BIOINFORMATICS

Vol. 30 ECCB 2014, pages i505–i511
doi:10.1093/bioinformatics/btu457

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

Gabriel Studer[1,2], Marco Biasini[1,2] and Torsten Schwede[1,2,*]

[1]Biozentrum, University of Basel, Basel, 4056, Switzerland and [2]SIB 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).

*To whom correspondence should be addressed.

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; Luthy *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 (lDDT) score (Mariani *et al.*, 2013) assessing the differences in interatomic

© The Author 2014. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at INRIA Paris on September 7, 2014

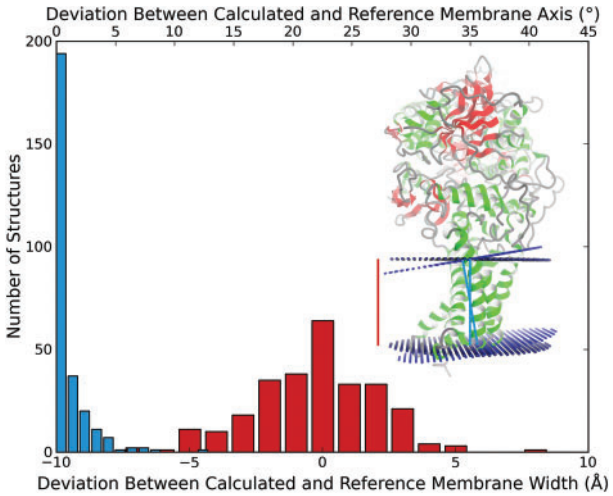


Fig. 1. Difference between membrane predictions of our algorithm and the predictions of OPM on the 200 high-resolution structures used to train membrane-specific statistical potentials

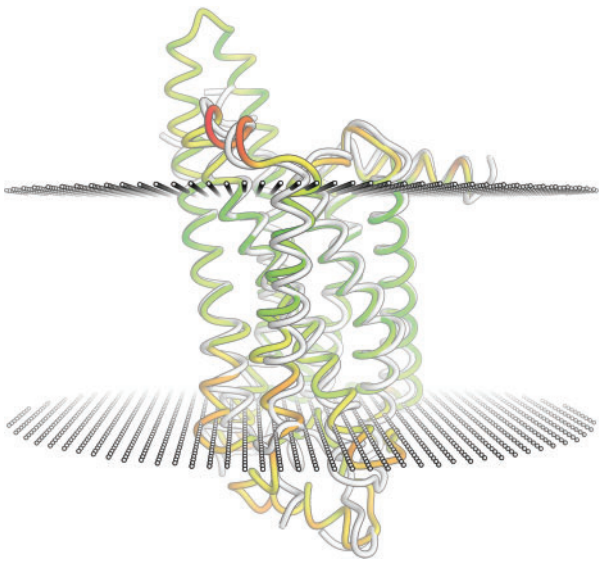


Fig. 2. Local QMEANBrane scores mapped on the best performing model (mod9jk) regarding RMSD of the GPCR Dock experiment 2008. Reference structure (2.6 Å crystal structure of a human A2A adenosine receptor bound to ZM241385, PDB: 3eml) and membrane-defining planes are shown in white

distances between model and reference structure. In this work, the lDDT inclusion radius is set to 10Å to ensure local behaviour. See Supplementary Figure S2 for a comparison of different structural similarity measures (Cα-distance, dRMSD, lDDT and CAD score; Olechnovic *et al.*, 2013).

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

$$\Delta G = \sum_i \sigma^{wat \to bil} f(z_i) ASA_i \quad (1)$$

with $\sigma^{wat \to bil}$ representing the transfer energy from water to decadiene for atom i per Å^2 (Lomize *et al.*, 2004), $f(z_i)$ the hydrophobicity as a function of the distance to the membrane centre z_i and ASA_i the accessible surface area of atom i in Å^2 as calculated with NACCESS (www.bioinf.manchester.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 (Sanner *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 Cα atom positions within 5Å of the membrane defining planes, a core membrane part consisting of all residues with their Cα 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 ϕ/ψ 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).

Downloaded from http://bioinformatics.oxfordjournals.org/ at INRIA Paris on September 7, 2014

The torsion term trained on soluble structures is applied to the whole membrane protein model. Conversely, solvation and interaction terms are specifically trained for and applied to the soluble, membrane and interface segments with different potentials for α-helical and β-barrel transmembrane structures. A residue belonging to one of these parts 'interacts' with all atoms in the full model, and a final score is assigned by averaging all scores originating from interactions associated with this specific residue. For the solvation and torsion terms, we use a formalism closely related to the statistical potentials of mean force (Sippl, 1990). However, instead of referring to an energy expression, we rather look at the problem as a log odds score between the probability of observing a particular interaction between partner s with conformation c relative to some reference state:

$$S(c|s) = -ln\left(\frac{p(c|s)}{p(c)}\right) \tag{2}$$

In case of sparse data, $p(c|s)$ cannot be expected to be saturated. Sippl and co-workers have proposed to use a combination of the extracted sequence-specific probability density function (pdf) and the reference state. The influence of the reference state vanishes at a rate determined by the newly introduced parameter σ towards large numbers of interactions (N) with sequence s:

$$p(c|s) \approx \frac{1}{1+N\sigma}p(c) + \frac{N\sigma}{1+N\sigma}p(c|s) \tag{3}$$

Using the aforementioned formalism, this leads to

$$S(c|s) \approx ln(1+N\sigma) - ln\left(1+N\sigma\frac{p(c|s)}{p(c)}\right) \tag{4}$$

Because of the increased abundance of structural information for soluble protein structures during the last decades, the use of the σ parameter has become largely unnecessary. However, for membrane proteins, data scarcity is still an issue and needs to be handled accordingly. In the Supplementary Materials, an analysis of the saturation behaviour of the different statistical potential terms is provided, suggesting a sufficient amount of training data for the solvation term, whereas the two interaction terms require more data to be fully saturated (Supplementary Fig. S1). For these cases, we introduced a treatment for sparse data by assuming that the statistics for soluble proteins are fully saturated. In other words, if there are no sufficient data available from membrane structures, we refer to the information we have from all protein structures to get a hybrid score:

$$\begin{aligned} HS(c|s) &= -\ln\left(\frac{1}{1+N\sigma}f_1 + \frac{N\sigma}{1+N\sigma}f_2\right) \\ &= \ln(1+N\sigma) - \ln(f_1 + N\sigma f_2) \end{aligned} \tag{5}$$

With f_1 representing the fraction of the probabilities of sequence-specific interactions and a reference state, where the pdfs of the specific interactions are saturated, and f_2 the fraction between the probabilities of sequence-specific interactions and a reference state, where the pdfs of the specific interactions are not necessarily saturated, as it may occur for membrane- and interface-specific cases.

For regions of the pdf with zero probability as they, for example, occur at low distances in pairwise interaction terms, we applied a constant cap value to avoid infinite scores.

2.4 Training datasets for statistical potentials

The pdfs to calculate the statistical potentials for the soluble part are built using statistics extracted from a non-redundant set of high resolution X-ray structures. PISCES (Wang and Dunbrack, 2003) has been used with the following parameters: sequence identity threshold 20%, resolution threshold 2 Å and R-factor threshold 0.25. Because only standard amino acids can be handled by QMEANBrane, a prior curation of the training structures is necessary. Non-standard amino acids such as phospho-serine or seleno-methionine have therefore been mapped to their standard parent residues. For the selection of appropriate membrane protein structures, we rely on the OPM database (Lomize *et al.*, 2006b). As of October 2013, OPM contained 746 unique PDB IDs of structures with transmembrane segments. Applying a resolution threshold of 2.5 Å, removing all chains with <30 membrane-associated residues and considering only one chain in case of homo-oligomers results in 283 remaining chains from 200 structures. Clustering the chains based on their SEQRES sequences with kClust (Hauser *et al.*, 2013) using a sequence identity threshold of 30% resulted in 187 clusters, 140 of them from helical transmembrane structures and 47 from β-barrel structures. All entries are used in the calculation of the pdfs, where a chain originating from a cluster with n members is downweighted and contributes with a weight of $1/n$ to the final distributions. These final distributions have then been extracted by considering the corresponding chains, using the full protein structure in the oligomeric state as assigned by OPM as environment.

2.5 Datasets for training linear combinations

A set of 3745 models for soluble proteins was generated by selecting a set of non-redundant high-resolution reference structures from the PDB using PISCES (maximum 20% sequence identity, resolution better 2Å, X-ray only), extracting their amino acid sequences, and building models using the automated SWISS-MODEL pipeline (Kiefer *et al.*, 2009) by excluding templates with a sequence identity >90% to the target (P. Benkert, personal communication). OPM was used to identify reference structures (resolution <3.0 Å) to generate membrane protein models. Structures with <30 membrane-associated residues and hetero-oligomeric complexes were excluded. In all, 132 unique PDB IDs, which had more than one suitable template, have been selected as targets for modelling. Templates identified with HHBlits (Remmert *et al.*, 2012) showing a sequence alignment coverage >50% served as input for MODELLER (Sali and Blundell, 1993) and resulted in 3226 models with oligomeric states equivalent to the template structure. Removal of redundancy, i.e. models originating from templates with same sequence, and removal of obvious incorrect oligomeric states upon visual inspection resulted in a set of 557 models, 386 with helical transmembrane parts and 171 β-barrels.

2.6 Spherical smoothing for noise reduction

Averaging/smoothing can reduce noise introduced by quality predictors on a per-residue level, resulting in single residue scores, which more accurately reflect the local model quality. Smoothing in space tends to outperform sequential smoothing. In the proposed algorithm, every residue gets represented by its Cα position. The final quality predictor score for a residue is calculated as a weighted mean of its own value and the values associated to surrounding residues:

$$s_i = \sum_j w_j s_j \tag{6}$$

with s_i representing the final score at position i, w_j the weight of score at position j and s_j the score at position j. The weights are calculated in a Gaussian-like manner and normalized, so they sum up to one:

$$w_j = \begin{cases} \dfrac{1}{Z\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}\left(\frac{d_{ij}}{\sigma}\right)^2} & d_{ij} < 3\sigma \\ 0 & else \end{cases} \tag{7}$$

with w_j representing the weight of score at position j, d_{ij} the distance from position i to position j, σ the standard deviation of the Gaussian-like formalism to control how fast the influence of a neighbouring score vanishes as a function of the distance (5 Å turned out to be a reasonable σ) and Z as normalization factor.

Downloaded from http://bioinformatics.oxfordjournals.org/ at INRIA Paris on September 7, 2014

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:

$$s_i = \sum_j w_j s_{ij} \quad (8)$$

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.

Table 1. Performances of single quality predictors and their combination on membrane-associated residues in our test set, measured as Pearsons' r between predicted score and actual local lDDT

Quality predictor	Helical structures	β-barrel structures
Exposed	0.39	0.15
Torsion	0.43	0.47
Cβ interaction	0.51	0.49
Solvation	0.55	0.51
All atom interaction	0.63	0.58
All predictors combined	0.71	0.67

Note: Even for single predictors, an amino acid-specific linear model has been trained to remove amino acid-specific biases.

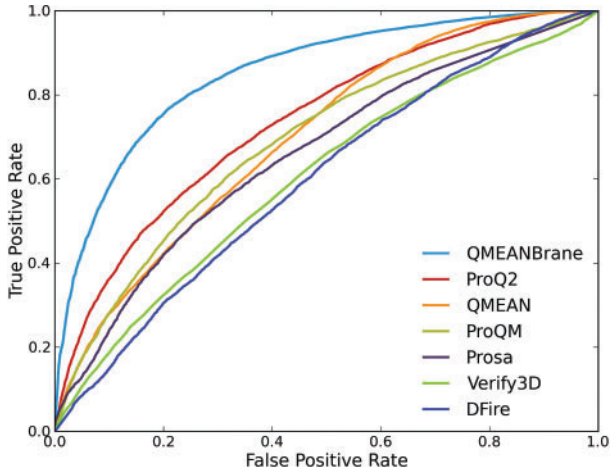


Fig. 3. ROC analysis of all membrane-associated residues of the models of the GPCR Dock experiments with local lDDT as target value and a class cutoff of 0.6

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*

Downloaded from http://bioinformatics.oxfordjournals.org/ at INRIA Paris on September 7, 2014

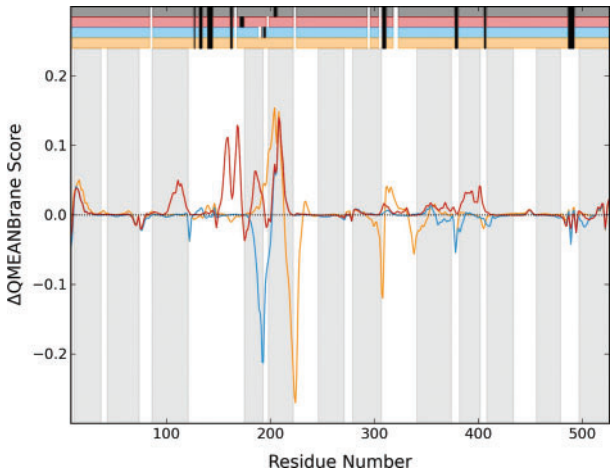


Fig. 4. Difference of QMEANBrane scores of three dopamine transporter models with modified alignments versus the model built with the initial HHBlits alignment, represented by the first horizontal bar. Insertions are marked black, and deletions are marked white. Second bar: shift of the insertion towards the N-terminus in front of helix 4, third bar: shift of insertion towards the N-terminus in between helices 4 and 5, fourth bar: shift of the insertion towards the C-terminus

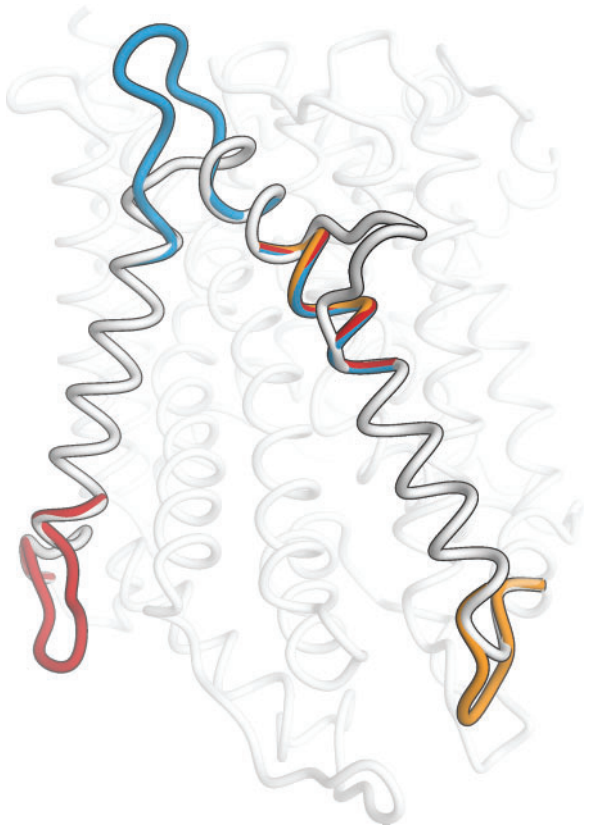


Fig. 5. Structural effects of the alignment modifications shown in Figure 4. The model based on the initial HHBlits alignment is coloured white; the other models are coloured according to the horizontal bar alignment representation in Figure 4

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

Downloaded from http://bioinformatics.oxfordjournals.org/ at INRIA Paris on September 7, 2014

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 Tjaart 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/[BC]^2 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

Altschul,S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Baker,D. and Sali,A. (2001) Protein structure prediction and structural genomics. *Science*, **294**, 93–96.

Barbato,A. *et al.* (2012) Improving your target-template alignment with MODalign. *Bioinformatics*, **28**, 1038–1039.

Benkert,P. *et al.* (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, **27**, 343–350.

Bernsel,A. *et al.* (2008) Prediction of membrane-protein topology from first principles. *Proc Natl Acad Sci USA*, **105**, 7177–7181.

Biasini,M. *et al.* (2010) OpenStructure: a flexible software framework for computational structural biology. *Bioinformatics*, **26**, 2626–2628.

Biasini,M. *et al.* (2013) OpenStructure: an integrated software framework for computational structural biology. *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 701–709.

Biasini,M. *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252–W258.

Fasnacht,M. *et al.* (2007) Local quality assessment in homology models using statistical potentials and support vector machines. *Protein Sci.*, **16**, 1557–1568.

Forrest,L.R. *et al.* (2006) On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys. J.*, **91**, 508–517.

Garman,E.F. (2014) Developments in x-ray crystallographic structure determination of biological macromolecules. *Science*, **343**, 1102–1108.

Haas,J. *et al.* (2013) The protein model portal—a comprehensive resource for protein structure and model information. *Database (Oxford)*, **2013**, bat031.

Hauser,M. *et al.* (2013) kClust: fast and sensitive clustering of large protein sequence databases. *BMC Bioinformatics*, **14**, 248.

Heim,A.J. and Li,Z. (2012) Developing a high-quality scoring function for membrane protein structures based on specific inter-residue interactions. *J. Comput. Aided Mol. Des.*, **26**, 301–309.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kiefer,F. *et al.* (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.*, **37**, D387–D392.

Kryshtafovych,A. *et al.* (2014) Assessment of the assessment: evaluation of the model quality estimates in CASP10. *Proteins*, **82** (**Suppl. 2**), 112–126.

Kufareva,I. *et al.* (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure*, **19**, 1108–1126.

Lomize,A.L. *et al.* (2004) Quantification of helix-helix binding affinities in micelles and lipid bilayers. *Protein Sci.*, **13**, 2600–2612.

Lomize,A.L. *et al.* (2006a) Positioning of proteins in membranes: a computational approach. *Protein Sci.*, **15**, 1318–1333.

Lomize,M.A. *et al.* (2006b) OPM: Orientations of Proteins in Membranes database. *Bioinformatics*, **22**, 623–625.

Luthy,R. *et al.* (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83–85.

Mariani,V. *et al.* (2013) lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, **29**, 2722–2728.

Michino,M. *et al.* (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.*, **8**, 455–463.

Moult,J. *et al.* (2014) Critical assessment of methods of protein structure prediction (CASP)—round x. *Proteins*, **82** (**Suppl. 2**), 1–6.

Olechnovic,K. *et al.* (2013) CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins*, **81**, 149–162.

Ray,A. *et al.* (2010) Model quality assessment for membrane proteins. *Bioinformatics*, **26**, 3067–3074.

Ray,A. *et al.* (2012) Improved model quality assessment using ProQ2. *BMC Bioinformatics*, **13**, 224.

Remmert,M. *et al.* (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, **9**, 173–175.

Roche,D.B. *et al.* (2014) Assessing the quality of modelled 3D protein structures using the ModFOLD server. *Methods Mol. Biol.*, **1137**, 83–103.

Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

Sanner,M.F. *et al.* (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305–320.

Schwede,T. (2013) Protein modeling: what happened to the "protein structure gap"? *Structure*, **21**, 1531–1540.

Schwede,T. *et al.* (2009) Outcome of a workshop on applications of protein models in biomedical research. *Structure*, **17**, 151–159.

Sippl,M.J. (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.*, **213**, 859–883.

Sippl,M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins*, **17**, 355–362.

Skwark,M.J. and Elofsson,A. (2013) PconsD: ultra rapid, accurate model quality assessment for protein structure prediction. *Bioinformatics*, **29**, 1817–1818.

Solis,A.D. and Rackovsky,S. (2006) Improvement of statistical potentials and threading score functions using information maximization. *Proteins*, **62**, 892–908.

Downloaded from http://bioinformatics.oxfordjournals.org/ at INRIA Paris on September 7, 2014

Taylor,T.J. *et al.* (2014) Definition and classification of evaluation units for CASP10. *Proteins*, **82** (**Suppl. 2**), 14–25.

Wang,G. and Dunbrack,R.L. Jr (2003) PISCES: a protein sequence culling server. *Bioinformatics*, **19**, 1589–1591.

White,S.H. (2004) The progress of membrane protein structure determination. *Protein Sci.*, **13**, 1948–1949.

White,S.H. (2009) Biophysical dissection of membrane proteins. *Nature*, **459**, 344–346.

White,S.H. *et al.* (2001) How membranes shape protein structure. *J. Biol. Chem.*, **276**, 32395–32398.

Wiederstein,M. and Sippl,M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, **35**, W407–W410.

Zhou,H. and Zhou,Y. (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.*, **11**, 2714–2726.

Downloaded from http://bioinformatics.oxfordjournals.org/ at INRIA Paris on September 7, 2014