


RosettaDDGPrediction for high-throughput mutational scans: From stability to binding

Valentina Sora^{1,2} | Adrian Otamendi Laspiur² | Kristine Degn² |
Matteo Arnaudi^{1,2} | Mattia Utichi^{1,2} | Ludovica Beltrame^{1,2} |
Dayana De Menezes² | Matteo Orlandi² | Ulrik Kristoffer Stoltze^{3,4,5} |
Olga Rigina² | Peter Wad Sackett² | Karin Wadt^{3,5} | Kjeld Schmiegelow^{4,5} |
Matteo Tiberti¹ | Elena Papaleo^{1,2} 

¹Cancer Structural Biology, Danish Cancer Society Research Center, Copenhagen, Denmark

²Cancer Systems Biology, Section for Bioinformatics, Department of Health and Technology, Technical University of Denmark, Lyngby, Denmark

³Department of Clinical Genetics, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

⁴Department of Pediatrics and Adolescent Medicine, University Hospital Rigshospitalet, Copenhagen, Denmark

⁵Institute of Clinical Medicine, Faculty of Medicine, University of Copenhagen, Copenhagen, Denmark

Correspondence

Elena Papaleo, Cancer Structural Biology, Danish Cancer Society Research Center, 2100 Copenhagen, Denmark.
Email: elpap@dtu.dk; elenap@cancer.dk

Funding information

Carlsberg Foundation Distinguished Fellowship, Grant/Award Number: CF18-0314; Danmarks Grundforskningsfond, Grant/Award Number: DNR125; Hartmanns Fond, Grant/Award Number: R241-A33877; LEO Foundation, Grant/Award Number: LF17006; NovoNordisk Fonden Bioscience and Basic Biomedicine, Grant/Award Number: NNF20OC0065262; European Regional Development Fund; Danish Cancer Society, Grant/Award Number: R-257-A14720; Danish Childhood Cancer Foundation, Grant/Award Numbers: 2020-5769, 2019-5934

Review Editor: Nir Ben-Tal

Abstract

Reliable prediction of free energy changes upon amino acid substitutions ($\Delta\Delta G$ s) is crucial to investigate their impact on protein stability and protein-protein interaction. Advances in experimental mutational scans allow high-throughput studies thanks to multiplex techniques. On the other hand, genomics initiatives provide a large amount of data on disease-related variants that can benefit from analyses with structure-based methods. Therefore, the computational field should keep the same pace and provide new tools for fast and accurate high-throughput $\Delta\Delta G$ calculations. In this context, the Rosetta modeling suite implements effective approaches to predict folding/unfolding $\Delta\Delta G$ s in a protein monomer upon amino acid substitutions and calculate the changes in binding free energy in protein complexes. However, their application can be challenging to users without extensive experience with Rosetta. Furthermore, Rosetta protocols for $\Delta\Delta G$ prediction are designed considering one variant at a time, making the setup of high-throughput screenings cumbersome. For these reasons, we devised RosettaDDGPrediction, a customizable Python wrapper designed to run free energy calculations on a set of amino acid substitutions using Rosetta protocols with little intervention from the user. Moreover, RosettaDDGPrediction assists with checking completed runs and aggregates raw data for multiple variants, as well as generates publication-

Valentina Sora and Adrian Otamendi Laspiur equally contributed to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

ready graphics. We showed the potential of the tool in four case studies, including variants of uncertain significance in childhood cancer, proteins with known experimental unfolding $\Delta\Delta G$ s values, interactions between target proteins and disordered motifs, and phosphomimetics. RosettaDDGPrediction is available, free of charge and under GNU General Public License v3.0, at <https://github.com/ELELAB/RosettaDDGPrediction>.

KEYWORDS

binding free energy, folding free energy, free energy calculations, Rosetta

1 | INTRODUCTION

Predicting the impact of amino acid substitutions in a protein or at a protein–protein interface is becoming more and more relevant as high-throughput sequencing data reveal a high rate of sequence polymorphisms of uncertain functional significance in protein-coding regions (Federici & Soddu, 2020). In this context, multiplex-based assays provide massive data that can be complemented by structural studies on the effects of protein variants (Anderson et al., 2022; Cagiada et al., 2021; Gasperini et al., 2016; Ollodart et al., 2021; Weile & Roth, 2018). Furthermore, saturation mutagenesis is experimentally accessible thanks to the advances in multiplex technologies. Therefore, molecular modeling approaches must keep the same pace and continue developing toward high-throughput applications.

A convenient and quantitative manner for assessing the impact of amino acid substitutions related to coding variants is based on estimating the changes in Gibbs free energy of folding/unfolding or binding. In this context, several computational approaches based on the analysis of protein structures are available to predict free energy changes upon mutation ($\Delta\Delta G$ s) in protein structures (Barlow et al., 2018; Delgado et al., 2019; Frenz et al., 2020; Geng et al., 2019; Kortemme & Baker, 2002; Kumari et al., 2014; Park et al., 2016; Schymkowitz et al., 2005; Seeliger & de Groot, 2010; Smith & Kortemme, 2008). These measurements can be used to classify the effect of disease-related variants on protein structural stability. As a consequence, they provide predictions of potential alterations in the protein cellular level, propensity to aggregation or proteasomal degradation (Gerasimavicius et al., 2020; Stein et al., 2019). In addition, they can also pinpoint functional effects due to local changes in the interactions with other proteins or biomolecules (Degn et al., 2022; Fas et al., 2020; Jepsen et al., 2020).

Rosetta provides a variety of protocols to estimate changes in free energy in terms of binding and folding/unfolding (Barlow et al., 2018; Frenz et al., 2020; Kellogg et al., 2011; Kortemme & Baker, 2002; Park et al., 2016). Most of these protocols estimate the change in free energy as an average over the free energy changes calculated in an ensemble of paired wild-type/mutated structures.

Three features generally characterize Rosetta protocols for the prediction of free energy changes upon mutation: (i) the sampling method employed to generate the structural ensemble, (ii) the energy function used to quantify the free energy associated with each structure, and (iii) the degree of flexibility allowed in the structure to accommodate the mutation.

Currently, three state-of-the-art strategies are available in Rosetta to estimate the change in either folding or binding free energy upon mutation. The first one, presented by Park and coworkers (Park et al., 2016) and referred to as *cartddg*, is designed to work on monomeric proteins. In this protocol, a sampling in the Cartesian space (as opposed to internal dihedrals sampling) is carried out, allowing small local backbone movements in a three-residue window around the mutation site, together with side-chains movements within a 6 Å radius from the mutation site. The second protocol, *cartddg2020*, represents an updated variant of *cartddg* (Frenz et al., 2020). The third protocol, developed by Barlow and coworkers (Barlow et al., 2018) and named here *flexddg*, deals with estimating the changes in binding free energy upon mutation in a protein complex. It applies the “backrub” sampling method (Smith & Kortemme, 2008) to recapitulate local backbone motions observed in crystal lattices. The *flexddg* protocol seems to perform better with the *talaris2014* energy function (Barlow et al., 2018). This protocol for binding free energies relies on a local sampling of backbone and side chains for residues within an 8 Å radius from the mutation, followed by global optimization of the side chains.

Rosetta is a feature-rich software suite under active development, backed by a sizable community of users, and built over roughly 20 years. Running the protocols mentioned above directly with Rosetta requires an extensive computational background and prior exposure to several Rosetta features. These requirements may discourage users with a more biology-oriented skillset, despite the benefit that accurate predictions of free energy changes upon mutations may bring to their research. Furthermore, Rosetta protocols for $\Delta\Delta G$ prediction are designed to be run considering one mutation at a time exclusively, making high-throughput screenings cumbersome to set up. We recently faced a similar challenge with implementing high-throughput scans based on the FoldX free energy function to make them parallelizable, more easily approachable, and applicable to structural ensembles. This led to the development of MutateX (Tiberti et al., 2022). FoldX, however, is known to suffer from limitations due to backbone stiffness during the sampling (Usmanova et al., 2018) and often low accuracy in predicting mutations with stabilizing effects, even though most prediction methods are biased toward destabilizing mutations (Buß et al., 2018; Usmanova et al., 2018). Rosetta-based calculations could offer a valuable complement to the $\Delta\Delta G$ estimates currently accessible with MutateX. Thus, we developed RosettaDDGPrediction, a Python wrapper to perform Rosetta-based protocols for $\Delta\Delta G$ prediction. RosettaDDGPrediction's outputs can also be converted to a format compatible with the MutateX plotting system, allowing for an expanded visualization toolkit. Here, we illustrate the applications and limits of the approach to four different cases of study, covering both methodological and biological applications. We focused on the comparison with experimentally determined unfolding $\Delta\Delta G$ values (Case Study 1). We showed an example of the application of RosettaDDGPrediction to the study of protein-protein interactions and posttranslational modifications (PTMs) (Case Study 2). We then evaluated the influence of using AlphaFold2 models as starting structures for the calculations (Case Study 3). We then used models from AlphaFold2 to assess the functional impact of mutations identified by whole genome sequencing to address cancer predisposition (Case Study 4).

2 | RESULTS

2.1 | Overview of the package

RosettaDDGPrediction is a pure Python package providing a uniform and easily accessible command-line interface to *flexddg*, *cartddg*, and *cartddg2020* protocols for

calculating free energy changes upon mutation. It is devised to help users unfamiliar with the Rosetta suite perform mutational scans and collect, aggregate, and visualize data from those scans in an intuitive fashion. In RosettaDDGPrediction, a “protocol” is intended as a set of Rosetta runs and Python-based processing steps. Each protocol takes as inputs the three-dimensional (3D) structure of the protein of interest and a list of mutations to be performed, finally returning the predicted free energy changes associated with each input mutation. The *flexddg* protocol consists of only one call to the *rosetta_scripts* executable for each mutation, which performs all the necessary calculations as defined by Barlow and coworkers (Barlow et al., 2018). On the other hand, the *cartddg* protocol first energetically relaxes the input structure by using the Rosetta *relax* program to generate an ensemble of relaxed conformations, followed by the selection of the most suitable one. Finally, it uses the *cartesian_ddg* application to relax the structure further and perform the free energy calculations. The *cartddg2020* protocol represents an updated version of the original *cartddg* protocol. Here, the relaxation is performed by a Rosetta script passed to the *rosetta_scripts* executable, and then *cartesian_ddg* is run on the lowest energy structure produced by the relaxation. It is worth noting that the relaxation procedure produces only one structure, as per the original files provided with the work first describing the *cartddg2020* protocol (Frenz et al., 2020). However, if the user decides to produce several relaxed structures, the most suitable one (according to user-selected criteria) will then be passed to *cartesian_ddg*. The standard protocols are described in specific YAML files provided with the package. With these files, expert users can still tap into the full potential of the Rosetta interface by providing virtually any Rosetta-compatible option to the executables used by each protocol.

RosettaDDGPrediction consists of four main executables (*rosetta_ddg_run*, *rosetta_ddg_check_run*, *rosetta_ddg_aggregate*, *rosetta_ddg_plot*) performing different tasks (Figure 1). Their behavior is controlled by a set of configuration files, which can be fully customized to fine-tune the parameters of each protocol, aggregation options, and plot aesthetics.

rosetta_ddg_run is the executable responsible for running a Rosetta protocol to predict free energy changes upon mutation over a set of selected mutations. Given a protein structure in PDB format and a set of mutations, it generates all the data structures and configuration files to perform several runs in parallel, making them straightforward to perform and making the most of modern many-cores computing infrastructures.

In *rosetta_ddg_run*, the user can specify the amino acid substitutions to be performed in two different ways.

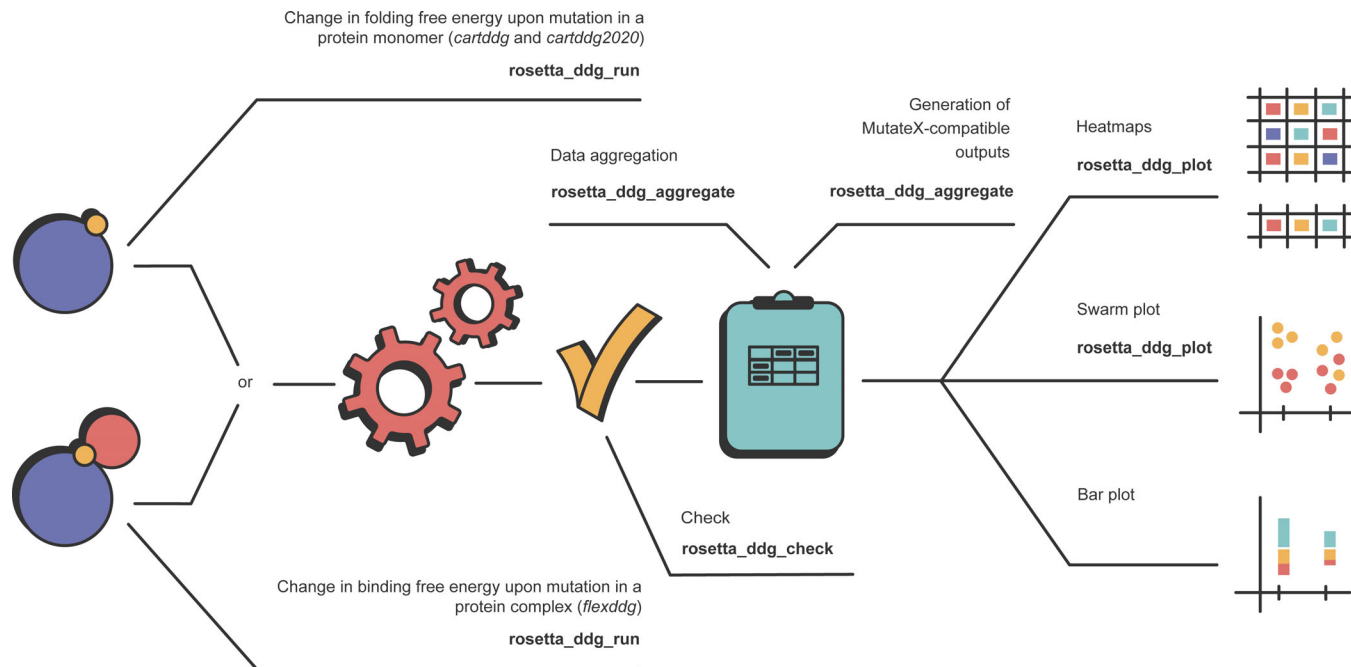


FIGURE 1 The RosettaDDGPrediction workflow and schematized plot types. The first step consists in running the *rosetta_ddg_run* executable to obtain the predicted $\Delta\Delta G$ values for the changes in folding free energy (for monomeric proteins) or binding free energy (for protein complexes). Then, *rosetta_ddg_check* can be used to ensure that all runs have been completed successfully. Data aggregation can then be performed with *rosetta_ddg_aggregate*, and aggregate data can finally be visualized in different ways (heatmaps, bar plots, swarm plots) using *rosetta_ddg_plot*.

In details, it is possible to provide a list of mutations containing both the positions and the residues to which such positions should be mutated (i.e., every single mutation is uniquely identified by the protein chain, the wild-type residue, the position of the residue in the chain, and the mutated residue). As an alternative, the user can specify multiple mutations to be performed simultaneously on different protein residues. On the other hand, the user can also pass a list of residues (each one identified by the protein chain, the wild-type residue, and the position of the residue in the chain) and a list of residue types. In the latter case, all the residues specified in the first list will be mutated, one at a time, to each residue type specified in the second list. This allows RosettaDDGPrediction to implement saturation mutagenesis scans alongside scans of specific mutations.

The *rosetta_ddg_run* executable can optimize the workload distribution over the available resources to ensure efficient scheduling of the runs, thanks to the Dask Python package operating under the hood. *rosetta_ddg_run* easily handles multistep protocols, requiring sequential Rosetta calls and possibly processing the output data between the steps. For example, for the aforementioned *cartddg* and *cartddg2020* protocols, *rosetta_ddg_run* takes care of both Rosetta calls and the processing steps.

Once the runs are completed, users can perform a sanity check on the calculations using *rosetta_ddg_check_run*, which identifies problematic runs by scraping the Rosetta output files. Finally, *rosetta_ddg_aggregate* can aggregate raw data from the large numbers of collected mutation runs into easily readable CSV table files for successfully completed runs. These aggregate files contain, together with the calculated differences in free energy, additional information about each mutation, the Rosetta energy function used, and the number of structures generated. *rosetta_ddg_aggregate* also allows generating aggregate outputs compatible with the MutateX plotting system. Indeed, MutateX offers additional visualization tools, including density plots, logo plots, distribution plots, and summary tables that can be easily navigated (Tiberti et al., 2022).

Finally, *rosetta_ddg_plot* provides plotting utilities to explore the aggregated data through several visualization types, such as one-dimensional or two-dimensional heatmaps. The latter is particularly convenient when a saturation mutagenesis scan is run on a set of positions. In addition, the contribution of each term of the energy function to the final $\Delta\Delta G$ values can be visualized as stacked bar plots, where positive and negative contributions add up on the corresponding semiaxes. Finally, since all protocols implemented so far in

TABLE 1 Examples of performances of RosettaDDGPrediction for different protein sizes, the number of cores, and protocols applied

Protein size	Number of cores	Protocol	Time (h)
120	16	cartddg (ref2015)	33
250	24	cartddg (ref2015)	67
250	8	cartddg (talaris2014)	89
340	16	cartddg (ref2015)	160
340	16	cartddg (talaris2014)	70
340	1	Relax	16
600	1	Relax	40
900	1	Relax	65
120; 17 ^a	40	flexddg (talaris2014)	67

Note: In the case of complexes, the “protein size” column includes two values, that is, one value for each protein/peptide in the complex. Calculations were run on servers equipped with either dual Xeon 6142 processors or dual Xeon 6242 processors. Each processor features 32 cores. The estimate refers to calculations done with Rosetta version 3.12.

^aThe one for which the saturation mutational scan was carried out.

RosettaDDGPrediction determine the $\Delta\Delta G$ value associated with a mutation by averaging over the values produced by an ensemble of structures, the user may want to visualize the distribution of such values to investigate the source of potential outliers that may bias the average. In this case, a swarm plot displaying such values as separate data points constitutes a very insightful overview provided by *rosetta_ddg_plot*.

To guide the user on the number of cores and time required for calculation, depending on the RosettaDDGPrediction protocol, energy function, and protein size, we reported the results for different saturation scans in Table 1.

2.2 | Case Study 1: Prediction of changes in folding free energy upon mutations and comparison with experimental values from the ThermoMut database

To illustrate the performance of the *ref2015* energy function, we performed folding free energy calculations with both the *cartddg* (Figure 2) and the *cartddg2020* (Figure S1) protocols and compared them to experimentally determined unfolding $\Delta\Delta G$ values. The following section illustrates, as an example, our findings when using the *cartddg* protocol. We downloaded the ThermoMut database (ThermoMutDB) (Xavier et al., 2021) and selected four proteins as detailed in Section 4. In particular, we selected two bacterial enzymes with 117 and 597 mutations, respectively: Enterobacteria phage T4 Endolysin, ENLYS (UniProt ID: P00720), and *Staphylococcus aureus* Theronuclease, NUC (P00644). In addition, we performed the calculations on two human proteins of interest in health and disease, that is, TP53

(P04637) and FKBP1A (P62942), with 45 and 68 mutations with structural coverage, respectively. We applied the secondary structure definition of PDBe (Varadi, Anyango, Armstrong, et al., 2022) and annotated each position as either α -helix, β -sheet, or loop in the wild-type structures. This case study investigates the relationship between experimental and predicted values per-mutation when the data from all four proteins are pooled, allowing us to achieve better statistical power than considering each protein separately.

We performed a preliminary data exploration to understand the agreement between the experimentally determined and the predicted stability. Interestingly, data points from the experimental and prediction dataset were similarly distributed (Figure 2a), as corroborated by the Kolmogorov–Smirnov test ($p = 0.21$).

We then investigated the relationship between predicted and experimental data using a simple linear regression model (SLM), assuming that a perfect agreement between the experimental and predicted values would have an intercept of 0 and a coefficient of 1. The SLM regression line had an intercept of 0.81 and a slope of 0.719 (Figure 2b). The variance of the linear model (σ^2) is 3.95, and the model produced an R^2 of 0.44, a Pearson correlation coefficient (PCC) of 0.66, and a mean absolute error (MAE) between the predicted and experimental $\Delta\Delta G$ s (MAE) of 1.39. The residuals plot for this model showed how the poor R^2 value was at least partially due to systematic bias (Figure S2). This illustrates that a linear model does not entirely explain the variance in the data.

To better understand this behavior, we tried to fit the data using a generalized additive model (GAM) (Figure 2d). The resulting model had a roughly linear behavior in the ~ 0 –5 kcal/mol range but becomes less so

at lower or larger $\Delta\Delta G$ values. Similarly, the confidence interval is very narrow in the linear regime interval, and it is wider for large and small $\Delta\Delta G$ values, for which we have fewer data points. This observation is in alignment

with Høie et al. (Høie et al., 2022), who found that $\Delta\Delta G$ predictions made with *ref2015* and the *cartesian2020* protocol in 29 proteins correlated with altered protein functions for $\Delta\Delta G > 4.5$ kcal/mol, but the severity of the

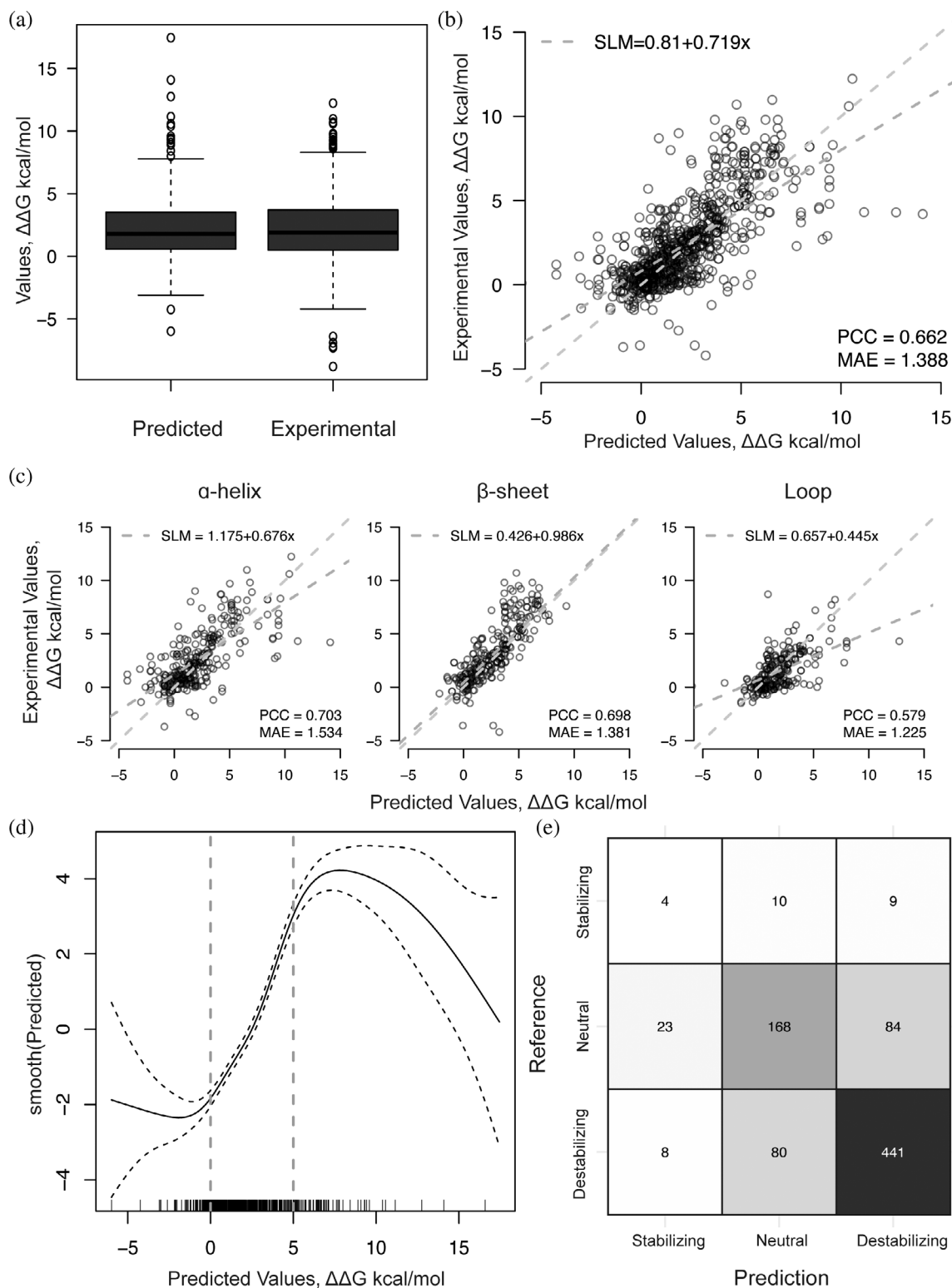


FIGURE 2 Legend on next page.

impact did not increase remarkably beyond this point. We then assessed the impact of the secondary structure on the performance of the prediction by building a SLM for each of the secondary structure groups, divided into α -helices, β -sheets, and loops (Figure 2c). Residues involved in structured regions are more likely to be part of the protein core, less flexible, and more sensitive to mutation with respect to solvent-accessible unstructured loops. According to the PCC, MAE slope, σ^2 , and R^2 values, the prediction was less consistent for the unstructured regions. Indeed, the loop subset featured a low PCC score (0.58), indicating a poor linear relationship. On the other side, the loop subset also resulted in the lowest MAE value (0.33) between predicted and experimental data, which could imply a better fit to the model. The low MAE is, however, an artifact of the comparatively low $\Delta\Delta G$ values observed in the loop regions. Loop regions are often flexible and less sensitive to changes in folding/unfolding $\Delta\Delta G$ upon mutation. This caused all the predicted or experimentally determined values to be grouped close together (Figure 2c), which caused lower MAE values than what observed for the α -helix or β -sheet subsets. In the loop dataset, we noticed several outliers in which amino acid substitutions are predicted to have a large destabilizing impact, whereas the experiments found the same variants to be neutral or mildly destabilizing. The experimental findings mostly align with the expectation that substitutions in flexible loops have mild effects on stability. However, some loop substitutions may extend or create secondary structure elements, for example, as a result of substitutions from proline (Pires et al., 2019). The difference witnessed in this dataset was likely due to Rosetta allowing local main chain flexibility, which might not be enough to represent the conformational heterogeneity that disordered regions experience in solution. We noticed similar behavior in applying

FoldX, which we could mitigate using ensembles of structures generated, for example, by molecular dynamics simulations (Fas et al., 2020; Nygaard et al., 2016; Tiberti et al., 2022). It should be noted that the α -helix mutations dataset also contains outliers. This dataset, however, had an overall better correlation with the experimental dataset, and the coefficient of its regression line is closer to 1. This suggests that changes in loops are more difficult to predict.

We then evaluated the performance of the predictions in classifying mutations into destabilizing, neutral, and stabilizing classes. We did so by classifying all mutations causing stability changes above 1 kcal/mol as destabilizing, all mutations causing stability changes between 1 and -1 kcal/mol as neutral, and all mutations causing stability changes below -1 kcal/mol as stabilizing (Frenz et al., 2020; Park et al., 2016) and constructing a confusion matrix (Figure 2e). This confusion matrix yielded an accuracy of 0.74. The prediction accuracy was best for the destabilizing class (0.76), with high sensitivity (0.83), while the accuracy of the stabilizing class was only 0.56, with a sensitivity of 0.17, indicating that the destabilizing class is more likely to be correctly identified as compared to the stabilizing class. The low accuracy could also be expected due to the imbalanced dataset available for the study, where a low number of stabilizing mutations is available. While this dataset is not balanced, this may not explain the bias in full, as the methodology itself could have been developed on a biased dataset (Bæk & Kepp, 2022; Pancotti et al., 2022). The neutral class performed similarly to the destabilizing class (Table S1).

We performed the same analyses on the dataset obtained using the *cartesian2020* protocol, which showed similar trends overall (Figure S1). Additionally, for comparison, we used a recent deep learning tool, which aims at simulating the *cartddg* protocol, that is, RaSP

FIGURE 2 Comparison of changes in structural stability predicted with the *ref2015 cartddg* protocol and experiments. (a) Distribution of the predicted and experimental stability changes in kilocalorie per mole. (b) Scatterplot of the $\Delta\Delta G$ values predicted by the *ref2015 cartddg* protocol and experimental values for the corresponding mutations. The blue line indicates a perfect correspondence between the variables. The green line is the fitted simple linear model. The model has an intercept of 0.81, a slope of 0.72, a variance (σ^2) of 3.95, and a R^2 of 0.44, a Pearson's correlation coefficient (PCC) of 0.66, and a mean absolute error between the predicted and experimental $\Delta\Delta G$ s (MAE) of 1.39. (c) Scatterplots dividing the data by the wild-type secondary structure of the mutated position. The blue line indicates a perfect correspondence between the variables for each plot. The green line is the fitted simple linear model. Here, it is evident how the structured sections have a better correlation when compared to the loops. This is likely due to the flexibility of the unstructured sections. α -helices: PCC = 0.70, slope = 0.68, σ^2 = 3.63, R^2 = 0.49, MAE = 1.53. β -sheets: PCC = 0.70, slope = 0.99, σ^2 = 4.98, R^2 = 0.49, MAE = 1.38. Loop: PCC = 0.58, slope = 0.45, σ^2 = 1.99, R^2 = 0.33, MAE = 1.22. (d) Generalized additive model (GAM) modeling the response variable, the experimental $\Delta\Delta G$ value, to a predictive variable, the predicted $\Delta\Delta G$ value, by estimating a smooth function, smooth (predict). The smooth function has an effective degree of freedom of 6.5, quantifying the complexity of the line. The dotted black lines indicate the confidence interval, which is sufficiently narrow in the $\Delta\Delta G$ interval 0–5 kcal/mol (indicated with red dotted lines) to indicate that a linear relationship is present in this interval. (e) Confusion matrix where the experimental values are annotated as the reference values. The threshold used to define the classes is a $\Delta\Delta G$ of < -1 kcal/mol for stabilizing mutations, $-1 < \Delta\Delta G < 1$ kcal/mol for neutral mutations and $\Delta\Delta G > 1$ kcal/mol for destabilizing mutations. The resulting accuracy is 0.74.

(Blaabjerg et al., 2022) (Figure S3). Here, we also noticed a remarkable performance both compared to both the experimental values and the Rosetta predictions.

In conclusion, the Case Study 1 showed a good linear correlation between predicted and experimental values, especially in the range of 0–5 kcal/mol, where the trend is generally conserved. Outside this range, the relationship between the predicted and the experimental values is less direct. We also show how predictions are more reliable for structured regions of the protein while correlation values are lower for unstructured regions.

2.3 | Case Study 2: Prediction of changes in binding free energy for protein-short linear motifs interactions

Within the protein–protein interaction landscape, intrinsically disordered proteins (IDPs) or regions (IDRs) have been proven to play an essential role in different biological events. IDPs and IDRs include functional motifs known as short linear motifs (SLiMs) that are important for the binding between IDPs and their target proteins (Davey, 2019; Davey et al., 2011; van Roey et al., 2014). An example is the LC3 interacting region (LIR), that is, a class of SLiMs involved in selective autophagy (Sora et al., 2020). One of the main features for regulating LIR binding to proteins of the LC3 family is through PTMs, especially through phosphorylation (Sora et al., 2020). Here, we aim to show an application of the *flexddg* protocol to capture the changes in binding free energy upon phosphorylation or mutations in the core region of LIR-containing proteins.

First, we selected two examples of experimentally characterized phospho-regulated LIRs for which the structures were available on the Protein Data Bank, that is, FUNDC1 in complex with LC3B (PDB entry 2N9X; Kuang et al., 2016) and PIK3C3 in complex with GABARAP (PDB entry 6HOG; Birgisdottir et al., 2019). Experimental data from isothermal titration calorimetry (ITC) or peptide arrays are available for these two complexes and include phosphorylations or phospho-mimetics (Birgisdottir et al., 2019; Kuang et al., 2016; Lv et al., 2017). We applied the *flexddg* protocol with the *talaris2014* Rosetta energy function to investigate the effects of single and multiple phospho-mimetic mutations at the known phosphosites (see Section 4). Indeed, Rosetta does not provide parameters for phosphorylated residues. In addition, we included a comparison with the estimates provided by FoldX using the binding free energy protocol implemented in MutateX. The results are described in detail below and reported in Figure 3.

FUNDC1 is a mitophagy receptor that mediates the selective removal of damaged mitochondria. It contains a canonical LIR (core region, 18-YEVL-21), which is necessary for interacting with LC3 and its role in mitophagy (Kuang et al., 2016). FUNDC1 presents three experimentally validated phosphosites in the surroundings of its LIR motif: S13, S17, and Y18 (Figure 3a). ITC experiments with different FUNDC1 LIR peptides and LC3B reported a K_d of $0.40 \pm 0.06 \mu\text{M}$ for the wild-type variant. Phosphorylation at the S13 site resulted only in a slight decrease of the LC3B affinity ($K_d = 0.60 \pm 0.05 \mu\text{M}$) with respect to the wild type. On the other hand, Y18 phosphorylation caused a five-fold K_d increase ($K_d = 1.72 \pm 0.30 \mu\text{M}$). This increase is slightly augmented if both phosphorylations are combined ($K_d = 2.00 \pm 0.37 \mu\text{M}$) (Kuang et al., 2016). Additionally, another work reported that S17 phosphorylation has an opposite effect and increases the binding affinity for LC3B by three folds (Lv et al., 2017). The *flexddg* protocol predicted the S13D and S13E substitutions to have neutral effects on the binding, in agreement with experiments (i.e., average $\Delta\Delta G < 0.25$ kcal/mol). However, the average $\Delta\Delta G$ s for the S17E and S17D mutations are also low, suggesting that, in this other case, the prediction cannot capture the changes in the binding affinity observed experimentally (Figure 3a). In the case of the single phospho-mimetic mutations at Y18 and S13, the predicted $\Delta\Delta G$ sign was in overall agreement with the effect measured experimentally. Although, we noticed that, in this case, to use tryptophan as a phospho-mimetic residue for phosphorylated tyrosine does not efficiently capture the destabilizing effects of the PTM.

Surprisingly, the combination of phospho-mimetic mutations at S13 and Y18 sites (i.e., S13E_Y18E and S13E_Y18W) resulted in negative $\Delta\Delta G$ values, suggesting a stabilizing effect in disagreement with what observed experimentally (Figure 3b). Nevertheless, we observed that the associated standard deviations are very high not allowing for quantitative conclusions.

We then studied PIK3C3, a class III phosphoinositide 3-kinase enzyme of the PtdIns3K complexes, involved in autophagy initiation. PIK3C3 presents a canonical F-type LIR (250-FELV-253) required for the interaction with GABARAP and GABARAPL1 (Birgisdottir, et al., 2019). The effect of a double phosphorylation at S244 and S249 was studied with ITC. In these experiments, the substitution of both the phosphosites with glutamate caused a 17-fold increase in GABARAP binding ($K_d = 2.9 \pm 0.1 \mu\text{M}$) compared to the wild-type variant ($K_d = 49.5 \pm 3.9 \mu\text{M}$). Moreover, peptide array experiments showed an increase in the binding affinity of the LIR peptide with all the LC3 family members for the S249E variant (Birgisdottir et al., 2019).

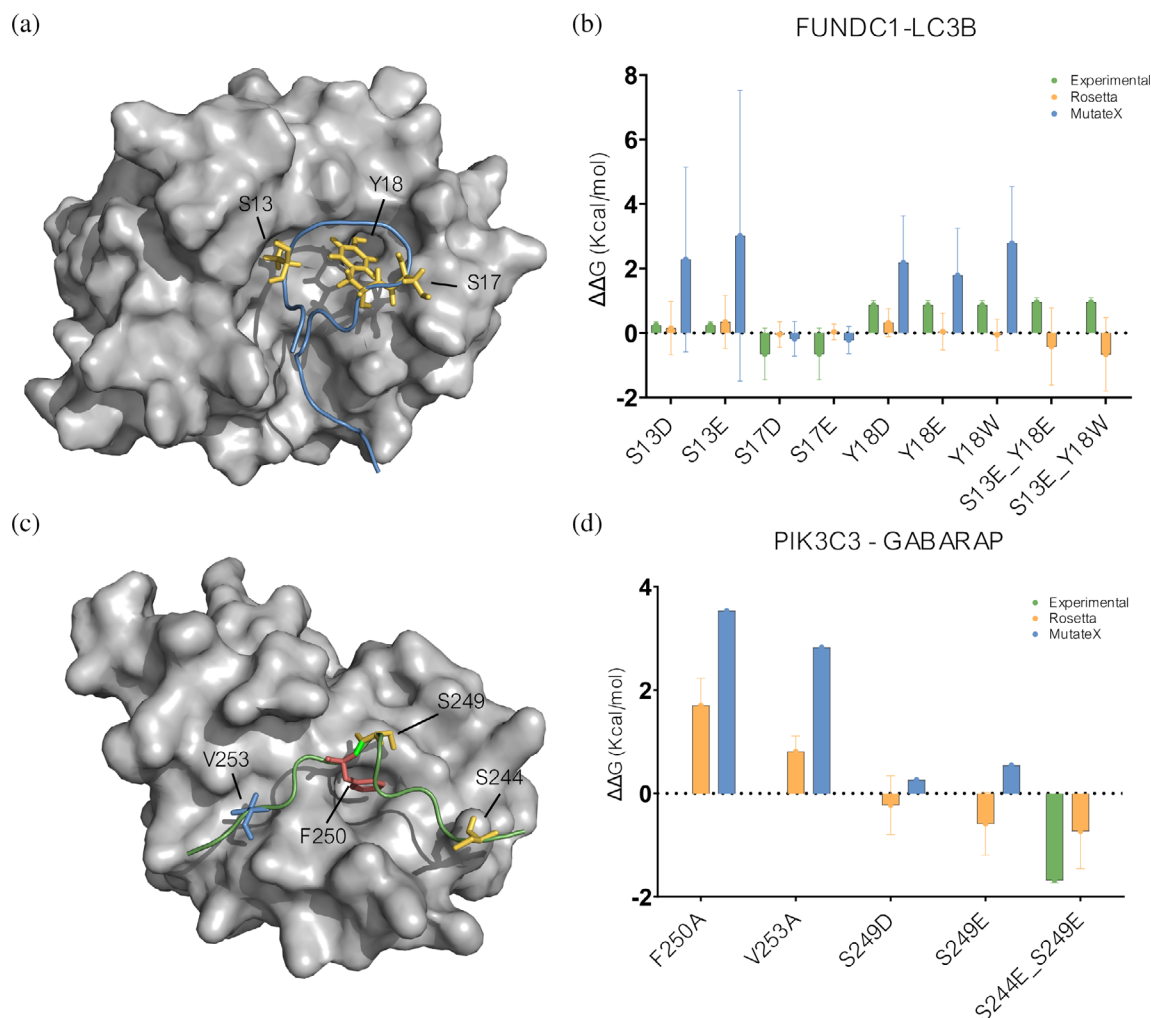


FIGURE 3 Prediction of changes in binding free energy using the *flexddg* protocol for protein interactions mediated by short linear motifs. (a) FUNDC1 LIR peptide (blue) in complex with LC3B (gray) in the structure associated with the PDB entry 2N9X. The S13, Y18, and S17 phosphosites are shown as sticks and colored in yellow. (b) We report the predicted binding $\Delta\Delta G$ s for the single and double phosphomimetic mutations for the FUNDC1 LIR phosphosites for which experimental data are available for comparison, along with the same single mutations predicted with MutateX. For every variant, we also included the $\Delta\Delta G$ values obtained from experimental K_d of the phosphorylated variants (pS13, pS17, pY18, and pS13_pY18). (c) PIK3C3 LIR peptide (blue) in complex with GABARAP (gray) in the structure associated with the PDB entry 6HOG. The S244 and S249 phosphosites are shown as sticks and colored in yellow, while the residues for binding to the GABARAP HP1 and HP2 pockets are shown as red and blue sticks, respectively. (d) We report the predicted binding $\Delta\Delta G$ s for single and double phosphomimetic mutations, along with mutations to alanine in the core motif of the PIK3C3 LIR, for which experimental data are available for comparison, along with the same single mutations predicted with MutateX. We also included the $\Delta\Delta G$ value obtained from the experimental K_d of a S244E_S249E variant.

To assess the potential of the *flexddg* protocol in capturing the effects induced by the phosphorylations of the PIK3C3 LIR, we modeled the S249E variant and a variant including phosphomimetic mutations at both the S244 and S249 sites (i.e., S244E_S249E; Figure 3c). We also tested the effect of the S249D substitution as a possible phosphomimetic, even if no experimental data are available for this mutation. We noticed that using S249D as a phosphomimetic provides different result than introducing a glutamate (Figure 3d). This supports the notion that aspartate and glutamate cannot always be used as

phosphomimetics in an interchangeable manner. The S249E variant had a slightly stabilizing effect on the binding (average $\Delta\Delta G = -0.59$ kcal/mol) and, in general, values of $\Delta\Delta G$ lower than 0 across the 35 independent runs (Figure 3d). This is in partial agreement with the peptide array results mentioned above. The results for the double mutant variant S244E_S249E are in agreement with the expected increase in binding affinity observed experimentally, even if with large deviations across measurements, suggesting that the *flexddg* protocol could provide, in some cases, a qualitative

understanding of the effects of multiple amino acid substitution if they are located at structural proximity. We noticed that FoldX failed in identifying the stabilizing effect of the mutation on the binding and predicted slightly destabilizing effects despite the possibility to model the phosphorylated variant of the residue.

Furthermore, we evaluated whether the *flexddg* protocol could provide insights into the effects of mutations in SLiMs where PTMs are not involved. In the case of LIRs, the interaction between an LIR-containing protein and an LC3 family member is mainly driven by two residues of the LIR motif, which bind to the hydrophobic pocket 1 (HP1) and the hydrophobic pocket 2 (HP2) residues of the LC3 protein, respectively (Sora et al., 2020). Thus, we tested the capability of the *flexddg* protocol with the *talaris2014* energy function to predict the impact of the known detrimental mutations F250A (residue for interaction with HP1 pocket) and V253A (HP2 pocket) of PIK3C in complex with GABARAP (PDB entry, 6HOG; Birgisdottir et al., 2019). We observed a good agreement between Rosetta- and FoldX-based calculations in identifying these mutations as detrimental for binding to GABARAP (Figure 3d) (Sora et al., 2020).

Overall, Case Study 2 illustrates the potential and limitations of the RosettaDDGPrediction workflow. We identified as main challenges the prediction of increased binding affinity (i.e., stabilizing mutations), to study combined mutations and the usage of phosphomimetic mutations instead of phosphorylated residues. Moreover, we noticed that the results from RosettaDDGPrediction are more consistent with the $\Delta\Delta G$ s from the ITC experiments with respect to the ones obtained from the MutateX protocol with the FoldX free energy function despite FoldX allows to include phosphorylated residues. The estimates provided by FoldX seem to capture variants destabilizing for the binding to the target protein but with highest $\Delta\Delta G$ than experimentally measured.

2.4 | Case Study 3: Influence of the source of initial structures for the calculations

Using structural models to perform prediction of $\Delta\Delta G$ s is a tantalizing perspective because of intrinsic limitations in the availability of experimental structures. This has been shown to be reliable to a good extent—for instance, using homology models with Rosetta allowed to achieve similar performance when comparing predictions with experimental $\Delta\Delta G$ s, as long as the sequence identity of the template to the target protein was at least of 40% (Valanciute et al., n.d.) and results obtained using Rosetta are relatively robust to the use of models (Blaabjerg

et al., 2022; Valanciute et al., n.d.). The advent of AlphaFold has revolutionized molecular modeling and structural biology (Jumper et al., 2021), resulting in models of 3D structures of proteins with quality comparable to that achievable with experimental approaches and useful in the context of computational biology, including the prediction of changes of free energy (Akdel et al., 2022). The current version (release 4) of the AlphaFold Protein Structure Database contains over 214 million predicted protein structures, corresponding to most proteins in Uniprot 2021_4 and including 48 complete proteomes (Varadi, Anyango, Deshpande, et al., 2022), providing a rich source of structures for in silico mutational scans.

Here, we evaluated the influence of using a model based on AlphaFold2 with respect to an x-ray structure of the same protein with good resolution. For this goal, we used as a case study the DNA binding domain (DBD) of p53, for which experimental data are also available on 31 mutant variants from ThermoMutDB (Xavier et al., 2021). We evaluated the agreement between our calculated and experimentally available data using the same parameters and energy functions, either the *cartddg* or *cartddg2020* protocol, and the two different starting structures. We also included a variant of *cartddg* in which we increased the number of runs per mutation up to 10 to determine whether it would improve our results. As the final $\Delta\Delta G$ depends on the values obtained by the single runs, we expect that increasing the number of samples might lead to better converged final $\Delta\Delta G$ values. We measured the agreement through several metrics, such as the Pearson correlation coefficient, MAE, and a ROC curve. We performed most of our comparison considering runs performed with the *cartddg* protocol. Therefore, in this section, we will refer to the *cartddg* protocol unless stated otherwise.

We obtained a similar pattern when comparing predictions and experiments using the experimental structure and the AlphaFold2 model (Figure 4a) with a positive linear correlation, as quantified by the Pearson correlation coefficient (Figure 4b). The highest Pearson correlation coefficient obtained was 0.79 using the scoring function *talaris2014* with the AlphaFold2 model and 10 runs (Figure 4b). However, all runs, including the ones using the *cartddg2020* protocol, achieved a correlation in the range of 0.57–0.79. Values ranging from 0.74 to 0.79 were obtained by all runs using the x-ray structure and by *talaris2014* with AlphaFold2 using 3 or 10 runs. Using *ref2015* with the AlphaFold2 model led to a slightly worse correlation of 0.57 for 3 runs and 0.68 for 10 runs. The runs with *ref2015* energy function and the *cartddg* protocol (x-ray structure) using 10 runs had the smallest MAE of 0.90 kcal/mol (Figure 4c). This result suggests that this combination featured the lowest average

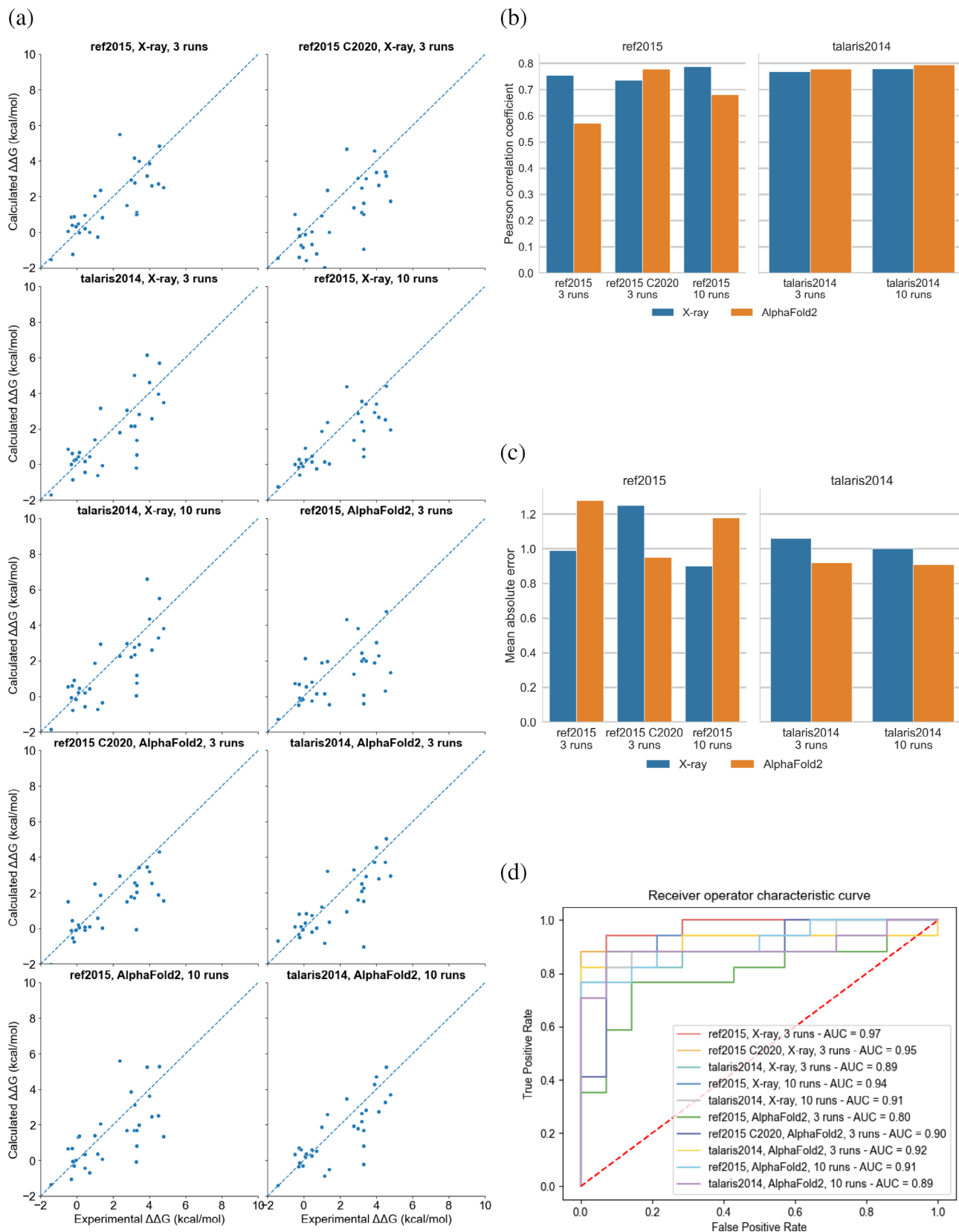


FIGURE 4 Legend on next page.

distance between predicted and target values among all the different tested methods. It was closely followed by *talaris2014* with the AlphaFold2 structure using 3 and 10 cycles with 0.92 and 0.91, respectively. The rest of the combinations had a MAE between 0.95 and 1.28.

Considering the ROC curve, we used the experimental free energy changes from ThermoMutDB as ground truth. We partitioned our dataset into destabilizing and nondestabilizing mutations depending on whether our prediction or ground truth had $\Delta\Delta G > = 1.2$ kcal/mol (Degn et al., 2022). The best area under the curve (AUC) was achieved by using the scoring function *ref2015* using the *cartdgg* protocol and the x-ray structure, yielding a value of 0.97 (Figure 4d). In general, the different scoring functions and structures behaved similarly.

The usage of the experimental x-ray structure or the AlphaFold2 model did not affect the prediction performance. The only exception was the combination of the *ref2015* energy function and the *cartddg* protocol with the AlphaFold2 model, which had a lower correlation and ROC AUC with respect to the other cases. Increasing the number of runs also slightly improved the performance, but with the trade-off of a considerably increased computing time. Finally, we obtained mixed results when comparing the *ref2015* x-ray three-cycles with *cartddg* with the corresponding *cartddg2020* run. We did not see any appreciable improvement when using *cartddg2020* on the x-ray structure, as the *cartddg2020* run has a slightly lower correlation (0.74 vs. 0.76), higher MAE (1.25 vs. 0.99 kcal/mol), and lower AUC (0.95 vs. 0.97) considering the experimental data. Nonetheless, using *cartddg2020* with the AlphaFold2 model rescued the subpar performance of *ref2015*, as all its performance measures are more similar to those of the other cases.

It should be noted that the DNA-binding domain in the p53 AlphaFold2 model, ranging from residues 91 to 289, features a good per-residue confidence score (pLDDT) score, mostly above 70. This implies that more tests on models or regions with lower quality should be carried out in the future to determine whether our findings can be generalized. It has been shown that protein regions predicted with a confidence score lower than 50 are less in agreement with experimental data, possibly

due to the low-confidence regions including more often disordered regions with a higher tolerance to mutations affecting stability (Akdal et al., 2022).

2.5 | Case Study 4: Variants predisposing to childhood cancer

In a recent study, 198 samples from different childhood cancer types were analyzed in terms of germline variation and cancer predisposition (Byrjalsen et al., 2020). Among these, different variants of uncertain significance (VUS) have been found with a frequency of <1% in the healthy population. Approximately, 20% of the patients investigated had VUS in DNA repair pathway genes. In addition, we carried out new analyses on a larger dataset accounting for more than 550 germline samples from Danish children. The selection criteria for the proteins and the variants included in the study are described in detail in Section 4 and in Figures S4 and S5. We retained 14 proteins, that is, ERCC4, BLM, FANCA, FANCE, FANCF, FANCG, FANCI, FANCL, MLH1, MSH2, MSH6, NBN, RAD51C, and RFD3 for structure-based calculations of the changes in folding $\Delta\Delta G$ s for the VUS. All these genes are classified as tumor suppressor genes in the COSMIC Cancer Gene Census v96 (Sondka et al., 2018) or from the literature, in the case of FANCI (Zhang et al., 2016) and RAD51C (Somyajit et al., 2010).

Since mutations in tumor suppressor genes are generally causing loss-of-function in cancer (Wang et al., 2018), we were interested in identifying VUS that could destabilize the protein structure and result in positive predicted $\Delta\Delta G$ values upon mutation. These variants could be relevant to investigate further in terms of genomic alterations predisposing to cancer. To this aim, we retained the variants with structural coverage in AlphaFold2 and high confidence scores for a total of 126 variants analyzed (Figures 5–7 and Table S2). According to searches in ClinVar (Landrum et al., 2014, 2020), some of the variants were annotated as benign or likely benign but not related to childhood cancer. On the other hand, only T1131A in FANCA was found as pathogenic. The remaining were not deposited in ClinVar or annotated as

FIGURE 4 Comparison of experimental and predicted $\Delta\Delta G$ s using p53 as a case study. $\Delta\Delta G$ values were predicted using Rosetta version 3.12 with the *ref2015* and *talaris2014* scoring functions, and the *cartddg* and *cartddg2020* protocols (referred to as “C2020” in the figure). We used the x-ray structure (PDB entry 2XWR) and a model from the AlphaFold2 database for the residues 91–289 of p53 as initial structures, using our default number of runs (3) or 10 runs. (a) Experimental versus predicted $\Delta\Delta G$ values. (b) Pearson's correlation coefficient between experimental and predicted values. (c) Mean absolute error (MAE) between experimental and predicted values. (d) Receiver operator characteristic (ROC) curve. The classification for this curve was done by considering the changes of free energy values reported in ThermoMutDB as ground truth, using 1.2 kcal/mol as $\Delta\Delta G$ cut-off to distinguish between destabilizing and nondestabilizing mutations (see Section 4). The same criterion was used for the predicted mutations.

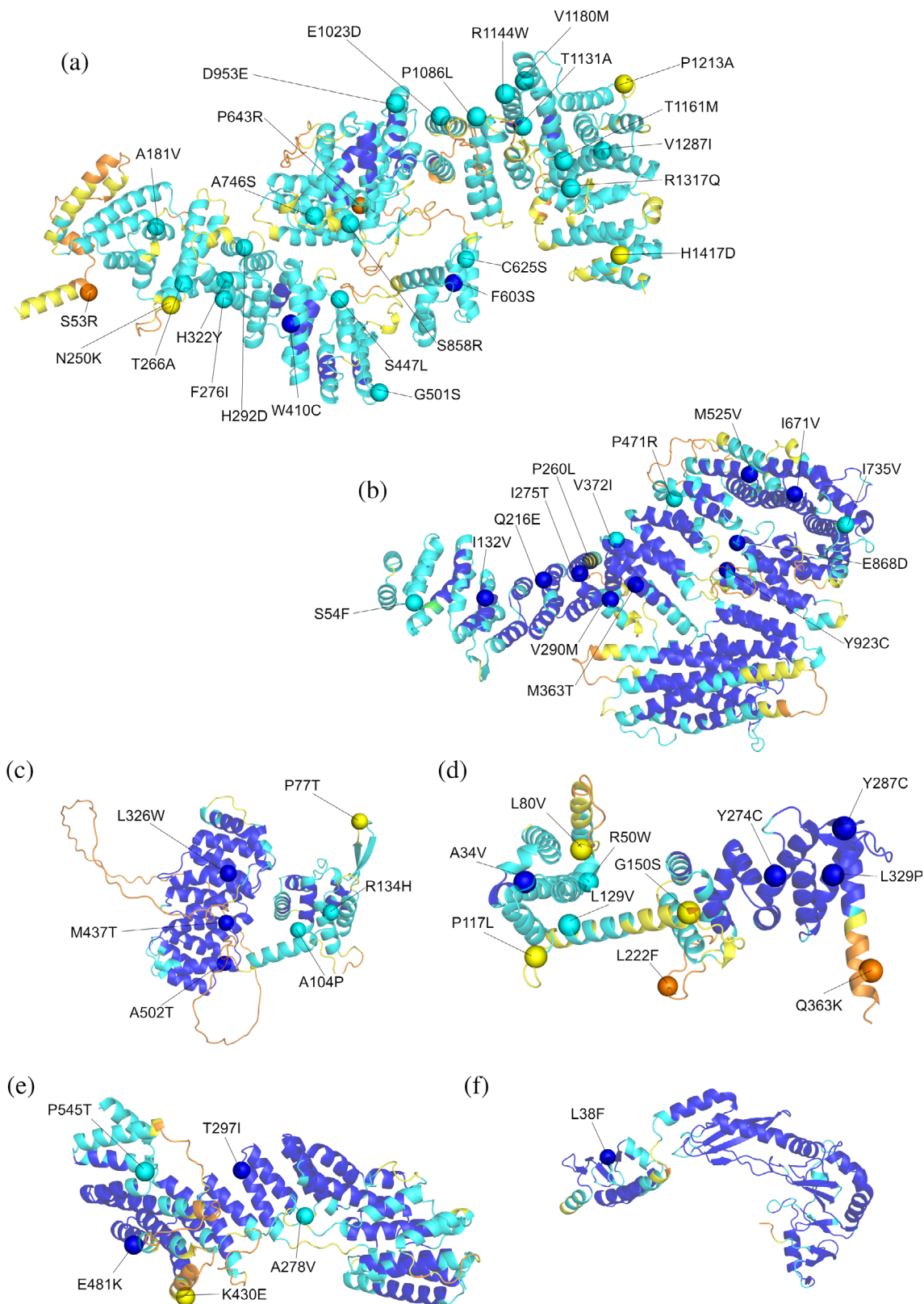


FIGURE 5 Trimmed AlphaFold structures of the FA (Fanconi Anemia) proteins selected for the Case Study 4. Cartoon representation of (a) FANCA₃₇₋₁₄₄₁, (b) FANCI₁₋₁₂₇₉, (c) FANCE₁₂₋₅₃₄, (d) FANCF₂₋₃₆₉, (e) FANCG₁₂₋₆₁₆, and (f) FANCL₁₋₃₇₅. The proteins are colored according to the AlphaFold2 pLDDT score: very low (orange, pLDDT < 50), low (yellow, 50 < pLDDT < 70), confident (light blue, 70 < pLDDT < 90), and very high (blue, pLDDT > 90). The Ca of the residues found mutated in pediatric cancer patients are shown as spheres and labeled.

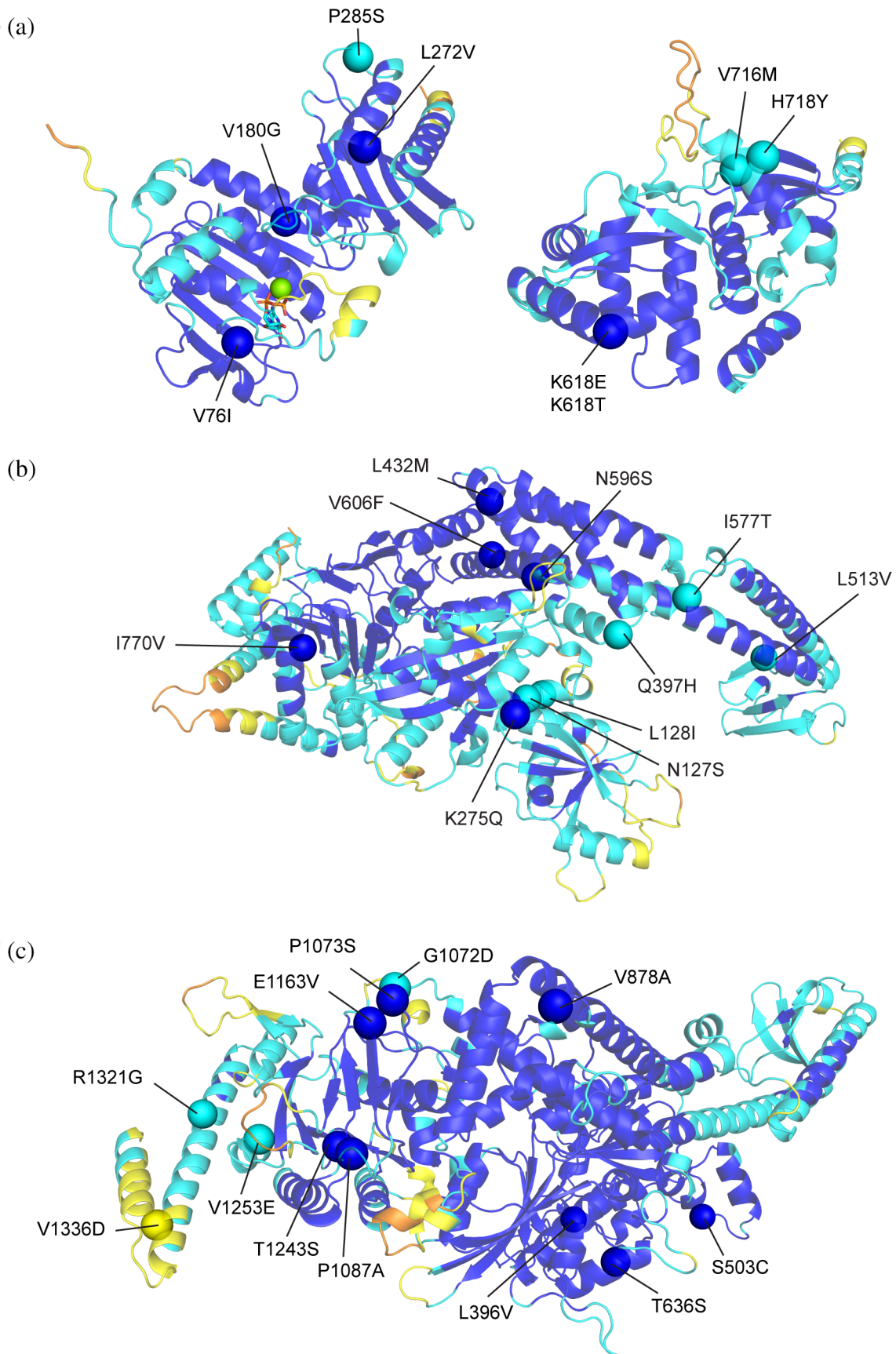


FIGURE 6 Legend on next page.

uncertain significance or with conflicting evidence, emphasizing the importance of additional analyses to understand the effects at the protein level.

In this example, we applied the *cartddg2020* protocol, which considers the $\Delta\Delta G$ value referring to the mutant structure with the lowest total energy. At first, we retained, as predicted destabilizing, the variants with $\Delta\Delta G$ values >1 kcal/mol (see Section 4) and confirmed destabilizing by calculations with MutateX (Table 2). Indeed, the *foldX5* energy function, which is applied in the MutateX protocol, is effective in capturing loss-of-function mutations (Gerasimavicius et al., 2022; Nielsen et al., 2017; Scheller et al., 2019). Of note, the pathogenic variant T1131A is not predicted to destabilize the structure of FANCA by both Rosetta and FoldX calculations. We hypothesize that the detrimental effects triggered by this variant could be due to other properties such as impaired activity, interactions, or PTMs at the cellular level. Experimental studies at the cellular level confirm that the T1131A substitution does not affect the protein levels, in agreement with a neutral effect on the folding $\Delta\Delta G$ s (Wilkes et al., 2017) and that the phenotype reflects a functional impairment that has a mild impact on drug sensitivity and the monoubiquitination of another protein (Adachi et al., 2002; Wilkes et al., 2017). T1131A could be further investigated using our recently proposed multilayered structural framework for variant annotations in proteins, that is, Multilayered Assessment of Variants by Structure for proteins (MAVISp) (Arnaudi et al., 2022).

We also observed that one variant annotated as benign in ClinVar (i.e., L605F FANCI) has predicted changes in folding $\Delta\Delta G$ higher than 3 kcal/mol and is, therefore, classified as destabilizing for the structural stability by our analysis. The variant has been characterized with cellular assays, showing decreased protein levels when compared to the wild type, which confirms our prediction (Fierheller et al., 2021). On the other hand, the variant P55L (predicted folding $\Delta\Delta G < 2.0$ kcal/mol) was expressed at the same level as the wild-type variant. In addition, other benign variants in ClinVar resulted in changes in free energy in the range of 1–2 kcal/mol (Table 2). This observation suggests that variants for which the predicted changes in stability are within 1–3 kcal/mol should be further investigated to evaluate whether they could result in neutral effects in terms of protein levels in the cell or propensity for degradation. In

the case of MSH2 and MLH1, for example, it has been shown that a predicted destabilization of more than 3 kcal/mol is sufficient to cause cellular degradation of the proteins (Abildgaard et al., 2019; Nielsen et al., 2017). Similar observations have been recently done in another recent work on different proteins with benign variants featuring predicted changes in stability in the range of 0.9–2.7 kcal/mol (Blaabjerg et al., 2022).

According to the results in Table 2 and the observation above, if we consider folding $\Delta\Delta G$ values higher than 3 kcal/mol, our analyses suggest a number of VUS that could predispose to loss-of-function through destabilization of the protein structure and have a high REVEL score which further support their possible pathogenic impact (i.e., A797T in BLM, I706T in ERCC4, W410C and F603S in FANCA, L329P in FANCF, V180G in MLH1, V606F in MSH2, and G1072D in MSH6). Of note, L329P in FANCE has been suggested to disrupt the stability of the catalytic module of the protein in a previous structural study (Shakeel et al., 2019).

3 | DISCUSSION

We developed RosettaDDGPrediction moved by the need to provide easy and scalable access to Rosetta-based approaches to predict free energy changes in proteins upon mutations. The possibility to perform mutational scans in an efficient and scalable manner allows to have a new systematic and large-scale approach at such data. The fact that other implementations of this process have been released in recent years (e.g., https://github.com/KULL-Centre/PRISM/tree/main/software/rosetta_ddG_pipeline) is a testament to its utility.

RosettaDDGPrediction takes care of the whole process by performing a large number of $\Delta\Delta G$ predictions in an efficient and scalable manner, making a high-throughput calculations with Rosetta accessible, which is helpful for both extensive mutational scans and structured benchmarks.

RosettaDDGPrediction is, to our knowledge, the first wrapper devised to integrate state-of-the-art Rosetta-based protocols for the predictions of free energy changes upon mutation on binding and stability under a uniform framework.

Furthermore, the software checks the success of the runs, aggregates the data in CSV tables that are easy to

FIGURE 6 Trimmed AlphaFold structures of the of the DNA mismatch repair proteins selected for the Case Study 4. Cartoon representation of (a) MLH1_{1–341} and MLH1_{501–756}, (b) MSH2_{1–934} and (c) MSH6_{362–1360}. The proteins are colored according to the AlphaFold2 pLDDT score: very low (orange, pLDDT < 50), low (yellow, 50 < pLDDT < 70), confident (light blue, 70 < pLDDT < 90), and very high (blue, pLDDT > 90). The C α of the residues found mutated in pediatric cancer patients are shown as spheres and labeled.

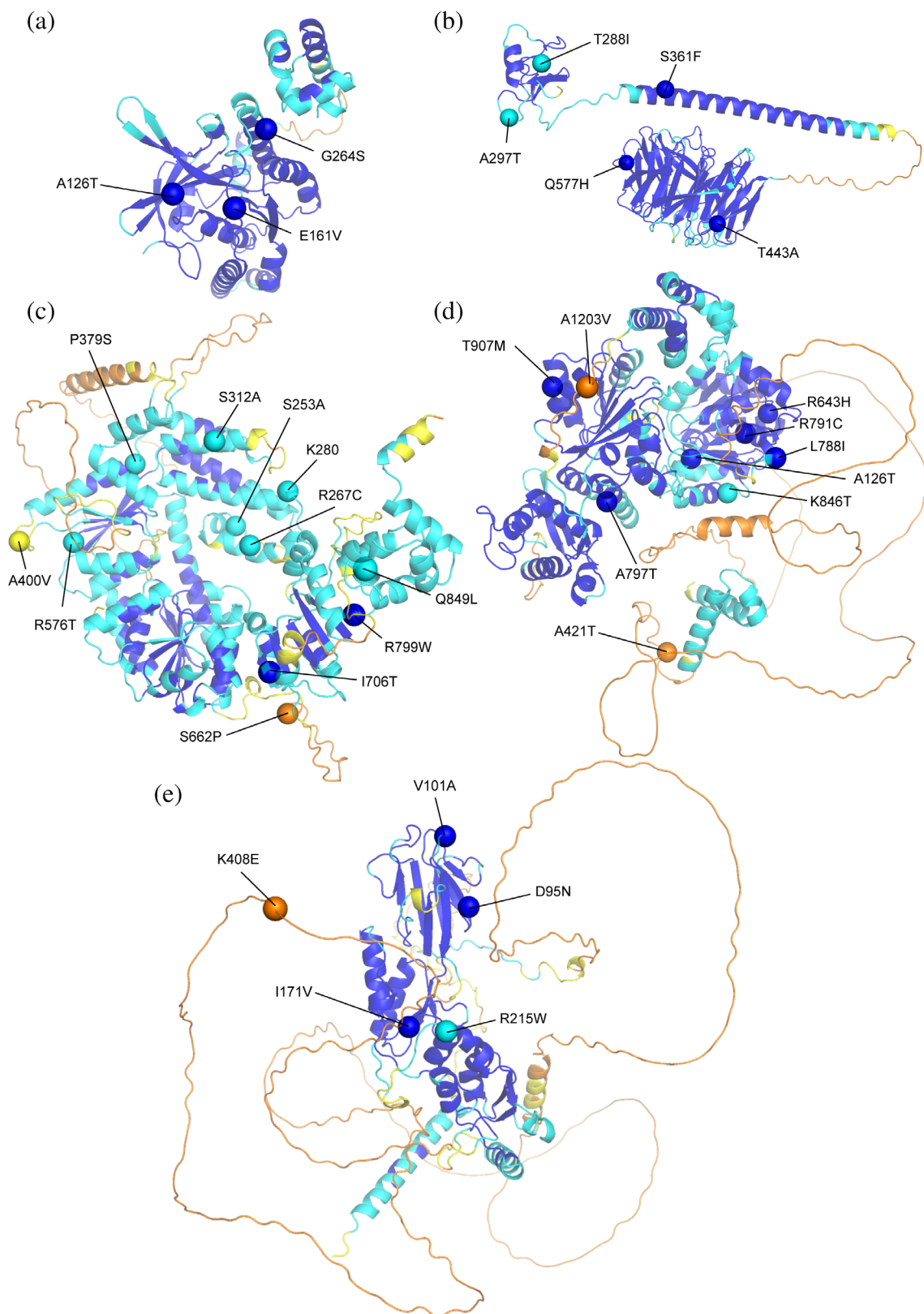


FIGURE 7 Trimmed AlphaFold structures of the proteins promoting the double-strand break (DBS) repair (RAD51C, RFWD3, ERCC4, and NBN) and RECQ helicase (BLM) selected for proteins used for the Case Study 4. Cartoon representation of (a) RAD51C₁₃₋₃₅₀, (b) RFWD3₂₈₄₋₇₇₄, (c) ERCC4₁₂₋₉₁₄, (d) BLM₃₆₈₋₁₂₉₀, and (e) NBN₁₋₇₄₉. The proteins are colored according to the AlphaFold2 pLDDT score: very low (orange, pLDDT < 50), low (yellow, 50 < pLDDT < 70), confident (light blue, 70 < pLDDT < 90), and very high (blue, pLDDT > 90). The Ca of the residues found mutated in pediatric cancer patients are shown as spheres and labeled.

TABLE 2 Summary of predicted $\Delta\Delta G$ s for predicted destabilizing variants in childhood cancer

Variant	ClinVar	Predicted folding $\Delta\Delta G$ (kcal/mol)	REVEL
BLM- L788I	Conflicting interpretations of pathogenicity	2.233	0.503
BLM -A797T	Entry N.A.	3.045	0.893
BLM - K846T	Uncertain significance	1.227	0.136
BLM - Y1024C	Uncertain significance	1.776	0.590
ERCC4 - R267C	Uncertain significance	1.495	0.470
ERCC4 - P379S	Conflicting interpretations of pathogenicity	2.051	0.526
ERCC4 - R576T	Uncertain significance	1.012	0.274
ERCC4 - I706T	Conflicting interpretations of pathogenicity	3.242	0.609
FANCA - F276I	Entry N.A.	1.922	0.160
FANCA - W410C	Entry N.A.	4.169	0.622
FANCA - F603S	Uncertain significance	6.111	0.631
FANCA - A746S	Benign/likely benign	1.696	0.374
FANCA - P1086L	Entry N.A.	1.012	0.775
FANCE - A104P	Entry N.A.	5.691	0.319
FANCE -L326W	Uncertain significance	1.620	0.154
FANCE - M437T	Conflicting interpretations of pathogenicity	1.930	0.134
FANCF - L329P	Uncertain significance	8.170	0.417
FANCF - Y287C	Uncertain significance	2.732	0.195
FANCF - Y274C	Uncertain significance	3.526	0.193
FANCF- L129V	Uncertain significance	1.086	0.069
FANCF - L80V	Entry N.A.	2.001	0.104
FANCG - P545T	Entry N.A.	1.918	0.465
FANCI - I275T	Uncertain significance	3.041	0.22
FANCI - M363T	Entry N.A.	2.593	0.242
FANCI - P471R	Uncertain significance	1.730	0.837
FANCI - M525V	Conflicting interpretations of pathogenicity	2.137	0.487
FANCI - L605F	Benign/Likely benign	3.773	0.238
FANCI - C742S	Benign	1.084	0.075
FANCI - Y923C	Uncertain significance	3.806	0.391
MLH1 - P285S	Uncertain significance	1.373	0.838
MLH1 - K618E	Benign/Likely benign	1.296	0.874
MLH1 - V180G	Uncertain significance	3.847	0.91
MSH2 - N127S	Benign	2.172	0.741
MSH2 - L128V	Conflicting interpretations of pathogenicity	2.309	0.613
MSH2 - L513V	Uncertain significance	2.595	0.829
MSH2 - I577T	Likely benign	2.067	0.928
MSH2 - V606F	Entry N.A.	5.096	0.889
MSH2 - I770V	Conflicting interpretations of pathogenicity	1.043	0.417
MSH6 - L396V	Benign	1.369	0.322
MSH6 - S503C	Entry N.A.	1.288	0.413
MSH6 - V878A	Benign	2.073	0.155
MSH6 - G1072D	Uncertain significance	6.349	0.623
MSH6 - V1253E	Uncertain significance	2.986	0.952

(Continues)

TABLE 2 (Continued)

Variant	ClinVar	Predicted folding $\Delta\Delta G$ (kcal/mol)	REVEL
NBN - D95N	Conflicting interpretations of pathogenicity	1.448	0.583
NBN - I171V	Conflicting interpretations of pathogenicity	1.133	0.398
RFWD3 - Q577H	Entry N.A.	1.469	0.162

Note: We did not report RAD51C and FANCL in the table since all the variants analyzed here for these proteins were predicted with neutral effects for stability. We reported the full list of VUS in Table S2. Here, we included those variants that are predicted destabilizing by both Rosetta- and FoldX-based estimates, using a threshold of folding $\Delta\Delta G$ of 1 kcal/mol. This is a threshold often used to discuss the effects of mutations on structural stability applying Rosetta- or FoldX-based methods. We observed that there are variants that resulted in changes of $\Delta\Delta G$ above the threshold but with a benign Clinvar classification. This suggests that a $\Delta\Delta G$ threshold of approximately 2–3 kcal/mol could be more suited to pinpoint pathogenic variants. NA indicates “not available.”

mine, and generates visual reports. As these steps are independent, the aggregation and visualization tools can be used on different datasets. In addition, we support additional output formats compatible with the MutateX plotting scheme (Tiberti et al., 2022). At the same time, raw or aggregated data can be easily manipulated externally. RosettaDDGPrediction also devotes particular attention to ensuring technical reproducibility by being controlled through configuration files. Further developments of RosettaDDGPrediction will focus on integrating its functionalities within MutateX, to provide a method-agnostic container to perform and collect high-throughput mutational scans in a reproducible, automatized, and sustainable manner.

In this context, the performances of RosettaDDGPrediction and MutateX are only as good as those of the Rosetta- and FoldX-based methods that they incorporate. Indeed, Rosetta-based protocols implemented so far rely on different sampling methods to obtain models of the mutant variant structures and on scoring the resulting structures via knowledge-based energy functions to predict changes in the folding and binding free energy upon mutation (O'Meara et al., 2015; Park et al., 2016). However, more rigorous strategies are available to predict both the effect of mutations on the folding free energy and the binding free energy (Benedix et al., 2009; Kumari et al., 2014b; Seeliger & de Groot, 2010; Siebenmorgen & Zacharias, 2020). For example, approaches leveraging enhanced sampling along reaction coordinates designed to study binding and unbinding events are available (Bertazzo et al., 2021; Raniolo & Limongelli, 2020; Wingbermühle & Schäfer, 2020).

The time and computational resources needed by these methods still prevent their usage for investigations going beyond a few mutations. In these contexts, which include, for instance, saturation mutagenesis scans, Rosetta- and FoldX-based protocols represent a good trade-off between accuracy and speed.

Nevertheless, Rosetta still presents a challenge when noncanonical residue types are considered. Indeed, while

most noncanonical amino acids are supported, mutations to phosphorylated residues cannot be performed in either protocol to predict free energy changes. For this reason, including strategies circumventing this issue would greatly expand the application of RosettaDDGPrediction.

Furthermore, a milestone in structural bioinformatics has been reached lately, with the release of AlphaFold2 and its outstanding performance in the CASP14 challenge (Jumper et al., 2021). Originally developed to solve the long-standing protein folding problem, AlphaFold2 has already seen many spin-off studies to assess its potential (Evans et al., 2022; Porta-Pardo et al., 2022; Robertson et al., 2021; Ruff & Pappu, 2021; Tsaban et al., 2022). So far, evidence suggests that AlphaFold2 cannot effectively predict changes in folding free energy upon mutation (Buel & Walters, 2022; McBride et al., 2022; Pak et al., 2021). However, more studies are needed to explore this possibility fully.

Our wrappers have been devised to be inherently extensible. As stated earlier, a long-term perspective may include transforming them into a more general platform for structure-based methods to predict free energy changes upon mutation based on freely accessible, open-source software. This will also allow us to support other energy functions or schemes for free energy calculations, as well as to include support to transmembrane proteins including protocols as the one developed by Tiemann et al. (Tiemann et al., n.d.). Moreover, the results obtained here and in the original publication (Blaabjerg et al., 2022) with the deep-learning method RaSP pinpoints this approach as an additional candidate to include in a unified framework together with the support to FoldX- and Rosetta-based calculations.

The efforts of centralizing the development of software for in silico deep mutational scans using free energy functions will help to move a step forward toward a unified framework for high-throughput structure-based calculations of free energy changes upon mutation.

4 | METHODS

The data and documentation on the case studies are reported in the OSF repository, <https://osf.io/84kwe/> (10.17605/OSF.IO/84KWE).

4.1 | Case Study 1

The ThermoMut database (Xavier et al., 2021) was downloaded on April 22, 2022, as a JSON file. We processed the database following four main steps: (i) For each reported protein, we retained only the entries including single mutations with an experimental value of $\Delta\Delta G$ discarding entries with multiple mutations with a combined $\Delta\Delta G$; (ii) we reversed the sign of all the $\Delta\Delta G$ values to fit the sign provided by the outputs of RosettaDDGPrediction, (iii) we retained information on pH values and experimental methods as metadata, and (iv) we removed protein entries for which <10 mutations were reported. Upon processing, we identified 133 proteins. We then searched for 3D structures available for each protein in the Protein Data Bank. In this step, we retained matches that covered at least one mutation of interest. We retained only protein structures in their free state (i.e., not in a complex with other interactors) for a total of 121 target proteins, effectively removing 12 proteins where no structure or free state was found. We selected two enzymes that included a large number of amino acid substitutions with structural coverage (i.e., ENLYS and NUC as represented by the PDB structures 1P7S; Mooers et al., 2003 and 1EY0; Chen et al., 2000, and two human proteins of interest in health and disease; p53 and FKB1A as represented by the PDB structures 2XWR; Natan et al., 2011 and 2PPN; Szep et al., 2009 as case studies for this work). All are used as simplistic monomeric structures and chosen based on the coverage, quality, and lack of interactors. The experimental values obtained in an acidic or alkaline experimental setting (pH < 6 and pH > 8) were excluded, as the *ref2015* Rosetta energy function (Cartesian space version) is simulating an environment at pH 7.

This leaves 845 observations across the four proteins for pH values 6, 7, and 8 and three methodologies, two chemically denaturant-induced protein unfolding experimental protocols, guanidine hydrochloride (GdnHCl), Urea Denaturation (Urea), and one thermal denaturation protocol (Thermal). We modeled the experimental and predicted values using a simple linear model, analyzed the contribution of secondary structures, and built a generalized additive model, thereby defining the limitations of the model. Furthermore, we constructed a confusion

matrix based on the thresholds $\Delta\Delta G < -1$ kcal/mol for the stabilizing group, -1 kcal/mol $> \Delta\Delta G < 1$ kcal/mol for the neutral group, and $\Delta\Delta G > 1$ kcal/mol for the destabilizing group. Calculations were carried out with Rosetta 3.12.

In addition, we applied the rapid protein stability prediction (RaSP) (Blaabjerg et al., 2022) for comparison. We used the Colab version of the software (<https://colab.research.google.com/github/KULL-Centre/papers/blob/main/2022/ML-ddG-Blaabjerg-et-al/RaSPLab.ipynb#scrollTo=Z8nUmHI5rgjy>). We used the resulting *score_ml* as a proxy for the $\Delta\Delta G$ values. We applied a simple linear model to compare RaSP predicted $\Delta\Delta G$ values to the experimentally derived $\Delta\Delta G$ s and to the $\Delta\Delta G$ values predicted by the Rosetta-based protocols implemented in RosettaDDGPrediction. Additionally, to explore the model limitations, we built a generalized additive model.

4.2 | Case Study 2

We started from the phospho-regulated LIRs reported in our previous review article (Sora et al., 2020) and other literature search, and, for each of them, we verified if a complex with one of the LC3/GABARAP family members was available to use as starting structure for the mutational scan. We retained for the analyses the following complexes: LC3B:FUNCD1 (PDB entry 2N9X; Kuang et al., 2016) and GABARAP: PIK3C3 (PDB entry 6HOG; Birgisdottir et al., 2019).

We reconstructed missing coordinates in the structures using MODELER version 10.1 (Webb & Sali, 2016).

We used the *flexddg* protocol, as implemented in RosettaDDGPrediction, with the *talaris2014* energy function and Rosetta 3.12. Rosetta energy units (REUs) were converted to kilocalorie per mole with the conversion factors provided for this energy function (Park et al., 2016). We modeled the phosphorylated residues using phosphomimetic mutations to aspartic acid and glutamic acid for each phosphosite and included also tryptophan for phospho-tyrosine to identify possible effects due to steric hindrance. In the calculations, we used 35,000 backrub trials and an absolute score threshold for minimization convergence of 1 REUs. We generated an ensemble of 35 structures for each mutant variant and calculated the average $\Delta\Delta G$ s and the standard deviation among the individual binding free energies.

For the MutateX runs, we calculated changes in binding free energy using the Build Model and Analyze Complex functions of FoldX5 suite and averaging over five runs. The standard deviation for the $\Delta\Delta G$ values predicted with RosettaDDGPrediction and MutateX have been calculated using the GraphPad Prism 9 software.

We derived the $\Delta\Delta G$ values for each experimental K_d by using the following Gibbs free energy and constant equilibrium equations:

$$\Delta\Delta G = -RT\ln K_{\text{eq}}$$

$$K_{\text{eq}} = \frac{1}{K_d}$$

We combined the equations in order to compute the ΔG for the mutant and the WT as follows:

$$\Delta G = RT\ln K_d$$

The $\Delta\Delta G$ has been calculated subtracting the ΔG of the WT to the ΔG of the mutant:

$$\Delta\Delta G = RT\ln K_{d_mutant} - RT\ln K_{d_WT}$$

$$\Delta\Delta G = RT\ln \frac{K_{d_mutant}}{K_{d_WT}}$$

The standard deviations associated with the K_d measurement have been propagated for the $\Delta\Delta G$ calculations by using Uncertainty Calculator (<https://uncertaintycalculator.com>) and Propagation-of-Uncertainty-Calculator (<https://nicoco007.github.io/Propagation-of-Uncertainty-Calculator/>).

4.3 | Case Study 3

We retrieved experimental $\Delta\Delta G$ values from point mutations of the p53 DNA-binding domain from ThermoMutDB. Since ThermoMutDB stores $\Delta\Delta G_u$ values, they were converted to $\Delta\Delta G_f$ by changing the sign to make them easily comparable with Rosetta output values. A total of 31 mutations were selected, and when multiple experimental values were reported for the same variant, the average of their $\Delta\Delta G_f$ was used.

We used two different structures. The first one consists of the x-ray crystallography of the PDB entry 2XWR, with a resolution of 1.68 Å, which covers the DNA-binding domain from residues 91 to 289 and includes the zinc ion. The water molecules were removed using PyMOL (<http://www.pymol.org/pymol>). We also used the model from the AlphaFold2 database, which was trimmed to cover the same residues as the experimental x-ray structure, from 91 to 289. The missing zinc ion was added using PyMOL, identifying its coordinates by rigid body superimposition with the original structure. Before, we verified that the residues, which coordinate the zinc ion (C176, H179, C238, and C242), had a good alignment

and similar rotamer conformations between the two structures.

For the $\Delta\Delta G$ predictions, we mostly used the *cartddg* protocol with the *ref2015* and *talaris2014* scoring functions, each with 3 and 10 sampling runs and Rosetta 3.12. We also used the *cartddg2020* protocol on both structures, but only using the *ref2015* scoring function and three runs.

The performance was measured by Pearson correlation coefficient, MAE, and area under the curve (AUC) of the ROC curve. For the ROC curve, we used a threshold of 1.2 kcal/mol for the ThermoMutDB averaged values, meaning that mutations associated with a free energy change higher than 1.2 kcal/mol were considered destabilizing, according to the threshold selection proposed for p53 in our previous study (Degn et al., 2022). In comparison, mutations associated with a free energy change lower than the threshold were classified as nondestabilizing. The MAE was calculated using the following equation:

$$\xi = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

Here, y_i is the experimental values and \hat{y}_i is the predicted values.

4.4 | Case Study 4

We retrieved relevant VUSs from germline WGS data on 566 children with cancer included in an in-house dataset and published, in part, in a previous study (Byrjalsen et al., 2020). In addition, we analyzed a dataset including 566 samples from Danish children with different cancer types and whole genome sequencing data. An illustration of the workflows for analyzing the sequencing data and annotating the called variants is provided in Figures S4 and S5. The sequencing data have been processed with a pipeline based on Sentieon using the reference genome reference genome Grch38/hg38 from GATK resource bundle (ELELAB/sentieon_wgs_pipeline). Reads were aligned with BWA-MEM to the reference genome, and duplicate reads were removed. Reads were realigned around indels, and we applied Base Quality Score Recalibration together with the Haplotype algorithm for variant calling (equivalent to the GATK Haplotype; van der Auwera & O'Connor, n.d.). Then, as suggested by GATK best practices, we used Variant Quality Score Recalibration, which is an advanced filtering technique used on the variant call set that models the technical profile of variants in a training set using machine learning and filters

out potential artifacts from the callset. The filtered variants were uploaded to an in-house MySQL database where we linked them with information about genomic context (ENSEMBL v95), ENSEMBL consequences, deleteriousness-scores (CADD 1.6, REVEL, SIFT, PolyPhen) and variant frequency in the healthy population (gnomAD v3; Karczewski et al., 2020) based on their genomic position and alternate allele (Figure S3). To this purpose, we annotated our variants with the gnomAD popmax allele frequency, that is, the maximum allele frequency in all the continental populations (Gudmundsson et al., 2022), along with the allele frequency of the non-Finnish European population.

In particular, we annotated the REVEL score (Ioannidis et al., 2016) associated with the genomic mutation by using the publicly available dataset of pre-computed scores, by matching genomic coordinates, annotated transcript for the mutation and alternate nucleotide. We could not annotate a REVEL score for four of the identified variants, most likely as they were not missense. Two of them caused early translation termination by introducing a stop codon in the reference BRCA2 transcript (13:g.32337185A>T and 13:g.32398489A>T, corresponding to p.Lys944* and p.Lys3326* at protein level). The other two (2:g.47607407G>A and 2:g.47607446G>A, corresponding to p.Arg923Gln and p.Gly936Asp in MSH2 at protein level for the reference transcript) were annotated as both missense and nonsense-mediated decay in our dataset, meaning they are annotated as nonsense-mediated decay for at least some of the MSH2 transcripts, and this is probably the reason they were not available in the REVEL database.

We retained, as VUS to investigate, those variants located in the coding regions and found with an allele frequency in the non-Finnish European population lower than 1% in gnomAD v3 (build 38) as a proxy for a healthy population. This threshold has been selected according to the guidelines for clinical VUS studies (Richards et al., 2015). An illustration of the workflow for analyzing the sequencing data is provided in Figure S4.

We searched each variant in the selected 14 genes for the study in ClinVar (Landrum et al., 2014; Landrum et al., 2020) and retrieved annotations on them to verify if they are VUS, variants with conflicting evidence, or not reported yet in the database. To select the proteins and variants that can be investigated with RosettaDDGPrediction, we then searched in the AlphaFold2 database (Varadi, Anyango, Deshpande, et al., 2022) for the corresponding protein structures and retained those that had structural coverage for the variants in regions with high confidence (pLDDT > 70) trimming the N-terminal or C-terminal tails. For MLH1, we used the structure of the

two protein domains, for the other proteins, we retained cases in which the pLDDT score was low but located in loops that connect structured regions of folded domains. These regions are often very flexible in a protein structure, and it is thus expected that they could have a lower pLDDT score. We analyzed 14 proteins and 126 variants in total.

We excluded mutations either not covered by our trimmed models or derived from an isoform different from the one available in the AlphaFold2 database. Concerning MSH2, we did not analyze G936D since our isoform had 934 residues, while R293Q refers to the A0A2R8Y-G02_HUMAN isoform (Hillier et al., 2005). In the case of MSH6, T1125M was removed since derived from the A0A494C0M1_HUMAN transcript (Hillier et al., 2005). Furthermore, the following seven variants found in FANCL were also disregarded: S356N, S356N, G322V, F257C, T229A, I199V, and V181I. These variants were generated from FANCL isoform 2 (ENST00000402135.8, Q9NW38-2, 380aa), which did not match the AlphaFold model for FANCL (ENST00000233741.9, Q9NW38, 375aa).

AUTHOR CONTRIBUTIONS

Valentina Sora: Conceptualization (equal); methodology (equal); software (equal); visualization (supporting); writing – original draft (equal); writing – review and editing (equal). **Adrian Otamendi Laspiur:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (supporting); methodology (equal); software (supporting); validation (supporting); visualization (equal); writing – original draft (supporting); writing – review and editing (supporting). **Kristine Degn:** Conceptualization (supporting); data curation (supporting); formal analysis (supporting); investigation (supporting); methodology (supporting); supervision (supporting); visualization (equal); writing – original draft (supporting). **Matteo Arnaudi:** Conceptualization (supporting); formal analysis (supporting); investigation (supporting); visualization (equal); writing – original draft (supporting). **Mattia Utichi:** Conceptualization (supporting); formal analysis (supporting); investigation (supporting); visualization (supporting); writing – original draft (supporting). **Ludovica Beltrame:** Formal analysis (supporting); investigation (supporting); validation (supporting); visualization (supporting); writing – original draft (supporting). **Dayana De Menezes:** Formal analysis (supporting); investigation (supporting); visualization (supporting); writing – original draft (supporting). **Matteo Orlandi:** Data curation (supporting); formal analysis (supporting). **Ulrik Kristoffer Stoltze:** Data curation (supporting); validation (supporting); writing – review and editing (supporting). **Olga Rigina:** Data curation

(supporting); software (supporting). **Peter Wad Sackett:** Resources (supporting); software (supporting). **Karin Wadt:** Data curation (supporting); funding acquisition (supporting); supervision (supporting); writing – review and editing (supporting). **Kjeld Schmiegelow:** Funding acquisition (equal); supervision (supporting); writing – review and editing (supporting). **Matteo Tiberti:** Conceptualization (supporting); investigation (supporting); methodology (supporting); supervision (supporting); validation (equal); visualization (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Elena Papaleo:** Conceptualization (lead); data curation (equal); formal analysis (supporting); funding acquisition (lead); investigation (equal); methodology (supporting); project administration (lead); resources (lead); supervision (lead); validation (equal); visualization (supporting); writing – original draft (lead); writing – review and editing (lead).

ACKNOWLEDGMENTS

Our research has been supported by Carlsberg Foundation Distinguished Fellowship (CF18-0314), Danmarks Grundforskningsfond (DNRF125), Hartmanns Fond (R241-A33877), LEO Foundation (LF17006), and Novo-Nordisk Fonden Bioscience and Basic Biomedicine (NNF20OC0065262). This work is part of Interregional Childhood Oncology Precision Medicine Exploration (iCOPE), a cross-Oresund collaboration between University Hospital Copenhagen, Rigshospitalet, Lund University, Region Skåne, and Technical University Denmark (DTU), supported by the European Regional Development Fund. This work is also part of the Danish nationwide research program Childhood Oncology Network Targeting Research, Organisation & Life expectancy (CONTROL) and supported by the Danish Cancer Society (R-257-A14720) and the Danish Childhood Cancer Foundation (2019-5934 and 2020-5769).

DATA AVAILABILITY STATEMENT

The software and data presented in this work are available at <https://github.com/ELELAB/RosettaDDGPrediction>, <https://osf.io/84kwe/> and https://github.com/ELELAB/sentieon_wgs_pipeline.

ORCID

Elena Papaleo  <https://orcid.org/0000-0002-7376-5894>

REFERENCES

- Abildgaard AB, Stein A, Nielsen SV, Schultz-Knudsen K, Papaleo E, Shrikhande A, et al. Computational and cellular studies reveal structural destabilization and degradation of MLH1 variants in Lynch syndrome. *Elife*. 2019;8:e49138. <https://doi.org/10.7554/eLife.49138.001>
- Adachi D, Oda T, Yagasaki H, Nakasato K, Taniguchi T, D'Andrea AD, et al. Heterogeneous activation of the Fanconi anemia pathway by patient-derived FANCA mutants. *Hum Mol Genet*. 2002;11:3125–34.
- Akdal M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, et al. A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol*. 2022;29:1056–67. <https://doi.org/10.1038/s41594-022-00849-w>
- Anderson CL, Munawar S, Reilly L, Kamp TJ, January CT, Delisle BP, et al. How functional genomics can keep pace with VUS identification. *Front Cardiovasc Med*. 2022;9:90043.
- Arnaudi M, Beltrame L, Degn K, Utichi M, Pettenella A, Scrima S, et al. MAVISp: multi-layered assessment of Variants by structure for proteins. *bioRxiv*. 2022. <https://doi.org/10.1101/2022.10.22.513328>
- Bæk KT, Kepp KP. Data set and fitting dependencies when estimating protein mutant stability: toward simple, balanced, and interpretable models. *J Comput Chem*. 2022;43:504–18.
- Barlow KA, Ó Conchúir S, Thompson S, Suresh P, Lucas JE, Heinonen M, et al. Flex ddG: Rosetta Ensemble-based estimation of changes in protein–protein binding affinity upon mutation. *J Phys Chem B*. 2018;122:5389–99.
- Benedix A, Becker CM, de Groot BL, Cafilisch A, Böckmann RA. Predicting free energy changes using structural ensembles. *Nat Methods*. 2009;6:3–4.
- Bertazzo M, Gobbo D, Decherchi S, Cavalli A. Machine learning and enhanced sampling simulations for computing the potential of mean force and standard binding free energy. *J Chem Theory Comput*. 2021;17:5287–300.
- Birgisdottir ÁB, Mouilleron S, Bhujabal Z, Wirth M, Sjøttem E, Evjen G, et al. Members of the autophagy class III phosphatidylinositol 3-kinase complex I interact with GABARAP and GABARAPL1 via LIR motifs. *Autophagy*. 2019;15:1333–55.
- Blaabjerg LM, Kassem MM, Good LL, Jonsson N, Cagiada M, Johansson KE, et al. Rapid protein stability prediction using deep learning representations. *bioRxiv*. 2022. <https://doi.org/10.1101/2022.07.14.500157>
- Buel GR, Walters KJ. Can AlphaFold2 predict the impact of missense mutations on structure? *Nat Struct Mol Biol*. 2022;29(1):1–2.
- Buß O, Rudat J, Ochsenreither K. FoldX as protein engineering tool: better than random based approaches? *Comput Struct Biotechnol J*. 2018;16:25–33.
- Byrjalsen A, Hansen TVO, Stoltze UK, Mehrjouy MM, Barnkob NM, Hjalgrim LL, et al. Nationwide germline whole genome sequencing of 198 consecutive pediatric cancer patients reveals a high incidence of cancer prone syndromes. *PLoS Genet*. 2020;16:e1009231.
- Cagiada M, Johansson KE, Valanciute A, Nielsen SV, Hartmann-Petersen R, Yang JJ, et al. Understanding the origins of loss of protein function by analyzing the effects of thousands of variants on activity and abundance. *Mol Biol Evol*. 2021;38:3235–46.
- Chen J, Lu Z, Sakon J, Stites WE. Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. *J Mol Biol*. 2000;303:125–30.

- Davey NE. The functional importance of structure in unstructured protein regions. *Curr Opin Struct Biol.* 2019;56:155–63.
- Davey NE, van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. *Mol Biosyst.* 2011;8:268–81.
- Degn K, Beltrame L, Dahl Hede F, Sora V, Nicolaci V, Vabistsevits M, et al. Cancer-related mutations with local or long-range effects on an allosteric loop of p53. *J Mol Biol.* 2022; 434:167663.
- Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics.* 2019;35:1–2. <https://doi.org/10.1093/bioinformatics/btz184>
- Evans R, O'Neill M, Pritzel A, Senior NAA, Green T, Židek A, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv.* 2022. <https://doi.org/10.1101/2021.10.04.463034>
- Fas BA, Maiani E, Sora V, Kumar M, Mashkooor M, Lambrugh M, et al. The conformational and mutational landscape of the ubiquitin-like marker for autophagosome formation in cancer. *Autophagy.* 2020;17:1–24.
- Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *J Exp Clin Cancer Res.* 2020;39:1–12.
- Fierheller CT, Guitton-Sert L, Alenezi WM, Revil T, Oros KK, Gao Y, et al. A functionally impaired missense variant identified in French Canadian families implicates FANCI as a candidate ovarian cancer-predisposing gene. *Genome Med.* 2021;13: 1–26.
- Frenz B, Lewis SM, King I, DiMaio F, Park H, Song Y. Prediction of protein mutational free energy: benchmark and sampling improvements increase classification accuracy. *Front Bioeng Biotechnol.* 2020;8:558247.
- Gasparini M, Starita L, Shendure J. The power of multiplexed functional analysis of genetic variants. *Nat Protoc.* 2016;11(10): 1782–7.
- Geng C, Xue LC, Roel-Touris J, Bonvin AMJJ. Finding the $\Delta\Delta G$ spot: are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it? *Wiley Interdiscip Rev Comput Mol Sci.* 2019;9:e1410.
- Gerasimavicius L, Liu X, Marsh JA. Identification of pathogenic missense mutations using protein stability predictors. *Sci Rep.* 2020;10(1):1–10.
- Gerasimavicius L, Livesey BJ, Marsh JA. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nature Commun.* 2022; 13(1):1–15.
- Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, et al. Variant interpretation using population databases: lessons from gnomAD. *Hum Mutat.* 2022;431:1012–30. <https://doi.org/10.1002/humu.24309>
- Hillier LDW, Graves TA, Fulton RS, Fulton LA, Pepin KH, Minx P, et al. Generation and annotation of the DNA sequences of human chromosomes 2 and 4. *Nature.* 2005;434: 724–31.
- Høie MH, Cagiada M, Beck Frederiksen AH, Stein A, Lindorff-Larsen K. Predicting and interpreting large-scale mutagenesis data using analyses of protein stability and conservation. *Cell Rep.* 2022;38:110207.
- Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet.* 2016;99:877–85.
- Jepsen MM, Fowler DM, Hartmann-Petersen R, Stein A, Lindorff-Larsen K. Classifying disease-associated variants using measures of protein activity and stability. *Protein Homeostasis Diseases;* 2020. p. 91–107. <https://doi.org/10.1016/B978-0-12-819132-3.00005-1>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583–9.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581:434–43.
- Katarina Sooe Tiemann J, Zschach H, Lindorff-Larsen K, Stein A. Interpreting the molecular mechanisms of disease variants in human membrane proteins. *bioRxiv.* <https://doi.org/10.1101/2022.07.12.499731>
- Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins.* 2011;79:830–8.
- Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–protein complexes. *Proc Natl Acad Sci U S A.* 2002;99(14):116–14121.
- Kuang Y, Ma K, Zhou C, Ding P, Zhu Y, Chen Q, et al. Structural basis for the phosphorylation of FUNDC1 LIR as a molecular switch of mitophagy. *Autophagy.* 2016;12:2363–73.
- Kumari R, Kumar R, Lynn A. G-mmpbsa – a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model.* 2014;54:1951–62.
- Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. *Nucleic Acids Res.* 2020;48:D835–44.
- Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42:D980–5.
- Lv M, Wang C, Li F, Peng J, Wen B, Gong Q, et al. Structural insights into the recognition of phosphorylated FUNDC1 by LC3B in mitophagy. *Protein Cell.* 2017;8:25–38.
- McBride JM, Polev K, Reinharz V, Grzybowski BA, Tlustý T. AlphaFold2 can predict structural and phenotypic effects of single mutations. *bioRxiv.* 2022. <https://doi.org/10.1101/2022.04.14.488301>
- Mooers BHM, Datta D, Baase WA, Zollars ES, Mayo SL, Matthews BW. Repacking the core of T4 lysozyme by automated design. *J Mol Biol.* 2003;332:741–56.
- Natan E, Baloglu C, Pagel K, Freund SMV, Morgner N, Robinson CV, et al. Interaction of the p53 DNA-binding domain with its n-terminal extension modulates the stability of the p53 tetramer. *J Mol Biol.* 2011;409:358–68.
- Nielsen SV, Stein A, Dinitzen AB, Papaleo E, Tatham MH, Poulsen EG, et al. Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. *PLoS Genet.* 2017;13:e1006739.
- Nygaard M, Terkelsen T, Vidas Olsen A, Sora V, Salamanca Vitoria J, Rizza F, et al. The mutational landscape of the

- oncogenic MZF1 SCAN domain in cancer. *Front Mol Biosci.* 2016;3:1–18.
- Ollodart AR, Yeh CLC, Miller AW, Shirts BH, Gordon AS, Dunham MJ. Multiplexing mutation rate assessment: determining pathogenicity of Msh2 variants in *Saccharomyces cerevisiae*. *Genetics.* 2021;218:iyab058.
- O'Meara MJ, Leaver-Fay A, Tyka MD, Stein A, Houlihan K, DiMaio F, et al. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J Chem Theory Comput.* 2015;11:609–22.
- Pak MA, Markhieva KA, Novikova MS, Petrov DS, Vorobyev IS, Maksimova ES, et al. Using AlphaFold to predict the impact of single mutations on protein stability and function. *bioRxiv.* 2021. <https://doi.org/10.1101/2021.09.19.460937>
- Pancotti C, Benevenuto S, Birolo G, Alberini V, Repetto V, Sanavia T, et al. Predicting protein stability changes upon single-point mutation: a thorough comparison of the available tools on a new dataset. *Brief Bioinform.* 2022;23:bbab555.
- Park H, Bradley P, Greisen P Jr, Liu Y, Mulligan VK, Kim DE, et al. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J Chem Theory Comput.* 2016;12:6201–12.
- Pires DEV, Rodrigues CHM, Albanaz ATS, Myung MKY, Xavier J, Michanetzi E-M, et al. Exploring protein supersecondary structure through changes in protein folding, stability, and flexibility. *Methods Mol Biol.* 2019;1958:173–85.
- Porta-Pardo E, Ruiz-Serra V, Valentini S, Valencia A. The structural coverage of the human proteome before and after AlphaFold. *PLoS Comput Biol.* 2022;18:e1009818.
- Raniolo S, Limongelli V. Ligand binding free-energy calculations with funnel metadynamics. *Nat Protoc.* 2020;15(9):2837–66.
- Richards S, Aziz N, Bale S, Bick D, das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17:405–24.
- Robertson AJ, Courtney JM, Shen Y, Ying J, Bax A. Concordance of X-ray and AlphaFold2 models of SARS-CoV-2 Main protease with residual dipolar couplings measured in solution. *J Am Chem Soc.* 2021;143(19):306–19310.
- Ruff KM, Pappu R v. AlphaFold and implications for intrinsically disordered proteins. *J Mol Biol.* 2021;433:167208.
- Scheller R, Stein A, Nielsen SV, Marin FI, Gerdes AM, Marco M, et al. Toward mechanistic models for genotype–phenotype correlations in phenylketonuria using protein stability calculations. *Hum Mutat.* 2019;40:444–57. <https://doi.org/10.1002/humu.23707>
- Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res.* 2005;33:W382–8.
- Seeliger D, de Groot BL. Protein thermostability calculations using alchemical free energy simulations. *Biophys J.* 2010;98:2309–16.
- Shakeel S, Rajendra E, Alcón P, O'Reilly F, Chorev DS, Maslen S, et al. Structure of the Fanconi anaemia monoubiquitin ligase complex. *Nature.* 2019;575(7781):234–7.
- Siebenmorgen T, Zacharias M. Computational prediction of protein–protein binding affinities. *Wiley Interdiscip Rev Comput Mol Sci.* 2020;10:e1448.
- Smith CA, Kortemme T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J Mol Biol.* 2008;380:742–56.
- Somyajit K, Subramanya S, Nagaraju G. RAD51C: a novel cancer susceptibility gene is linked to Fanconi anemia and breast cancer. *Carcinogenesis.* 2010;31:2031–8.
- Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nature Rev Cancer.* 2018;18(11):696–705.
- Sora V, Kumar M, Maiani E, Lambrughini M, Tiberti M, Papaleo E. Structure and dynamics in the ATG8 family from experimental to computational techniques. *Front Cell Dev Biol.* 2020;8:420.
- Stein A, Fowler DM, Hartmann-Petersen R, Lindorff-Larsen K. Biophysical and mechanistic models for disease-causing protein variants. *Trends Biochem Sci.* 2019;44:575–88.
- Szep S, Park S, Boder ET, van Duyne GD, Saven JG. Structural coupling between FKBP12 and buried water. *Proteins.* 2009;74:603–11.
- Tiberti M, Terkelsen T, Degn K, Beltrame L, Cremers TC, da Piedade I, et al. MutateX: an automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles. *Brief Bioinform.* 2022;23:bbac074.
- Tsaban T, Varga JK, Avraham O, Ben-Aharon Z, Khramushin A, Schueler-Furman O. Harnessing protein folding neural networks for peptide-protein docking. *Nat Commun.* 2022;13:176.
- Usmanova DR, Bogatyreva NS, Ariño Bernad J, Eremina AA, Gorshkova AA, Kanevskiy GM, et al. Self-consistency test reveals systematic bias in programs for prediction change of stability upon mutation. *Bioinformatics.* 2018;34:3653–8.
- Valanciute A, Nygaard L, Zschach H, Jepsen MM, Lindorff-Larsen K, Stein A. Accurate protein stability predictions from homology models. *Comput Struct Biotech J.* 2022;21:66–73.
- van der Auwera G, O'Connor B. *Genomics in the Cloud.* O'Reilly Media, Inc. 2020, p. 300.
- van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* 2014;114:6733–78.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022;50:D439–44.
- Varadi M, Anyango S, Armstrong D, Berrisford J, Choudhary P, Deshpande M, et al. PDBE-KB: collaboratively defining the biological context of structural data. *Nucleic Acids Res.* 2022;50:D534–42.
- Wang LH, Wu CF, Rajasekaran N, Shin YK. Loss of tumor suppressor gene function in human cancer: an overview. *Cell Physiol Biochem.* 2018;51:2647–93.
- Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics.* 2016;54:5.6.1–5.6.37.
- Weile J, Roth FP. Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. *Hum Genet.* 2018;137(9):665–78.

- Wilkes DC, Sailer V, Xue H, Cheng H, Collins CC, Gleave M, et al. A germline FANCA alteration that is associated with increased sensitivity to DNA damaging agents. *Cold Spring Harb Mol Case Stud.* 2017;3:a001487.
- Wingbermhle S, Schäfer L v. Capturing the flexibility of a protein-ligand complex: binding free energies from different enhanced sampling techniques. *J Chem Theory Comput.* 2020;16:4615–30.
- Xavier JS, Nguyen TB, Karmarkar M, Portelli S, Rezende PM, Velloso JPL, et al. ThermoMutDB: a thermodynamic database for missense mutations. *Nucleic Acids Res.* 2021;49:D475–9.
- Zhang X, Lu X, Akhter S, Georgescu MM, Legerski RJ. FANCI is a negative regulator of Akt activation. *Cell Cycle.* 2016;15:1134–43.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sora V, Laspiur AO, Degn K, Arnaudi M, Utichi M, Beltrame L, et al. RosettaDDGPrediction for high-throughput mutational scans: From stability to binding. *Protein Science.* 2023;32(1):e4527. <https://doi.org/10.1002/pro.4527>