BIOINFORMATICS ORIGINAL PAPER

Vol. 29 no. 1 2013, pages 54–61
doi:10.1093/bioinformatics/bts640

Structural bioinformatics

Advance Access publication October 30, 2012

MP-T: improving membrane protein alignment for structure prediction

Jamie R. Hill and Charlotte M. Deane*

Department of Statistics, University of Oxford, Oxford, OX1 3TG, UK

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Membrane proteins are clinically relevant, yet their crystal structures are rare. Models of membrane proteins are typically built from template structures with low sequence identity to the target sequence, using a sequence-structure alignment as a blueprint. This alignment is usually made with programs designed for use on soluble proteins. Biological membranes have layers of varying hydrophobicity, and membrane proteins have different amino-acid substitution preferences from their soluble counterparts. Here we include these factors into an alignment method to improve alignments and consequently improve membrane protein models.

Results: We developed Membrane Protein Threader (MP-T), a sequence-structure alignment tool for membrane proteins based on multiple sequence alignment. Alignment accuracy is tested against seven other alignment methods over 165 non-redundant alignments of membrane proteins. MP-T produces more accurate alignments than all other methods tested (δF_M from +0.9 to +5.5%). Alignments generated by MP-T also lead to significantly better models than those of the best alternative alignment tool (one-fourth of models see an increase in GDT_TS of $\geq$4%).

Availability: All source code, alignments and models are available at http://www.stats.ox.ac.uk/proteins/resources

Contact: deane@stats.ox.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on July 7, 2012; revised on September 26, 2012; accepted on October 22, 2012

1 INTRODUCTION

There are only ~350 unique membrane protein crystal structures (http://www.blanco.biomol.uci.edu/mpstruc). For comparison, in July 2012 there were >45 000 unique crystal structures in the Protein Data Bank (Berman *et al.*, 2000). This disparity is not due to importance: the location of membrane proteins on the surface of cells and their involvement in signalling pathways make them targets for approximately one-half of current and future drugs (Overington *et al.*, 2006). The disparity arises because membrane proteins are inherently hard to crystallize—the hydrophilic conditions required for crystallization are often incompatible with the hydrophobic conditions in which a membrane protein assumes its native structure.

In the absence of crystal structures, a model of a target sequence may be built using as a template a related protein of known structure (homology modelling). Conceptually, homology modelling can be divided into three phases: (i) identifying a template ('fold recognition'), (ii) aligning the residues of the target on to the template structure, (iii) completion of the model implicit from this alignment (e.g. by modelling missing loops). Many methods perform fold-recognition and alignment simultaneously. In this work, we maintain a distinction and focus solely on alignment.

Alignment methods differ in how they use information about protein structure and sequence. Sequence information is derived from a set of sequences that are homologous to the target and template. These sequences are often combined into a profile. Structure information is commonly incorporated either by the use of statistical potentials (e.g. Yang *et al.*, 2011) or by annotation of the template sequence, e.g. with secondary structure strings (Shi *et al.*, 2001).

The alignment of membrane proteins presents unique challenges. Most soluble proteins are globular with a hydrophilic surface, whereas membrane proteins possess neither of these properties. Alignment methods incorporating statistical potentials derived from soluble proteins are thus expected to perform poorly on membrane proteins. At least two such methods, pGenThreader (Lobley *et al.*, 2009) and SPARKS-X (Yang *et al.*, 2011), recognize this and filter transmembrane sections from their results.

Alignment methods that annotate a template include HHsearch (Söding, 2005) and PROMALS (Pei and Grishin, 2007). These align pairs of profiles that are annotated with secondary structure strings. PROMALS may also be used as a homology-extension method. In this approach, a multiple sequence alignment (MSA) is performed, but with each sequence replaced by a profile.

Most MSA methods, such as MAFFT (Katoh and Toh, 2008), MSAProbs (Liu *et al.*, 2010) and T-Coffee (Notredame *et al.*, 2000) do not use any structure information, but instead derive their accuracy from a 'consistency' criterion and/or iterative optimization. Consistency-based approaches aim to generate a MSA that accords best with a library of pairwise alignments between the sequences being aligned.

A small number of alignment methods have been designed specifically for membrane proteins. The AlignMe program has been developed to study LeuT-fold transporters (Khafizov *et al.*, 2010). Praline^{TM} (Pirovano *et al.*, 2008) and TM-Coffee (Chang *et al.*, 2012) are homology-extension (multiple profile alignment) methods that have been found to perform well on alignments of transmembrane proteins from the BALiBASE benchmark. Neither method uses secondary structure to aid alignment, but

*To whom correspondence should be addressed.

Praline^{TM} uses different scoring in regions annotated as being transmembrane.

Owing to the scarcity of known structures, the best template for a membrane protein sequence is likely to have extremely low sequence identity—potentially beneath the ~30% 'twilight zone'. Fortunately, even at this low level of sequence identity, accurate alignment may be possible, as biological membranes have a sandwich structure with a hydrophobic middle (lipid tail layer) and hydrophilic edges (lipid head layer). These differing hydrophobicities constrain the amino acids likely to be found in each environment (Forrest *et al*., 2006).

In previous work, we demonstrated that environment-specific substitution tables improved the pairwise alignment of membrane proteins (Hill *et al*., 2011). Here, we incorporate these tables into a consistency-based MSA program, Membrane Protein Threader (MP-T). The evolutionary information provided by homologous sequences greatly increases alignment accuracy at low sequence identities compared with pairwise alignments.

Our approach involves no sequence weighting or iterative refinement, and yet it is as accurate as leading profile–profile aligners such as HHsearch (Söding, 2005) and PROMALS (Pei and Grishin, 2007) while introducing fewer misaligned pairs of residues. MP-T leads to models with significantly higher GDT_TS (Zemla *et al*., 2001) than the best alternative alignment method tested—one-fourth of models see an increase in GDT_TS of at least 4%.

2 APPROACH

We develop MP-T, a progressive multiple alignment method specifically for use on membrane protein sequences. Sequences homologous to the target and template are collected and used to construct a phylogenetic 'guide' tree. The sequences are then aligned in the order dictated by the branching of the tree, working from the leaves and aligning towards the root.

In previous work, we performed pairwise alignment of a sequence on to a template (Hill *et al*., 2011). Our pairwise alignment method used structural information by annotating residues in the template with the region of the membrane that they contacted, their secondary structure type and their accessible surface area. Different scoring systems were used for each combination of annotations (hereafter termed an 'environment') and gap penalties were also varied with environment as, for example, a gap is unlikely to be opened in the middle of a transmembrane helix.

This pairwise method cannot be translated directly into a MSA method as only one sequence—the template—has the structural annotation required for accurate alignments. To remedy this, pairwise alignments are made between the template sequence and each homologue, and structural annotation is transferred between aligned pairs of residues. After annotation transfer, a guide tree is constructed and used to select homologues for the multiple-alignment phase: only sequences judged by the guide tree to be descendants of the most recent common ancestor of the target and template are selected. Multiple-alignment then proceeds using our implementation of the T-Coffee objective criterion (Notredame *et al*., 2000). The criterion requires a library of pairwise alignments, which we construct by applying our pairwise alignment method to all pairs of selected homologues.

The above-outlined approach merges our accurate pairwise aligner into a consistency-based membrane multiple sequence aligner. The approach uses phylogeny to exclude spurious homology information from the alignment.

3 METHODS

3.1 Scoring system of our method

Our approach requires each template to be annotated with environment information. To do this we used iMembrane (Kelm *et al*., 2009), which determines membrane positioning based on coarse-grained molecular-dynamics simulations available in the CGDB database (Scott *et al*., 2008).

Substitution tables were built for each environment by counting substitutions between 416 annotated structures and a large number of sequence homologues in a process analogous to that used in the construction of BLOSUM matrices (Henikoff and Henikoff, 1992). MAFFT L-INS-i (Katoh and Toh, 2008) was used to align each structure with its homologues. In contrast to our previous work, substitutions for an environment were counted symmetrically, e.g. a substitution from alanine in a helix structure to glycine in a sequence $A_{helix} \rightarrow G_{sequence}$, implies a substitution $G_{helix} \rightarrow A_{sequence}$ even though the glycine could potentially not be in a helix. This change ensures consistency with our transfer of environment annotations (see Section 2) between the template and each sequence. The substitution tables used here are available at http://www.stats.ox.ac.uk/proteins/resources.

During multiple sequence alignment, no gap penalties were imposed other than a terminal gap penalty of −5. In constructing the pairwise library, a terminal gap penalty of −50 was used in addition to four pairs of affine gap penalties that depended on whether a gap occurred in the tail layer of the membrane or in a region of known secondary structure. For gaps in the tail layer disrupting secondary structure, $g_{open} = -360$, $g_{ext} = -60$; for other gaps in the tail layer, $g_{open} = -220$, $g_{ext} = -20$; for other gaps disrupting secondary structure, $g_{open} = -300$, $g_{ext} = -20$; all other gaps, $g_{open} = -200$, $g_{ext} = -10$. Gap penalty values were non-exhaustively optimized on the 72 pairwise alignments in Hill *et al*. (2011), by testing perturbations around plausible values. Gap penalties are most severe in transmembrane regions and in secondary structure elements. This agrees with the high conservation of secondary structure elements, and with the fact that transmembrane elements are constrained to be at least the width of the membrane.

3.2 Construction of training and test sets

Sequences were extracted from proteins with PDB codes accessible at the Membrane Proteins of Known 3D Structures Database on 10 January 2012. The sequences were ordered as follows:

(1) Sequences that formed the gap-penalty training set
(2) Sequences present in the CGDB molecular dynamics database
(3) Sequences from X-ray structures (sorted by resolution)
(4) Sequences from NMR structures

As the list of sequences was traversed in the above order, a sequence was 'accepted' if it had <60% sequence identity [as determined by a TM-align (Zhang and Skolnick, 2005) structural alignment] to all previously accepted sequences.

Pairwise structure alignments were made between all pairings of accepted sequences. These alignments were then filtered such that no alignments had RMSDs of ≥3.5 Å, had TM scores of <0.5, contained a sequence with length <60 residues, contained a co-crystallized soluble protein fragment or were aligned over <60% of the length of the shorter

sequence. Alignments involving sequences without an iMembrane annotation were discarded. This procedure prevents bias in the selected sequences, and ensures alignments are between proteins with sufficient structural similarity for use in modelling.

The alignments were then partitioned into a training set and a test set. The training set was composed of 73 alignments that satisfied these requirements and where both sequences were taken from the first category of the above list. The test set consisted of 165 alignments where neither sequence was part of the first category. A schematic of this procedure is provided in Supplementary Figure S1. Properties of the training and test sets are plotted in Supplementary Figure S2. All alignments can be found at http://www.stats.ox.ac.uk/proteins/resources.

3.3 Alignment input

Alignments were made with the profile–profile alignment programs PROMALS (Pei and Grishin, 2007) and HHsearch v2.0.5 (Söding, 2005), and with the MSA programs MP-T (this work), MUSCLE v3.8.31 (Edgar, 2004), MSAProbs v0.9.5 (Liu *et al.*, 2010), MAFFT v6.864 *b*, T-Coffee v9.01 (Notredame *et al.*, 2000) and clustalΩ v1.0.3 (Sievers *et al.*, 2011). We allowed both profile–profile programs to select their own homologues (see next section), but tested all MSA programs on identical input sequences, which were collected as described below.

Homologous sequences to each chain were extracted from the Uniref90 database (accessed 6 February 2012) (Suzek *et al.*, 2007) by running PSI-BLAST (Altschul *et al.*, 1997) for five iterations with e-value thresholds of 10^{-3} to keep a hit, and 10^{-5} to incorporate a hit into the search profile. Homologues with <15% sequence identity to the query over the aligned region were discarded, as were those with lengths <2/3 or >3/2 that of the query. The remaining homologues were randomly ordered and then made non-redundant by UCLUST (Edgar, 2010) at 80% sequence identity.

For each alignment, the homologues from the template and target were combined in equal numbers (alternating the order to first take a homologue from the template, then one from the target with any surplus being discarded), filtered to have lengths >2/3 and <3/2 that of the template, and again made redundant at 80% using UCLUST. The first 125 sequences were used as input to each aligner. This is similar to the way in which the general purpose PREFAB benchmark is constructed, but with a greater number of selected sequences (PREFAB alignments have at most 50 sequences) and different cut-offs (Edgar, 2004).

This protocol ensures that the MSA contains a reasonable number of sequences with little bias towards either the target or template, a wide range of sequence identity and sequences of approximately equal length.

3.4 Optimization of alignment programs

We attempted to find optimal settings for each alignment method based on its performance on our training set.

The default settings of MUSCLE and MSAProbs were found to perform well on the training set and so no changes were made. We tested the L-INS-i and G-INS-i modes of MAFFT and used the former as it had a higher accuracy. We tested the PSI-Coffee mode of T-Coffee (with the Uniref90 database for sequence search), as we found it to be more accurate than the default T-Coffee settings, while providing results directly comparable with those of TM-Coffee. By default, ClustalΩ constructs trees using a variant of the fast mBed algorithm and does not perform iteration. We found two rounds of iteration, and non-mBed tree construction improved performance (options: –full –full-iter –iter 2) and so ran these settings.

Although PROMALS (Pei and Grishin, 2007) is capable of performing a multiple profile alignment, we found that a pairwise profile–profile alignment was more accurate on our training set. PROMALS comes with its own sequence database (Uniref90 dated Feb 2007) against which it finds homologues using its own copy of BLAST.

HHsearch (Söding, 2005) is part of HHsuite v. 2.0.5. The suite contains the HHblits search tool (Remmert *et al.*, 2012), which we used to generate HMMs for target and template. Searches were conducted against the nr20 database (version dated 11 January 2010). Both HMMs were annotated by PSIPRED v. 2.5 (Jones, 1999), and the template was further annotated with DSSP secondary structure states by JOY (Mizuguchi *et al.*, 1998). Best performance on the training set was obtained by combining a local initial Viterbi algorithm with a global MAC realignment.

We note that HHsearch and PROMALS enjoy the advantages of secondary structure annotation and homologue selection aimed at optimizing their performance. Our method, MP-T, also has these advantages. PSI-Coffee is a homology extension method and so uses more sequences in each alignment than are available to other methods. MSAProbs, MAFFT L-INS-i, MUSCLE and ClustalΩ are directly comparable with each other.

3.5 Assessment of alignments and models

Alignment accuracy was assessed with reference to a TM-align structure alignment between the target and template. Two separate assessments were made: one over the transmembrane domain only, and the other over the entire sequence. Transmembrane domains were defined, using the 'membrane layer' annotation from iMembrane, to be regions that contain <30 consecutive not-in-membrane layer residues, >15 tail layer residues and are at least 40-residues long. Putative domains were trimmed to begin and end with at most 15 consecutive not-in-membrane layer residues. In the test set, 115 of the 165 alignments contained TM domains: most of the other alignments form soluble parts of TM complexes.

Alignment accuracy is commonly assessed by two related measures: the modeller score F_M and the developer score F_D. These scores are different normalizations of the number of correctly aligned residues in an alignment. In F_M, this number is divided by the number of aligned residues in the alignment being assessed; in F_D, it is divided by the number of aligned residues in the reference alignment (Sauder *et al.*, 2000). An alignment with high F_M includes few mistakenly aligned residues, but may not reproduce much of the reference alignment; one with high F_D reproduces a lot of the reference alignment but may misalign residues.

A third measure of alignment accuracy is the number of alignments where one method aligns 10 residues more correctly than another. Compared with F_D and F_M, this measure avoids the problems of normalization and is robust against anomalously good or bad alignments.

Models of transmembrane domains were built with MEDELLER (Kelm *et al.*, 2010) and assessed using the GDT_TS measure calculated by the TM-score program (Zhang and Skolnick, 2004) over aligned pairs of residues. GDT_TS is the average fraction of residues in the model that are within 1, 2, 4 and 8 Å of their position in the experimental structure.

4 DISCUSSION

4.1 Homologue selection

It is unclear how homologues should be chosen to help create a good sequence alignment between two distantly related proteins. In this section, we discuss variations on the procedure used throughout the rest of this work, which is described in Section 3.3.

MSA accuracy is thought to deteriorate as the number of aligned sequences increases above ~100 (Thompson *et al.*, 2011), but thousands of homologues can be returned in a BLAST search. We tested MP-T on our training set with 25–150 homologues in steps of 25, and found that the accuracy tended to a fixed upper limit in this range, with little improvement from 75–150 homologues (percentage of the reference structure alignment that was reproduced, F_D: 86.1–87.2, see Supplementary Fig. S4).

A feature of our homologue selection procedure is randomization. One alternative approach is to order homologues by their BLAST score, favouring more nearly related sequences. The percentages of the reference alignment reproduced when homologues were selected by BLAST score were 82.9–86.4 for 25–150 homologues in steps of 25. For MP-T–randomized selections always yield improved accuracy, suggesting that homologues should be diverse and spread over the entire range of sequence identity.

4.2 Tree building

Progressive multiple alignment requires a 'guide tree' to determine the order of alignment. Counterintuitively, simple tree-building heuristics such as single-linkage clustering and UPGMA have been found to lead to more accurate alignments than phylogenetically more precise methods such as neighbour joining (NJ) (Wheeler and Kececioglu, 2007; Plyusnin and Holm, 2012).

In MP-T, tree building determines not just the order in which sequences are aligned, but also which sequences are used to align the target and template. For example, a tree with the target and template on adjacent leaves would give a standard pairwise alignment, whereas a tree with the target and template meeting at the root would involve all other sequences. We tested the tree-building procedures NJ, BIONJ (Gascuel, 1997), single-linkage clustering and UPGMA on our training set to determine whether simpler algorithms still led to more accurate alignments in this case.

The algorithms differed in how many sequences they used to align the target and template. For BIONJ and NJ, on average ~70% of the possible number of homologues were used; for UPGMA and single-linkage clustering, ~90% of homologues were used. Despite this, changes in tree building did not greatly affect accuracy. Single-linkage clustering appeared to be most accurate and NJ least accurate. We use UPGMA in the remainder of this work.

4.3 Alignment accuracy

MP-T alignments have a significantly higher average F_M (lower error rate, see Section 3.5) than all other methods, both over the transmembrane domain and the full alignment (Table 1). Little variation is seen between the top methods in F_D, the fraction of the reference structure alignment that is reproduced, but methods that make use of secondary structure annotation (MP-T, PROMALS and HHsearch) achieve higher scores over the TM domain. In addition to the methods discussed in Section 3.3, Table 1 contains results for a version of MP-T that uses no environment information [*MP-T (1 table)*], and a version that uses no homologous sequences [*MP-T (pair)*].

The small variation in F_D over full alignments allows the intrinsic error rates of the different methods to be compared. The low error rates of MP-T appear to be a consequence of the T-Coffee objective criterion—use of a single environment-independent substitution table in MP-T drastically reduces performance but still leads to fewer incorrect pairings than for other methods [MP-T (1 table), Table 1]. After MUSCLE, the two profile–profile methods PROMALS and HHsearch incorrectly align the most residues. We find that, in general, MSA programs incorrectly align a smaller proportion of residues between the target and template than profile–profile methods. This is probably because in a MSA, there are more ways of incorrectly pairing residues without matching a residue in the target with one in the template.

Table 1. Performance of different methods over the transmembrane domain and full alignment

	TM domain		Full alignment	
Method	F_M	F_D	F_M	F_D
MP-T	66.3	65.8	69.6	70.1
MP-T (1 table)	65.3	63.7*	69.2	68.5*
MP-T (pair)	59.1*	60.5*	61.5*	63.8*
MSAProbs	62.5**	62.1**	68.2**	69.0
MAFFT L-INS-I	64.1*	63.8*	68.4**	69.1
HHsearch	64.1*	65.3	67.5*	69.6
PROMALS	64.4*	66.0	67.4*	69.9
PSI-Coffee	65.4**	64.6**	68.7**	69.1
MUSCLE	60.8*	61.3*	65.1*	66.6*
clustalΩ	63.9[a],*	63.8**	67.9[a],*	68.5*

MP-T has significantly higher values of F_M than all other methods. The top-scoring methods differ little in F_D.
[a]In 2 cases, clustalΩ does not align any residues in the target with those in the template. We set F_M for these alignments to 0. Entries marked with asterisks are significantly different from MP-T ($P<0.01$ are marked *, $P<0.1$ are marked **). Significance is assessed by a Wilcoxon signed-rank test.

For 15 alignments, no method aligned more than 25% of the residues correctly. These 'unalignable' proteins typically have low sequence identity (average 7%) and are either short or have the majority of their residues in β-strands (eight are shorter than 150 residues; five are majority β-strand). The short proteins paired here may not be homologous, but merely share a large secondary structure feature. The over-representation of β-strands may highlight difficulties in deriving a reference alignment from a β-barrel structure alignment. For example, models that are built based on an alignment that differs significantly from the reference should be poor, but in the case of the eight-stranded β-barrels 3QRA and 3DZM, a model is built with a GDT_TS of 67.5% despite the MP-T alignment including only 26% of the residue pairings in the reference alignment.

Table 2 shows the number of alignments for which each column method correctly aligns at least 10 more residues than the corresponding row method. The first entry in the table is assessed over the 115 transmembrane domains, and the entry in parentheses is assessed over the whole target sequence (165 alignments). MP-T consistently outperforms all other methods more often than they outperform it, with the advantage being largest over transmembrane regions. The MP-T (pair) column performs a pairwise alignment; its poor performance illustrates the importance of homologues in creating an accurate alignment.

The alignments in the test set are diverse—spanning an order of magnitude in length and sequence identity (Fig. 1a). Secondary structure content varies from >75% α-helix to >75% β-strand. Sequences forming soluble parts of transmembrane complexes are also represented, with 50 of the 165 alignments having sequences without a transmembrane domain under our definition (Section 3.5).

Table 2. Number of times column method beats row method by at least 10 residues

	MP-T	MSAProbs	MAFFT L-INS-i	HHsearch	PROMALS	PSI-Coffee	MUSCLE	clustalΩ	MP-T (pair)
MP-T	—	6 (12)	7 (13)	17 (21)	17 (25)	13 (18)	8 (9)	12 (15)	9 (8)
MSAProbs	**20 (24)**	—	**14** (16)	17 (20)	**25 (30)**	**15 (17)**	10 (10)	12 (12)	13 (14)
MAFFT L-INS-I	**22 (25)**	13 **(17)**	—	19 (23)	**19** (23)	**16 (21)**	8 (9)	**17** (18)	14 (13)
HHsearch	**28 (38)**	**20 (33)**	**20 (35)**	—	**25 (34)**	**25 (34)**	18 (25)	20 **(30)**	14 (19)
PROMALS	**21 (31)**	11 (28)	15 **(30)**	16 (27)	—	11 (22)	10 (20)	16 (27)	12 (13)
PSI-Coffee	**20 (25)**	8 (13)	12 (19)	16 (21)	**20 (25)**	—	5 (6)	15 (21)	9 (12)
MUSCLE	**37 (52)**	**30 (50)**	**21 (37)**	**28 (43)**	**32 (43)**	**28 (41)**	—	**23 (34)**	20 (25)
clustalΩ	**20 (27)**	**14 (24)**	13 **(24)**	20 (26)	**21 (29)**	15 **(24)**	5 (7)	—	14 (20)
MP-T (pair)	**40 (70)**	**39 (73)**	**41 (74)**	**34 (68)**	**38 (68)**	**36 (65)**	**32 (60)**	**38 (72)**	—

The first entry is assessed over the transmembrane region only, whereas parenthetical entries are assessed over the whole target sequence. For example, there are 20 alignments in which HHsearch beats clustalΩ in the transmembrane region, and 26 where HHsearch beats clustalΩ overall.
An entry is in bold if the column method beats the row method more often than the row method beats the column method. Better performing methods have more bold entries in their column. Details of how each program was run are provided in Section 3.4.

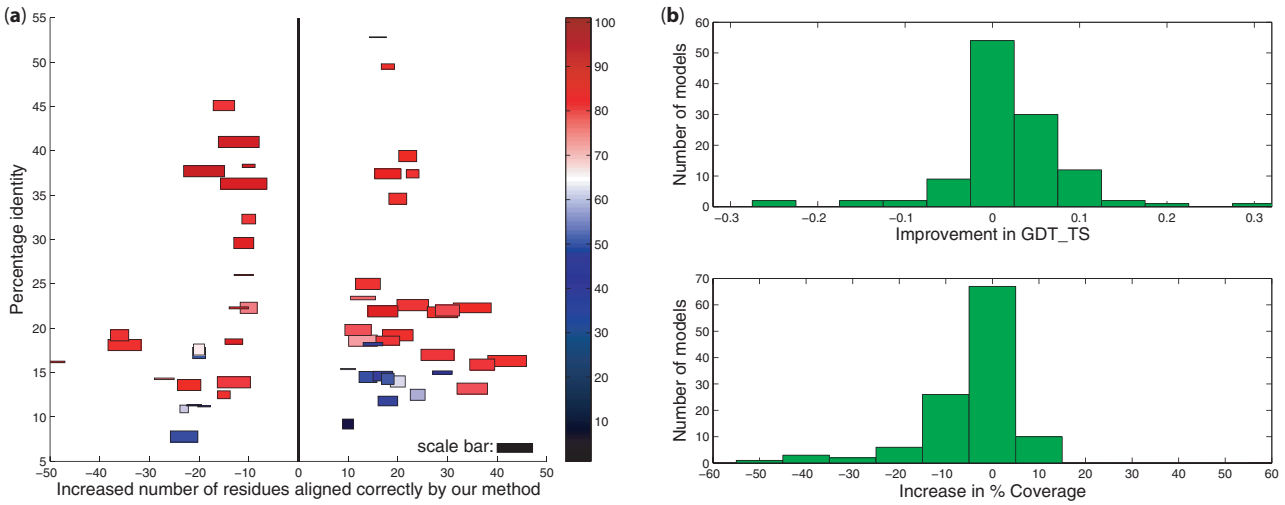

Fig. 1. (**a**) Alignments for which PROMALS or MP-T aligned at least 10 residues more correctly than the other. The assessment is made over the full target sequence. The size of each rectangle is a proxy for the size of the MSA for each input: rectangle height is proportional to the number of sequences in the MP-T alignment, rectangle length is proportional to the length of the target sequence. The scale bar shows a rectangle corresponding to an alignment made using 100 sequences and with a target length of 500 residues. Rectangles are coloured by the fraction of the structure alignment that the better performing method reproduced. One point at (−72, 9) falls outside the axes of the graph: PROMALS aligns 63% of the residues correctly. (**b**) Distribution of improvements in model accuracy (top) and model coverage (bottom) from using MP-T rather than PROMALS. Models produced from MP-T alignments provide lower coverage than PROMALS (bars are higher to the left of the origin in the lower panel) but are significantly more accurate (bars are higher to the right of the origin in the upper panel, $P < 10^{-4}$ by Wilcoxon signed-rank test)

As PROMALS came nearest to parity with MP-T (Table 2), we selected it to investigate the types of alignments each performs best. Figure 1a shows all alignments where PROMALS and MP-T outperform each other by at least 10 residues over the full alignment. Several cases where MP-T performs less well (rectangles to the left of the black line) may be attributable to the use of a small number of homologues to make the alignment (thin rectangles in the figure). This can arise either because the procedure in Section 3.3 returned few homologues or because the guide tree excluded many homologues (see Section 2).

MP-T appears to perform better on longer proteins (longer bars to the right of the black line), which is to be expected: differences between a single substitution table and environment-specific tables accumulate with sequence length. However, the improved performance may also reflect the composition of the training set, which contained only one alignment with a target sequence shorter than 200 residues (Supplementary Fig. S2). PROMALS may perform better at making distant alignments <15% sequence identity, whereas MP-T has an advantage in the 15–30% identity range, which is arguably more useful for modelling.

Alignment methods show a greater range of accuracy over majority β-strand targets than over majority α-helical targets. On average, β-strand targets are shorter than other targets, and have lower sequence identity reference alignments that

contain a higher density of gap open events. These factors may differentiate methods by their gap-penalty schemes.

The above results are obtained when only the template sequence is structurally annotated. However, MP-T is capable of accepting more annotations, derived either from crystal structures or from prediction programs such as PSIPRED. MP-T's internal transfer of structural annotation between sequences is an implicit prediction of secondary-structure, solvent accessibility and membrane-positioning. Additional annotations should lead to improved alignments if they have a lower error rate than these implicit predictions. An approximate bound to the accuracy gains from adding more annotations can be found by providing structural annotations for the target sequence as well as for the template structure. This increased F_M and F_D by ~0.7% over both TM domains and full alignments.

4.4 Model accuracy in the transmembrane region

We built models for the 115 transmembrane domains in the test set, which in most cases, such as ion channels, is the domain of interest. Models were built using MP-T, PROMALS and MUSCLE alignments with the high-accuracy mode of MEDELLER (Kelm *et al.*, 2010), which incorporates the FREAD loop-modelling method (Choi and Deane, 2010). The loops for most targets are present in FREAD's database, meaning the models built here should be better than those for blind structure prediction.

PROMALS and MUSCLE were chosen for model building alongside MP-T as they respectively had the best and worst alignment accuracy of the methods against which we compared. PROMALS has also been successful in the CASP competition (Moult *et al.*, 2011). Model accuracy was assessed by calculating the model GDT_TS. This is a number in the range (0, 1), with higher values corresponding to more accurate models. To isolate the contribution of the alignment method to the model, GDT_TS was calculated only over pairs of residues that were aligned in the input sequence alignment. Models were built using the full sequence alignment, as in the MEDELLER procedure this cannot worsen the transmembrane model, but can improve loop modelling.

MP-T produces models of significantly higher GDT_TS than PROMALS or MUSCLE ($P < 10^{-4}$, Wilcoxon signed-rank test). The upper panel of Figure 1b shows the distribution of improvements in model GDT_TS gained by using MP-T rather than PROMALS. For example, for 30 of the 115 models, MP-T alignments led to a 2.5–7.5% improvement in model GDT_TS compared with models made with PROMALS. The small height of bars to the left of the origin shows that MP-T rarely generates significantly worse models than PROMALS. The lower panel of Figure 1b shows that PROMALS alignments lead to models with higher coverage than MP-T. On average, MUSCLE's models had the lowest GDT_TS and the lowest coverage, demonstrating that less accurate alignments lead to less accurate models.

Figure 2 illustrates improved model building at low sequence identity. The target (PDB code: 1IZL, chain D gray) and template (PDB code: 1OGV, chain M) are taken from photosynthetic reaction centres. Models from MP-T (left) and PROMALS (right) are colour coded per residue by the distance to the corresponding residue in the crystal structure (0 Å = blue, > 5Å = red). Numbering transmembrane helices 1–5 from left to right, MP-T models more of the loop region above helix 1, and more of helix 4 than PROMALS. MP-T assigns the correct residues to helix 1, whereas the PROMALS model introduces a

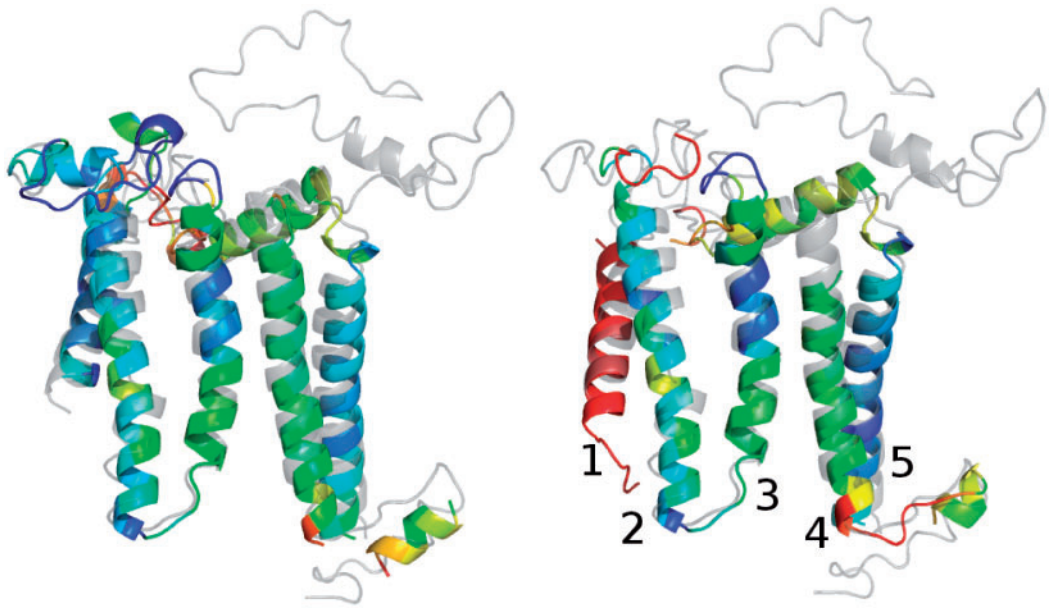

Fig. 2. Part of a photosynthetic reaction centre PDB code: 1IZL, chain D (gray) is modelled using PDB code: 1OGV, chain M as a template. The target and template share 15% sequence identity. Residues in the model are colored by the distance to their position in the crystal structure: blue residues are close to their native position, red residues are ≥5Å away from their native position. The model on the left is built using MP-T, whereas that on the right is built using PROMALS. The MP-T model provides a more accurate representation of helix 1, and models the region above this helix

large shift (in the red helix, each residue is at least 5Å from its true position).

5 CONCLUSION

We have created a program, MP-T, to perform sequence to structure alignment for the homology modelling of membrane proteins. Tested on a set of 165 diverse structure alignments, MP-T is found to introduce the fewest misaligned residues (high F_M) while aligning as accurately as the state-of-the-art methods HHsearch and PROMALS (Table 1). MP-T performs particularly well on longer membrane proteins (a category that includes drug targets such as GPCRs and ion channels) and in the twilight zone (15–30% sequence identity).

The accuracy of MP-T is derived from its effective use of information about accessible surface area, membrane positioning and secondary structure. This information is obtained from the template structure and used to make predictions about the same properties of homologous sequences. Gap penalties and substitution scores are adjusted on a per-residue basis depending on this information—for example, the creation of gaps in the middle of the membrane is discouraged, and substitutions within a helix are scored according to the propensity for an amino acid type to be in a helix.

Choice of homologues is found to be the single greatest factor affecting alignment accuracy for our method. Selecting homologues to favour sequences that are more closely related to the target and template leads to substantially less accurate alignments than random selection. It is likely that improvements in accuracy can be gained by further refining the selection process: for example, by loosening cut-offs if too few sequences are returned.

In agreement with previous studies, we find that guide tree construction does not strongly affect alignment accuracy. However, the general trend is that single-linkage clustering is a better method for building guide trees than neighbour joining, BIONJ or UPGMA.

We perform no sequence weighting, no iterative refinement and use a simpler consistency criterion than those used in leading multiple sequence aligners such as MSAProbs. Despite this, the use of environment-aware gap penalties and substitution tables allows us to produce more accurate models of transmembrane domains than other methods tested. Incorporation of environment awareness into a more sophisticated aligner may yield even larger improvements in the quality of membrane protein models.

ACKNOWLEDGEMENT

The authors thank Sebastian Kelm for help running the iMembrane and MEDELLER programs, James Dunbar for assistance with Pymol and the members of the Oxford Protein Informatics Group for useful discussions.

Funding: This work was supported by the Engineering and Physical Sciences Research Council (to J.R.H. and C.M.D.), the Biological Sciences Research Council and the University of Oxford Doctoral Training Centres (to C.M.D.).

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Berman,H.M. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

Chang,J.M. *et al.* (2012) Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *BMC Bioinformatics*, **13** (**Suppl. 4**), S1.

Choi,Y. and Deane,C.M. (2010) FREAD revisited: accurate loop structure prediction using a database search algorithm. *Proteins*, **78**, 1431–1440.

Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.

Edgar,R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

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

Gascuel,O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.*, **14**, 685–695.

Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.

Hill,J.R. *et al.* (2011) Environment specific substitution tables improve membrane protein alignment. *Bioinformatics*, **27**, i15–i23.

Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.

Katoh,K. and Toh,H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.*, **9**, 286–298.

Kelm,S. *et al.* (2009) iMembrane: homology-based membrane-insertion of proteins. *Bioinformatics*, **25**, 1086–1088.

Kelm,S. *et al.* (2010) MEDELLER: homology-based coordinate generation for membrane proteins. *Bioinformatics*, **26**, 2833–2840.

Khafizov,K. *et al.* (2010) A study of the evolution of inverted-topology repeats from LeuT-fold transporters using AlignMe. *Biochemistry*, **49**, 10702–10713.

Liu,Y. *et al.* (2010) MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. *Bioinformatics*, **26**, 1958–1964.

Lobley,A. *et al.* (2009) pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics*, **25**, 1761–1767.

Mizuguchi,K. *et al.* (1998) JOY: protein sequence-structure representation and analysis. *Bioinformatics*, **14**, 617–623.

Moult,J. *et al.* (2011) Critical assessment of methods of protein structure prediction (CASP)—round IX. *Proteins*, **79** (**Suppl. 1**), 1–5.

Notredame,C. *et al.* (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.

Overington,J.P. *et al.* (2006) How many drug targets are there? *Nat. Rev. Drug Discov.*, **5**, 993–996.

Pei,J. and Grishin,N.V. (2007) PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*, **23**, 802–808.

Pirovano,W. *et al.* (2008) PRALINE^{TM}: a strategy for improved multiple alignment of transmembrane proteins. *Bioinformatics*, **24**, 492–497.

Plyusnin,I. and Holm,L. (2012) Comprehensive comparison of graph-based multiple protein sequence alignment strategies. *BMC Bioinformatics*, **13**, 64.

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

Sauder,J.M. *et al.* (2000) Large-scale comparison of protein sequence alignment algorithms with structure alignments. *Proteins*, **40**, 6–22.

Scott,K.A. *et al.* (2008) Coarse-grained md simulations of membrane protein-bilayer self-assembly. *Structure*, **16**, 621–630.

Shi,J. *et al.* (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.*, **310**, 243–257.

Sievers,F. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.

Söding,J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.

Suzek,B.E. *et al.* (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, **23**, 1282–1288.

Thompson,J.D. *et al.* (2011) A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives. *PLoS One*, **6**, e18093.

Wheeler,T.J. and Kececioglu,J.D. (2007) Multiple alignment by aligning alignments. *Bioinformatics*, **23**, i559–i568.

Yang,Y. *et al*. (2011) Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics*, **27**, 2076–2082.

Zemla,A. *et al*. (2001) Processing and evaluation of predictions in CASP4. *Proteins*, **(Suppl 5)**, 13–21.

Zhang,Y. and Skolnick,J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins*, **57**, 702–710.

Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*., **33**, 2302–2309.