONCLUSION

Investigating function and interactions in membrane proteins is an active field of research, with modelling techniques as an important tool to bridge the gap when structural data are missing. Comparative modelling methods automatically profit from the increased number of available experimental membrane structures, which can be used to build models for membrane proteins (Forrest et al., 2006). However, most knowledge-based
approaches fail in assigning reliable local quality estimates when confronted with the unique structural features and interactions resulting from direct contact with the phospholipid bilayer.

With QMEANBrane, we present a framework that widely covers the aspects of membrane protein model quality assessment. In a first step, our membrane detection method allows to reliably locate the transmembrane part of the model. We introduce an interface region to account for the non-isotropy of protein properties along the z-axis. Statistical potential terms were trained specifically for these three regions, introducing a new hybrid potential formalism to circumvent problems arising from a lack of sufficient training data. The final local scores are then calculated using linear models trained for all 20 standard amino acids. We could show a clear improvement in accuracy over widely used quality assessment methods when considering alpha-helical transmembrane structures. It is possible to detect errors introduced in the modelling procedure such as incorrect alignments, which would facilitate the visual exploration of alternative alignments, e.g. as suggested previously in (Barbato et al., 2012).

Despite similar observed overall performance for β-barrel structures, problems arise with shifted alignments, as they can occur when aligning sequences from remote homologues. The low number of pairwise atomic interactions in combination with the regular hydrophobicity pattern often observed in alignment shifts by two residues hamper the reliable detection of such errors and require further investigation.

ACKNOWLEDGEMENTS

We would like to thank all members of the Torsten Schwede group at the University of Basel for helpful discussions. We are grateful to Pascal Benkert for providing the training model dataset. Special thanks go to Jürgen Haas, Lorenza Bordoli and Tjaard de Beer for valuable inputs and helping in setting up the manuscript and Alessandro Barbato for his help in performing the performance analysis using different quality assessment methods. The authors gratefully acknowledge the computational resources provided by the sciCORE/BC1 Basel Computational Biology Center at the University of Basel and the support by the system administration team.

Funding: This work was supported by SIB Swiss Institute of Bioinformatics. G.S. has been supported by a PhD fellowship of the SIB by the ‘Swiss Foundation for Excellence and Talent in Biomedical Research’.

Conflict of interest: none declared.

REFERENCES

Mariani,V. et al. (2013) iDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics, 29, 2722–2728.
Assessing the local structural quality of transmembrane protein models