



## A base measure of precision for protein stability predictors: structural sensitivity

**Caldararu, Octav; Blundell, Tom L; Kepp, Kasper Planeta**

*Published in:*  
BMC Bioinformatics

*Link to article, DOI:*  
[10.1186/s12859-021-04030-w](https://doi.org/10.1186/s12859-021-04030-w)

*Publication date:*  
2021

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Caldararu, O., Blundell, T. L., & Kepp, K. P. (2021). A base measure of precision for protein stability predictors: structural sensitivity. *BMC Bioinformatics*, 22(1), Article 88. <https://doi.org/10.1186/s12859-021-04030-w>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE

Open Access



# A base measure of precision for protein stability predictors: structural sensitivity

Octav Caldararu<sup>1</sup> , Tom L. Blundell<sup>2</sup> and Kasper P. Kepp<sup>1\*</sup>

\*Correspondence:

kpj@kemi.dtu.dk

<sup>1</sup> DTU Chemistry, Technical University of Denmark, Building 206, 2800 Kgs. Lyngby, Denmark

Full list of author information is available at the end of the article

## Abstract

**Background:** Prediction of the change in fold stability ( $\Delta\Delta G$ ) of a protein upon mutation is of major importance to protein engineering and screening of disease-causing variants. Many prediction methods can use 3D structural information to predict  $\Delta\Delta G$ . While the performance of these methods has been extensively studied, a new problem has arisen due to the abundance of crystal structures: How precise are these methods in terms of structure input used, which structure should be used, and how much does it matter? Thus, there is a need to quantify the structural sensitivity of protein stability prediction methods.

**Results:** We computed the structural sensitivity of six widely-used prediction methods by use of saturated computational mutagenesis on a diverse set of 87 structures of 25 proteins. Our results show that structural sensitivity varies massively and surprisingly falls into two very distinct groups, with methods that take detailed account of the local environment showing a sensitivity of  $\sim 0.6$  to  $0.8$  kcal/mol, whereas machine-learning methods display much lower sensitivity ( $\sim 0.1$  kcal/mol). We also observe that the precision correlates with the accuracy for mutation-type-balanced data sets but not generally reported accuracy of the methods, indicating the importance of mutation-type balance in both contexts.

**Conclusions:** The structural sensitivity of stability prediction methods varies greatly and is caused mainly by the models and less by the actual protein structural differences. As a new recommended standard, we therefore suggest that  $\Delta\Delta G$  values are evaluated on three protein structures when available and the associated standard deviation reported, to emphasize not just the accuracy but also the precision of the method in a specific study. Our observation that machine-learning methods deemphasize structure may indicate that folded wild-type structures alone, without the folded mutant and unfolded structures, only add modest value for assessing protein stability effects, and that side-chain-sensitive methods overstate the significance of the folded wild-type structure.

**Keywords:** Protein stability, Mutation, Computation, Protein structure, Structural sensitivity



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The accurate prediction of the change in protein fold stability ( $\Delta\Delta G$ ) upon amino-acid substitution is a central challenge in modern biology, the solution to which would enable efficient rational engineering of stable proteins for industry and medicine [1–3], help us to understand protein evolution where stability effects play a major role [4–9], and improve our understanding of many protein-stability related genetic diseases driven by point mutations [10–12].

Many programs can predict  $\Delta\Delta G$  by utilizing diverse prediction models from machine-learning to energy-based force-fields [13–22]. An important distinction can be made between those methods that only use the protein amino-acid sequence to predict stability and those that use a three-dimensional wild-type structure as input. Logically, one expects methods that use a 3D structure to perform better, since interactions between amino acids cannot be fully deduced from sequence alone. However, the structure-based methods importantly miss both the unfolded wild-type structure, and the folded and unfolded mutants structures, i.e. three of the four structures of the thermodynamic cycle of  $\Delta\Delta G$  values; perhaps partly for this reason, structure-based methods perform only slightly better than their sequence-based counterparts [23].

The worse-than-expected performance of structure-based methods can also relate directly to the quality of the structures used. Indeed, it has been long debated whether crystal structures reproduce the native structures of proteins in solution and cells, as structures could be affected by crystal packing effects [24, 25]. Thus, while we can say whether a structure is more precisely determined, it is difficult to say which protein structure is more realistic. Databases such as ProTherm [26] and VariBench [27] annotate each experimental data point with a Protein Data Bank (PDB) [28] code that may not represent the best structure if more structures are available, and this could affect the computed  $\Delta\Delta G$  value.

It is well-known that energy-based methods such as FoldX [29] or Rosetta [30] can be quite affected by the structure used, and the authors of the methods recommend to minimize the structures used as input before prediction. On the other hand, machine learning methods, which arguably deemphasize the protein structure relative to such methods (as shown clearly below), have been previously indicated to be less structurally sensitive, at least for certain proteins [31, 32].

In addition to the accuracy (as estimated from the performance for a balanced data set) [33], it is therefore also important to know how sensitive a method is to the input structure used, i.e. its “precision”. Method accuracy has been extensively studied and the moderate accuracy of the prediction methods was attributed to failings in the scoring functions [17], inability to correctly predict stabilizing mutations [23] or biases originating from the datasets used for training [33]. An important set of principles for evaluating protein stability prediction methods of the prediction methods was reported by Niroula and Vihinen [34], but the impact of structure choice on the computed outcome was not discussed, yet, as shown below, adds an additional criteria for evaluating these methods. To our knowledge, structural sensitivity has only been considered when studying select proteins such as superoxide dismutase [31] and myoglobin [32].

A more generally confirmed structural sensitivity measure would represent a base measure for the precision of the methods, which seems required from a scientific

point of view, since accuracy without precision does not suffice to establish predictive power, and any output should ideally be seen in the context of such a precision measure, i.e. the noise expected due to choice of structural input. However, it should be noted that a method showing no structural sensitivity at all is not desirable either, as it would imply that the method underemphasizes the wild type structure, leaving it practically if not formally a sequence-based method.

In this paper we studied the structural sensitivity of six widely used protein stability prediction methods, including both energy and knowledge-based methods and machine learning methods. The structural sensitivity was determined from computational saturated mutagenesis [35] applied on 25 different proteins, each having multiple possible structures published in the PDB. Our results show that structural sensitivity varies greatly, with some methods showing high sensitivity ( $\sim 0.6$  to  $0.8$  kcal/mol) whereas others, notably all studied machine-learning methods, are very insensitive to structure choice ( $\sim 0.1$  kcal/mol). Furthermore, this sensitivity was rather constant across the proteins studied, showing that the models themselves cause the behaviour regardless of the structural heterogeneity itself. Our results provide a good baseline for the precision of protein stability calculators useful for future studies. As a consequence, we recommend the use of triplicate  $\Delta\Delta G$  evaluations on three distinct structures and reporting of the associated standard deviation whenever this is possible.

## Methods

### Protein structures studied

23 proteins were selected from an exhaustive search of PDB structures that fulfilled the following conditions:

- At least three crystal structures (3–5 used) in the PDB at different resolutions (at least  $0.15$  Å standard deviation of the resolution among the structures) but with the same amino-acid sequence;
- All structures monomeric in the crystal form;
- All structures apo (no ligands) in the crystal form;
- No missing CA atoms, except for N-terminal or C-terminal residues.

Furthermore, one protein with a metal centre (carbonic anhydrase) and one protein in a tetrameric form (deoxy-haemoglobin) were included to test the influence of such substantial features on structural sensitivity, for a total of 25 proteins. All structures were renumbered so that the same number corresponds to the same residue in all structures of a protein. More than three structures were selected for proteins with a wide spread of resolutions, to a maximum of five structures for Lyz, CAH, Rnase and UBQ. This resulted in a varied list of long and short proteins and proteins belonging to all secondary structure classes in CATH, which should make our results more general; however, importantly it turns out that these variations have modest effect, i.e. the sensitivities are quite generic to the methods. The full list of proteins and their PDB IDs is given in Table 1.

**Table 1** List of proteins studied in this work

Name	Abbreviation	CATH	Length (N)	PDB IDs	Average resolution	STDEV resolution
Human acidic fibroblast growth factor	FGF	$\beta$	513	2AFG, 3UD7, 1JQZ, 1RG8	1.88	0.71
Barnase	Bar	$\alpha+\beta$	110	1BNI, 1B2X, 1A2P, 1X1X	1.32	0.35
Hen Lysozyme	Lyz	$\alpha$	129	2LYZ, 2VB1, 3LZT, 5LIN, 5HNL	1.50	0.73
Beta Lactamase	BLac	$\alpha+\beta$	263	1ZG4, 1BTL, 4OQG	1.31	0.44
Alpha Amylase	AAmy	$\alpha+\beta$	453	1AQM, 1AQH, 1B0I	2.08	0.28
Ubiquitin	UBQ	$\alpha+\beta$	76	1UBQ, 6Q00, 4XOK, 4XOL, 4XOF	1.78	0.83
Bovine pancreatic trypsin inhibitor	BPTI	NoSS	58	1BPI, 3LDJ, 9PTI	1.34	0.32
Carbonic anhydrase	CAH	$\alpha+\beta$	258	1CAH, 3KS3, 6PEA, 1CA2, 1HCA	1.69	0.56
Hemoglobin	Hem	$\alpha$	574	4HHB, 2DN2, 2HBS, 1COH	1.39	0.69
Ribonuclease A	Rnase	$\alpha+\beta$	108	6ETK, 1KF4, 2BPL, 1C0B, 1JWV	1.43	0.56
Thaumatin	Thm	$\beta$	207	5AVG, 1THI, 1PP3	1.32	1.16
Human serum albumin	HSA	$\alpha$	578	4G03, 1AO6, 4K2C	2.65	0.53
Turkey lysozyme	MLyz	$\alpha$	129	2LZ2, 3LZ2, 135L	2.00	0.62
T-cell glycoprotein	TGly	$\beta$	363	1WIQ, 1WIP, 1WIO	4.3	0.61
Uracil-DNA Glycosylase	UGlc	$\alpha+\beta$	223	2OYT, 2OXM, 1EMH, 1AKZ	1.37	0.40
Beta-secretase 1	BACE1	$\beta$	370	2ZHS, 2ZHT, 2ZHV	2.30	0.42
Toxic shock toxin 1	TST1	$\beta$	194	5TSS, 2TSS, 2QIL	2.34	0.48
Human Procathepsin	HPC	$\alpha+\beta$	317	1PBH, 2PBH, 3PBH	3.01	0.44
Gamma-B Crystallin	GBC	$\beta$	174	4GCR, 1GCS, 1AMM	1.56	0.41
Thermomyces Lanuginosa Lipase	TLL	$\alpha+\beta$	269	1TIB, 1DTE, 1DT3	2.26	0.39
Spasmolytic polypeptide	SPP	NoSS	105	1POS, 1PSP, 2PSP	1.35	0.35
Human neonatal Fc receptor	FCR	$\beta$	263	3M17, 3M1B, 5BJT	2.37	0.32
Cyclophilin A	CycA	$\beta$	164	4YUJ, 5F66, 5KUL	1.42	0.28
Lysin	Lys	$\alpha$	131	1LIS, 2LIS, 5I19	1.79	0.39
Tryparedoxin I	TRR	$\alpha+\beta$	143	1EWX, 1O7U, 1QK8	1.53	0.16

CATH secondary structure class, PDB IDs, and the average and standard deviation among all the resolution values

### Structural sensitivity calculations

In order to calculate the structural sensitivity, a saturated computational mutagenesis [35] (all amino-acids were mutated to all other 19 amino-acids to produce 19  $N$  mutants) was performed starting from all PDB structures with all six prediction methods, with a total of 87 structures subjected to saturated computational mutagenesis.

Structural sensitivity per mutation ( $SS_{mut}$ ) was defined as the standard deviation of the predicted  $\Delta\Delta G$  values (in kcal/mol) for one mutation among all PDB structures of the same protein. Structural sensitivity per site ( $SS_{site}$ ) was defined as the average  $SS_{mut}$  (in kcal/mol) for all mutations from a specific residue of one protein, i.e. the average of 19 mutations. Structural sensitivity per protein ( $SS_{prot}$ ) was defined as the average  $SS_{mut}$  (kcal/mol) for all mutations in a specific protein.

### Prediction methods studied

The studied prediction methods were selected based on their ability to model any mutation, to give a quantitative  $\Delta\Delta G$  prediction (rather than just qualitatively; destabilizing or stabilizing) and to work at high computational speed so that saturated computational mutagenesis was feasible. Structural sensitivity should depend on the model used for  $\Delta\Delta G$  prediction, and accordingly a diverse group of methods was desired to assess sensitivity broadly. Six publicly available predictors were explored in this study: FoldX [36], I-Mutant 3.0 [37], PoPMuSiC 2.1 [38], Maestro [39], mCSM [19] and CUPSAT [40].

The chosen methods use a variety of algorithms to compute the change in protein stability upon mutation: FoldX uses an empirical force field to calculate the free energy of folding for the wild-type and mutant structures. As the force field is rather sensitive to structure [41], a minimization of the wild-type structure was performed before prediction using the FoldX command *RepairPDB*. The differences between the minimized structures from *RepairPDB* and the original PDB structures are very small, with the maximum all-atom root mean square deviation (RMSD)  $< 0.01 \text{ \AA}$  (Additional file 1: Table S1), thus FoldX's structural sensitivity can be compared with the other methods. CUPSAT uses atomic potentials from chemical properties and empirically derived torsion potentials. I-Mutant 3.0 uses support-vector machines that account for amino acid substitution and structural environments. Similarly, Maestro combines support vector machines with a random-forest approach to obtain a consensus free energy. mCSM uses graph-based signatures that encode distance patterns between atoms. PoPMuSiC uses a statistical potential calculated from contact probabilities of amino acids close to the mutated residue. Thus, we studied one energy-based method (FoldX), two knowledge-based methods (PoPMuSiC and CUPSAT) and three machine learning methods (mCSM, I-Mutant 3.0 and Maestro).

Unless specified otherwise, all prediction programs were run with default parameters.

### Calculating global and local structural variables of amino-acids

The secondary-structure composition of each protein was taken from CATH [42]. The length of the proteins ( $N$ ) was considered as the number of amino-acids in the structure. Pairwise root-mean-square deviation (RMSD) was calculated with the PyMol [43] command *rms\_cur* after alignment of the structures. The average RMSD is reported as the average of all pairwise RMSD values for one protein.

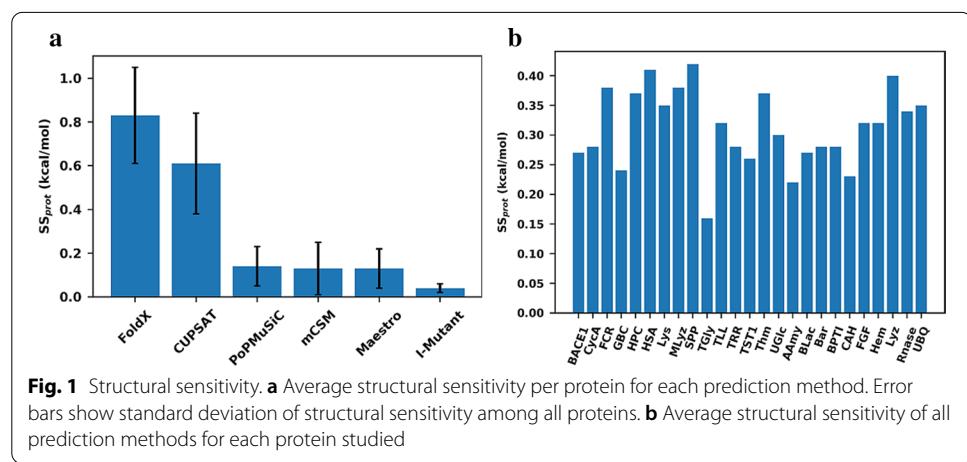
The secondary structure of each amino-acid was calculated using the *dssp* program [44] and then converted to a four-category secondary structure. The secondary structure per residue is reported as the consensus *dssp* calculation for all PDB structures used. The RMSD per residues is the average pairwise RMSD per residue between all structures of the same protein, calculated with the PyMol script *RmsdByResidue* after alignment. Relative solvent accessibility (RSA) was calculated with Naccess [45, 46], using default van der Waals atomic radii, and is reported as the average accessibility for all PDB structures of the same protein. Cystine bridges in each structure were predicted using the DisulfideByDesign 2.0 server [47, 48].

## Results

### Structural sensitivity measured for the full proteins

The 25 proteins were subjected to computational saturated mutagenesis, started from each of the selected structures for each protein. Figure 1a shows the average  $SS_{prot}$  for all six methods and the standard deviation. The individual data for each protein can also be found in Additional file 1: Table S1. Of the six methods, CUPSAT and FoldX displayed much larger structural sensitivity than the other four methods, i.e.  $SS_{prot} = 0.83$  kcal/mol and 0.61 kcal/mol, respectively. FoldX is an energy-based method that applies many terms in its energy function, such as electrostatic and van der Waals forces, whereas CUPSAT is a knowledge-based method that bases its energy calculation on statistics of torsional angles, which are very sensitive to differences in side-chain structure. We conclude that the magnitude of the structural sensitivity for these two methods is quite worrying, since, in perspective, the average  $\Delta\Delta G$  for any typical, random mutation is perhaps +1.0 kcal/mol on average. Thus, CUPSAT and FoldX come with an intrinsic imprecision that approaches the actual predicted value. We note that this does not necessarily imply lower trend accuracy, but it will certainly affect the predictive capacities of the methods. CUPSAT and FoldX also displayed the largest differences in structural sensitivity across the protein structures used, with a standard deviation of more than 0.2 kcal/mol in total for the 25 studied proteins.

In contrast, the other four studied methods displayed quite modest structural sensitivity, ranging from 0.04 kcal/mol for I-Mutant 3.0 to 0.14 kcal/mol for PoPMuSiC 2.0. I-Mutant, mCSM and Maestro are all machine-learning methods that do not take into account detailed features of the local environment, whereas PoPMuSiC is a knowledge-based method that captures interactions between close and distant amino-acids, but it is heavily parametrized, which can perhaps explain the much lower structural sensitivity compared to CUPSAT. Interestingly, PoPMuSiC, mCSM and Maestro all showed almost the same average structural sensitivity. Conversely, I-Mutant was very structurally insensitive, and produced very little differences between each protein. Although this means that any 3D structure can be used for prediction with I-Mutant, it also suggests that I-Mutant underreports structural information, i.e. is practically very close to a sequence-based method. Our analysis thus inspires a more quantitative view on sequence-versus



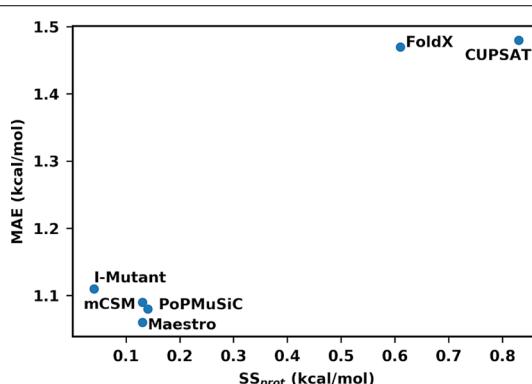
structure-based methods than just qualitative yes/no, with methods lying on a spectrum, as seen from their actual structure sensitivity.

Figure 1b shows the average  $SS_{prot}$  for each protein among all six methods. Surprisingly, the structural sensitivity did not vary much between proteins, with all showing averages around  $0.3 \pm 0.1$  kcal/mol, except one case, TGly, which has the lowest average resolution of the structures used ( $4.3 \text{ \AA}$ ), indicating that structural sensitivity is low when all structures are of poor quality. This low variance is confirmed by the low standard deviation of the structural sensitivity of the methods (Fig. 1a and Additional file 1: Table S1) and by a single-factor analysis of variance (ANOVA) performed for all 25 protein (Additional file 1: Table S3). This tendency is observed regardless of the variable quality of the structures used, regardless of the different spreads in resolution for each protein, and largely regardless of the presence of metal sites, as for Hem and CAH. Furthermore, the proteins for which more than three structures were used did not exhibit distinct structural sensitivities. These results indicate that structural sensitivity depends mostly on the prediction model itself and not so much on the features and differences between the protein structures selected for our study.

Comparing the computed precision of the six methods with the accuracy (mean absolute error) calculated in our previous study for a mutation-type balanced data set [33] reveals that the two methods with higher structural sensitivity also display lower accuracy for balanced data (Fig. 2). This suggests that accuracy and precision are correlated, and that structural sensitivity may play a part in the accuracy of the methods, along with the data set bias previously identified. It is therefore important to carefully select the protein structures used for training the models and assess the impact of structural sensitivity during training. We note that CUPSAT and FoldX still work quite well for trend predictions, as they retain accuracy and precision locally for certain mutations, as shown in previous benchmarks.

#### Factors that affect structural sensitivity

In order to understand and control the structural sensitivity issue identified in Fig. 1, we must first understand the factors that influence the structural sensitivity of each method, for example whether certain proteins are more structurally sensitive than others.



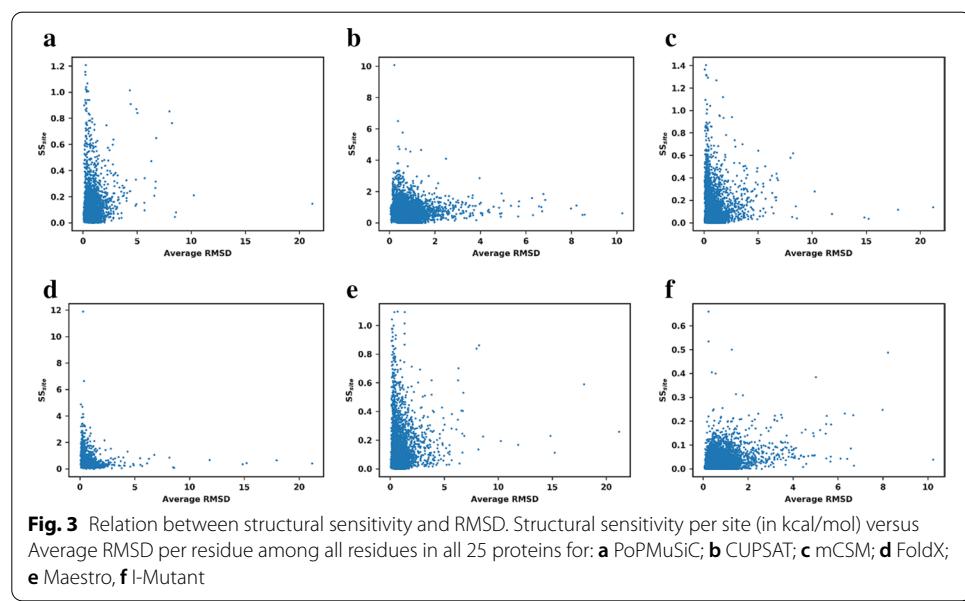
**Fig. 2** Accuracy and precision of studied methods. Mean absolute deviation (MAE, in kcal/mol) of the six methods against the balanced O2567 data set, as calculated in Caldararu et al. [33] versus average structural sensitivity per protein (in kcal/mol) for each prediction method

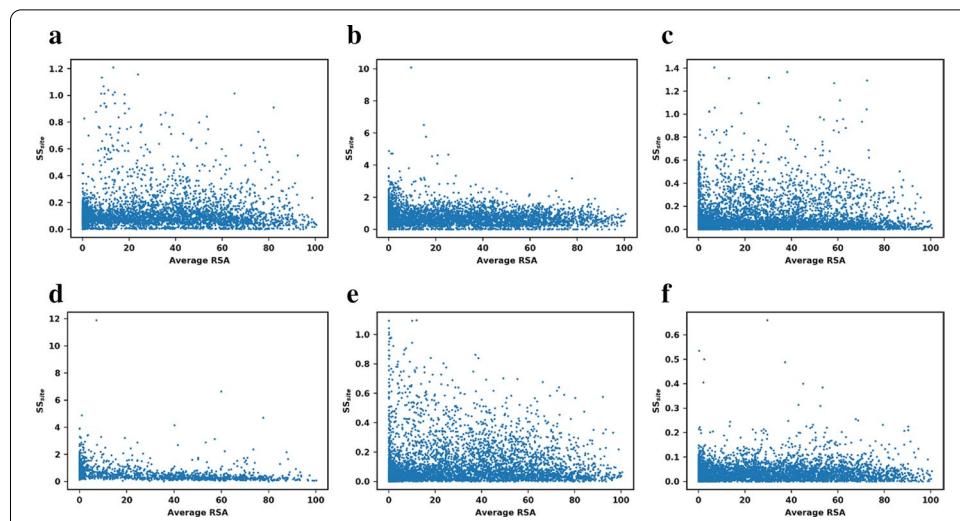
Moreover, mutation studies often focus on certain sites and residues in a protein, and thus it is important to determine which type of residues are mostly contributing structural sensitivity during prediction.

The most logical cause of structural sensitivity is the structural difference between the PDB structures used, which can be measured by the average RMSD between all structures. To determine whether amino-acids that are in different conformations in different structures display larger differences than amino-acids in the same conformation we plotted  $SS_{site}$  for all the residues in all 25 proteins against the average RMSD per residue (Fig. 3). Strikingly, no methods showed any correlation between RMSD and structural sensitivity. Actually, most sites that had high structural sensitivity also had an RMSD relatively close to 0, i.e. the conformation of the residues was the same in all structures used. Furthermore, rigid residues (residues with a B-factor close to the average B-factor of the protein) were also found to have higher structural sensitivity than flexible residues (Additional file 1: Fig. S1). Although this might at first seem counter-intuitive, it implies that prediction methods are more structurally sensitive to buried residues, which usually have the same conformation in all structures of the same protein and are more rigid, and usually associated with larger energy effects. More flexible residues on the other hand typically reflect low-energy modes such as rotations, which may not affect  $\Delta\Delta G$  as much.

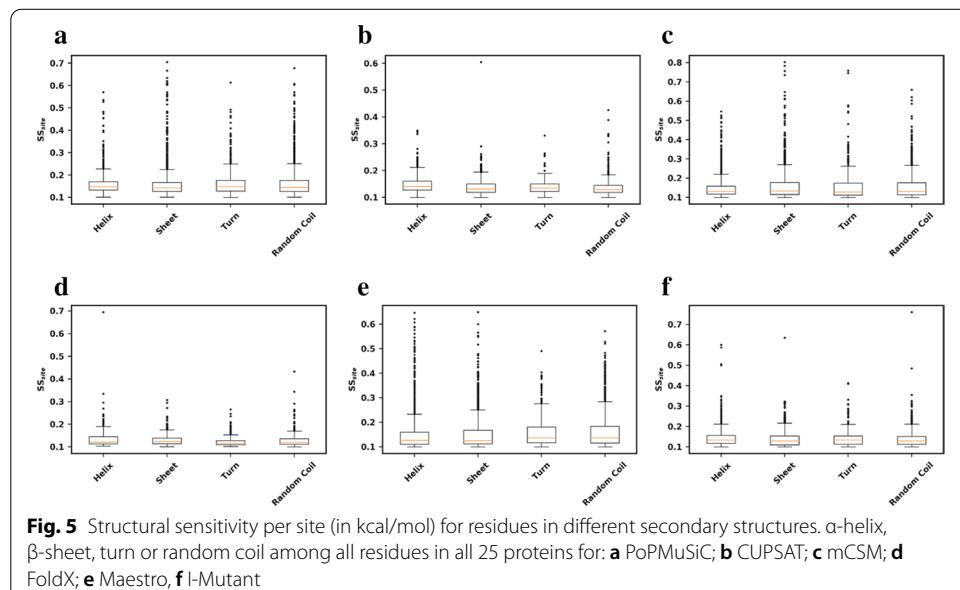
For these reasons we also studied the relationship between the relative solvent accessibility (RSA) of each residue and  $SS_{site}$ . The results, illustrated in Fig. 4, again revealed no correlation between the RSA and the structural sensitivity for any of the methods. Several outliers for FoldX and I-Mutant are indeed buried residues, but we observed no general trend.

Next, we investigated whether the secondary structure of each residue influenced structural sensitivity, as many prediction methods contain a term for the secondary structure of the wild-type residue during  $\Delta\Delta G$  calculation. As shown in Fig. 5, the average  $SS_{site}$  for all residues in the 25 proteins was similar for all four types of





**Fig. 4** Structural sensitivity per site (in kcal/mol) versus Average RSA per residue. The plots show all residues in all 25 proteins for the methods: **a** PoPMuSiC; **b** CUPSAT; **c** mCSM; **d** FoldX; **e** Maestro; **f** I-Mutant

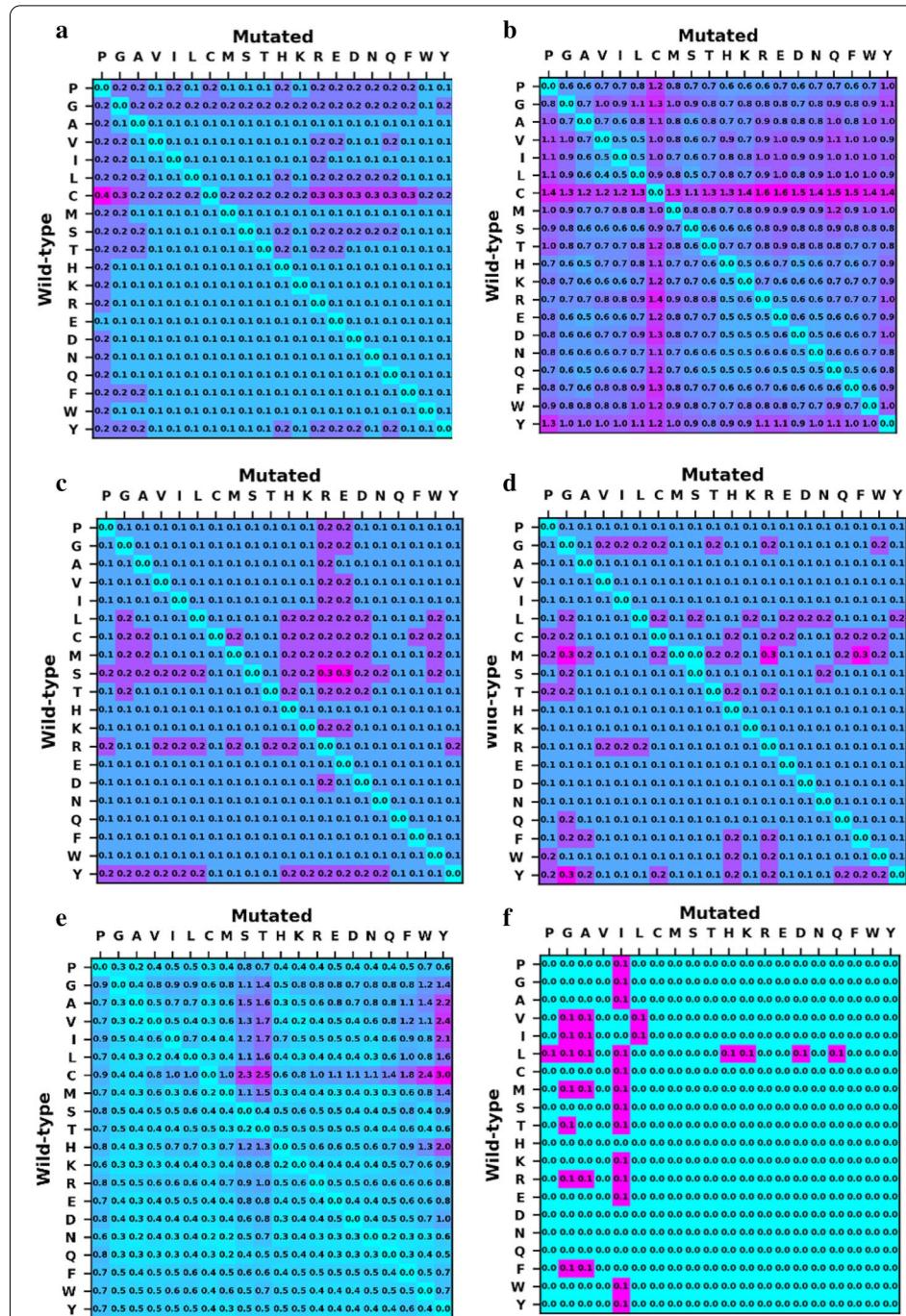


**Fig. 5** Structural sensitivity per site (in kcal/mol) for residues in different secondary structures.  $\alpha$ -helix,  $\beta$ -sheet, turn or random coil among all residues in all 25 proteins for: **a** PoPMuSiC; **b** CUPSAT; **c** mCSM; **d** FoldX; **e** Maestro; **f** I-Mutant

secondary structure in all six prediction methods, with outliers also being similarly shared across the four types of secondary structure.

In a previous study, we showed that mutation type greatly affects method accuracy, and none of the studied methods were generally transferable and balanced in this context [33]. Therefore, we studied in the present work whether mutation type also affects the precision (i.e. structural sensitivity) of the methods. A mutation type was considered to have high sensitivity if its average SS<sub>mut</sub> was significantly higher than the average for the method, as shown in Fig. 1.

Figure 6 shows the average SS<sub>mut</sub> for each of the 380 mutation types in all 25 proteins and for each of the prediction methods. We observe that each prediction



**Fig. 6** Structural sensitivity per mutation. Values shown in kcal/mol for all mutations of a certain mutation type among all mutations studied in all 25 proteins for: **a** PoPMuSiC; **b** CUPSAT; **c** mCSM; **d** FoldX; **e** Maestro, **f** I-Mutant

method is more structurally sensitive for certain mutation types. PoPMuSiC is sensitive to mutations from C (Fig. 6a), whereas CUPSAT has higher sensitivity for all mutations involving C (Fig. 6b). mCSM presented higher sensitivity for mutations involving charged residues, especially R and E (Fig. 6c). FoldX showed the most

substantial differences, with mutations to hydroxyl-containing amino-acids (S, T, Y) having significantly higher structural sensitivity than other mutations (Fig. 6d). Maestro (Fig. 6e) and I-Mutant (Fig. 6f) were more balanced, although Maestro had a slightly higher structural sensitivity for mutations from M and I-Mutant showed a similar behaviour for mutations to I.

Although none of the structures used in this study had any explicit cystine bridges in their PDB files, we investigated more closely if cystine bridges could occur in the selected proteins, since two methods showed particularly high structural sensitivity for mutations from C. We separated the data into two separate data sets, proteins with predicted cystine bridges and proteins without any predicted cystine bridges (Additional file 1: Table S4) and recalculated the average  $SS_{mut}$  for each mutation type for CUP-SAT and PoPMuSiC (Additional file 1: Fig. S2). Our results indicate that the sensitivity of PoPMuSiC towards mutations from C may be caused by possible cystine bridges, whereas CUP-SAT was not affected by the presence of cystine bridges.

In summary, we conclude that among the properties potentially contributing to structural sensitivity studied here, the mutation type has the highest effect, and thus the precision of a method may be quite dependent on the mutation type studied, as we saw previously in terms of accuracy.

As a final note, we also studied several global parameters of the proteins: CATH structure (Additional file 1: Table S5), length and average global RMSD (Additional file 1: Table S6) but none showed any significant correlation to structural sensitivity. As  $SS_{site}$  has shown more variation than  $SS_{protein}$  it is expected that these global parameters will have little effect on the precision of the methods.

## Discussion

In this study, we have tried to evaluate the precision of commonly used protein stability prediction methods, defined as their structural sensitivity evaluated for all possible mutations in 25 proteins. Our results show that structural sensitivity varies substantially among the six studied methods and intriguingly cluster in two groups—those that are highly trained and those that are highly dependent on local environment of the mutation (FoldX and CUP-SAT), which displayed high structural sensitivity (0.6 and 0.8 kcal/mol, respectively). This is a matter of concern since the average  $\Delta\Delta G$  of a typical random mutation is of the order of ~ 1.0 kcal/mol. The methods still work to some extent because they retain accuracy locally in any structure used, but they are not very precise in our definition. On the other hand, machine-learning methods (mCSM, Maestro and I-Mutant) and parametrized knowledge-based methods (PoPMuSiC) are very insensitive to structure choice.

To deepen our understanding of structural sensitivity, we also studied what factors cause it. Our results indicate that there is no correlation between the structural sensitivity of residues and the RMSD between the structures used, probably because low-energy modes define much of the structural RMSD. Furthermore, neither the solvent accessibility, secondary structure nor B-factor values of the amino-acids had any significant correlation with structural sensitivity. Instead, the only factor that seems to affect precision was the mutation type, with different methods being more sensitive to different mutation types. Thus, structural sensitivity is caused by either the modelling of the wild-type

structure or the parametrization bias towards some mutation types as touched upon in our recent related work [33]. The correlation in Fig. 2 shows that mutation-type is the most salient feature of both accuracy and precision and that none of these should be evaluated on mutation-type biased data if wider conclusions on performance are to be drawn. More importantly, and perhaps controversially, it is very interesting that all the studied machine-learning methods tend to train away almost entirely the used wild-type structure.

The least structure-sensitive methods are also the most accurate for our balanced benchmark data, and their structural sensitivity ( $\sim 0.1$  kcal/mol) is within the experimental uncertainty in the data. This may controversially indicate that folded wild type structures provide modest value to  $\Delta\Delta G$  predictions, which, if true, could be because the three “other” structures of the thermodynamical cycle are missing. Some methods seem to overemphasize structure, probably side chain conformations, which may differ substantially in mutant and wild type unfolded states that produce the experimental data. We invite further studies to settle this question and also note that there is no “optimal” structural sensitivity except that which provides highest accuracy upon independent benchmarking.

## Conclusions

Users of stability prediction methods are often faced with a choice between many possible structures for a wild type protein of interest, and we wanted to explore in this study how this affects outcome. Our study provides a base measure of the precision of methods in relation to structure input used for calculations. This will be important both as a general tendency and for specific mutation types of distinct sensitivity in most studies using these methods, and thus we recommend that triplicate structures are used as input and the standard deviation of the  $\Delta\Delta G$  reported as a best-practice for these methods, of course only if several reasonable structures are available. The specific structural sensitivities for each method and mutation type reveal the aspects needed to be improved in each method in order to optimize structural sensitivity.

Finally, instead of just considering sequence- and structure-based methods as either/or, our base measure provides a spectrum of actual structural emphasis of the methods, noting that a machine-learning could in principle “train away” structure completely even if used formally. The fact that all studied machine-learning methods are rather insensitive to structure input raises a perhaps controversial question on the relevance of folded wild-type structures alone (without folded mutant and unfolded structures) to  $\Delta\Delta G$  prediction altogether. However, we show that a substantial reason for imprecision is, as for accuracy [33], bias from certain mutation types arising from training on imbalanced data sets, via their effect on local residue geometry. We do not claim to answer these questions in any completeness but given its importance suggest that this is explored further in future studies using other data sets.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-021-04030-w>.

**Additional file 1.** The supporting information file (SupplInfo.pdf) contains full data for structural sensitivity per protein, and additional information on relations between structural sensitivity and B-factors and mutation types.

**Abbreviations**

PDB: Protein Data Bank; SS<sub>mut</sub>: Structural sensitivity per mutation; SS<sub>site</sub>: Structural sensitivity per wild-type residue; SS<sub>prot</sub>: Structural sensitivity per protein; RMSD: Root mean square deviation; RSA: Relative solvent accessibility.

**Acknowledgements**

Not applicable.

**Authors' contributions**

All authors designed the study together. OC performed all computations and analyzed and discussed the data with help from KPK and TLB. OC and KPK wrote the manuscript, with input from TLB. All authors read and approved the final manuscript.

**Funding**

The Danish Council for Independent Research (Grant Case: 8022-00041B) is gratefully acknowledged for providing funding for OC. The funding agency had no role in the actual study.

**Availability of data and materials**

The 3D protein structures that have been used in this study are publicly available online in the PDB, [www.rcsb.org](http://www.rcsb.org). The prediction programs used are freely available as servers or stand-alone programs. The website of each prediction program is listed below. FoldX: <http://foldxsuite.crg.eu/>; Maestro: <https://pbwww.che.sbg.ac.at/maestro/web/>; PoPMuSiC v 2.1: <https://soft.dezyme.com/>; CUPSAT: <http://cupsat.tu-bs.de/>; mCSM: <http://biosig.unimelb.edu.au/mcsm/>; I-Mutant: <http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi>. Raw prediction data are available from the corresponding author upon request.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

Professor Tom L. Blundell is one of the authors of the mCSM method studied in this work.

**Author details**

<sup>1</sup> DTU Chemistry, Technical University of Denmark, Building 206, 2800 Kgs. Lyngby, Denmark. <sup>2</sup> Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK.

Received: 30 September 2020 Accepted: 15 February 2021

Published online: 25 February 2021

**References**

- Huang P-S, Boyken SE, Baker D. The coming of age of de novo protein design. *Nature*. 2016;537:320.
- Street AG, Mayo SL. Computational protein design. *Structure*. 1999;7:R105–9.
- Yeung N, Lin Y-W, Gao Y-G, Zhao X, Russell BS, Lei L, et al. Rational design of a structural and functional nitric oxide reductase. *Nature*. 2009;462:1079–82.
- Zeldovich KB, Chen P, Shakhnovich EI. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc Natl Acad Sci USA*. 2007;104:16152–7. <https://doi.org/10.1073/pnas.0705366104>.
- Dasmeh P, Kepp KP. Superoxide dismutase 1 is positively selected to minimize protein aggregation in great apes. *Cell Mol Life Sci*. 2017;74:3023–37. <https://doi.org/10.1007/s00018-017-2519-8>.
- Dasmeh P, Serohijos AWR, Kepp KP, Shakhnovich EI. The influence of selection for protein stability on dN/dS estimations. *Genome Biol Evol*. 2014;6:2956–67. <https://doi.org/10.1093/gbe/evu223>.
- Kepp KP, Dasmeh P. A model of proteostatic energy cost and its use in analysis of proteome trends and sequence evolution. *PLoS ONE*. 2014;9:e90504. <https://doi.org/10.1371/journal.pone.0090504>.
- Godoy-Ruiz R, Perez-Jimenez R, Ibarra-Molero B, Sanchez-Ruiz JM. Relation between protein stability, evolution and structure, as probed by carboxylic acid mutations. *J Mol Biol*. 2004;336:313–8.
- Goldstein RA. The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins*. 2011;79:1396–407. <https://doi.org/10.1002/prot.22964>.
- Kucukkal TG, Petukh M, Li L, Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr Opin Struct Biol*. 2015;32:18–24.
- Petukh M, Kucukkal TG, Alexov E. On human disease-causing amino acid variants: statistical study of sequence and structural patterns. *Hum Mutat*. 2015;36:524–34. <https://doi.org/10.1002/humu.22770>.
- Yue P, Li Z, Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol*. 2005;353:459–73.
- Topham CM, Srinivasan N, Blundell TL. Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng*. 1997;10:7–21.
- Kulshreshtha S, Chaudhary V, Goswami GK, Mathur N. Computational approaches for predicting mutant protein stability. *J Comput Aided Mol Des*. 2016;30:401–12.
- Montanucci L, Savojardo C, Martelli PL, Casadio R, Fariselli P. On the biases in predictions of protein stability changes upon variations: the INPS test case. *Bioinformatics*. 2019;35:2525–7.
- Pucci F, Bernaerts KV, Kwasigroch JM, Rooman M. Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics*. 2018;34:3659–65.

17. Potapov V, Cohen M, Schreiber G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel.* 2009;22:553–60. <https://doi.org/10.1093/protein/gzp030>.
18. Worth CL, Preissner R, Blundell TL. SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucl Acids Res.* 2011;39(Web Server issue):W215–22. <https://doi.org/10.1093/nar/gkr363>.
19. Pires DEV, Ascher DB, Blundell TL. MCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics.* 2014;30:335–42.
20. Gilis D, Rooman M. PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins. *Protein Eng.* 2000;13:849–56.
21. Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, Rooman M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics.* 2009;25:2537–43.
22. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucl Acids Res.* 2005;33:W306–10.
23. Khan S, Vihtinen M. Performance of protein stability predictors. *Hum Mutat.* 2010;31:675–84. <https://doi.org/10.1002/humu.21242>.
24. Chen YW, Dodson EJ, Kleywegt GJ. Does NMR mean “not for molecular replacement”? Using NMR-based search models to solve protein crystal structures. *Structure.* 2000;8:R213–20.
25. Hinsen K. Structural flexibility in proteins: impact of the crystal environment. *Bioinformatics.* 2008;24:521–8.
26. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucl Acids Res.* 2004;32(Suppl\_1):D120–1.
27. Sasidharan Nair P, Vihtinen M. VariBench: a benchmark database for variations. *Hum Mutat.* 2013;34:42–9.
28. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Wessig H, et al. The protein data bank. *Nucl Acids Res.* 2000;28:235–42. <https://doi.org/10.1093/nar/28.1.235>.
29. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol.* 2002;320:369–87.
30. Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins Struct Funct Bioinforma.* 2011;79:830–8.
31. Kepp KP. Computing stability effects of mutations in human superoxide dismutase 1. *J Phys Chem B.* 2014;118:1799–812. <https://doi.org/10.1021/jp4119138>.
32. Kepp KP. Towards a “Golden Standard” for computing globin stability: stability and structure sensitivity of myoglobin mutants. *Biochim Biophys Acta Proteins Proteomics.* 2015;1854:1239–48.
33. Caldararu O, Mehra R, Blundell TL, Kepp KP. Systematic investigation of the data set dependency of protein stability predictors. *J Chem Inf Model.* 2020;60:4772–84.
34. Niroula A, Vihtinen M. Variation interpretation predictors: principles, types, performance, and choice. *Hum Mutat.* 2016;37:579–97.
35. Fischer A, Seitz T, Lochner A, Sterner R, Merkl R, Bocola M. A fast and precise approach for computational saturation mutagenesis and its experimental validation by using an artificial  $\beta$ -barrel protein. *ChemBioChem.* 2011;12:1544–50.
36. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucl Acids Res.* 2005;33:W382–8. <https://doi.org/10.1093/nar/gki387>.
37. Capriotti E, Fariselli P, Rossi I, Casadio R. A three-state prediction of single point mutations on protein stability changes. *BMC Bioinform.* 2008;9(Suppl 2):S6. <https://doi.org/10.1186/1471-2105-9-S2-S6>.
38. Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinform.* 2011;12:151.
39. Laimer J, Hofer H, Fritz M, Wegenkittl S, Lackner P. MAESTRO—multi agent stability prediction upon point mutations. *BMC Bioinform.* 2015;16:116.
40. Gromiha MM. Prediction of protein stability upon point mutations. *Biochem Soc Trans.* 2007;35(Pt 6):1569–73.
41. Christensen NJ, Kepp KP. Accurate stabilities of laccase mutants predicted with a modified FoldX protocol. *J Chem Inf Model.* 2012;52:3028–42.
42. Knudsen M, Wiuf C. The CATH database. *Hum Genomics.* 2010;4:207–12.
43. DeLano WL. PyMol: an open-source molecular graphics tool. *CCP4 News/ Protein Crystallogr.* 2002;40:82–92.
44. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983;22:2577–637.
45. Hubbard S, Thornton J. NACCESS; 1993.
46. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol.* 1971;55:379–400.
47. Craig DB, Dombrowski AA. Disulfide by design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinform.* 2013;14:1–7.
48. Dombrowski AA. Disulfide by Design<sup>TM</sup>: a computational method for the rational design of disulfide bonds in proteins. *Bioinformatics.* 2003;19:1852–3.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.