Structural bioinformatics

Co-evolving residues in membrane proteins

Angelika Fuchs[1], Antonio J. Martin-Galiano[1], Matan Kalman[2], Sarel Fleishman[2],
Nir Ben-Tal[2] and Dmitrij Frishman[1,*]

[1]Department of Genome Oriented Bioinformatics, Technische Universität München, Wissenschaftszentrum
Weihenstephan, 85350 Freising, Germany and [2]Department of Biochemistry, George S. Wise Faculty of Life Sciences,
Tel-Aviv University, 69978 Ramat Aviv, Israel

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: The analysis of co-evolving residues has been exhaustively evaluated for the prediction of intramolecular amino acid contacts in soluble proteins. Although a variety of different methods for the detection of these co-evolving residues have been developed, the fraction of correctly predicted contacts remained insufficient for their reliable application in the construction of structural models. Membrane proteins, which constitute between one-fourth and one-third of all proteins in an organism, were only considered in few individual case studies.

Results: We present the first general study of correlated mutations in α-helical membrane proteins. Using seven different prediction algorithms, we extracted co-evolving residues for 14 membrane proteins having a solved 3D structure. On average, distances between correlated pairs of residues lying on different transmembrane segments were found to be significantly smaller compared to a random prediction. Covariation of residues was frequently found in direct sequence neighborhood to helix–helix contacts. Based on the results obtained from individual prediction methods, we constructed a consensus prediction for every protein in the dataset that combines obtained correlations from different prediction algorithms and simultaneously removes likely false positives. Using this consensus prediction, 53% of all predicted residue pairs were found within one helix turn of an observed helix–helix contact. Based on the combination of co-evolving residues detected with the four best prediction algorithms, interacting helices could be predicted with a specificity of 83% and sensitivity of 42%.

Availability: http://webclu.bio.wzw.tum.de/helixcorr/

Contact: d.frishman@wzw.tum.de

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

1 INTRODUCTION

In order to maintain protein function, mutations that tend to destabilize a particular protein structure may provoke other positions to mutate concurrently in order to compensate for the loss of stability. The first examples of such compensatory changes were described by analyzing individual families having a solved structure (Altschuh *et al.*, 1987, 1988). Amino acid contacts were initially thought to be primary spots of these compensatory processes, making the detection of sequence positions with correlated mutational behavior an interesting feature for contact prediction methods. Accordingly, the first approach to detect co-evolving residues in a multiple sequence alignment was published in 1994 (Gobel *et al.*, 1994) and a variety of additional detection algorithms have been reported since then. However, prediction accuracies for structural contacts in globular proteins hardly exceeded 20% (Fodor and Aldrich, 2004), strongly limiting the practical utility of the predicted contacts as structural constraints in *ab initio* structure prediction. Despite this general drawback for the intensive use of correlated mutations, several successes in individual applications have been reported. For example, the global fold could be successfully predicted for 20 non-homologous proteins with less than 100 amino acids with a root-mean-square deviation (r.m.s.d.) between 3.0 and 6.5 Å in an *ab initio* folding simulation using co-evolving residues as constraints (Ortiz *et al.*, 1998a, b). Besides their application in fold recognition or structure prediction, compensatory mutations were also found instrumental in detecting both interacting proteins and interaction regions between two proteins (Filizola *et al.*, 2002; Pazos and Valencia, 2002; Pazos *et al.*, 1997). It was also shown that networks of correlated mutations tend to appear near binding regions and active sites (Gloor *et al.*, 2005) allowing for the identification of functionally important regions in combination with entropic measures (Saraf *et al.*, 2003). Lately, WW domain sequences that fold into stable structures and function like natural WW domains could be artificially engineered based on information obtained from an evolutionary coupling analysis (Russ *et al.*, 2005; Socolich *et al.*, 2005).

The broad range of reported applications for co-evolving residues is accompanied by an even more comprehensive number of approaches proposed for their detection in a multiple alignment. The strategy most commonly used relies on the calculation of a Pearson correlation coefficient to detect alignment positions with similar patterns of amino acid change (Gobel *et al.*, 1994; Kundrotas and Alexov, 2006; Neher, 1994; Olmea and Valencia, 1997; Pollock and Taylor, 1997; Taylor and Hatrick, 1994). In order to reduce the number

---

*To whom correspondence should be addressed.

of false positives, several authors proposed to additionally include phylogenetic information about the analyzed sequences in the prediction process (Chelvanayagam *et al.*, 1997; Fares and Travers, 2006; Fleishman *et al.*, 2004a, b; Noivirt *et al.*, 2005; Shindyalov *et al.*, 1994). Other prediction algorithms try to detect significant co-evolution based on a χ^2 goodness-of-fit test comparing the observed co-occurrence of two residues with their expected co-occurrence (Kass and Horovitz, 2002; Larson *et al.*, 2000; Noivirt *et al.*, 2005), by using a maximum-likelihood approach (Pollock *et al.*, 1999) or through the application of information theory (Clarke, 1995; Gloor *et al.*, 2005; Martin *et al.*, 2005). An alternative approach is constituted by perturbation-based methods such as SCA (Statistical Coupling Analysis) (Dekker *et al.*, 2004; Lockless and Ranganathan, 1999; Suel *et al.*, 2003) where co-evolution of residues is identified by the analysis of statistical coupling of amino acid distributions. Subalignments having a changed amino acid distribution at certain positions are used to evaluate the effect of this perturbation on the residue compositions at other positions of the alignment.

Independent of the prediction strategy used, most studies on co-evolving residues so far were conducted with soluble proteins. Membrane proteins, which constitute between one-fourth and one-third of all proteins in an organism (Frishman and Mewes, 1997; Wallin and von Heijne, 1998), were only considered in few individual case studies (Oliveira *et al.*, 2003; Filizola *et al.*, 2002; Fleishman *et al.*, 2004a, b). Due to the paucity of 3D structures, a general analysis of co-evolving residues in a large non-redundant dataset of membrane proteins is still missing. Approaches to predict residues participating in helix–helix contacts are rather based on the idea to identify membrane-exposed and buried residues (Adamian and Liang, 2006; Chen and Xu, 2006; Hildebrand *et al.*, 2006). However, pairs of contacting residues cannot be identified with these methods. With the accumulation of more membrane protein structures in recent years, it is now possible to analyze the utility of correlated mutation detection methods specifically for membrane proteins.

Here, we present the first large-scale analysis of co-evolving residues in membrane proteins with special emphasis on their possible application for the prediction of interacting helices. Using nine different prediction algorithms, we extracted co-evolving residues for membrane protein clusters taken from the CAMPS database of membrane proteins (Martin-Galiano and Frishman, 2006), which include at least one protein with solved structure. While residue contact prediction accuracies generally did not exceed 10% with any of the employed methods, a remarkable fraction of up to 49% of all correlations were found to be as close as one helical turn within an actual helix–helix contact. Using a consensus prediction approach, where correlations detected with different prediction algorithms were combined, this value could be further increased to 53%, and interacting helices could be predicted at a specificity of 83% and sensitivity of 42%.

2 METHODS

2.1 Data

Protein sequences were taken from the most recent version of the CAMPS database of membrane proteins (Martin-Galiano and

Frishman, 2006) covering 120 prokaryotic genomes. Transmembrane segments (TMS) were predicted by TMHMM 2.0 (Krogh *et al.*, 2001). For our study of co-evolving residues, we extracted all SC-clusters from CAMPS. At this clustering level, the generated groups of proteins roughly correspond to structural folds. Cluster sequences were aligned by CLUSTALW using default parameters (Thompson *et al.*, 1994). For the prediction of co-evolving residues in TMS, conserved transmembrane regions (TMS cores) were extracted from CAMPS and concatenated to form sequences representing only the transmembrane parts of each protein.

From the set of pre-aligned TMS sequences, all sequences considered inappropriate for the analysis were discarded. Since highly similar sequences might result in few correlations due to a lack of variability, sequences with a pair-wise identity above a pre-set threshold were considered redundant and removed. Different thresholds were used in individual predictions ranging from 95% pair-wise identity down to 50% identity. The thresholds were chosen dependent on the total number of sequences in the alignment to allow for an optimal trade-off between a minimal number of required sequences and sufficient sequence diversity for a successful prediction. In addition, sequences with 25% or more gaps within at least one TMS were removed. According to these rules, the number of removed sequences varied between 3 in clusters with well-aligned sequences uniformly covering the cluster sequence space and 346 in the case of a cluster with several tightly connected sub-clusters. The final number of sequences in individual alignments ranged from 20 to 228.

During an initial prediction step, co-evolving residues were predicted for all SC-clusters of CAMPS. In a subsequent selection step, clusters found to be suboptimal due to high sequence diversity were either discarded in case the number of valid sequences in the final multiple alignments was below 15 or replaced by sub-clusters with higher similarity among their members. Inappropriate clusters were identified by either an average pair-wise identity of below 15% or an extremely small number of obtained correlations (less than one per TMS). This procedure was repeated until a cluster was either appropriate for the analysis of correlated mutations or had to be removed due to an insufficient number of sequences. In total, starting from 266 SC-clusters currently available in CAMPS 91 optimal clusters were obtained, 14 of which contained a representative structure (Supplementary Table 1).

2.2 Prediction of co-evolving residues

Co-evolving residues were predicted using seven different prediction algorithms: McBASC (Olmea and Valencia, 1997), OMES (Fodor and Aldrich, 2004; Kass and Horovitz, 2002), CORRMUT (Fleishman *et al.*, 2004a), CAPS (Fares and Travers, 2006), MI (Gloor *et al.*, 2005), SCA (Lockless and Ranganathan, 1999) and ELSC (Dekker *et al.*, 2004). For the McBASC algorithm two different substitution matrices [the Miyata matrix (Miyata *et al.*, 1979) and the McLachlan matrix (McLachlan, 1971)] were evaluated and the OMES algorithm was applied in two different versions, as originally introduced by Kass and Horovitz (OMES-KASS) (Kass and Horovitz, 2002) and in its modified version presented by Fodor and Aldrich (OMES-FODOR) (Fodor and Aldrich, 2004).

McBASC. We used the original algorithm of Gobel *et al.* (1994) with its refinements as introduced by Olmea and Valencia (1997). To select significantly correlated sequence positions, we applied a length-dependent threshold by choosing no more of the highest correlated pairs than one-fifth of the protein length (L/5 criterium).

OMES. We implemented the original version of the algorithm (OMES-KASS) (Kass and Horovitz, 2002) where the statistical significance of the difference between observed and expected frequencies is calculated using the χ^2 goodness-of-fit test. All covariations with *P*-values of less than 0.001 were considered to be significantly correlated and the L/5 most significant correlations were selected.

Additionally, co-evolving residues were predicted with the modified OMES algorithm as provided by A. Fodor (www.afodor.net). Based on the calculated correlation score, the L/5 highest correlated residues were selected.

CORRMUT. For predictions using the CORRMUT algorithm (Fleishman *et al.*, 2004a), a phylogenetic tree of all sequences in the obtained multiple alignment was calculated and ancestor sequences at internal nodes of the tree were reconstructed with the program FASTML (Pupko *et al.*, 2000). The Miyata matrix (Miyata *et al.*, 1979) was chosen as substitution matrix. The significance of the derived correlation coefficient was estimated by confidence intervals obtained from a bootstrap procedure using a sample size of 400. Correlation coefficients calculated for each sample were used to derive a mean Pearson correlation coefficient (r) as well as the 95% confidence intervals (r_{low}, r_{high}). To identify significantly correlated residue pairs, we applied a minimal threshold of 0.4 for the mean correlation coefficient and a minimal r_{low}-value of 0.05 and selected then the length/5 highest correlated sequence positions. The thresholds for the mean correlation coefficient and the lower confidence boundary were established in preliminary experiments where they were found to permit the best trade-off between number of detected correlations and prediction accuracy. However, with the exception of one protein (1PW4 chain A) the number of correlations satisfying these thresholds was clearly lower in comparison to other prediction algorithms such as McBASC or OMES.

CAPS. Predictions with the CAPS algorithm (Fares and Travers, 2006) were executed using the provided program (http://bioinf.gen.tcd.ie/~faresm/software/caps/) and recommended standard parameters. Again a minimal correlation coefficient of 0.4 was applied as threshold and the length/5 highest correlated pairs were selected. Using these selection criteria predictions for only five proteins could be obtained (1XQE chain A, 1IWG chain A, 1RC2 chain A, 1L7V chain A and 1RHZ chain A). In all other cases, the number of sequences in the multiple alignment (generally less than 50) seemed to be insufficient for this prediction algorithm.

MI. The MI algorithm was implemented as described earlier in comparative studies on correlated mutations (Halperin *et al.*, 2006). The L/5 pairs with the highest MI score were selected.

SCA/ELSC. Both algorithms are used in the implementation provided by A. Fodor (www.afodor.net). The L/5 correlations with the highest score were again chosen in both cases.

2.3 Structural validation

Observed distances between residue pairs were extracted by calculating the minimal distance between side-chain or backbone atoms of the two residues. Two residues were considered in contact if their minimal distance was <5.5 Å. The 5.5 Å cutoff was chosen as the maximal distance between a pair of heavy (i.e. non-hydrogen) atoms that is indicative of a direct contact; at larger distances, a third atom may fit in between the atom pair. Other studies on correlated mutations have often used a contact definition based on C_β-distances and an 8 Å cutoff. However, due to the regular backbone conformation of α-helical membrane proteins we consider a contact criterion incorporating side-chain atoms better suited for the analysis of helix–helix interactions. Contact prediction accuracies based on C_β-distances and a contact threshold of 8 Å are provided in the Supplementary Table 3.

From the number of predicted contacts and the number of observed contacts, the prediction accuracy (fraction of correctly predicted contacts out of all correlations found) was calculated for only those correlated pairs lying on different transmembrane helices (accuracy$_{sep}$). A *P*-value was determined based on the hypergeometric distribution and the probability to pick a residue pair in contact by random to estimate the significance of the obtained prediction.

Since correlated positions may contain information beyond mere physical contacts between individual residues, two additional quality measures were used to investigate the prediction outcome. The harmonic average Xd was introduced by Pazos *et al.* (1997) as a measure of relative proximity rather than direct contact. For the calculation of the Xd value, the distribution of C_β-distances between correlated residues is compared to the distribution of distances for all pairs of positions. Distances from both distributions are grouped into bins of 4 Å and the differences between the two distributions are calculated for each bin. The differences are weighed with the inverse of the normalized distance of the corresponding bin and are added. A value of $Xd = 0$ indicates no separation between the two distributions, while $Xd > 0$ indicates a shift of correlated residues towards smaller distances. The larger a positive Xd-value the more efficient is the corresponding prediction:

$$Xd = \sum_{i=1}^{n} \frac{P_{ic} - P_{ia}}{d_i n} \tag{1}$$

where P_{ic} and P_{ia} are the percentages of correlated and all residue pairs with distance between d_i and d_{i-1}, d_i is the upper limit of each bin (normalized to 60) and n is the number of distance bins (15 for the range from 4 to 60 Å).

Additionally we used a 'δ-Analysis' (Ortiz *et al.*, 1999) to investigate the position of found correlations with respect to observed helix–helix contacts. Within this analysis, the fraction of correlations with residues i and j is calculated that have an observed contact between residues in the interval $\{i - \delta, i + \delta\}$ and $\{j - \delta, j + \delta\}$. With $\delta = 4$, the fraction of correlations where both participating residues lie within one turn of residues forming an interhelical contact is detected. Again a *P*-value was calculated based on the hypergeometric distribution to estimate the significance of the prediction.

2.4 Prediction of helix–helix interactions

From 14 proteins with available crystal structure, a dataset of helix–helix pairs was extracted. After removing helix pairs from proteins with only three transmembrane segments, a total of 370 predicted helix pairs remained. Predicted transmembrane segments were compared against transmembrane helix positions determined using structural information as obtained from the Protein Data Bank of Transmembrane Proteins (Tusnady *et al.*, 2005). The comparison revealed a total number of 3 missing helices in our dataset, 3 additional predicted helices and 3 cases where segments were either joined or split (Supplementary Table 1). These transmembrane segments were not included in our analysis resulting in 325 helix pairs, of which 166 were considered to be in contact since they contained at least one residue pair having a minimal distance of <5.5 Å. Several consensus helix pair prediction methods were evaluated where both the required number of correlated mutations for a certain helix pair was varied and combinations of different prediction methods were tested. In each case, the sensitivity and specificity of the observed prediction was calculated and compared to a random prediction obtained by calculating the expected number of correctly predicted helix pairs based on the probability for a contacting helix pair. The significance of each prediction was evaluated using a χ^2-test.

3 RESULTS AND DISCUSSION

3.1 Selection of optimal CAMPS protein clusters for the prediction of co-evolving residues

For the analysis of co-evolving residues in membrane proteins, we developed a procedure to extract optimal protein clusters from the CAMPS database of membrane proteins

(Martin-Galiano and Frishman, 2006). Starting with the 266 clusters corresponding to structural folds (SC-clusters), clusters whose sequences were too diverse to allow for reliable predictions were discarded or replaced by sub-clusters. On the other hand, since highly similar sequences might result in few correlations due to a lack of variability, sequences with a pair-wise identity above a pre-set threshold were considered redundant and removed. Based on this selection procedure, 91 optimal protein clusters were selected, of which 14 contained at least one representative protein structure (Supplementary Table 1). Multiple alignments for these clusters were obtained by concatenating transmembrane core sequences extracted from CAMPS. Co-evolving residues were extracted using seven different prediction algorithms [McBASC (Olmea and Valencia, 1997), OMES (Fodor and Aldrich, 2004; Kass and Horovitz, 2002), CORRMUT (Fleishman *et al.*, 2004a), CAPS (Fares and Travers, 2006), MI (Gloor *et al.*, 2005), SCA (Lockless and Ranganathan, 1999) and ELSC (Dekker *et al.*, 2004)], which broadly cover the range of prediction approaches known from literature. Additionally two different substitution matrices (the Miyata matrix and the McLachlan matrix) were evaluated in combination with the McBASC algorithm (McBASC-Miyata, McBASC-McLachlan) and the OMES algorithm was applied in two different versions, both as originally introduced by Kass and Horovitz (OMES-KASS) (Kass and Horovitz, 2002) and in its modified version presented by Fodor and Aldrich (OMES-FODOR) (Fodor and Aldrich, 2004), resulting in a total of nine different predictions for every multiple alignment. The number of significantly correlated residue pairs was chosen proportional to the length of the multiple alignment by extracting the top L/5 correlations, with L being the alignment length. However, in the case of the two prediction algorithms CORRMUT and CAPS the number of obtained correlations with a minimal correlation coefficient of 0.4 was less than L/5 in most proteins. Figure 1 shows the sequence separation between all co-evolving residue pairs obtained by this procedure. A clearly resolved peak corresponding to a sequence separation of four residues (one turn of an α-helix) is observed, which confirms that the obtained multiple alignments are indeed well suited for the prediction of co-evolving residues in membrane proteins.

3.2 Helix–helix contact predictions obtained with different prediction algorithms

In order to evaluate the ability of individual algorithms to predict structural contacts in membrane proteins, contact prediction accuracies (fraction of correctly predicted contacts out of all correlations found) were calculated for all correlations with residues lying on separate TMS. Depending on the prediction algorithm, between 3% (SCA) and 9% (McBASC-McLachlan) of these correlations were found to be helix–helix interactions (Table 1). According to a hypergeometric distribution, significant predictions with *P*-values of less than 0.001 were obtained for all prediction methods except SCA.

Similar results were obtained when the number of selected correlations was not chosen proportional to the length of the multiple alignment used for the prediction but varied over
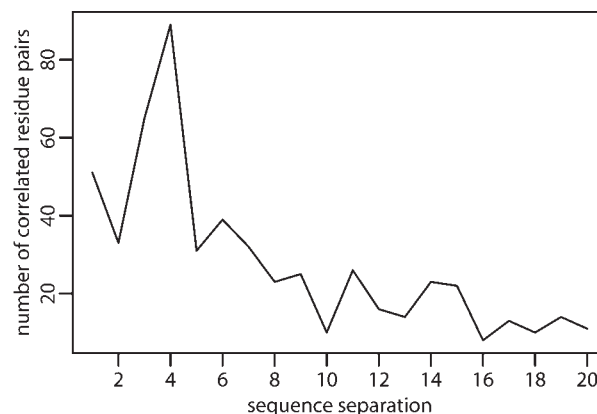

Fig. 1. Sequence separation of co-evolving residues detected in membrane proteins with nine different prediction algorithms.

Table 1. Contact prediction accuracies for different prediction algorithms applied to 14 membrane proteins with transmembrane segments predicted by TMHMM2.0. CONSENSUS-14 and CONSENSUS-R-5 correspond to two consensus predictions, where predicted correlated mutations, obtained with different prediction algorithms are combined

Method	Proteins	Accuracy [%]	*P*-value	*Xd*	Accuracy ($|\delta|=4$) [%]	*P*-value
McBASC-Miyata	14	8	2.59*E*–14	5.6	49	3.91*E*–30
McBASC-McLachlan	14	9	3.050*E*–17	5.0	42	1.36*E*–16
OMES-KASS	14	8	6.37*E*–13	4.9	43	2.55*E*–17
OMES-FODOR	14	7	1.38*E*–11	4.0	38	4.73*E*–11
CORRMUT	13	7	3.26*E*–05	4.4	38	9.25*E*–06
CAPS	5	7	8.06*E*–04	4.4	42	5.09*E*–05
MI	14	5	2.08*E*–04	−1.8	19	0.998
SCA	14	3	0.032	0.48	26	0.152
ELSC	14	7	4.42*E*–09	3.2	37	2.24*E*–09
CONSENSUS-14	14	11	1.08*E*–54	8.5	53	4.1*E*–100
CONSENSUS-R-5	14	10	4.35*E*–47	6.7	51	5.18*E*–82

a broad range independent of the proteins' lengths (Fig. 2A). Again McBASC used with the McLachlan matrix performed slightly better than the other prediction algorithms. From the two OMES variations we found the original version (OMES-KASS) to be slightly superior to its variation introduced by Fodor and Aldrich. MI and SCA were found to be the least powerful algorithms in the prediction of helix–helix contacts independent of the number of significantly correlated residue pairs selected. The algorithms CAPS and CORRMUT were excluded from this analysis since the number of significantly correlated residues obtained with these two algorithms was in most proteins clearly smaller than with the other prediction algorithms.
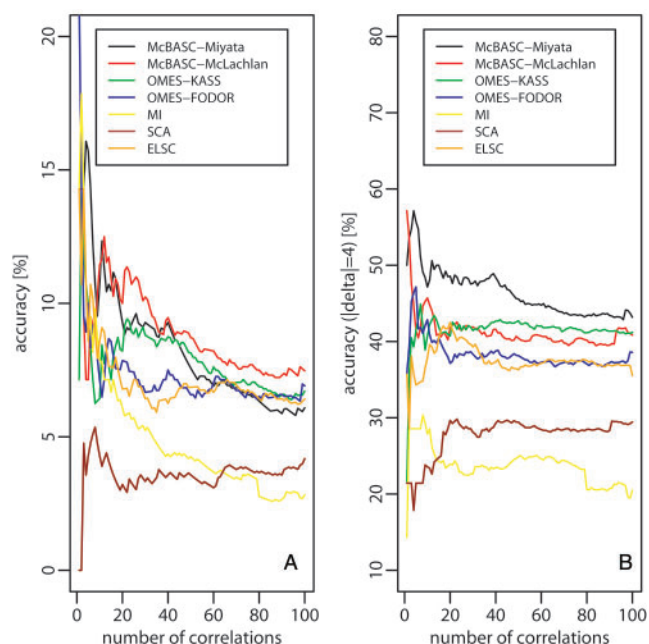
Fig. 2. Predictive performance of seven prediction methods for transmembrane proteins in relation to the number of predicted correlations. (**A**) Prediction accuracy for helix–helix contacts. (**B**) Fraction of correlations lying within one helix turn of a helix–helix contact (accuracy with $|\delta| = 4$).

3.3 Co-evolving residues are frequently found within close neighborhood to helix–helix contacts

Despite the low percentage of correctly predicted contacts, we observed a high fraction of detected correlations to be in direct neighborhood of helix–helix interactions. Starting with a general analysis of residue–residue distances we found distances between correlated residues lying on different transmembrane segments clearly shifted towards smaller values compared to the distance distribution observed for all possible pair of amino acids, as was already described for soluble proteins (Olmea et al., 1999). When analyzing the results of every prediction method individually, this shift towards smaller residues was observed for all algorithms except MI and SCA. In the case of MI, distances between correlations were even shifted towards larger distances compared to the overall distances distribution. (Supplementary Fig. 1). The difference between the two distance distributions can be quantified using the harmonic average Xd as introduced by Valencia and co-workers (Pazos et al., 1997), where $Xd > 0$ indicates a shift of the population of predicted residue pairs to smaller distances with respect to the population of all pairs. For our dataset, we obtained a maximal Xd-value of 5.6 (McBASC-Miyata) (Table 1). Intermediate Xd-values between 3.2 and 5.0 were obtained with the methods ELSC, McBASC-McLachlan, OMES-KASS, OMES-FODOR, CAPS and CORRMUT. Besides the results obtained with MI and SCA (negative Xd-value or Xd close to zero, respectively), these results are comparable to those obtained for soluble proteins, where a contact prediction accuracy of 9% and an Xd of 4.31 was reported for a dataset of 173 proteins using the

McBASC algorithm in combination with the McLachlan matrix (Fariselli et al., 2001). However, it is noteworthy that all-α soluble proteins are known to be the most difficult targets for contact prediction using correlated mutations.

Using a neural network approach, incorporating also other information such as sequence conservation or predicted secondary structure, an average prediction accuracy of 7% and Xd-values for individual proteins between -5.0 and 2.7 were reported (Fariselli et al., 2001). While the contact prediction accuracy is comparable to the results obtained here for α-helical membrane proteins, the Xd-values are clearly smaller for all-α soluble proteins. The better results obtained for membrane proteins suggest that their typical structural arrangement, with several contacting residues between parallel interacting helices, is more inclined to prompt residues lying in close structural distance to co-evolve in comparison to soluble all-α proteins.

Using a 'δ-evaluation' (Ortiz et al., 1999), where the fraction of correlated positions i and j is calculated with an observed helix–helix contact between residues in the intervals $\{i - \delta, i + \delta\}$ and $\{j - \delta, j + \delta\}$, we found, on average, up to 49% (McBASC-Miyata) of all detected correlated pairs to be situated within the same helical turn as an actual contact (accuracy with $|\delta| = 4$) (Table 1). The exact fraction differed strongly, depending on the protein and the applied prediction algorithm (Supplementary Table 2). In individual cases, such as the mechanosensitive channel protein (1MXM), the best prediction was obtained with the MI algorithm, which, on average, performed worse than all other prediction algorithms. As presented earlier for helix–helix contact prediction accuracies, we also analyzed the influence of the selected number of correlated residues on the obtained accuracy with $|\delta| = 4$ (Fig. 2B). In contrast to the prediction of helix–helix contacts, where the McLachlan matrix performed better than the Miyata matrix, in this analysis best results were obtained using the McBASC algorithm in combination with the Miyata matrix. OMES-KASS on the other hand outperformed also in this analysis its variant OMES-FODOR. Results using SCA and MI were again clearly inferior to results from all other prediction algorithms.

In publications on co-evolving residues in soluble proteins, the low contact prediction accuracy using correlated mutations has often been attributed to methodological problems in separating real correlated mutational behavior from random noise as well as to co-evolution of distant residues due to long-range interactions (Fleishman et al., 2004a; Lockless and Ranganathan, 1999) or functional reasons (Gloor et al., 2005). In membrane proteins, pairs or networks of compensatory mutations seem to affect the packing context of transmembrane helices rather than the contacts themselves, as can be concluded from the high fraction of co-evolving residues found in direct neighborhood to helix–helix contacts (see also Supplementary Fig. 2). The surrounding residues of helix–helix contacts might be generally more amenable to mutational change than the residues in actual contacts, but are still sufficiently important for proper helix interactions to make the compensation of destabilizing amino acid substitutions beneficial for protein stability. This finding is compatible with experimental evidence that helix–helix

interactions mediated both by polar residues and interaction motifs are dependent on the sequence context (Schneider and Engelman, 2004).

3.4 Prediction accuracies can be improved by using a consensus approach combining several prediction methods

Based on the observation that results with different prediction algorithms vary remarkably for individual proteins (Supplementary Table 2), we aimed on constructing consensus predictions for every protein combining detected co-evolving residues from different prediction methods. Since the two prediction methods SCA and MI were shown to perform worse regardless of the prediction quality measure used these two algorithms were excluded of any consensus approach. The results of the remaining seven predictions were combined to retrieve a consensus prediction (termed CONSENSUS). To further improve the obtained prediction by reducing likely false positives, all correlations lying on a pair of helices with a total number of detected correlations less than a given threshold N were removed.

Figure 3 shows that both the helix–helix contact prediction accuracy and the accuracy ($|\delta| = 4$) indeed increase with an increase in this threshold N. With $N = 14$ (CONSENSUS-14), the helix–helix contact prediction accuracy could be improved to close to 11% (compared to 9% as best result for a single algorithm) and the accuracy ($|\delta| = 4$) could be elevated to 53% (compared to 49% again as best results for an individual algorithm) (Table 1).

For a second consensus approach, correlations detected with the four best performing prediction methods (McBASC-Miyata, McBASC-McLachlan, OMES-KASS and CAPS, selected based on their accuracy with $|\delta| = 4$) were combined and again all correlations on helix pairs with less than N correlations in total were removed (reduced consensus or CONSENSUS-R). With $N = 5$ (CONSENSUS-R-5), helix–helix contacts could be predicted with 10% accuracy and the fraction of correlations lying within one helical turn of an actual helix–helix contact was found to be 51%. The contact map of the AcrB bacterial multidrug efflux transporter (Supplementary Fig. 2) shows an example of how false positive predictions can be removed by applying this threshold $N = 5$ in comparison to a mere combination of individual prediction methods.

3.5 Prediction accuracies based on experimentally determined TMS

All results presented so far were based on multiple alignments consisting of transmembrane segments predicted with TMHMM. As shown in Supplementary Table 1, these predicted TMS positions may differ slightly from those determined based on the PDB structure. To evaluate whether these differences have a noticeable effect on the detection of co-evolving residues, multiple alignments consisting of experimentally determined transmembrane segments were obtained. Correlated mutations were predicted using the same procedure as used earlier for the multiple alignments consisting of predicted TMS. We found that helix–helix contact prediction
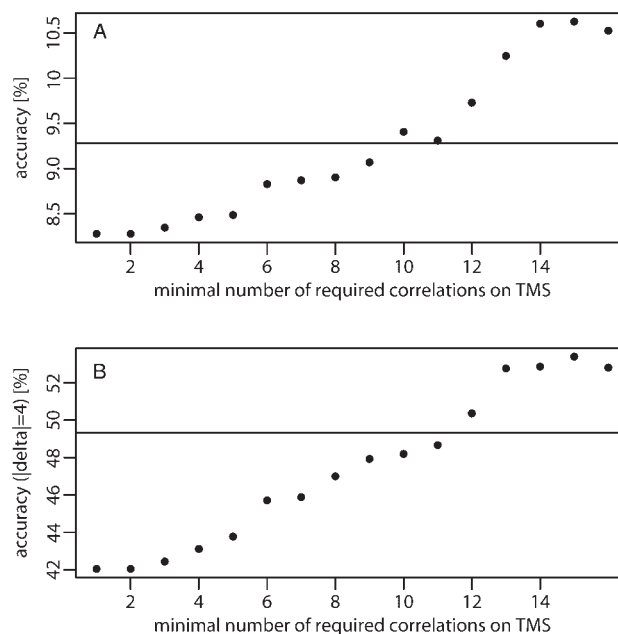

Fig. 3. Improvement of helix–helix contact prediction accuracy (**A**) and accuracy ($|\delta| = 4$) (**B**) by applying a consensus approach. The line indicates for comparison the maximal value obtained with a single prediction algorithm.

Table 2. Contact prediction accuracies for different prediction algorithms applied to 14 membrane proteins with experimentally determined transmembrane segments

Method	Proteins	Accuracy [%]	P-value	Accuracy ($|\delta| = 4$) [%]	P-value
McBASC-Miyata	14	9	6.22E–14	49	7.82E–28
McBASC-McLachlan	14	10	6.48E–17	46	4.92E–21
OMES-KASS	14	9	1.17E–14	50	2.55E–28
OMES-FODOR	14	8	7.78E–12	41	1.73E–14
CORRMUT	14	7	1.65E–05	33	0.0016
CAPS	5	12	1.29E–06	47	8.31E–07
MI	12	8	3.07E–10	32	0.0014
SCA	14	3	0.0702	27	0.1131
ELSC	14	10	6.48E–17	41	1.73E–14
CONSENSUS-14	14	12	1.41E–53	55	2.48E–95
CONSENSUS-R-5	14	11	3.64E–42	56	6.26E–94

accuracies increased between 1% and 5% while the fraction of correlations within one helix turn of a helix–helix contact increased even by 13% in the case of the MI algorithm (Table 2). The analysis of co-evolving residues in membrane proteins, where solved protein structures as well as experimentally determined topologies are generally only available in rare cases, is therefore also significantly dependent on the quality of the predicted TMS positions.

Table 3. Prediction of interacting helices using a consensus approach

Required correlations	Accuracy [%]	Sensitivity [%]	Specificity [%]	P-value
1	56.8	78.3	37.7	$2.30E-03$
2	61.5	66.2	56.6	$5.54E-05$
3	64.1	56.0	67.3	$3.85E-05$
4	65.0	45.8	74.2	$2.74E-04$
5	71.9	41.6	83.0	$2.19E-06$
6	72.5	30.1	88.0	$1.09E-04$
7	78.9	27.1	92.5	$7.14E-06$
8	78.3	21.7	93.7	$1.33E-04$
9	84.8	16.9	96.9	$9.20E-05$
10	88.5	13.9	98.1	$1.62E-04$

3.6 Prediction of interacting helices using correlated mutations

Membrane proteins show a relatively high structural simplicity compared to soluble proteins due to the severe structural constraints imposed by the lipid bilayer. Thus, in a first approximation, structure prediction of transmembrane domains is basically reduced to the question of how transmembrane segments interact along the membrane. Since co-evolving residues were found to frequently appear in close neighborhood to helix–helix contacts, we wanted to evaluate the applicability of those co-evolving residues for the prediction of interacting helix pairs in membrane proteins. To minimize the number of incorrectly predicted interactions without loosing too much valuable information, the CONSENSUS-R approach was used for this analysis, which combines results from four prediction algorithms that we earlier found to perform best on membrane proteins (McBASC-Miyata, McBASC-McLachlan, OMES-KASS and CAPS). Out of 325 helix pairs in our dataset, the 166 actual interacting pairs were predicted with varying specificity and sensitivity, depending on the number N of correlated mutations required for a positive prediction (Table 3). For example, with $N = 5$, i.e. where a helix pair is predicted as interacting in case at least five correlations are found for these helix pair, interacting helices could be predicted with a sensitivity of 42% and a specificity of 83%. A prediction accuracy of 71.9% could be achieved in this case. According to a χ-test, this prediction is significant with a P-value of $2.19E-06$. By raising the threshold N to higher values, the specificity of the prediction rises at the expense of a smaller number of predicted interactions. For example, with $N = 7$ about one-fourth (27.1%) of all interacting helices in the dataset can be predicted with a specificity of around 92.5% (P-value $7.14E-06$). The prediction accuracy increases to 78.9%.

4 CONCLUSION

According to the two-stage model, membrane protein folding involves the packing of preformed α-helices (Popot and Engelman, 2000). Therefore, the number of required

structural restraints for the construction of a structural model is clearly reduced compared to soluble proteins. Although in our dataset of 14 proteins of known crystal structure only a small fraction of predicted correlations involved pairs of residues in physical contact, a sizeable fraction of the correlations was found to be in close vicinity to interhelical contacts. This allows for the prediction of helix pairs that are likely to be in direct contact, possibly along with the approximate interacting region. Together with the analysis of interaction motifs (Walters and DeGrado, 2006) and evolutionarily conserved residues (Fleishman et al., 2004a) this should be a valuable tool for the prediction of membrane protein structures.

REFERENCES

Adamian,L. and Liang,J. (2006) Prediction of transmembrane helix orientation in polytopic membrane proteins. *BMC Struct. Biol.*, **6**, 13.

Altschuh,D. *et al.* (1987) Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *J. Mol. Biol.*, **193**, 693–707.

Altschuh,D. *et al.* (1988) Coordinated amino acid changes in homologous protein families. *Protein Eng.*, **2**, 193–199.

Chelvanayagam,G. *et al.* (1997) An analysis of simultaneous variation in protein structures. *Protein Eng.*, **10**, 307–316.

Chen,Z. and Xu,Y. (2006) Structure prediction of helical transmembrane proteins at two length scales. *J. Bioinform. Comput. Biol.*, **4**, 317–333.

Clarke,N.D. (1995) Covariation of residues in the homeodomain sequence family. *Protein Sci.*, **4**, 2269–2278.

Dekker,J.P. *et al.* (2004) A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. *Bioinformatics (Oxford, England)*, **20**, 1565–1572.

Fares,M.A. and Travers,S.A. (2006) A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics*, **173**, 9–23.

Fariselli,P. *et al.* (2001) Progress in predicting inter-residue contacts of proteins with neural networks and correlated mutations. *Proteins*, **45** (Suppl. 5), 157–162.

Filizola,M. *et al.* (2002) Prediction of heterodimerization interfaces of G-protein coupled receptors with a new subtractive correlated mutation method. *Protein Eng.*, **15**, 881–885.

Fleishman,S.J. *et al.* (2004a) A Calpha model for the transmembrane alpha helices of gap junction intercellular channels. *Mol. Cell*, **15**, 879–888.

Fleishman,S.J. *et al.* (2004b) An evolutionarily conserved network of amino acids mediates gating in voltage-dependent potassium channels. *J. Mol. Biol.*, **340**, 307–318.

Fodor,A.A. and Aldrich,R.W. (2004) Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins*, **56**, 211–221.

Frishman,D. and Mewes,H.W. (1997) Protein structural classes in five complete genomes. *Nat. Struct. Biol.*, **4**, 626–628.

Gloor,G.B. *et al.* (2005) Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. *Biochemistry*, **44**, 7156–7165.

Gobel,U. *et al.* (1994) Correlated mutations and residue contacts in proteins. *Proteins*, **18**, 309–317.

Halperin,I. *et al.* (2006) Correlated mutations: advances and limitations. A study on fusion proteins and on the Cohesin-Dockerin families. *Proteins*, **63**, 832–845.

Hildebrand,P.W. *et al.* (2006) Analysis and prediction of helix-helix interactions in membrane channels and transporters. *Proteins*, **64**, 253–262.

Kass,I. and Horovitz,A. (2002) Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins*, **48**, 611–617.

Krogh,A. *et al.* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.

Kundrotas,P.J. and Alexov,E.G. (2006) Predicting residue contacts using pragmatic correlated mutations method: reducing the false positives. *BMC Bioinformatics*, **7**, 503.

Larson,S.M. *et al.* (2000) Analysis of covariation in an SH3 domain sequence alignment: applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions. *J. Mol. Biol.*, **303**, 433–446.

Lockless,S.W. and Ranganathan,R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, **286**, 295–299.

Martin-Galiano,A.J. and Frishman,D. (2006) Defining the fold space of membrane proteins: the CAMPS database. *Proteins*, **64**, 906–922.

Martin,L.C. *et al.* (2005) Using information theory to search for co-evolving residues in proteins. *Bioinformatics (Oxford, England)*, **21**, 4116–4124.

McLachlan,A.D. (1971) Tests for comparing related amino-acid sequences. Cytochrome c and cytochrome c 551. *J. Mol. Biol.*, **61**, 409–424.

Miyata,T. *et al.* (1979) Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.*, **12**, 219–236.

Neher,E. (1994) How frequent are correlated changes in families of protein sequences? *Proc. Natl Acad. Sci. USA*, **91**, 98–102.

Noivirt,O. *et al.* (2005) Detection and reduction of evolutionary noise in correlated mutation analysis. *Protein Eng. Des. Sel.*, **18**, 247–253.

Oliveira,L. *et al.* (2003) Sequence analysis reveals how G protein-coupled receptors transduce the signal to the G protein. *Proteins*, **52**, 553–560.

Olmea,O. and Valencia,A. (1997) Improving contact predictions by the combination of correlated mutations and other sources of sequence information. *Fold. Des.*, **2**, S25–S32.

Olmea,O. *et al.* (1999) Effective use of sequence correlation and conservation in fold recognition. *J. Mol. Biol.*, **293**, 1221–1239.

Ortiz,A.R. *et al.* (1998a) Fold assembly of small proteins using monte carlo simulations driven by restraints derived from multiple sequence alignments. *J. Mol. Biol.*, **277**, 419–448.

Ortiz,A.R. *et al.* (1998b) Nativelike topology assembly of small proteins using predicted restraints in Monte Carlo folding simulations. *Proc. Natl Acad. Sci. USA*, **95**, 1020–1025.

Ortiz,A.R. *et al.* (1999) Ab initio folding of proteins using restraints derived from evolutionary information. *Proteins*, **37** (Suppl. 3), 177–185.

Pazos,F. and Valencia,A. (2002) In silico two-hybrid system for the selection of physically interacting protein pairs. *Proteins*, **47**, 219–227.

Pazos,F. *et al.* (1997) Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.*, **271**, 511–523.

Pollock,D.D. and Taylor,W.R. (1997) Effectiveness of correlation analysis in identifying protein residues undergoing correlated evolution. *Protein Eng.*, **10**, 647–657.

Pollock,D.D. *et al.* (1999) Coevolving protein residues: maximum likelihood identification and relationship to structure. *J. Mol. Biol.*, **287**, 187–198.

Popot,J.L. and Engelman,D.M. (2000) Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.*, **69**, 881–922.

Pupko,T. *et al.* (2000) A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.*, **17**, 890–896.

Russ,W.P. *et al.* (2005) Natural-like function in artificial WW domains. *Nature*, **437**, 579–583.

Saraf,M.C. *et al.* (2003) Using multiple sequence correlation analysis to characterize functionally important protein regions. *Protein Eng.*, **16**, 397–406.

Schneider,D. and Engelman,D.M. (2004) Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *J. Mol. Biol.*, **343**, 799–804.

Shindyalov,I.N. *et al.* (1994) Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Eng.*, **7**, 349–358.

Socolich,M. *et al.* (2005) Evolutionary information for specifying a protein fold. *Nature*, **437**, 512–518.

Suel,G.M. *et al.* (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.*, **10**, 59–69.

Taylor,W.R. and Hatrick,K. (1994) Compensating changes in protein multiple sequence alignments. *Protein Eng.*, **7**, 341–348.

Thompson,J.D. *et al.* (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

Tusnady,G.E. *et al.* (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res.*, **33**, D275–D278.

Wallin,E. and von Heijne,G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, **7**, 1029–1038.

Walters,R.F. and DeGrado,W.F. (2006) Helix-packing motifs in membrane proteins. *Proc. Natl Acad. Sci. USA*, **103**, 13658–13663.