Author/s:
Pan, Q;Nguyen, TB;Ascher, DB;Pires, DEV

Title:
Systematic evaluation of computational tools to predict the effects of mutations on protein stability in the absence of experimental structures

Date:
2022-03-01

Citation:
Pan, Q., Nguyen, T. B., Ascher, D. B. & Pires, D. E. V. (2022). Systematic evaluation of computational tools to predict the effects of mutations on protein stability in the absence of experimental structures. Briefings in Bioinformatics, 23 (2), https://doi.org/10.1093/bib/bbac025.

Persistent Link:
https://hdl.handle.net/11343/315868

OXFORD

Briefings in Bioinformatics, 2022, **23(2)**, 1–13
https://doi.org/10.1093/bib/bbac025
Review

Systematic evaluation of computational tools to predict the effects of mutations on protein stability in the absence of experimental structures

Qisheng Pan, Thanh Binh Nguyen, David B. Ascher and Douglas E.V. Pires

Corresponding authors: Douglas E.V. Pires. E-mail: douglas.pires@unimelb.edu.au; David B. Ascher. Tel.: +61 90354794; E-mail: d.ascher@uq.edu.au

Abstract

Changes in protein sequence can have dramatic effects on how proteins fold, their stability and dynamics. Over the last 20 years, pioneering methods have been developed to try to estimate the effects of missense mutations on protein stability, leveraging growing availability of protein 3D structures. These, however, have been developed and validated using experimentally derived structures and biophysical measurements. A large proportion of protein structures remain to be experimentally elucidated and, while many studies have based their conclusions on predictions made using homology models, there has been no systematic evaluation of the reliability of these tools in the absence of experimental structural data. We have, therefore, systematically investigated the performance and robustness of ten widely used structural methods when presented with homology models built using templates at a range of sequence identity levels (from 15% to 95%) and contrasted performance with sequence-based tools, as a baseline. We found there is indeed performance deterioration on homology models built using templates with sequence identity below 40%, where sequence-based tools might become preferable. This was most marked for mutations in solvent exposed residues and stabilizing mutations. As structure prediction tools improve, the reliability of these predictors is expected to follow, however we strongly suggest that these factors should be taken into consideration when interpreting results from structure-based predictors of mutation effects on protein stability.

Keywords: AlphaFold2, mutation effects on protein stability, homology modelling, performance evaluation

Introduction

Proteins are considered metastable, with their stability intricately linked to their structure and function. Small changes in the protein sequence can lead to large effects on overall protein stability and dynamics, and have been associated with a range of genetic diseases [1–15] and even drug resistance [16–33]. The ability to accurately predict these effects has broad potential applications, not only in interpreting the molecular mechanisms of novel variants [34–36], but also in the industry, where the design of more stable enzymes is of significant importance [37–40].

Experimental approaches to measuring the impact of changes in protein sequence to protein stability, for example thermal melts and urea denaturation, have focussed on measuring the change in Gibbs Free Energy (ΔΔG, expressed in Kcal/mol) of folding by comparing the stability of the purified wild-type and mutant proteins [41]. Although these approaches provide direct experimental insight into protein stability, they are costly and time consuming. This makes it prohibitive to test the effects of every possible mutation and combination of mutations experimentally, and hence, has driven interest in computational approaches to guide more rational mutation analysis and design.

Over the last 20 years, a range of computational approaches have been developed for large-scale studies of the effects of mutations on protein thermodynamics stability. Although they have used a range of different approaches, including statistical [42,43], machine learning [44–54] and energy calculations [55–57], the vast majority have relied upon experimentally solved 3D structures and biophysical measurements in their development [58].

Determination of protein structures, however, is not always straightforward, with a significant number of protein structures yet to be determined experimentally. In the absence of experimental information, homology modelling [59,60] has been widely used to build a 3D model of a protein from its amino acid sequence

Qisheng Pan is a PhD candidate at the School of Chemistry and Molecular Bioscience of the University of Queensland. His research interests include characterising and predicting mutation effects on protein function and structure.
Thanh Binh Nguyen is a postdoctoral research fellow at the Baker Institute of Heart and Diabetes and the University of Queensland. Her interest is in computational biology and understanding protein structures, function and their interactions.
David B. Ascher is Deputy Director of Biotechnology at The University of Queensland, and head of Computational Biology and Clinical Informatics at the Baker Institute and Systems and computational Biology at Bio21 Institute. He is interested is in developing and applying computational tools to assist leveraging clinical and omics data for drug discovery and personalised medicine.
Douglas E.V. Pires is a Senior Lecturer in Digital Health with the School of Computing and Information Systems at the University of Melbourne and group leader at Bio21 Institute. He is a computer scientist and bioinformatician specialising in machine learning and AI and the development of tools to analyse omics data.
Received: July 19, 2021. **Revised:** January 13, 2022. **Accepted:** January 30, 2022

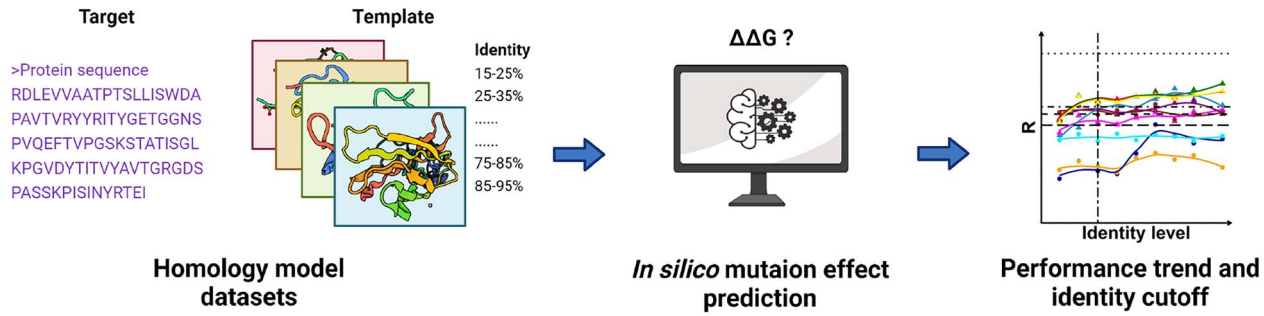

Figure 1. Analysis workflow for assessing the performance of mutation effect predictors using homology models in different sequence identity ranges.

based on an alignment with a similar protein with known structure, or template. In general, the higher the sequence identity to the template, the more reliable the homology model is likely to be [61]. Although homology models have been widely used to guide interpretation of the effects of mutations using these structure-based tools, it has not been well-established how inaccuracies introduced during the homology modelling affect their reliability and accuracy. In addition to template-based approaches for protein structure prediction, more recently, the development of AlphaFold2 [62] has revolutionized the field with a significant increase in performance, promising to bridge the gap in protein 3D information with high-quality models [63]. Understanding how the use of these protein models affect predictive performance is important to ensure that they are used and interpreted appropriately.

To address this, we have systematically evaluated the effect of homology models on the performance of ten publicly available computational tools for predicting the effects of missense mutations on protein stability.

Materials and methods

The methodology for the present work can be divided into four main steps (depicted in Figure 1): (i) data acquisition, including collecting experimentally solved protein structures and effects of missense mutation on protein stability; (ii) generating homology models for different identity ranges; (iii) predicting effects of mutations using generated models for a range of available tools; and (iv) comparative analysis of predictive tools.

Mutation dataset linking effects on stability to experimental protein structures

Over the years, considerable efforts have been dedicated to extracting and manually curating experimentally derived protein thermodynamics information from the literature, including the effects of mutations on stability [64–67] and interactions [68,69], and linking these to high-resolution protein structures. The dataset used in this work was derived from ProTherm [64], and links experimentally measured thermodynamics effects of missense mutations to a diverse set of protein structures. A subset of ProTherm, the S2648 dataset [70], was selected and is composed of 2648 single point missense mutations across 132 unique globular proteins, with a range of mutation effects (Supplementary Figure S1) expressed as the difference in Gibbs Free Energy (ΔΔG) between wild-type (ΔG_{WT}) and mutant (ΔG_{MT}) as follows:

$$\Delta\Delta G = \Delta G_{WT} - \Delta G_{MT}$$

Where positive values denote mutations leading to increased protein stability, while negative values denote destabilizing mutations. Mutations on S2648 are mapped to protein structures solved by either Crystallography with X-ray diffraction or nuclear magnetic resonance (NMR), with 77% of mutations leading to a decrease in protein stability, as observed previously and in other datasets [49,71]. This dataset has been extensively used over the past decade as a benchmark for computational method development aiming to assess mutation effects on stability [44,46,47,70] and, therefore, has been extensively curated and manually inspected. Although this dataset has been used for development purposes before by different methods, the overarching goal of this work is not to assess global performance of the available methods, but rather evaluate how they cope in the absence of experimental structures, when presented with homology models built at different identity thresholds, and how this would impact their relative performance.

Generating homology models at varied identity levels

In homology modelling, it is well-established that there is a correlation between template identity level and the reliability and quality of models generated [61], despite recent advances in the field pushing the boundaries of both *de novo* [72] and template-free modelling [73]. In order to assess the robustness of currently available predictive methods to input uncertainty and noise, we compared their performance presenting them with the same mutation dataset mapped to homology models built using templates at different sequence identity levels. To achieve this, template candidates for the 132 proteins contained in the S2648 dataset were divided into eight groups, with target-template sequence identities in the following ranges: 15–25%, 25–35%, 35–45%, 45–55%, 55–65%, 65–75%, 75–85% and 85–95%. In addition, performance on the experimental structures was used as upper baseline (e.g. 100% identity dataset). A summary of the developed datasets is shown in Table 1.

Table 1. Distributions of the structure- and sequence-based properties of homology model and experimental structure datasets

Identity	# PDBs	# mutations	RSA		Residue depth		Secondary structure types				Secondary structure composition based on CATH			Change of residue volume		
			Buried (%)	Exposed (%)	Deep (%)	Shallow (%)	Alpha helix (%)	Beta sheet (%)	Turn (%)	Random coil (%)	Mainly alpha (%)	Mainly beta (%)	Mixed alpha/beta (%)	L2S (%)	S2L (%)	S2S (%)
15–25	58	819	50.06	49.94	54.70	45.30	35.90	27.35	5.62	21.12	20.27	27.84	51.89	38.1	14.41	47.50
25–35	81	1143	59.14	40.86	57.66	42.34	30.45	39.28	6.39	15.31	15.84	25.11	56.78	36.83	15.05	48.12
35–45	86	1421	53.27	46.73	54.61	45.39	31.60	31.39	8.94	17.38	32.09	21.25	44.83	35.61	16.40	47.99
45–55	65	1078	53.62	46.38	55.29	44.71	28.01	34.14	9.18	15.40	28.39	21.11	49.75	35.06	16.33	48.61
55–65	64	995	57.79	42.21	52.76	47.24	27.14	34.47	8.74	18.09	28.64	21.11	49.75	29.55	19.70	50.75
65–75	47	931	51.34	48.66	50.27	49.73	29.75	29.22	9.67	17.62	35.45	26.10	35.66	36.31	15.04	48.66
75–85	42	918	53.38	46.62	56.10	43.90	38.24	30.17	5.34	17.21	32.24	20.48	46.73	33.66	16.23	50.11
85–95	56	1573	47.36	52.64	47.23	52.77	33.76	29.94	8.26	17.36	23.9	44.88	28.10	37.19	15.58	47.23
100	124	2492	50.76	49.24	61.16	38.84	32.95	30.82	10.23	15.25	26.12	33.15	38.56	38.32	14.69	46.99

Homology modelling was performed using MODELLER version 9.24 [74]. Potential templates were searched in the *pdb_95.bin* database (a cluster at 95% sequence identity in MODELLER) using *blastp* with the target sequence as input. Target-template alignments were manually inspected and a representative template selected per protein/per identity range based on the following criteria:

1. Target-template coverage >75%;
2. Template structure determined by crystallography with X-ray diffraction;
3. Best quality models were selected based on DOPE score.

Models generated were submitted to FoldX [55] for minimization and refinement. The structural similarity between homology models and protein experimental structures was then inspected using the root mean square deviation (RMSD) calculated by the *align* and *rms_cur* command in Pymol [75–77] and TM-score calculated by TM-align [78]. The full list of templates and targets used is available as Supplementary Materials.

Generating high-quality protein models via AlphaFold2

The AlphaFold2 program developed by DeepMind dominated the 14th Critical Assessment of protein Structure Prediction (CASP14) [79], representing a significant advance in the field. The performance of predictive methods on the protein models generated via AlphaFold2 were also used as a benchmark in this study. The Uniprot sequences of the proteins in the S2648 dataset were used for construction of AlphaFold2 models. The installation of AlphaFold2 was introduced in their manuscript and is available in their *github* page. As suggested, the same database version *—max_template_date = 2020-05-14* was used in this work. Model quality was recorded in output files and determined by pLDDT values [80].

Methods to predict effects of mutations on protein stability

Increased availability of high-quality mutation data and advances in computational approaches, particularly in machine learning, have supported and enabled the development of a range of computational tools aiming to understand how missense mutations affect protein folding and stability, which have been fundamental to unravelling molecular mechanisms of mutations leading to protein malfunction and diseases [81], also playing a role in cancer [82] and cancer risk [83], as well as drug resistance [17,18,84]. These developed methods can be divided into three main groups, without loss of generality: (i) tools based on energy function and dynamics simulations; (ii) knowledge-based and statistical; and (iii) machine learning methods. For this study, a representative set of ten structure-based methods was selected for assessment, including representatives of these three

groups, for which either standalone packages or web-server interfaces were publicly available. These include:

Energy-based and dynamics

1. FoldX [55] is a well-established tool that uses empirical force fields and energy term calculations to estimate the effects of mutations on protein stability.
2. ENCoM [56] uses a coarse-grained normal mode analysis (NMA) approach to simulate effects of mutations on the conformational repertoire, and therefore dynamics, of protein structures.

Knowledge-based and statistical

1. SDM [42] adopts environment-specific substitution tables to generate a structure-based score for mutation propensity and effects, considering structurally aligned protein families.
2. DDGun [85] is a prediction method using a linear combination of sequence and structural features.

Machine learning

1. MAESTRO [46] uses statistical scoring functions and a multi-agent system to predict effects of mutations on protein stability.
2. I-Mutant 2.0 [51] is a Support Vector Machine (SVM)-based method to predict the change of protein stability upon mutation.
3. mCSM-Stability [44] uses the concept of graph-based signature to describe residue environments in protein structures and then train and test predictive models via supervised learning.
4. DUET [45] is an integrated computational approach to predict the ΔΔG values upon mutation that combines the prediction power of mCSM-Stability and SDM.
5. DynaMut1 [47] is a method that combines the graph-based signatures and NMA to give a consensus prediction of ΔΔG values upon mutation via supervised learning.
6. DynaMut2 [50] is an optimized version that also considers the global environment of the wild-type residues to estimate the conformational change upon residue substitution and train supervised learning methods.

We have also included the following currently available sequence-based methods as a baseline for comparison purposes and assess potential situations where performance deterioration by using homology models would suggest that sequence-based methods would be more adequate:

1. SAAFEC-SEQ [86] is a sequence-based method that applies a gradient boosting decision tree algorithm on protein sequence features descriptors, different physico-chemical factors and evolutionary knowledge to make predictions.
2. MUpro [87] is a method that considers a small window size on the neighbour of targeted residue to train a SVM-based predictive model.
3. I-Mutant 2.0 [51] is a sequence-based version of the I-Mutant package using SVM to predict mutation effects on protein stability.
4. DDGun-Seq [85] is a sequence-based version of DDGun using evolutionary information to predict ΔΔG values upon variants.

All the predictions were run at the same experimental pH and temperature as described in the S2648 dataset (available as Supplementary Materials), when these parameters were available. All the prediction tools were otherwise run with default settings. The *nr* database [88] was used for predictions of SAAFEC-SEQ. A detailed introduction of all methods used in this work is available in Table S1, including the implementation, relevant datasets and sources.

Mutations and homology models obtained in previous steps were systematically used and provided to the methods to predict effects of mutation on protein stability, with the experimental effect used as ground truth. Performance metrics were calculated including root mean square error (RMSE) and Pearson's Correlation Coefficient (R), for regression purposes and Matthew's Correlation Coefficient (MCC) and F1-score, for classification purposes. A description of metrics used can be found in Supplementary Materials.

Characterizing method performance based on structural and sequence properties

To better characterize the performance of different methods, for different identity ranges, the mutation dataset was further divided for analysis purposes. This involved generating subsets of mutations based on structure- and sequence-based properties as follows:

Structure-based properties

Mutation subsets were divided for analysis purposes based on:

i. Residue relative solvent accessibility (RSA—*buried versus exposed* residues), calculated using Biopython [89];
ii. Residue depth, calculated by the *msms* program in Biopython;
iii. Secondary structure type (SST), obtained from the DSSP algorithm [90,91]. Four main SST, namely alpha helix, beta sheet, turn and random coil, were considered in this work;
iv. CATH structural classification of proteins [92]. Three main types of structure classes, namely mainly alpha, mainly beta, and mixed alpha/beta, were considered in this work.

Sequence-based properties

Mutation groups were further obtained by grouping residues based on:

i. Polarity, where wild-type residues are assigned to either polar (referred as P) including Q, N, S, T, R, H, K, D, E and C or hydrophobic/apolar (referred as H) including A, L, M, I, V, F, Y and W. Glycine residues were considered as a separate class. No mutations on Proline residues were identified in the S2648 dataset;
ii. Residue volume difference [93,94], classified based on the difference between wild-type and mutant residue volumes. Three groups were constructed, Large to Small (L2S), Same to Same (S2S) and Small to Large (S2L). Only when the difference was greater or equal to 30 $Å^3$ would the mutation be inserted into the L2S or S2L groups [71];
iii. Stability of mutations was classified based on the experimentally measured effects. These were labelled as either stabilizing or destabilizing (based on the $\Delta\Delta G$ sign).

Results

Property distribution of the homology model datasets

Eight homology model datasets at different identity levels, namely Iden15–25, Iden25–35, Iden35–45, Iden45–55, Iden55–65, Iden65–75, Iden75–85 and Iden85–95, were built and presented no significant differences in distributions of $\Delta\Delta G$ values (Figure 2A and Table S2, *P*-value = 0.38 via one-way ANOVA) when compared with the experimental dataset (Iden100), with significance decreasing with identity levels. The proportion of destabilizing ($\Delta\Delta G < 0$) and stabilizing mutations ($\Delta\Delta G \geq 0$) in these datasets were consistent, with averages of 26% (sd = 0.03) and 74% (sd = 0.03), respectively, reflecting a natural bias towards detabilizing mutations in the S2648 dataset (Figure S1).

Distribution of structure-based properties

All datasets presented similar RSA distributions (Figure 2B). A general cutoff of 20% was used to define whether a residue was buried (53%) or exposed (47%) (Table 1), consistent with previous studies [95–97]. Residue depth distributions were also similar, ranging from 1.7 to 6.6 Å. In this study, we used 2.2 Å as a cutoff to distinguish deep (54%) and shallow (46%) residues, to achieve a relatively balanced split (Table 1). The distribution of mutations per SST is listed in Table 1. For all homology model datasets, two-thirds of mutations were located in alpha helix (32%) and beta sheet (32%) structures, with a smaller proportion of mutations in turns (8%), random coil (17%) or other secondary structures (11%). When looking at the distribution of protein structural classes, based on the CATH database, the majority of mutations were in proteins belonging to the alpha/beta class (45%), followed by mainly alpha (27%) and mainly beta (26%), with a small fraction of proteins (2%) labelled as 'few secondary structures' in CATH (Table 1).

Distribution of sequence-based properties

The distribution of mutation types showed that most mutations fall into the apolar-to-apolar (HH: 37%) category, followed by polar-to-polar (PP: 20%), and polar-to-apolar (PH: 20%) categories (Figure 2C). Eight percent of mutations were to Glycine and 27% of mutations to Alanine, as a reflection of experimental mutagenesis efforts. Most mutations involved wild-type and mutant residues of similar volumes (S2S: 49%), with smaller proportions involving the introduction of smaller (L2S: 35%) or larger residues (S2L: 16%) (Table 1).

Distribution of RMSD and TM-score

We observed that higher sequence identities led in general to lower RMSDs and higher TM-score (Table 2, Figure 2D and E), consistent with previous analyses [61]. When sequence identity reaches around 40%, almost two-thirds of the models in the dataset obtained RMSD of around 1 Å and TM-score of above 0.75. Only one model in the dataset Iden55–65 had a high RMSD value of around 21 Å. However, this has been kept to mimic the real-world scenario of application of homology models.

Distribution of target-template identity, coverage and quality of models

Target-template identity distributions for each dataset are depicted in Supplementary Figure S2A, with average values consistently in the middle of the range. Most models presented target-template coverage higher than 85% (Supplementary Figure S2B), suggesting that the target-template coverage was less limiting than target-template identity when electing a template for homology modelling. The quality of all homology models shared similar distribution in each identity range (Supplementary Figure S2C).

Predictive performance trends on homology model datasets

Overall performance

We observed that, in general, predictive performance of the evaluated methods increases with target-template identity, which was consistent for both regression and classification tasks (Figure 3 and Supplementary Figure S3). Alternatively, we observed a consistent performance deterioration on the task of predicting mutation effects on stability for all structure-based methods, particularly in machine learning based methods and FoldX, when the sequence identity of the homology modelling template dropped. Interestingly, the predictive performance of most methods studied on AlphaFold2 models is close to those obtained on experimental structures (Figure 3B).

Based on the performance trend shown in Figure 3A, a proposed sequence identity cutoff for DynaMut2, DUET and mCSM-Stability was around 40%. The regression performances of these three prediction methods presented a sharp decreasing trend when the sequence identity

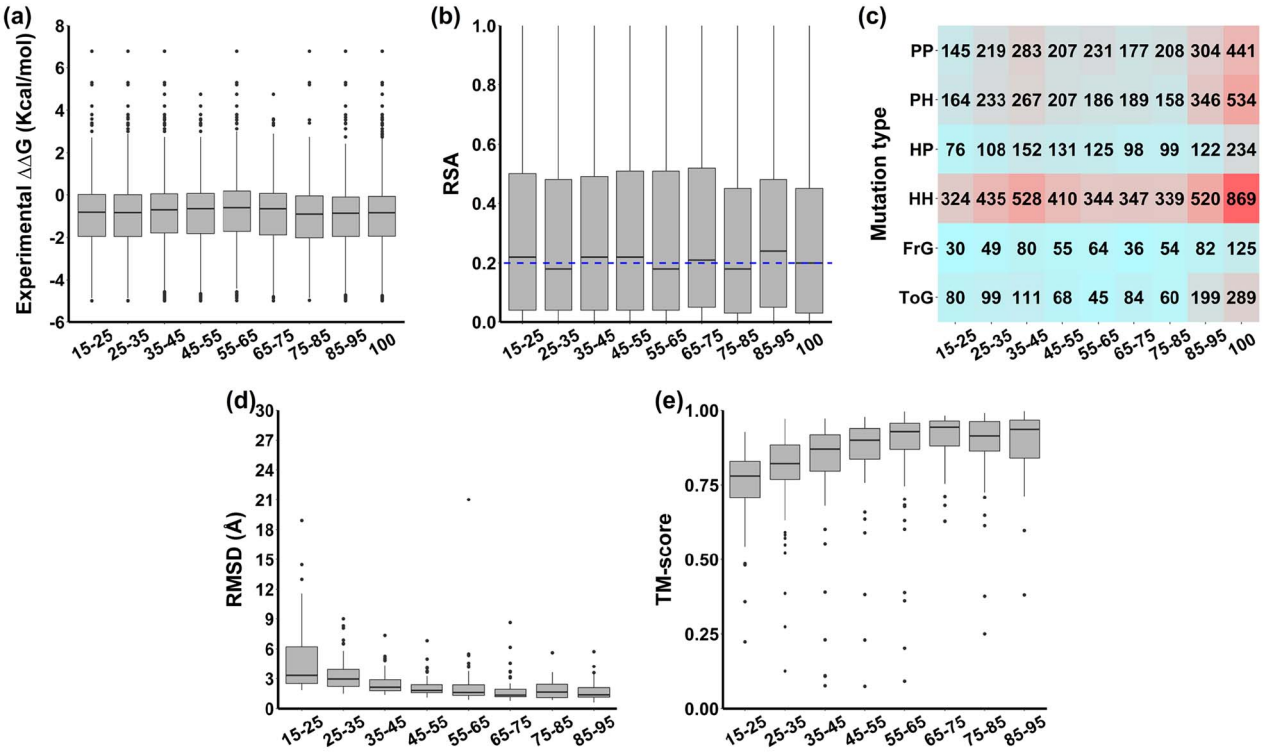

Figure 2. Distribution of (A) experimental ΔΔG values, (B) relative solvent accessibility (RSA), (C) mutation types, (D) root mean square deviation (RMSD) and (E) TM-score between homology models and experimental structures in eight homology model datasets of different identity levels. The RSA cutoff to define buried or exposed residues was 20% and shown as a blue dashed line in (B).

Table 2. Template description for homology model datasets.

Identity level (%)	Mean RMSD (Å)	Mean TM-score	Mean actual identity (%)	Mean coverage (%)	Mean template resolution (Å)
15–25	4.90	0.75	21.19	86.49	1.91
25–35	3.68	0.79	30.73	88.13	1.93
35–45	2.50	0.82	39.86	91.01	1.85
45–55	2.19	0.86	49.11	92.76	1.87
55–65	2.29	0.86	59.95	94.51	1.78
65–75	1.89	0.90	69.43	96.48	1.74
75–85	1.86	0.87	79.74	95.96	1.92
85–95	1.74	0.89	89.82	95.11	1.78

was less than 40% (with *R* values dropping from 0.63 to 0.53 and RMSE increasing from 1.22 to 1.34 Kcal/-mol), while keeping relatively stable performance for identity levels higher than 40%. Interestingly, below 40% identity the performance of structure-based methods deteriorated as low as that of sequence-based methods, indicating that the latter would be recommended in the absence of higher identity templates for homology modelling. The same result can be observed for classification tasks (MCC = 0.21–0.36, F1-score = 0.32–0.45). The SAAFEC-SEQ, as a sequence-based benchmark, showed the highest correlation among all methods ($R = 0.89$).

DynaMut1, Maestro and I-Mutant presented a similar behaviour when sequence identity reaches the 40% mark, with *R* values decreasing from 0.62 to 0.46 and RMSE increasing from 1.21 to 1.44 Kcal/mol (Figure 3 and Supplementary Figure S3). The performance of these three methods also deteriorated below the baseline from sequence-based methods when identity is under this threshold. FoldX had the largest degree of variation on performance when sequence identity changes and did not show robust performance for models built with templates with sequence identity below 70%. As for the performance trends of DDGun, SDM and ENCoM, there was no clear sequence identity cutoff for these three methods, with performance varying substantially. Only DDGun exceeded the baseline performance of sequence-based methods. To remove the bias caused by outliers and homologs in the original S2648 dataset, a detailed performance report is showcased in Supplementary Table S3. Similar identity cutoffs for machine learning

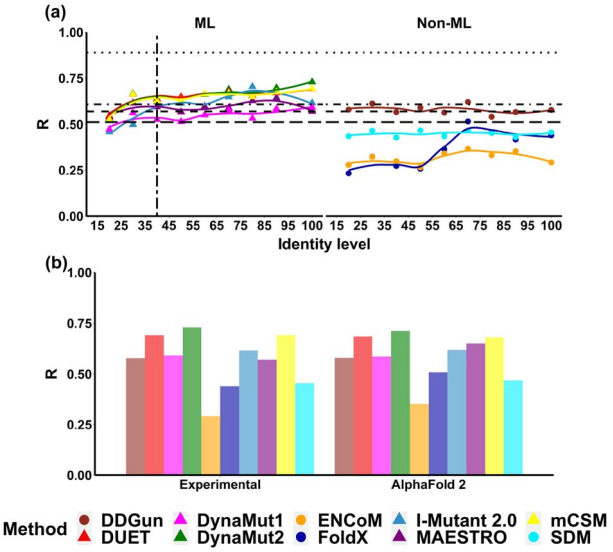

Figure 3. Overall performance trends based on Pearson's correlation coefficient (*R*) of ten methods predicting mutation effects on protein stability, namely DDGun (brown), DUET (red), DynaMut1 (pink), DynaMut2 (green), ENCoM (orange), FoldX (blue), I-Mutant 2.0 (light blue), MAESTRO (purple), mCSM-Stability (yellow) and SDM (cyan). The *R* values and their trends on homology models are represented in dots and lines, respectively. A vertical long-dashed line indicates the proposed identity cutoff for homology modelling, whereas the horizontal lines are the baseline performance of four sequence-based methods, namely SAAFEC-SEQ (dotted), MUpro (dot-dashed), I-Mutant (dashed) and DDGun (long-dashed).

methods were determined and no significant difference between the results before and after removing homolog structures was observed.

Performance trends based on structural properties

We further assessed how the performance of different predictive methods vary based on the structural properties of proteins and residues involved. In this study, we considered buried versus exposed residues, residues in different secondary structures and in proteins of different structural classes derived from CATH.

Exposed versus buried residues

The assessed methods tended to perform better on buried than exposed residues. When sequence identity was lower than 40%, I-Mutant, DUET, mCSM-Stability and DynaMut2 presented a larger drop in performance on buried residues (R dropped from 0.67 to 0.55) (Figure 4A) than on exposed residues (R dropped from 0.48 to 0.45), even though overall performance on the former mutation group was still higher. Classification performance on buried residues for these methods showed a sharp reduction when sequence identity was less than 50% (MCC dropped from 0.41 to 0.20) (Supplementary Figure S4). Among these four methods, I-Mutant showed the best classification performance on exposed mutations (MCC up to 0.55), whereas most could not reach the baseline performance of sequence-based methods. This trend was also observed for the remaining methods. When it comes to the statistical/energy function methods, only FoldX showed large performance deterioration when sequence identity dropped.

Shallow versus deep residues

For globular proteins, residue depth can distinguish between buried residues just under the surface and those near the protein core region [98–100]. As expected, prediction performance trends on deep and shallow residues correlated with buried and exposed residues (Figure 4A and B). DynaMut2, mCSM-Stability, DUET and I-Mutant shared a similar trend, with performance deteriorating below identity cutoff of 40% and 50% for deep and shallow residue mutations, respectively (consistent with F1-score for classification tasks—Supplementary Figure S5). FoldX shows the larger performance variation below 70% identity for both deep and shallow mutations (Figure 4B). Little performance variation until 40% identity was observed for DynaMut1, Maestro, SDM, ENCoM and DDGun.

Secondary structure

In general, methods performed better on structured regions (alpha helices, beta sheets and turns) than on unstructured ones (random coil). DynaMut2, mCSM-Stability and DUET shared similar performance trends, performing well up to 40% identity for mutations on alpha helix and beta sheet and 50% for turn and random coil (Figure 4C and S6), outperforming sequence-based methods. Similar performance trends were observed in DynaMut1, Maestro, as well as I-Mutant, revealing lower sequence identity demands on alpha helix and beta sheet, which were higher for FoldX. I-Mutant had the highest performance deterioration in turn conformations (R dropped from 0.62 to 0.36, Figure 4C). Similar identity cutoff on alpha helices and beta sheets can be observed in classification tasks, with MCC ranging from 0.27 to 0.47 (Figure S6). SDM, ENCoM and DDGun showed no distinguishable trend based on secondary structure.

CATH classification

When assessing performance based on protein structural classification, mainly alpha and mixed alpha/beta proteins presented clearer trends (Figure 4D), with a drop in performance below 40% identity. DynaMut2, mCSM-Stability and DUET showed a steadier performance on mainly alpha proteins, with larger drops for mainly beta and mixed alpha/beta. These were consistent for machine learning based methods and FoldX, with I-Mutant showing the most significant performance deterioration on classification tasks (Figure S7) and no discernible trend for ENCoM and DDGun. There was no obvious identity cutoff for DynaMut1 and Maestro in the mainly alpha group (Figure 4D). Performance deterioration was more pronounced in mainly beta and mixed alpha/beta proteins.

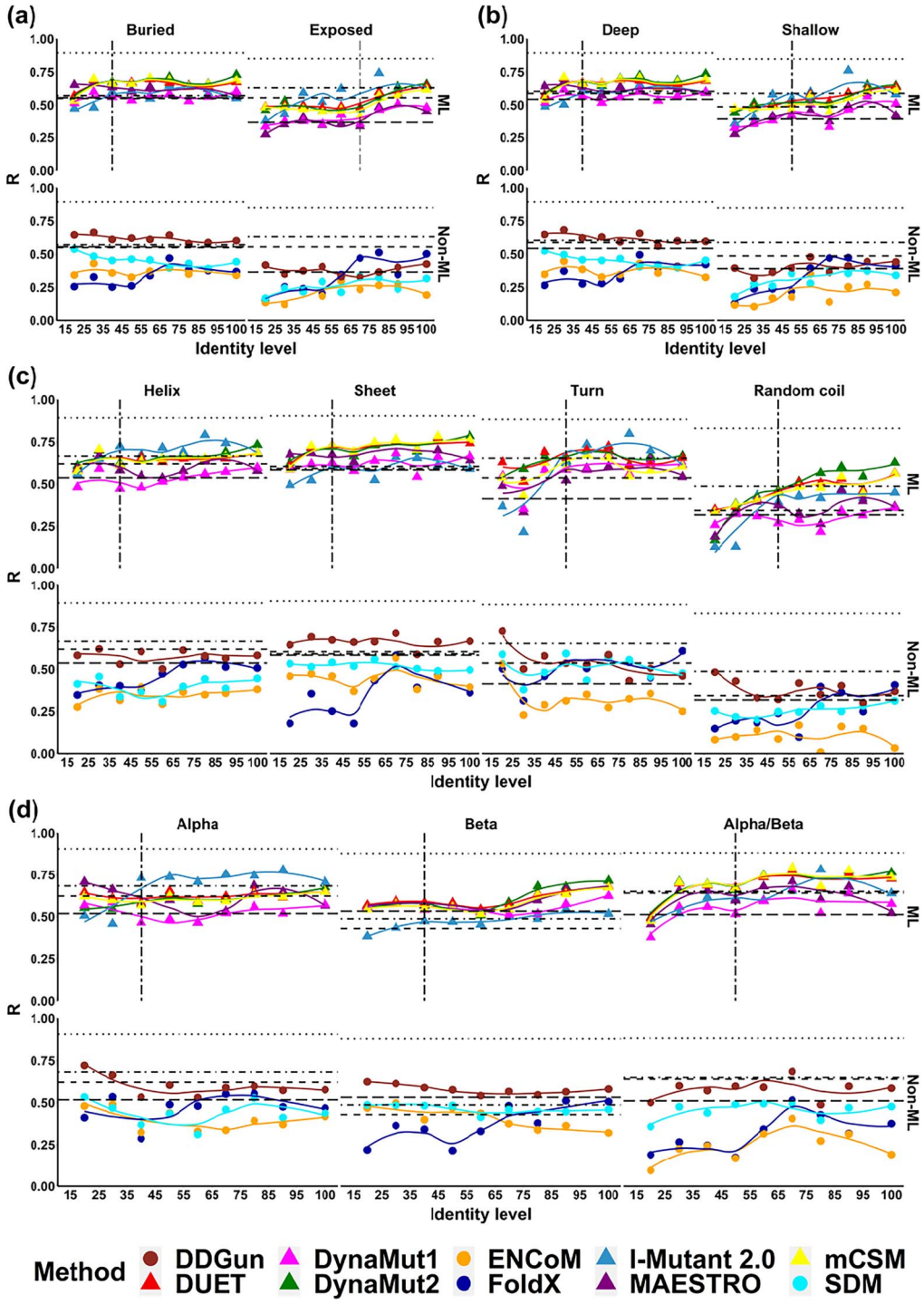

Figure 4. Performance trends based on Pearson's correlation coefficient (*R*) of ten methods with mutations grouped based on four structure-based properties, namely (A) relative solvent accessibility (RSA), (B) residue depth, (C) secondary structure types and (D) structural class based on CATH. The performance trends of two main types of methods, namely machine learning based (ML) and Statistical/Energy function based (Non-ML), were displayed respectively. The RSA cutoff of 20% was used to determine buried or exposed residues. The residue depth cutoff of 2.2 Å was used to determine deep or shallow residues. Four secondary structure types, namely alpha helix, beta sheet, turn and random coil, were considered in this study. Three structural classifications, namely mainly alpha, mainly beta and mixed alpha/beta, were analysed.

Performance trends based on sequence-based properties

For sequence-based properties, this study focused on mutations on different amino acid residue types, residue volume differences, and the direction of stability change upon mutation.

Mutation types

When assessing performance trends based on mutation types (to and from polar and apolar residues), inter-group mutations (HP and PH) presented, in general, a more pronounced performance deterioration below 40–50%, particularly for the PH type for all methods

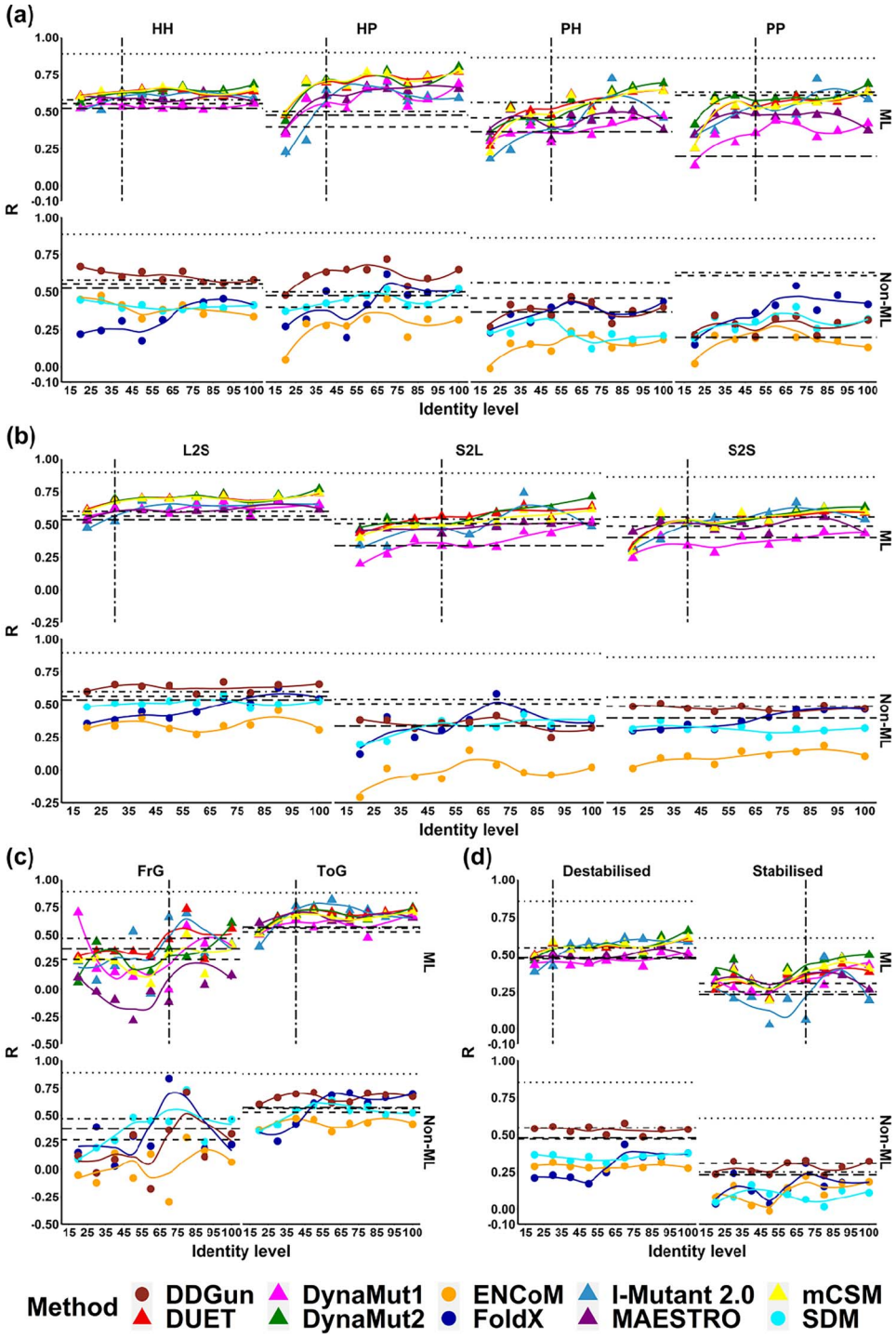

Figure 5. Performance trends based on Pearson's correlation coefficient (R) of ten methods with mutations grouped based on four sequence-based properties, namely (A) mutation type based on the change of polarity, (B) change of residue volume, (C) mutations related to Glycine and (D) mutation effects on protein stability.

(Figure 5A and S8). Only a small performance deterioration was observed for HH mutations for most methods, apart from FoldX. Better performance and reliability on inter-group mutations could be influenced by the natural distribution of mutation effects around mildly destabilizing mutations, which were enriched in S2648, and would influence particularly machine learning methods. Consistent with that, no trends for SDM, ENCoM and DDGun were observed.

Change of residue volume

Figure 5B and S9 depict the performance trends when categorizing mutations based on volume changes between wild-type and mutant residues. In general, performance deteriorated less for this category, with most methods being outperformed by sequence-based alternatives for L2S volume mutations below 40%. Methods seemed more robust for large to small changes (L2S), with DynaMut2, mCSM-Stability and

DUET, maintaining performance down to 30% identity. On the other hand, performance deteriorated more quickly for S2L mutations (around 50% identity). No clear trends were identified for SDM, ENCoM as well as DDGun.

Mutations to/from Glycine

Performance trends were consistent for mutations to Glycine (ToG), with most methods only showing a sharp decrease in correlation around 40% identity (Figure 5C and S10), while outperforming sequence-based methods even beyond this threshold. Alternatively, for mutations from Glycine (FrG), a substantial variation in performance was observed, with a peak in performance around 70%, which might be due to the limited number of mutations in this category (5% mutations), in comparison with mutations to Glycine, which were often used in mutagenesis experiments (8% mutations).

Effect of mutation on thermodynamics stability of proteins

In general, all methods performed better on destabilizing mutations, which is consistent with previous observations [71,101,102]. For destabilizing mutations (Figure 5D and S11), most machine learning methods only presented performance deterioration below 30% identity. These methods have been shown before to perform better on this mutation type, a bias that was introduced due to the natural distribution of mutation effects, which has been attempted to be corrected with the use of hypothetical reverse mutations [47,48,54,86]. No trends were observed for SDM, ENCoM or DDGun. Consistently, a sharper decrease in performance was observed for stabilizing mutations below 70% identity.

Discussion and conclusion

To better understand how *in silico* mutation analysis tools behave in the absence of high-resolution experimental protein structures, we used homology models at different identity levels to systematically test the performance of ten widely used computational tools to predict mutation effects on protein stability. We found that when target-template identity for homology modelling drops below 40%, there is an evident performance deterioration for structure-based methods, especially for the machine learning based approaches, a point where sequence-based methods might be preferable. It has been previously reported that in order to build reliable models of a protein of interest, the structure used as a template should share at least 30% sequence identity to the target [60]. The identity cutoff identified in this work is more conservative than this accepted rule-of-thumb and can be further validated in the future as more thermodynamics data becomes available. Although some small differences between the prediction using homology models and one using experimental structure were noticed, we think this mainly results from the prediction variance of the prediction methods themselves, which was reported in a test on structural sensitivity [103]. We also found that the predictive performance on AlphaFold2 models was highly consistent with that on experimental structures for most tools. It represents an important breakthrough in the field of protein structure prediction as the community actively seeks the explore its limitations and adapt it for other applications.

When assessing different mutation categories based on structural and sequence-based properties, we found that the identity cutoffs varied from the overall threshold described above. The reason for this may be a native bias of the prediction methods, with predictors performing better on residues that are not solvent exposed, deeper in the structures and for destabilizing mutations, consistent with previous observations [71]. Alternatively, structure-based methods performed worse on exposed residues, random coil conformations, less frequent mutations (e.g. from Glycine) and stabilizing mutations, requiring a higher identity cutoff (50–70%) when using homology models.

In brief, this work showed that, as sequence identity of the template decreased, the performance of the tools deteriorated beyond the performance of sequence-based tools. As expected, this was more pronounced for exposed residues and mutations in random coils, where the largest deviations in structure modelling are likely to be found. We found that a minimum target-template identity cutoff around 40% was necessary for robust performance of structure-based tools when using homology models as inputs, larger than the minimum 30% sequence identity threshold often used as a rule-of-thumb for homology modelling. We expect that these insights will help guide the accurate use and interpretation of these computational tools in the absence of experimental structures moving forward.

Key Points/Highlights

- We present the first systematic study assessing how methods to predict stability changes upon mutations cope in the absence of high-resolution experimental protein structures.
- This work provides a detailed guideline for *in silico* mutation analysis, which will assist users in appropriately using and interpreting prediction results, which could assist in the study of mutations in protein design and in genetic diseases.
- This work first applied protein structural models from traditional homology modelling and AlphaFold2 modelling to mutation effect analysis.

Acknowledgements

We thank Dr Moshe Olshansky for providing support for running the homology modelling pipeline and Dr Carlos

Rodrigues for technical support while running structure-based predictors.

Data Availability

Datasets used in the analyses described have been made available as Supplementary Materials.

Funding

An Investigator Grant from the National Health and Medical Research Council (NHMRC) of Australia (GNT1174405). Supported in part by the Victorian Government's Operational Infrastructure Support Program.

References

1. Protasevich I, Yang Z, Wang C, *et al.* Thermal unfolding studies show the disease causing F508del mutation in CFTR thermodynamically destabilizes nucleotide-binding domain 1. *Protein Sci* 2010;**19**:1917–31.
2. Jafri M, Wake NC, Ascher DB, *et al.* Germline mutations in the CDKN2B tumor suppressor gene predispose to renal cell carcinoma. *Cancer Discov* 2015;**5**:723–9.
3. Usher JL, Ascher DB, Pires DE, *et al.* Analysis of HGD gene mutations in patients with Alkaptonuria from the United Kingdom: identification of novel mutations. *JIMD Rep* 2015;**24**:3–11.
4. Nemethova M, Radvanszky J, Kadasi L, *et al.* Twelve novel HGD gene variants identified in 99 alkaptonuria patients: focus on 'black bone disease' in Italy. *Eur J Hum Genet* 2016;**24**:66–72.
5. Pires DE, Chen J, Blundell TL, *et al.* In silico functional dissection of saturation mutagenesis: interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. *Sci Rep* 2016;**6**:19848.
6. Casey RT, Ascher DB, Rattenberry E, *et al.* SDHA related tumorigenesis: a new case series and literature review for variant interpretation and pathogenicity. *Mol Genet Genomic Med* 2017;**5**:237–50.
7. Andrews KA, Ascher DB, Pires DEV, *et al.* Tumour risks and genotype-phenotype correlations associated with germline variants in succinate dehydrogenase subunit genes SDHB, SDHC and SDHD. *J Med Genet* 2018;**55**:384–94.
8. Hildebrand JM, Kauppi M, Majewski IJ, *et al.* A missense mutation in the MLKL brace region promotes lethal neonatal inflammation and hematopoietic dysfunction. *Nat Commun* 2020;**11**:3150.
9. Portelli S, Barr L, de Sa AGC, *et al.* Distinguishing between PTEN clinical phenotypes through mutation analysis. *Comput Struct Biotechnol J* 2021;**19**:3097–109.
10. Trezza A, Bernini A, Langella A, *et al.* A computational approach from gene to structure analysis of the human ABCA4 transporter involved in genetic retinal diseases. *Invest Ophthalmol Vis Sci* 2017;**58**:5320–8.
11. Hnizda A, Fabry M, Moriyama T, *et al.* Relapsed acute lymphoblastic leukemia-specific mutations in NT5C2 cluster into hotspots driving intersubunit stimulation. *Leukemia* 2018;**32**:1393–403.
12. Soardi FC, Machado-Silva A, Linhares ND, *et al.* Familial STAG2 germline mutation defines a new human cohesinopathy. *NPJ Genom Med* 2017;**2**:7.
13. Traynelis J, Silk M, Wang Q, *et al.* Optimizing genomic medicine in epilepsy through a gene-customized approach to missense variant interpretation. *Genome Res* 2017;**27**:1715–29.
14. Karmakar M, Cicaloni V, Rodrigues CHM, *et al.* HGDiscovery: an online tool providing functional and phenotypic information on novel variants of homogentisate 1,2- dioxigenase. *bioRxiv* 2021; 2021.2004.2026.441386.
15. Lai CY, Tsai IJ, Chiu PC, *et al.* A novel deep intronic variant strongly associates with Alkaptonuria. *NPJ Genom Med* 2021;**6**:89.
16. Patel AB, O'Hare T, Deininger MW. Mechanisms of resistance to ABL kinase inhibition in chronic myeloid leukemia and the development of next generation ABL kinase inhibitors. *Hematol Oncol Clin North Am* 2017;**31**:589–612.
17. Pires DE, Ascher DB. CSM-lig: a web server for assessing and comparing protein-small molecule affinities. *Nucleic Acids Res* 2016;**44**:W557–61.
18. Portelli S, Myung Y, Furnham N, *et al.* Prediction of rifampicin resistance beyond the RRDR using structure-based machine learning approaches. *Sci Rep* 2020;**10**:18120.
19. Ascher DB, Wielens J, Nero TL, *et al.* Potent hepatitis C inhibitors bind directly to NS5A and reduce its affinity for RNA. *Sci Rep* 2014;**4**:4765.
20. Silvino AC, Costa GL, Araujo FC, *et al.* Variation in human cytochrome P-450 drug-metabolism genes: a gateway to the understanding of plasmodium vivax relapses. *PLoS One* 2016;**11**:e0160172.
21. Hawkey J, Ascher DB, Judd LM, *et al.* Evolution of carbapenem resistance in Acinetobacter baumannii during a prolonged infection. *Microb Genom* 2018;**4**:e000165.
22. Holt KE, McAdam P, Thai PVK, *et al.* Frequent transmission of the mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat Genet* 2018;**50**:849–56.
23. Karmakar M, Globan M, Fyfe JAM, *et al.* Analysis of a novel pncA mutation for susceptibility to pyrazinamide therapy. *Am J Respir Crit Care Med* 2018;**198**:541–4.
24. Vedithi SC, Malhotra S, Das M, *et al.* Structural implications of mutations conferring rifampin resistance in mycobacterium leprae. *Sci Rep* 2018;**8**:5016.
25. Portelli S, Olshansky M, Rodrigues CHM, *et al.* Exploring the structural distribution of genetic variation in SARS-CoV-2 with the COVID-3D online resource. *Nat Genet* 2020;**52**:999–1001.
26. Tunstall T, Portelli S, Phelan J, *et al.* Combining structure and genomics to understand antimicrobial resistance. *Comput Struct Biotechnol J* 2020;**18**:3377–94.
27. Vedithi SC, Malhotra S, Skwark MJ, *et al.* HARP: a database of structural impacts of systematic missense mutations in drug targets of mycobacterium leprae. *Comput Struct Biotechnol J* 2020;**18**:3692–704.
28. Vedithi SC, Rodrigues CHM, Portelli S, *et al.* Computational saturation mutagenesis to predict structural consequences of systematic mutations in the beta subunit of RNA polymerase in mycobacterium leprae. *Comput Struct Biotechnol J* 2020;**18**:271–86.
29. Tunes LG, Ascher DB, Pires DEV, *et al.* The mutation G133D on Leishmania guyanensis AQP1 is highly destabilizing as revealed by molecular modeling and hypo-osmotic shock assay. *Biochim Biophys Acta Biomembr* 2021;**1863**:183682.
30. Karmakar M, Rodrigues CHM, Horan K, *et al.* Structure guided prediction of pyrazinamide resistance mutations in pncA. *Sci Rep* 2020;**10**:1875.

31. Portelli S, Phelan JE, Ascher DB, *et al.* Understanding molecular consequences of putative drug resistant mutations in mycobacterium tuberculosis. *Sci Rep* 2018;**8**:15356.
32. Zhou Y, Portelli S, Pat M, *et al.* Structure-guided machine learning prediction of drug resistance mutations in Abelson 1 kinase. *Comput Struct Biotechnol J* 2021;**19**:5381–91.
33. Karmakar M, Rodrigues CHM, Holt KE, *et al.* Empirical ways to identify novel Bedaquiline resistance mutations in AtpE. *PLoS One* 2019;**14**:e0217169.
34. Panchal V, Jatana N, Malik A, *et al.* A novel mutation alters the stability of PapA2 resulting in the complete abrogation of sulfolipids in clinical mycobacterial strains. *FASEB Bioadv* 2019;**1**:306–19.
35. Sanavia T, Birolo G, Montanucci L, *et al.* Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. *Comput Struct Biotechnol J* 2020;**18**:1968–79.
36. Gerasimavicius L, Liu X, Marsh JA. Identification of pathogenic missense mutations using protein stability predictors. *Sci Rep* 2020;**10**:15387.
37. Turner P, Mamo G, Karlsson EN. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microb Cell Fact* 2007;**6**:9.
38. Ferdjani S, Ionita M, Roy B, *et al.* Correlation between thermostability and stability of glycosidases in ionic liquid. *Biotechnol Lett* 2011;**33**:1215–9.
39. Xie Y, An J, Yang G, *et al.* Enhanced enzyme kinetic stability by increasing rigidity within the active site. *J Biol Chem* 2014;**289**:7994–8006.
40. Wu I, Arnold FH. Engineered thermostable fungal Cel6A and Cel7A cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures. *Biotechnol Bioeng* 2013;**110**:1874–83.
41. O'Fagain C. Protein stability: enhancement and measurement. *Methods Mol Biol* 2017;**1485**:101–29.
42. Pandurangan AP, Ochoa-Montano B, Ascher DB, *et al.* SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res* 2017;**45**:W229–35.
43. Chen CW, Lin MH, Liao CC, *et al.* iStable 2.0: predicting protein thermal stability changes by integrating various characteristic modules, Comput Struct. *Biotechnol J* 2020;**18**:622–30.
44. Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 2014;**30**:335–42.
45. Pires DE, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res* 2014;**42**:W314–9.
46. Laimer J, Hofer H, Fritz M, *et al.* MAESTRO–multi agent stability prediction upon point mutations. *BMC Bioinformatics* 2015;**16**:116.
47. Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res* 2018;**46**:W350–5.
48. Cao H, Wang J, He L, *et al.* DeepDDG: predicting the stability change of protein point mutations using neural networks. *J Chem Inf Model* 2019;**59**:1508–14.
49. Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics* 2016;**32**:2936–46.
50. Rodrigues CHM, Pires DEV, Ascher DB. DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Sci* 2021;**30**:60–9.
51. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* 2005;**33**:W306–10.
52. Masso M, Vaisman II. AUTO-MUTE 2.0: a portable framework with enhanced capabilities for predicting protein functional consequences upon mutation. *Adv Bioinformatics* 2014;**2014**:278385.
53. Li B, Yang YT, Capra JA, *et al.* Predicting changes in protein thermodynamic stability upon point mutation with deep 3D convolutional neural networks. *PLoS Comput Biol* 2020;**16**:e1008291.
54. Chen Y, Lu H, Zhang N, *et al.* PremPS: predicting the impact of missense mutations on protein stability. *PLoS Comput Biol* 2020;**16**:e1008543.
55. Schymkowitz J, Borg J, Stricher F, *et al.* The FoldX web server: an online force field. *Nucleic Acids Res* 2005;**33**:W382–8.
56. Frappier V, Chartier M, Najmanovich RJ. ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. *Nucleic Acids Res* 2015;**43**:W395–400.
57. Dehouck Y, Kwasigroch JM, Gilis D, *et al.* PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics* 2011;**12**:151.
58. Iqbal S, Li F, Akutsu T, *et al.* Assessing the performance of computational predictors for estimating protein stability changes upon missense mutations. *Brief Bioinform* 2021;**22**:bbab184.
59. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;**234**:779–815.
60. Bitencourt-Ferreira G, de Azevedo WF, Jr. Homology modeling of protein targets with MODELLER. *Methods Mol Biol* 2019;**2053**:231–49.
61. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* 1986;**5**:823–6.
62. Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9.
63. Akdel M, Pires DEV, Porta Pardo E, *et al.* A structural biology community assessment of AlphaFold 2 applications. *bioRxiv* 2021; 2021.2009.2026.461876.
64. Bava KA, Gromiha MM, Uedaira H, *et al.* ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res* 2004;**32**:D120–1.
65. Nikam R, Kulandaisamy A, Harini K, *et al.* ProThermDB: thermodynamic database for proteins and mutants revisited after 15 years. *Nucleic Acids Res* 2021;**49**:D420–4.
66. Xavier JS, Nguyen TB, Karmarkar M, *et al.* ThermoMutDB: a thermodynamic database for missense mutations. *Nucleic Acids Res* 2021;**49**:D475–9.
67. Pezeshgi Modarres H, Mofrad MR, Sanati-Nezhad A. ProtDataTherm: a database for thermostability analysis and engineering of proteins. *PLoS One* 2018;**13**:e0191222.
68. Pires DE, Blundell TL, Ascher DB. Platinum: a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. *Nucleic Acids Res* 2015;**43**:D387–91.
69. Jankauskaite J, Jimenez-Garcia B, Dapkunas J, *et al.* SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* 2019;**35**:462–9.
70. Dehouck Y, Grosfils A, Folch B, *et al.* Fast and accurate predictions of protein stability changes upon mutations using

statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* 2009;**25**:2537–43.
71. Caldararu O, Mehra R, Blundell TL, *et al.* Systematic investigation of the data set dependency of protein stability predictors. *J Chem Inf Model* 2020;**60**:4772–84.
72. Bradley P, Chivian D, Meiler J, *et al.* Rosetta predictions in CASP5: successes, failures, and prospects for complete automation. *Proteins* 2003;**53**(Suppl 6):457–68.
73. AlQuraishi M. AlphaFold at CASP13. *Bioinformatics* 2019;**35**: 4862–5.
74. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics* 2016;**54**:5 6 1–37.
75. Schrodinger LLC. The AxPyMOL molecular graphics plugin for Microsoft PowerPoint. *Version* 2015;**1**:8.
76. Schrodinger LLC. The JyMOL molecular graphics development component. *Version* 2015;**1**:8.
77. Schrodinger LLC. The PyMOL molecular graphics system. *Version* 2015;**1**:8.
78. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;**33**: 2302–9.
79. Simpkin AJ, Sanchez Rodriguez F, Mesdaghi S, *et al.* Evaluation of model refinement in CASP14. *Proteins* 2021;**89**:1852–69.
80. Mariani V, Biasini M, Barbato A, *et al.* lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 2013;**29**: 2722–8.
81. Ascher DB, Spiga O, Sekelska M, *et al.* Homogentisate 1,2-dioxygenase (HGD) gene variants, their analysis and genotype-phenotype correlations in the largest cohort of patients with AKU. *Eur J Hum Genet* 2019;**27**:888–902.
82. Lakshmana G, Baniahmad A. Interference with the androgen receptor protein stability in therapy-resistant prostate cancer. *Int J Cancer* 2019;**144**:1775–9.
83. Gossage L, Pires DE, Olivera-Nappa A, *et al.* An integrated computational approach can classify VHL missense mutations according to risk of clear cell renal carcinoma. *Hum Mol Genet* 2014;**23**:5976–88.
84. Pires DE, Blundell TL, Ascher DB. mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Sci Rep* 2016;**6**:29575.
85. Montanucci L, Capriotti E, Frank Y, *et al.* DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. *BMC Bioinformatics* 2019;**20**:335.
86. Li G, Panday SK, Alexov E. SAAFEC-SEQ: a sequence-based method for predicting the effect of single point mutations on protein thermodynamic stability. *Int J Mol Sci* 2021;**22**:606.
87. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* 2006;**62**:1125–32.
88. Coordinators NR. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2018;**46**:D8–13.
89. Cock PJ, Antao T, Chang JT, *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;**25**:1422–3.
90. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;**22**:2577–637.
91. Touw WG, Baakman C, Black J, *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res* 2015;**43**:D364–8.
92. Knudsen M, Wiuf C. The CATH database. *Hum Genomics* 2010;**4**: 207–12.
93. Zamyatnin AA. Protein volume in solution. *Prog Biophys Mol Biol* 1972;**24**:107–23.
94. Zamyatnin AA. Amino acid, peptide, and protein volume in solution. *Annu Rev Biophys Bioeng* 1984;**13**:145–65.
95. Savojardo C, Manfredi M, Martelli PL, *et al.* Solvent accessibility of residues undergoing pathogenic variations in humans: from protein structures to protein sequences. *Front Mol Biosci* 2020;**7**:626363.
96. Chen H, Zhou HX. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. *Nucleic Acids Res* 2005;**33**:3193–9.
97. Silk M, Pires DEV, Rodrigues CHM, *et al.* MTR3D: identifying regions within protein tertiary structures under purifying selection. *Nucleic Acids Res* 2021;**49**:W438–45.
98. Chakravarty S, Varadarajan R. Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure* 1999;**7**:723–32.
99. Jubb H, Blundell TL, Ascher DB. Flexibility and small pockets at protein-protein interfaces: new insights into druggability. *Prog Biophys Mol Biol* 2015;**119**:2–9.
100. Pandurangan AP, Ascher DB, Thomas SE, *et al.* Genomes, structural biology and drug discovery: combating the impacts of mutations in genetic disease and antibiotic resistance. *Biochem Soc Trans* 2017;**45**:303–11.
101. Pucci F, Bernaerts KV, Kwasigroch JM, *et al.* Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics* 2018;**34**:3659–65.
102. Marabotti A, Del Prete E, Scafuri B, *et al.* Performance of web tools for predicting changes in protein stability caused by mutations. *BMC Bioinformatics* 2021;**22**:345.
103. Caldararu O, Blundell TL, Kepp KP. A base measure of precision for protein stability predictors: structural sensitivity. *BMC Bioinformatics* 2021;**22**:88.