Web Server

MPTherm-pred: Analysis and Prediction of Thermal Stability Changes upon Mutations in Transmembrane Proteins

A. Kulandaisamy [1], Jan Zaucha [2], Dmitrij Frishman [2,3] and M. Michael Gromiha [1,]*

*1 - **Department of Biotechnology,** Bhupat and Jyoti Mehta School of BioSciences, Indian Institute of Technology Madras, Chennai 600 036, Tamilnadu, India*

*2 - **Department of Bioinformatics,** Technische Universität München, Wissenschaftszentrum Weihenstephan, Freising, Germany*

*3 - **Department of Bioinformatics,** Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation*

Correspondence to M. Michael Gromiha: *gromiha@iitm.ac.in*

https://doi.org/10.1016/j.jmb.2020.09.005
Edited by Michael Sternberg

Abstract

The stability of membrane proteins differs from globular proteins due to the presence of nonpolar membrane-spanning regions. Using a dataset of 929 membrane protein mutations whose effects on thermal stability (ΔT_m) were experimentally determined, we found that the average ΔT_m due to 190 stabilizing and 232 destabilizing mutations occurring in membrane-spanning regions are 2.43(3.1) °C and − 5.48 (5.5) °C, respectively. The ΔT_m values for mutations occurring in solvent-exposed regions are 2.56 (2.82) and − 6.8(7.2) °C. We have systematically analyzed the factors influencing the stability of mutants and observed that changes in hydrophobicity, number of contacts between Cα atoms and frequency of aliphatic residues are important determinants of the stability change induced by mutations occurring in membrane-spanning regions. We have developed structure- and sequence-based machine learning predictors of ΔT_m due to mutations specifically for membrane proteins. They showed a correlation and mean absolute error (MAE) of 0.72 and 2.85 °C, respectively, between experimental and predicted ΔT_m for mutations in membrane-spanning regions on 10-fold group-wise cross-validation. The average correlation and MAE for mutations in aqueous regions are 0.73 and 3.7 °C, respectively. These MAE values are about 50% lower than standard deviations from the mean ΔT_m values. The reliability of the method was affirmed on a test set of mutations occurring in evolutionary independent protein sequences. The developed MPTherm-pred server for predicting thermal stability changes upon mutations in membrane proteins is available at https://web.iitm.ac.in/bioinfo2/mpthermpred/. Our results provide insights into factors influencing the stability of membrane proteins and can aid in designing mutants that are more resistant to thermal stress.

Introduction

Transmembrane proteins have unique structural features that interact with both lipid bilayers and the aqueous environment. These proteins adopt either an α-helical (single or multi-pass) or a β-barrel (in the case of porins) structural conformation across the lipid bilayer and are involved in different biological processes including signal transduction, transporting ions and molecules across the cell or organelle membrane and intercellular communication [1,2]. The stability of membrane proteins is important for maintaining the ability to fulfill their biological roles. The native state of a protein is generally affected by several internal and external factors such as mutations, pH, temperature, ion concentration and many others [3–6]. Among them, missense mutations (also known as single amino acid variants) play a major role in altering folding, stability and functions of membrane proteins, thus leading to different diseases including cancer and cystic fibrosis, among many others [7–11]. For example, the F508S mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) protein, which is associated with the cystic fibrosis disease decreases the thermal stability of the protein by 1.8 °C [12,13].

The stability of membrane proteins and their mutants is experimentally determined by the combination of site-directed mutagenesis and equilibrium unfolding studies. Specifically, biophysical methods such as differential scanning calorimetry, circular dichroism spectroscopy and fluorescence spectroscopy are used to determine the thermodynamic stability parameters such as melting temperature (T_m), enthalpy change, heat capacity change and Gibbs free energy (ΔG) [6,14–18]. These experimentally measured stability data for globular and membrane proteins are accumulated in ProTherm, the thermodynamic database for proteins and mutants [19]. Recently, we have constructed a thermodynamic database specifically for membrane proteins and their mutants, MPTherm (https://www.iitm.ac.in/bioinfo/mptherm/) [20].

Several investigations have been carried out to understand the factors influencing the stability of proteins upon mutations as well as predicting their induced changes in stability [21–36]. These methods are based on physicochemical properties, secondary structure, solvent accessibility, conservation scores, position specific scoring matrices, non-covalent interactions, atomic contacts, statistical potentials, knowledge-based potentials and different force field parameters. Saraboji *et al.* [33] proposed an average assignment method for predicting the thermal stability (ΔT_m) of globular proteins upon mutations. HoTMuSiC [34] and AUTO-MUTE [35] use temperature-dependent statistical potentials and residue environment scores, respectively, for their prediction. Most of the currently available stability prediction methods are general-purpose, developed primarily for globular proteins, and large-scale studies of membrane protein mutations are scarce. Kroncke *et al.* [37] predicted the changes in free energy of 223 membrane protein mutations and reported that the performance all the methods is poor with the maximum correlation of 0.37. Recently, however, a notable advancement has been achieved. Pires *et al.*, [36] developed a method for predicting the Gibbs free energy change ($\Delta\Delta G$) of membrane protein upon mutations using sequence- and structure-based features, obtaining a correlation of 0.72. Nevertheless, up until now, there existed no methods for predicting the thermal stability changes of membrane proteins expressed in terms of altered melting temperatures ΔT_m. With our suite of machine learning predictors MPTherm-pred, we fill this gap and provide a complementary approach to predicting $\Delta\Delta G$.

In particular, we have systematically analyzed the thermal stability changes of missense mutations in different types of membrane proteins and developed topology-specific regression models for predicting the changes in stabilities upon mutations by utilizing various sequence- and structure-based features, including sequence conservation, physicochemical properties, contact potentials, atomic contacts and interaction energies. For mutations in membrane-spanning and aqueous regions, we achieved a correlation of 0.73 and 0.77 and a mean absolute error (MAE) of 3.1 °C and 2.2 °C, respectively using a blind test set of 43 and 51 mutants. Further, our method is able to distinguish between stabilizing and destabilizing mutants with an accuracy of 77%. We have developed a web server for predicting the change in thermal stability of membrane proteins upon mutations, which is freely available at https://web.iitm.ac.in/bioinfo2/mpthermpred/.

Materials and Methods

Data collection and curation

The thermal stability data for membrane protein mutations with known three-dimensional structures, measured in terms of the melting temperature (ΔT_m) were collected from the MPTherm database [20]. To ensure comparable physiological conditions, we considered the data obtained within the pH range of 6–9. For the same mutant with multiple ΔT_m measurements we have taken their average. A positive and negative sign of ΔT_m, represents stabilizing and destabilizing mutations, respectively.

Non-redundant dataset

To avoid biases in the model, we constructed a non-redundant dataset by removing redundancies according to the following conditions [38]: (i) using CD-HIT, we clustered protein sequences based on their sequence identity, with cut-off of <40%, and we included the representative protein mutations in the dataset (ii) for rest of the mutations, we computed the residue conservation at the mutated site by obtaining a multiple sequence alignments within a cluster and non-conservative mutations were included in the dataset (iii) in case of conservative mutations, we looked up the type of amino acid changes and only the mutations with different changes were included. The same strategy has been used in our previous study [38]. These stringent criteria yielded a final dataset of 929 missense mutations from 144 membrane proteins. The complete data set with details on protein name, UniProt ID, mutation, topology, function, affected secondary structure element and solvent accessibility, PubMed ID, experimental ΔT_m, MPTherm database accession number and variants used for group wise cross-validation/test are provided in Table S1 as well as at https://web.iitm.ac.in/bioinfo2/mptherm-pred/dataset.html.

Computation of features for model development

We have derived several sequence and structure-based features for developing prediction methods, which are discussed below.

Sequence-based features

Sequence-based features were grouped into four categories: (i) physicochemical properties, (ii) neighboring residues, (iii) substitution matrices and contact potentials, and (iv) conservation scores and PSSM profiles.

Physicochemical properties

A set of 253 physicochemical, energetic and conformational properties of amino acid residues was collected from the AAindex database and literature [38–40]. The impact of mutations on each property ($\Delta P_{\text{mutation}}$) was computed as the difference between the respective property values in mutant and wild-type residues. It is given by

$$\Delta P_{\text{mutation}} = P_{\text{mutant}} - P_{\text{wild-type}} \tag{1}$$

where P_{mutant} and $P_{\text{wild-type}}$ are the property values of the mutant and wild-type residue, respectively.

Neighboring residue-based features

We have included the information on neighboring residue features for the mutants using the equation:

$$\Delta P_{\text{local}} = P_i\,(\text{mutant}) - \frac{\sum_{n=i-j}^{i+j} P_n\,(\text{wild}-\text{type})}{2j+1} \tag{2}$$

where j = 1, 2 and 3 represent window lengths of 3, 5 and 7, respectively. $2j$ + 1 is the total window length, and 'i' is the mutant position in the sequence.

Further, the 20 amino acid residues were classified into six groups namely, aliphatic-small (G, A), aliphatic-large (I, L, V and P), aromatic (F, W, Y), sulfur-containing (C, M), polar (N, Q, S, T) and charged (D, E, K, R, H). We considered the distribution of amino acid residues belonging to these classes around the mutated site using a window length of 15 residues. Further, the information about the wild-type, mutant and two neighboring residues of the affected site were also considered in our prediction method.

Amino acid substitution matrices and contact potentials

We have collected a set of 94 amino acid substitution matrices from the AAindex2 database [39] and used these data directly for each mutation. We used the immediate N-/C-terminal neighbor residue information for calculating the changes in amino acid contact potentials due to the mutation from 47 contact potentials available in AAindex3 [41].

Conservation score and PSSM profile

We used a set of 18 types of conservation scores for each mutation site obtained from the AACon tool [42,43]. In addition, the position-specific scoring matrix (PSSM) profile was constructed by using PSI-BLAST with three iterations and an E-value cutoff of 0.001 [44]. From the PSSM matrix, we have extracted the PSSM profiles of the wild-type residue, mutant residue and the difference between them.

Structure-based features. Inter-atomic and inter-residue contacts. We have computed various parameters based on atomic and residue contacts of wild-type and mutant structures, which are listed below:

(i) Number of residue contacts between C_α and C_β atoms between the wild-type residue and other residues in the protein within a distance of 8 and 12 Å;
(ii) Normalized residue contacts by the length of the protein;
(iii) Number of contacts made by 12 different types of atoms used in AMBER [45] within a distance of 5 Å;
(iv) Number of heavy atom contacts;
(v) Number of contacts between polar-polar, polar-nonpolar, nonpolar-nonpolar and charged atoms;
(vi) Number of short, medium and long-range interactions [46];
(vii) Surrounding hydrophobicity of wild-type residue and the whole protein [47];
(viii) Long-range order of wild-type residue and whole protein [48];
(ix) Contact order of the protein [49];
(x) Total contact distance of the protein, normalized by the protein's size [50];

Impact of mutations due to surrounding residues at the mutation site

We have included the influence of surrounding residues using the following equation:

$$P_{\text{str (i)}} = P_{\text{mut}-}P_{\text{sur}}(\text{i}) \tag{3}$$

$$P_{\text{sur}}\,(\text{i}) = \sum\nolimits_j n_{ij}.P_j \tag{4}$$

where n_{ij} is the total number of residues of type j surrounded by wild-type residue (i) of the protein within a distance of 8 Å, and P_j is the property value of residue type *j*; P_{mut} is the property value of the mutant residue.

Contact potentials from protein structures

Generic contact potential matrices were collected from the AAindex3 database [39]. We have identified the contacts between the wild-type residue and other residues in the protein within a distance of 8 Å and for each contact, the respective value was retrieved from the contact potential matrix.

Energy terms

The FoldX program was used to compute energetic parameters including electrostatic, van der Waals, hydrogen bond and entropic solvation energies as well as the total energy for the wild-type and mutant protein structures [28]. We have also considered the energy of the interactions between different types of atoms and the total energy obtained with the dDFIRE platform [51].

Other structural features

The residue depth (R_d) and secondary structure annotation of the wild-type residue were computed using the DEPTH and DSSP programs, respectively [52,53]. The wild-type residue solvent accessibility was obtained from the NACCESS program using a probe size of 1.4 Å (H_2O) for water-soluble regions. Similarly, the lipid accessibility of mutations in the membrane regions was determined by setting the size of the probe to 2.0 Å (-CH2) [54].

Membrane protein-based features

Further, we have also obtained membrane protein-specific features including the number of membrane-spanning segments and topology information (inside/outside) from the PDBTM database [55].

Feature selection and model validation

We have used the following procedure to select the features and develop the prediction model [56]: (i) from a systematic search of all possible combinations of two features, we selected the top 50 unique pairs, which showed the highest correlation with the experimentally-determined thermal stability change; (ii) utilizing the top ranked features in these 50 pairs, we have carried out a systematic search across all possible combinations of 5 features; (iii) we implemented the forward feature selection method in which each single feature is included along with the best combination of other features and examined the performance (correlation and MAE; see below); and (iv) extended the forward selection until there was no increase in correlation or decrease in MAE during 10-fold group-wise cross-validation procedure.

During the model development, the dataset was grouped based on sequence identity and thus, we have used the evolutionary independent mutations in the training and test datasets. Further, the model was validated by 10-fold group-wise and leave-one-group-out (ensuring evolutionary independence of sequences in each fold) cross validation. We have considered each sequence cluster as an individual group in the 10-fold group-wise cross-validation procedure. Moreover, the performance of each model was tested with a blind test set of data.

Assessment of prediction performance

We have evaluated the performance of our prediction methods using the Pearson correlation coefficient (*r*) and MAE when comparing experimental and predicted ΔT_m values. Further, sensitivity, specificity and accuracy were used to assess the performance of distinguishing between stabilizing and destabilizing mutants.

$$\text{MAE} = \frac{\sum_{i=1}^{n} | x_{\text{pred}}(i) - x_{\text{exp}}(i) |}{n} \tag{5}$$

where $x_{\text{exp}}(i)$ and $x_{\text{pred}}(i)$ represent the experimental and predicted ΔT_m for i^{th} mutant, respectively; *n* is the total number of mutants.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}} \tag{6}$$

$$\text{Sensitivity (SN)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{Specificity (SP)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{8}$$

$$\text{Accuracy (ACC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{9}$$

$$\text{Balanced accuracy (BAC)} = \frac{\text{SN} + \text{SP}}{2} \tag{10}$$

where, TN, TP, FP and FN represent true negatives, true positives, false positives and false negatives, respectively.

Results and Discussion

Distribution of thermal stability data for membrane protein mutations

The distribution of experimental thermal stability values of 929 membrane protein mutations (ΔT_m) at different intervals is shown in Figure 1. We observed that 42% mutants stabilize membrane proteins, whereas an opposite trend was observed for 58% mutants. Further, 16% and 34% of the mutants have a change in thermal stability of less than ±1 and ±2 °C, respectively. Interestingly, 11% of mutants drastically alter the stability of membrane proteins yielding a ΔT_m of more than

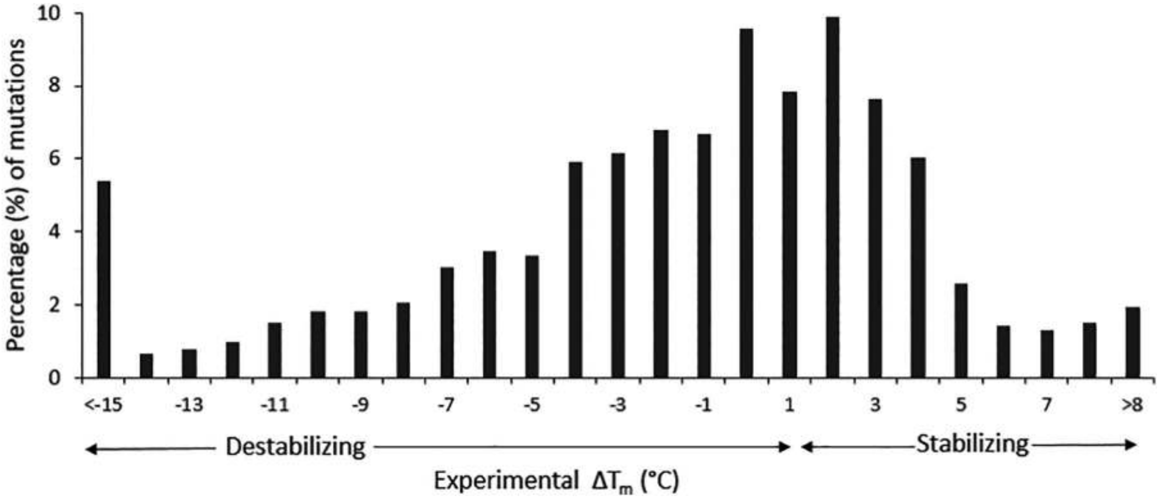

Figure 1. Distribution of thermal stability (ΔT_m) changes caused by membrane protein mutations.

10 °C. The two most extreme cases are for the mutant G149V in the Copper-transporting ATPase 2 protein, which decreases the stability by 38 °C [57] and the T61I in the Envelope glycoprotein gp160 from the Simian immunodeficiency virus, which increases the stability by 21 °C [58].

Stability of mutations in membrane spanning and aqueous regions of membrane proteins

We have classified the location of mutations into two groups, namely membrane and aqueous (including the both cytoplasmic and extracellular) regions based on the protein's topology. We observed that 44% of mutations in membrane-spanning regions and 40% in aqueous regions stabilize membrane proteins. Further, we have classified the mutations based on different kingdoms of life as well as structure and function of membrane proteins, and analyzed the stability. Table S2 presents the average ΔT_m of stabilizing and destabilizing mutations in membrane-spanning and aqueous regions of membrane proteins distinguished based on various classifications.

Mutations in membrane-spanning regions of eukaryotes have similar magnitudes of effects on stability, regardless of the direction of change: the average T_m for stabilizing and destabilizing mutations is 3.24 °C and −3.46 °C, respectively. Further, the T_m of stabilizing mutations is similar in both transmembrane and aqueous regions, within a deviation of 2 °C. In prokaryotes, the average increase in T_m of stabilizing mutations in membrane-spanning regions is half (1.82 °C) of the respective value for eukaryotic proteins (3.24 °C). Interestingly, the average ΔT_m for destabilizing mutations in eukaryotes and prokaryotes are −3.5 °C and −8.9 °C, respectively, suggesting that prokaryotic proteins are more prone to become destabilized upon mutations. Further, in aqueous regions, 33 mutations in viral proteins decrease the T_m by about 15°, which is 2- to 3-fold stronger than the respective effects in eukaryotes and prokaryotes. A similar tendency was also observed for stabilizing mutants.

Stability of mutations based on structural and functional classes

We have compared the effects of mutations on the stability of α-helical and β-barrel membrane protein structures. We observed that in α-helical proteins, 187 stabilizing mutations in membrane spanning regions increase the average T_m by 3 °C, whereas 223 destabilizing mutations showed an average decrease in T_m of −5.4 °C. Further, the effect of destabilizing mutations is higher in aqueous regions (−6.99 °C) than the membrane spanning regions (−5.41 °C). On the contrary, in β-barrel membrane proteins, the effect of destabilizing mutations in membrane spanning regions (−7.4 °C) is stronger than in the aqueous regions (−4.99 °C).

Based on functions, receptor mutations in transmembrane regions enhance the average stability by 4 °C, which is higher than the respective change in aqueous regions (2.8 °C). In addition, transporter mutations have similar average ΔT_m for both stabilizing (2.6 °C) and destabilizing mutations (6.5 to 6.8 °C) in membrane-spanning and aqueous regions. Finally, mutations of multi-pass proteins, on average, increase the protein stability by 3 and 2.5 °C in membrane and aqueous regions, respectively(Table S2).

Comparison of mutations in topologically similar regions of proteins

We compared the thermal stability of mutations in the transmembrane and aqueous regions of membrane proteins against buried and exposed regions of globular proteins, respectively. The thermal stability data of globular protein mutations were retrieved from the ProTherm database [56], and we have considered the mutations, which have at least 10 data on stability in the respective regions. The percentage of common stabilizing mutations in membrane-spanning regions *versus* buried mutations as well as exposed mutations in aqueous regions of membrane and globular proteins are presented in Table S3. Interestingly, we observed that more than 80% of V → A and I → A mutations in transmembrane regions stabilize membrane proteins, whereas this is true for less than 10% of cases in the buried regions of globular proteins. A similar tendency was observed for L → A (Table S3a).

We have analyzed the structural implications of these mutations and observed that membrane protein stability upon changes from large to small hydrophobic residues are attributed with increased packing at helix–helix interfaces, reduction in flexibility and enhanced energetic contribution due to hydrophobic and van der Waals interactions. Notably, this is in stark contrast to the effects observed in globular proteins. Specifically, (i) small aliphatic residues (Ala, Gly) and small hydroxyl-containing residues (Ser, Thr) exhibit a high packing tendency in membrane regions, primarily at helix–helix interfaces [59,60], (ii) mutation of Leu to Val and Ile to Ala diminishes steric interactions in transmembrane helix associations [61,62] while (iii) large hydrophobic (Leu, Ile, Val) and aromatic (Phe, Tyr, Trp) residues are major contributors of protein stability in the hydrophobic cores of globular proteins [59].

In aqueous regions, 75% of K → A mutations are stabilizing in membrane proteins, whereas only 25% of mutations in exposed regions of globular proteins show a similar trend (Table S3b). The effect of R → A is also more pronounced in membrane proteins than in globular proteins. E → A has a similar tendency in both membrane and globular proteins.

Prediction on thermal stability of membrane protein mutations

We have developed machine learning predictors of the change in stability of membrane proteins upon mutations, separately for two topological environments: (i) membrane and (ii) aqueous. Mutations in membrane-spanning regions were classified into two groups based on their functions: (i) receptors and (ii) others, which include transporters and enzymes. The functional information on membrane proteins were retrieved from the UniProt database. Saraboji *et al.* (2006) showed that the classification of mutations based on their secondary structure and solvent accessibility improved the prediction performance of the stability change of globular proteins upon mutations [33]. Hence, mutations in the aqueous environment were grouped into three categories: (i) buried mutations in regular secondary structures (helices and strands), (ii) exposed mutations in regular secondary structures and (iii) coil mutations. In each group, the experimental thermal stability change is related with sequence and structure-based features of membrane protein mutants using multiple linear regression and the details are provided in Tables S4–S8.

Mutations in membrane spanning regions of transporters and other proteins

We have developed a specific method for mutations, which are located in membrane-spanning regions of transporters and enzymes. This dataset comprises of 286 mutations from 34 proteins and the thermal stability changes encompass a wide range from −28.1 to 11.4 °C (Figure S1a). We developed a regression model using 12 sequence and structure based features, which showed a correlation and MAE of 0.75 with 3.2 °C, respectively, between the experimental and the predicted thermal stability change on 10-fold group-wise cross-validation (Figure 2(a)). Further, the method was also validated with leave-one-group-out and we obtained a correlation of 0.75 (Table S9).

The selected structure-based features are primarily based on the change in the backbone hydrogen bond energy and the total energy from the FoldX program, the difference in van der Waals contacts and the number of carbon atoms within a 5 Å distance, the surrounding hydrophobicity of the wild-type residue and the average number of Cα atom contacts of the wild-type residue. The selected sequence-based features include the bulkiness of the residue, transmembrane-helix forming tendency, inter-residue protein contact energies (MIYS990101), the conservation score obtained by the TAYLOR method, residue depth-dependent amino acid substitution matrices and the number of aliphatic small-sized residues (Gly and Ala) (Table S10).

Mutations in membrane spanning regions of receptors

In receptors, 136 mutations are located in transmembrane segments of 17 proteins and their induced thermal stability change varies from −9.4 to 13.8 °C. In this dataset, 52% of mutations are destabilizing and their average change in thermal stability is −3.6(2.5) °C. We used 12 sequence and structure-based features (Table S10) and developed a method for predicting the thermal stability change upon mutation. Our method showed a correlation of 0.68 with MAE of 2.5 °C in 10-fold group-wise cross-validation between experimental and predicted ΔT_m (Figure S1b).

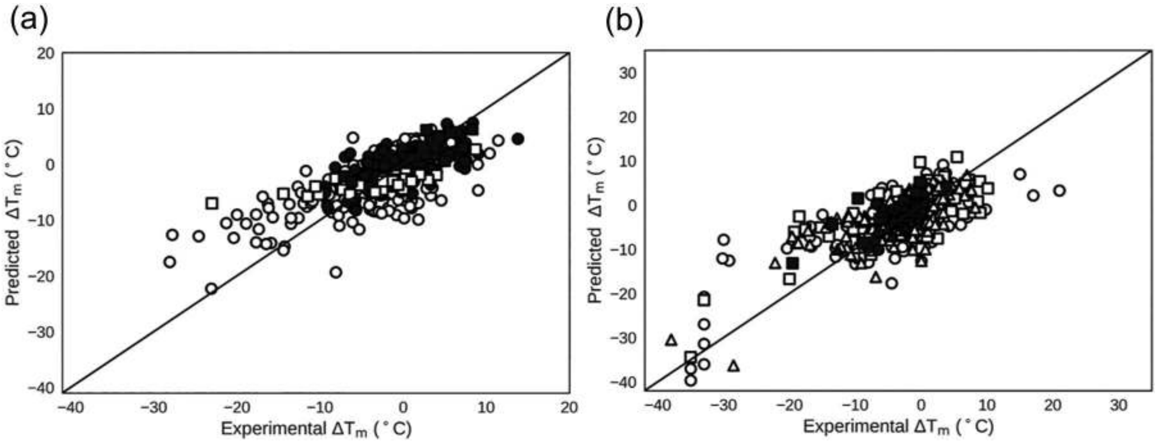

Figure 2. Performance of MPTherm-pred in 10-fold group-wise cross-validation and on the blind test dataset for each model. (a) Membrane-spanning; ○: 10-fold group-wise cross-validation, transporters; ●: 10-fold group-wise cross-validation, receptors; □: blind test, transporters; ■: blind test, receptors. (b) Aqueous regions: ○: 10-fold group-wise cross-validation, exposed; □: 10-fold group-wise cross-validation, buried; △: 10-fold group-wise cross-validation, coil; ●: blind test, exposed; ■: blind test, buried; ▲: blind test, coil.

Buried mutations in regular secondary structures

The dataset of buried mutations in helical and strand segments contains 127 mutations from 40 proteins. Among them, 43% are stabilizing with an average ΔT_m of 2.9(2.6) °C and 57% are destabilizing and ΔT_m vary from −35 to 10.13 °C. We have developed a multiple regression model using 11 sequence and structural features to predict the change in thermal stability. The selected features include van der Waals energy, solvent accessibility of non-polar atoms, torsional angles, contact potentials, polarizability, frequency of polar residues around the mutation site and conservation score (Table S10). Our method achieved a correlation of 0.74 and MAE of 3.6 °C between experimental and predicted change in thermal stability on leave-one-group out cross validation (Figure S1c).

Exposed mutations in helical and strand segments

One hundred eighty mutations occurred in secondary structures (helical or strand) of solvent-exposed segments across 58 membrane proteins. The changes in thermal stability (ΔT_m) imposed by these mutations are scattered between −35 to 21 °C and their mean is −3.8 °C. We employed 15 different features, including electrostatic energy, contact potentials, change in solvent accessibility of polar atoms, change in hydrophobicity, number of heavy atomic contacts of the wild-type residue and the PSSM profile, to predict the change in thermal stability (ΔT_m) of the proteins upon these mutations. These resultant model predicts the change in thermal stability within a MAE of 4.3 °C on 10-fold group-wise cross-validation (Figure S1d).

Coil mutations in aqueous regions

The thermal stability change (ΔT_m) of coil mutations in aqueous regions varies from −38 to 8.2 °C. The dataset consists of 199 mutations across 63 proteins. We have used 11 sequence and structure-based features for predicting the change in the thermal stability, which are presented in Table S10. We obtained a correlation of 0.73 and 0.79 on 10-fold group-wise cross-validation and test set, respectively (Figure S1e). The MAEs of 10-fold group-wise cross-validation are 3.2 °C (Figure 2(b)).

Validation on a blind test set of mutants

We have constructed a test set comprising 94 mutations that occurred in evolutionarily independent sequences. The number of mutants in five categories based on topology and location is presented in Table S11. The performance of our methods for predicting the change in thermal stability and distinguishing between stabilizing and destabilizing mutations is also included in this table. We observed that our method is capable of distinguishing between stabilizing and destabilizing mutants in membrane spanning regions of receptors and transporters at an accuracy of 100% and 76%, respectively. Further, our method predicted the change in thermal stability of receptors and transporters with a correlation and MAE of 0.7 and 2.4 °C, and 0.76 and 3.7 °C, respectively.

For buried and exposed mutations, which are located in regular secondary structural elements, we achieved an accuracy of 72% and 88%, respectively. Further, our method showed a correlation of 0.79 and MAE of 1.4 °C for a set of 20 mutations in coil regions with a classification accuracy of 85%. Overall, our method predicted the stability of 94 evolutionary independent membrane protein mutations with a correlation, MAE and accuracy of 0.70, 2.7 °C and 83%, respectively.

Comparison with existing methods

We have compared the performance of our method with other existing methods in the literature with respect to two aspects: (i) predicting the change in thermal stability and (ii) distinguishing between stabilizing and destabilizing mutations. We noticed that only two methods, namely, AUTOMUTE [35] and HoTMuSiC [34] are available in the literature for predicting the change in thermal stability upon mutation and both of the methods were developed as general-purpose predictors trained primarily on globular proteins. The correlation and MAE of our method on a test set of 94 mutations along with the results obtained with AUTOMUTE [35] and HoTMuSiC [34] are presented in Table 1. Our method predicted the change in thermal stability within a MAE of 2.7 °C, whereas it is 4.2 °C and 4 °C using AUTOMUTE and HoTMuSiC, respectively. This comparison showed that the performance of our method is higher than that of the other methods existing in the literature, most likely due to the fact that our method is specifically designed for membrane proteins. Further, we have examined the capability of distinguishing between stabilizing and destabilizing mutations using eight different methods available in the literature. All existing methods are developed for globular proteins except mCSM-membrane and for predicting the free energy change upon mutation [36]. The discrimination results are included in Table 1. We observed that mCSM-membrane distinguished between stabilizing and destabilizing mutants at an accuracy of 67%. In comparison, our method showed an accuracy of 83%, which makes it better than all other existing methods in the literature.

In addition, we have examined the performance of our method in predicting the effects of the previously discussed large to small residue replacements (Val to Ala, Leu to Ala and Ile to Ala) mutations. Our method showed a balanced accuracy of 76% in distinguishing between the stabilizing and destabilizing mutants. Further, 69% of mutants were predicted within the MAE of 3 °C. On the other hand, the performance of the methods FoldX, AUTOMUTE and HOTMuSiC, which were developed for globular proteins, showed a balanced accuracy of 47%, 50% and 57% respectively. Furthermore, only 9% and 46% of mutants were predicted within the MAE of 3 °C using AUTOMUTE and HOTMuSIC, respectively.

Web server development

We have developed a web server for predicting the change in thermal stability (ΔT_m) of membrane protein mutations and it is freely available at https://web.iitm.ac.in/bioinfo2/mpthermpred/. The web server takes the PDB code, mutation information and the functional class of the protein

Table 1 Comparison of MPTherm-pred with other existing methods on a test set of 94 mutations

Methods	Prediction performance	
	Correlation (*r*)	MAE (°C)
AUTOMUTE SVM [35]	−0.02	4.0
AUTOMUTE_REPT [35]	0.15	4.5
HoTMuSiC [34]	0.12	4.0
MPTherm-pred	0.73	2.7
	Discrimination performance	
	Accuracy (%)	**Balanced accuracy (%)**
DynaMut [24]	51.6	48.5
ENCoM (DynaMut) [24]	60.4	52.6
mCSM [26]	63.7	46.4
SDM [63]	54.9	55.2
DUET [27]	59.3	50.7
SDM2 [64]	54.9	50.0
I-MUTANT 3 [21]	72.8	53.1
FoldX [28]	50.0	59.5
MAESTRO [29]	55.3	51.3
MuPro [22]	72.3	51.1
CUPSAT_Denaturation [30]	39.8	51.8
CUPSAT_Thermal [30]	59.0	59.0
PoPMuSiC [31]	37.3	56.0
INSP [32]	67.8	53.4
mCSM-membrane [36]	67.0	52.2
MPTherm-pred	83.0	77.6

(receptor or others) as inputs. It automatically retrieves the membrane topology information from the PDBTM database [55]. Further, based on the membrane topology and structural conformation, it chooses the appropriate method and displays the output results in a table format, which contains the information about PDB accession, mutations and their effect on the membrane protein's stability. A detailed tutorial is also provided on the web page.

Conclusions

In this study, we have constructed a non-redundant dataset comprising changes in thermal stability (ΔT_m) of membrane proteins upon mutations, taken from the MPTherm database and analyzed the importance of sequence and structure-based features for predicting the change in stability. We have developed topology-specific multiple linear regression models for predicting ΔT_m, which showed a correlation and MAE of 0.73 and 3.2 °C, respectively, between experimental and predicted stabilities on 10-fold group-wise cross-validation. It could also successfully distinguish between stabilizing and destabilizing mutants at an accuracy of 77%. The performance of our method is better than other tools that are currently available. The developed models are provided in a web server, which can be used for large-scale stability predictions on membrane protein mutations. The tool is useful for designing new stable mutants, understanding the factors influencing the stability of membrane proteins and exploring the complex relationships between membrane protein structure, stability and function.

Acknowledgments

We thank the Bioinformatics facility, Department of Biotechnology and Indian Institute of Technology Madras for computational facilities. The work is partially funded by the Department of Science and Technology, Government of India (INT/RUS/RSF/P-09) to M.M.G. and of the Deutsche Forschungsgemeinschaft for a research grant (FR 1411/14) to D.F. A.K. is supported by the Ministry of Human Resource and Development, Government of India.

Declarations of Interest

None.

Credit Author Statement

M.M.G. and D.F. conceived the work; M.M.G., D. F., A.K. and J.Z. desgined experiments; A.K. carried out the computations; A.K., J.Z., D.F. and M.M.G. analyzed the results; A.K. drafted the manuscript; J.Z., D.F. and M.M.G. refined the manuscript; All authors read and finalized the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2020.09.005.

Received 15 August 2020;
Accepted 4 September 2020;

Keywords:
membrane proteins;
thermal stability;
missense mutations;
stabilizing and destabilizing;
disease-causing mutations

Abbreviations:
PSSM, position-specific scoring matrix; MAE, mean absolute error

References

1. Nugent, T., Jones, D.T., (2012). Membrane protein structural bioinformatics. *J. Struct. Biol.*, **179**, 327–337.
2. Almén, M.S., Nordström, K.J., Fredriksson, R., Schiöth, H. B., (2009). Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.*, **7**, 50.
3. Talley, K., Alexov, E., (2010). On the pH-optimum of activity and stability of proteins. *Proteins*, **78**, 2699–2706.
4. Tokuriki, N., Tawfik, D.S., (2009). Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.*, **19**, 596–604.
5. González Flecha, F.L., (2017). Kinetic stability of membrane proteins. *Biophys. Rev.*, **9**, 563–572.
6. Harris, N.J., Booth, P.J., (2012). Folding and stability of membrane transport proteins in vitro. *Biochim. Biophys. Acta*, **1818**, 1055–1066.
7. Stefl, S., Nishi, H., Petukh, M., Panchenko, A.R., Alexov, E., (2013). Molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol.*, **425**, 3919–3936.
8. Marinko, J.T., Huang, H., Penn, W.D., Capra, J.A., Schlebach, J.P., Sanders, C.R., (2019). Folding and misfolding of human membrane proteins in health and disease: from single molecules to cellular proteostasis. *Chem. Rev.*, **119**, 5537–5606.
9. Sanders, C.R., Myers, J.K., (2004). Disease-related misassembly of membrane proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 25–51.
10. Kulandaisamy, A., Priya, S.B., Sakthivel, R., Frishman, D., Gromiha, M.M., (2019). Statistical analysis of disease-causing and neutral mutations in human membrane proteins. *Proteins*, **87**, 452–466.
11. Zaucha, J., Heinzinger, M., Kulandaisamy, A., Kataka, E., Salvádor, Ó.L., Popov, P., Rost, B., Gromiha, M.M., et al., (2020). Mutations in transmembrane proteins: diseases,

evolutionary insights and prediction. *Brief. Bioinformatics*,. https://doi.org/10.1093/bib/bbaa132 (in press).

12. Rabeh, W.M., Bossard, F., Xu, H., Okiyoneda, T., Bagdany, M., Mulvihill, C.M., et al., (2012). Correction of both NBD1 energetics and domain interface is required to restore ΔF508 CFTR folding and function. *Cell*, **148**, 150–163.
13. Kulandaisamy, A., Binny Priya, S., Sakthivel, R., Tarnovskaya, S., Bizin, I., Hönigschmid, P., et al., (2018). MutHTP: mutations in human transmembrane proteins. *Bioinformatics*, **34**, 2325–2326.
14. Gao, K., Oerlemans, R., Groves, M.R., (2020). Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.*, **12**, 85–104.
15. Chang, Y.C., Bowie, J.U., (2014). Measuring membrane protein stability under native conditions. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 219–224.
16. Senisterra, G., Chau, I., Vedadi, M., (2012). Thermal denaturation assays in chemical biology. *Assay Drug Dev. Technol*, **10**, 128–136.
17. Liu, W., Hanson, M.A., Stevens, R.C., Cherezov, V., (2010). LCP-tm: an assay to measure and understand stability of membrane proteins in a membrane environment. *Biophys. J.*, **98**, 1539–1548.
18. Roman, E.A., González Flecha, F.L., (2014). Kinetics and thermodynamics of membrane protein folding. *Biomolecules*, **4**, 354–373.
19. Bava, K.A., Gromiha, M.M., Uedaira, H., Kitajima, K., Sarai, A., (2004). ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, **32**, D120–D121.
20. Kulandaisamy, A., Sakthivel, R., Gromiha, M.M., (2020). MPTherm: database for membrane protein thermodynamics for understanding folding and stability. *Brief. Bioinformatics*,. https://doi.org/10.1093/bib/bbaa064 (in press).
21. Capriotti, E., Fariselli, P., Casadio, R., (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, **33**, W306–W310.
22. Cheng, J., Randall, A., Baldi, P., (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, **62**, 1125–1132.
23. Masso, M., Vaisman, I.I., (2008). Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. *Bioinformatics*, **24**, 2002–2009.
24. Rodrigues, C.H., Pires, D.E., Ascher, D.B., (2018). DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.*, **46**, W350–W355.
25. Pandurangan, A.P., Blundell, T.L., (2020). Prediction of impacts of mutations on protein structure and interactions: SDM, a statistical approach, and mCSM, using machine learning. *Protein Sci.*, **29**, 247–257.
26. Pires, D.E., Ascher, D.B., Blundell, T.L., (2014). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, **30**, 335–342.
27. Pires, D.E., Ascher, D.B., Blundell, T.L., (2014). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.*, **42**, W314–W319.
28. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L., (2005). The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.
29. Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S., Lackner, P., (2015). MAESTRO—multi agent stability prediction upon point mutations. *BMC Bioinformatics*, **16**, 116.
30. Parthiban, V., Gromiha, M.M., Schomburg, D., (2006). CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.*, **34**, W239–W242.
31. Dehouck, Y., Kwasigroch, J.M., Gilis, D., Rooman, M., (2011). PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, **12**, 151.
32. Savojardo, C., Fariselli, P., Martelli, P.L., Casadio, R., (2016). INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics*, **32**, 2542–2544.
33. Saraboji, K., Gromiha, M.M., Ponnuswamy, M.N., (2006). Average assignment method for predicting the stability of protein mutants. *Biopolymers*, **82**, 80–92.
34. Pucci, F., Bourgeas, R., Rooman, M., (2016). Predicting protein thermal stability changes upon point mutations using statistical potentials: introducing HoTMuSiC. *Sci. Rep.*, **6**, 23257.
35. Masso, M., Vaisman, I.I., (2014). AUTO-MUTE 2.0: a portable framework with enhanced capabilities for predicting protein functional consequences upon mutation. *Adv. Bioinforma.*, **2014**, 278385.
36. Pires, D.E., Rodrigues, C.H., Ascher, D.B., (2020). mCSM-membrane: predicting the effects of mutations on transmembrane proteins. *Nucleic Acids Res.*, **48**, W147–W153.
37. Kroncke, B.M., Duran, A.M., Mendenhall, J.L., Meiler, J., Blume, J.D., Sanders, C.R., (2016). Documentation of an imperative to improve methods for predicting membrane protein stability. *Biochemistry*, **55**, 5002–5009.
38. Kulandaisamy, A., Zaucha, J., Sakthivel, R., Frishman, D., Michael Gromiha, M., (2020). Pred-MutHTP: prediction of disease-causing and neutral mutations in human transmembrane proteins. *Hum. Mutat.*, **41**, 581–590.
39. Kawashima, S., Pokarowski, P., Pokarowska, M., Kolinski, A., Katayama, T., Kanehisa, M., (2008). AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res.*, **36**, D202–D205.
40. Gromiha, M.M., (2005). A statistical model for predicting protein folding rates from amino acid sequence with structural class information. *J. Chem. Inf. Model.*, **45**, 494–501.
41. Anoosha, P., Huang, L.T., Sakthivel, R., Karunagaran, D., Gromiha, M.M., (2015). Discrimination of driver and passenger mutations in epidermal growth factor receptor in cancer. *Mutat. Res.*, **780**, 24–34.
42. Valdar, W.S., (2002). Scoring residue conservation. *Proteins*, **48**, 227–241.
43. Manning, J.R., Jefferson, E.R., Barton, G.J., (2008). The contrasting properties of conservation and correlated phylogeny in protein functional residue prediction. *BMC Bioinformatics*, **9**, 51.
44. Bhagwat, M., Aravind, L., (2007). PSI-BLAST tutorial. *Methods Mol. Biol.*, **395**, 177–186.
45. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz Jr., K.M., Ferguson, et al., (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

46. Gromiha, M.M., Selvaraj, S., (2004). Inter-residue interactions in protein folding and stability. *Prog. Biophys. Mol. Biol.*, **86**, 235–277.
47. Manavalan, P., Ponnuswamy, P.K., (1978). Hydrophobic character of amino acid residues in globular proteins. *Nature*, **275**, 673–674.
48. Gromiha, M.M., Selvaraj, S., (2001). Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J. Mol. Biol.*, **310**, 27–32.
49. Plaxco, K.W., Simons, K.T., Baker, D., (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.*, **277**, 985–994.
50. Zhou, H., Zhou, Y., (2002). Folding rate prediction using total contact distance. *Biophys. J.*, **82**, 458–463.
51. Yang, Y., Zhou, Y., (2008). Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins*, **72**, 793–803.
52. Chakravarty, S., Varadarajan, R., (1999). Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure*, **7**, 723–732.
53. Kabsch, W., Sander, C., (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
54. NACCESS, (1993). V2.1.1 . A Computer Program for Solvent Accessible Area Calculations. Department of Biochemistry and Molecular Biology, University College London, London, UK.
55. Kozma, D., Simon, I., Tusnády, G.E., (2013). PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.*, **41**, D524–D529.
56. Rawat, P., Prabakaran, R., Kumar, S., Gromiha, M.M., (2020). AggreRATE-Pred: a mathematical model for the prediction of change in aggregation rate upon point mutation. *Bioinformatics*, **36**, 1439–1444.
57. Kumar, R., Ariöz, C., Li, Y., Bosaeus, N., Rocha, S., Wittung-Stafshede, P., (2017). Disease-causing point-mutations in metal-binding domains of Wilson disease protein decrease stability and increase structural dynamics. *Biometals*, **30**, 27–35.
58. Ji, H., Bracken, C., Lu, M., (2000). Buried polar interactions and conformational stability in the simian immunodeficiency virus (SIV) gp41 core. *Biochemistry*, **39**, 676–685.
59. Eilers, M., Shekar, S.C., Shieh, T., Smith, S.O., Fleming, P. J., (2000). Internal packing of helical membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 5796–5801.
60. Heydenreich, F.M., Vuckovic, Z., Matkovic, M., Veprintsev, D.B., (2015). Stabilization of G protein-coupled receptors by point mutations. *Front. Pharmacol.*, **6**, 82.
61. Bowie, J.U., (2001). Stabilizing membrane proteins. *Curr. Opin. Struct. Biol.*, **11**, 397–402.
62. Mackenzie, K.R., (2006). Folding and stability of alpha-helical integral membrane proteins. *Chem. Rev.*, **106**, 1931–1977.
63. Worth, C.L., Preissner, R., Blundell, T.L., (2011). SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.*, **39**, W215–W222.
64. Pandurangan, A.P., Ochoa-Montaño, B., Ascher, D.B., Blundell, T.L., (2017). SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.*, **45**, W229–W235.