THE PROTEIN SOCIETY · WILEY

DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations

Carlos H.M. Rodrigues[1,2] | Douglas E.V. Pires[1,2,3] | David B. Ascher[1,2,4]

[1]Structural Biology and Bioinformatics, Department of Biochemistry, Bio21 Institute, University of Melbourne, Melbourne, Victoria, Australia

[2]Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

[3]School of Computing and Information Systems, University of Melbourne, Melbourne, Victoria, Australia

[4]Department of Biochemistry, University of Cambridge, Cambridge, UK

Correspondence
David B. Ascher and Douglas E.V. Pires, Structural Biology and Bioinformatics, Department of Biochemistry, Bio21 Institute, University of Melbourne, Victoria, Australia.
Email: david.ascher@unimelb.edu.au (D. B. A.) and douglas.pires@unimelb.edu.au (D. E. V. P.)

Abstract

Predicting the effect of missense variations on protein stability and dynamics is important for understanding their role in diseases, and the link between protein structure and function. Approaches to estimate these changes have been proposed, but most only consider single-point missense variants and a static state of the protein, with those that incorporate dynamics are computationally expensive. Here we present DynaMut2, a web server that combines Normal Mode Analysis (NMA) methods to capture protein motion and our graph-based signatures to represent the wildtype environment to investigate the effects of single and multiple point mutations on protein stability and dynamics. DynaMut2 was able to accurately predict the effects of missense mutations on protein stability, achieving Pearson's correlation of up to 0.72 (RMSE: 1.02 kcal/mol) on a single point and 0.64 (RMSE: 1.80 kcal/mol) on multiple-point missense mutations across 10-fold cross-validation and independent blind tests. For single-point mutations, DynaMut2 achieved comparable performance with other methods when predicting variations in Gibbs Free Energy ($\Delta\Delta G$) and in melting temperature (ΔT_m). We anticipate our tool to be a valuable suite for the study of protein flexibility analysis and the study of the role of variants in disease. DynaMut2 is freely available as a web server and API at http://biosig.unimelb.edu.au/dynamut2.

KEYWORDS
dynamics, graph-based signatures, missense mutations, stability changes

1 | INTRODUCTION

Proteins are highly dynamic, metastable molecular machines. Missense mutations are associated with more than half of all known inherited diseases, however, they are often associated with more subtle molecular effects than mutations that lead to larger changes to the mature peptide. These single amino acid changes can readily disrupt the intricate network of intramolecular interactions, affecting how a protein folds, its stability, dynamics, and ultimately protein function. Beyond phenotypic outcomes,[1–22] it also has direct implications for their experimental study, protein engineering,[23,24] drug design,[25–30] and use in industrial processes.[31]

A number of approaches have been developed to predict how missense mutations affect protein stability using either sequence[32–34] or structural information.[35–37] The information from both approaches is often complementary;

however, structural methods have generally assumed a protein is static and does not consider the implications of a mutation within its conformational landscape. We previously showed that by considering both the mutation environment and the protein dynamics, we could more accurately predict the effects of single-point missense mutations.[38]

Most predictive tools, however, have been limited to single point missense variants, and the inclusion of protein dynamics computationally scales poorly with protein size. Here we present DynaMut2, an enhanced server that combines normal mode analysis with our graph-based representation of protein structure, to accurately and quickly predict the effects of single and multiple point mutations on protein stability and dynamics.

2 | RESULTS AND DISCUSSION

The DynaMut2 development workflow is summarized in Figure 1. Data on single and multiple point mutations were derived from ProTherm.[39] Given the wide range of molecular mechanisms by which mutations can impact protein function, we modeled the effects of each mutation using a range of features, including protein dynamics (NMA), wild-type residue environment, substitution propensities and contact potential scores, interatomic interactions[40] and also our well-validated graph-based signatures approach.[35,41–48] These were then used to train and test machine learning algorithms. Our predictive models were further evaluated using independent blind test sets.

2.1 | Predicting the effects of single point mutations

We initially evaluated the performance of our approach to predict changes in stability caused by single point mutations. DynaMut2 was able to achieve a Pearson's correlation of $r = 0.72$ (RMSE = 1.02 kcal/mol) for the dataset S4022, under 10-fold cross-validation, and $r = 0.68$ on S611, our non-redundant independent test set (RMSE = 1.14 kcal/mol) (Figure 2), outperforming all other methods (Table 1). The comparable performance between cross-validation and non-redundant blind test supports the generalizability of the final model. After removing 10% of outliers, performance remained
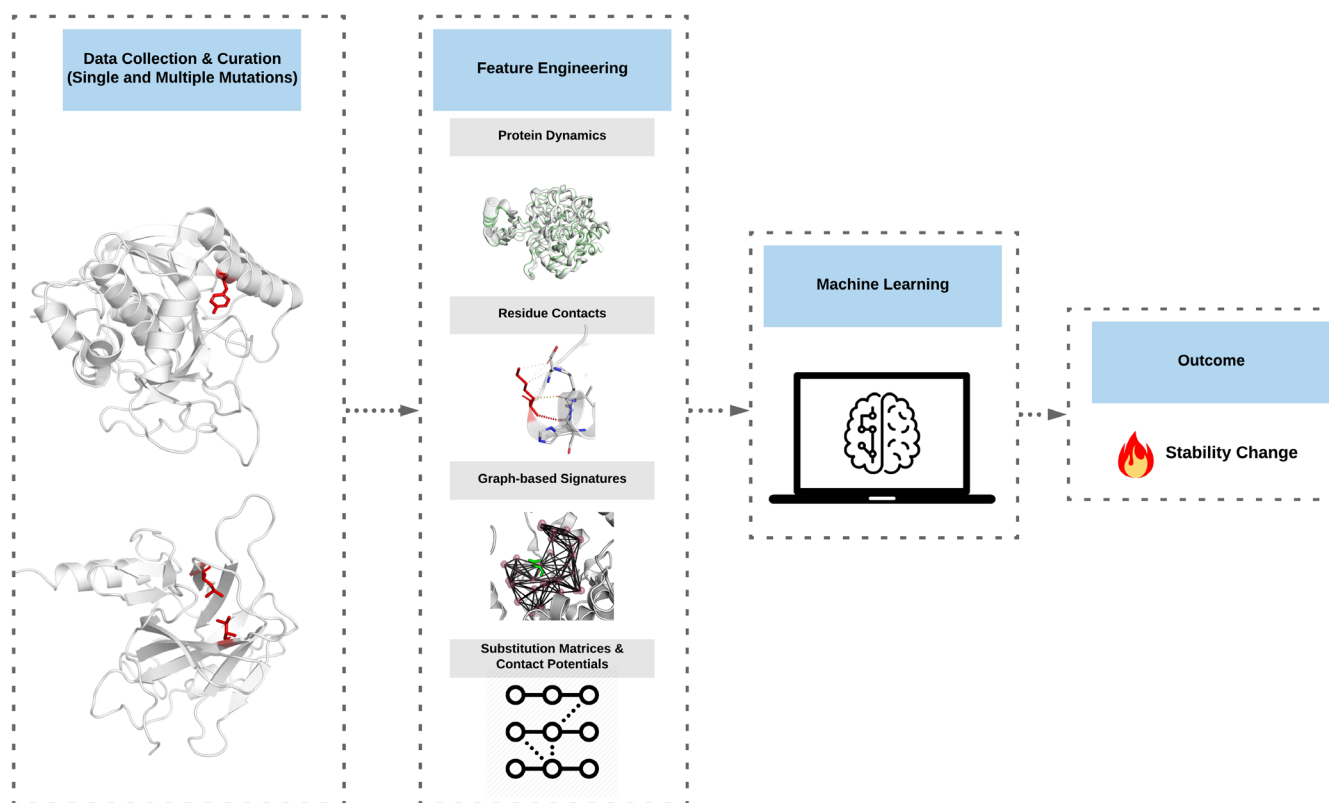
FIGURE 1 DynaMut2 workflow. The methodology for this work can be summarized into four steps: (1) data collection and curation of single and multiple mutations, (2) feature engineering to model the effects of mutations, (3) supervised machine learning, and (4) the predicted effects on stability and dynamics
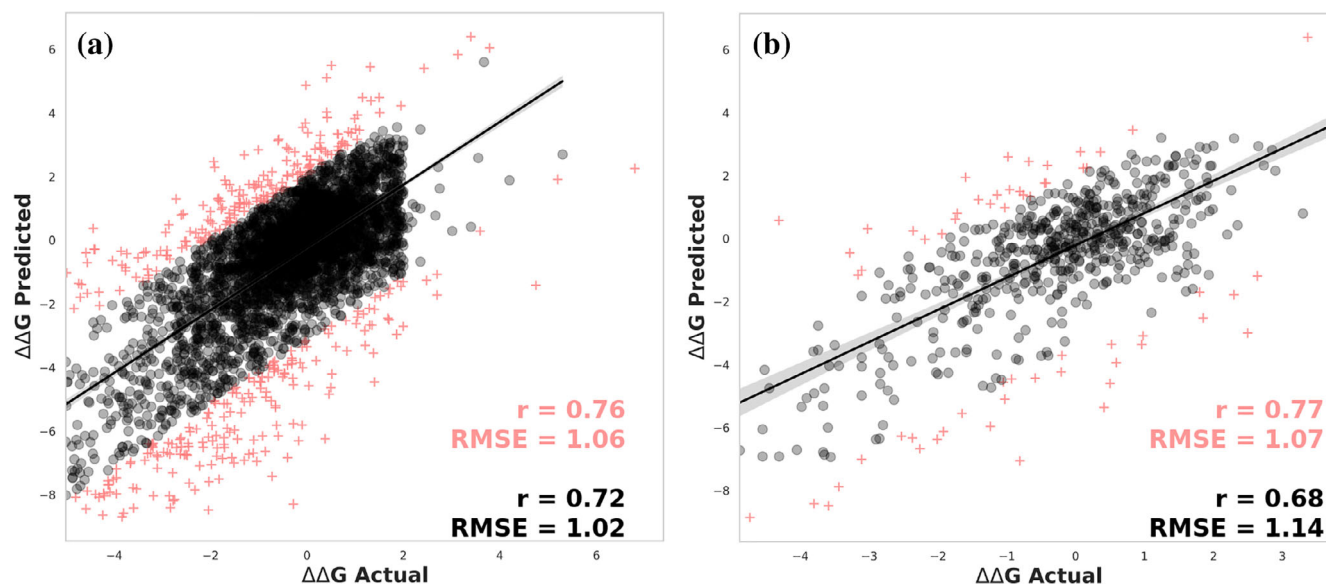
FIGURE 2 Predictive performance of DynaMut2 on 10-fold cross-validation (a) and non-redundant test sets (b) for single point mutations. 10% of outliers are shown as pink crosses

TABLE 1 Comparative performance across the non-redundant test set S611

Method	Overall		Stabilizing mutations		Destabilizing mutations		
	Pearson (r)	RMSE (kcal/Mol)	Pearson (r)	RMSE (kcal/Mol)	Pearson (r)	RMSE (kcal/Mol)	AUC
DynaMut2	**0.68**	**1.14**	**0.51**	**1.02**	**0.62**	**0.91**	**0.68**
DynaMut1	0.49[*]	1.38[+]	0.47	1.24	0.55	1.01	0.62
SDM	0.35[*]	1.93[+]	0.15[*]	2.00[+]	0.36[*]	1.86[+]	0.60[#]
mCSM	0.46[*]	1.42[+]	0.11[*]	1.81[+]	0.56	0.98	0.56[#]
DUET	0.48[*]	1.40[+]	0.09[*]	1.75[+]	0.58	1.00	0.56[#]
ENCoM	−0.14[*]	2.03[+]	−0.01[*]	1.94[+]	−0.18[*]	2.09[+]	0.41[#]
Maestro	−0.36[*]	1.55[+]	0.27[*]	1.17	0.43[*]	1.81[+]	0.46[#]
I-mutant	0.33[*]	1.47[+]	0.03[*]	1.83[+]	0.49[*]	1.09[+]	0.51[#]
MUpro[a]	0.15[*]	1.71[+]	−0.05[*]	2.15[+]	0.23[*]	1.21[+]	0.50[#]

[*]p Value $< .05$ compared with DynaMut2 using z test.
[#]p Value $< .05$ compared with DynaMut2 using t test.
[+]p Value $< .05$ compared with DynaMut2 using Diebold-Mariano test.
[a]48 mutations were left out due to input issues.

consistent for the training set with $r = 0.76$ (RMSE = 1.06), and increased to $r = 0.77$ (RMSE = 1.07) on the test set. No significant differences in the distributions of properties were observed for the outliers compared to the overall dataset.

Due to the natural imbalance between stabilizing and destabilizing mutations present in the training and evaluation data (Figure S1), we further analyzed performance on the respective classes separately. Across the non-redundant validation se S611, DynaMut2 achieved a Pearson's correlation of $r = 0.62$ (RMSE = 1.75) and $r = 0.51$ (RMSE = 1.88) on destabilizing and stabilizing

mutations, respectively. The slightly lower performance toward stabilizing mutations was expected due to the imbalanced distribution of data but was significantly improved compared to previous methods (Table 1). These results remained consistent when we compared the performances using rank coefficient scores Kendall and Spearman (Table S1). This was further reflected in the ability of DynaMut2 to correctly classify stabilizing and destabilizing mutations (AUC 0.68), outperforming previous approaches.

To further investigate potential biases in the predictive performance, we evaluated the performance of

DynaMut2 on subsets derived from the O2567 dataset.[49] DynaMut2 showed significantly better performance than all other approaches for mutations on buried residues (RSA ≤30%; Table S2). A small deterioration in performance is observed on mutations on exposed residues (RSA > 30%; Table S3), likely to be related to the smaller number features captured by the graph-based signatures in DynaMut2; however, our method still achieved comparable results to mCSM, MAESTRO and SDM, and outperformed other approaches. Evaluating the performance on different protein CATH classifications, DynaMut2 outperformed other approaches across β-sheet structures (Table S4), and α-helix and β-helix structures (Table S5). The size of the protein being mutated did not affect performance, with comparable performance between larger proteins (>150 residues; Table S6) and small proteins (<150 residues; Table S7), outperforming all other evaluated approaches. Similarly, DynaMut2 performance was similar to mutations from large to small residues (Table S8), from small to large residues (Table S9), or for mutations between residues of comparable sizes (Table S10). Encouragingly, DynaMut2 outperformed all other approaches on mutations leading to a change in volume and demonstrated comparable performance to the top approaches for mutations between residues of similar volume. Overall, this highlighted that DynaMut2 predictive performance across all single-point mutations was significantly more balanced and less biased than all other methods evaluated.

We further evaluated the performance of our model across an additional independent test set, S276. DynaMut2 achieved a Pearson's correlation of 0.52, comparable with the best-performing methods (Table 2) and significantly better than MUpro.[50] Although not directly comparable, as there is a correlation between changes upon mutation in stability (ΔG) and thermal stability (T_m),[51] the performance of DynaMut2 on predicting changes in melting temperature was assessed using the blind test set S173. Results were stratified by protein structure and summarized in Table S11. Overall, DynaMut2 ranks fourth among the methods evaluated; however, performances of all methods varied greatly between structures. These results indicate a possible challenge in accurately predicting the thermal stability effects of mutations on a more diverse set of proteins.
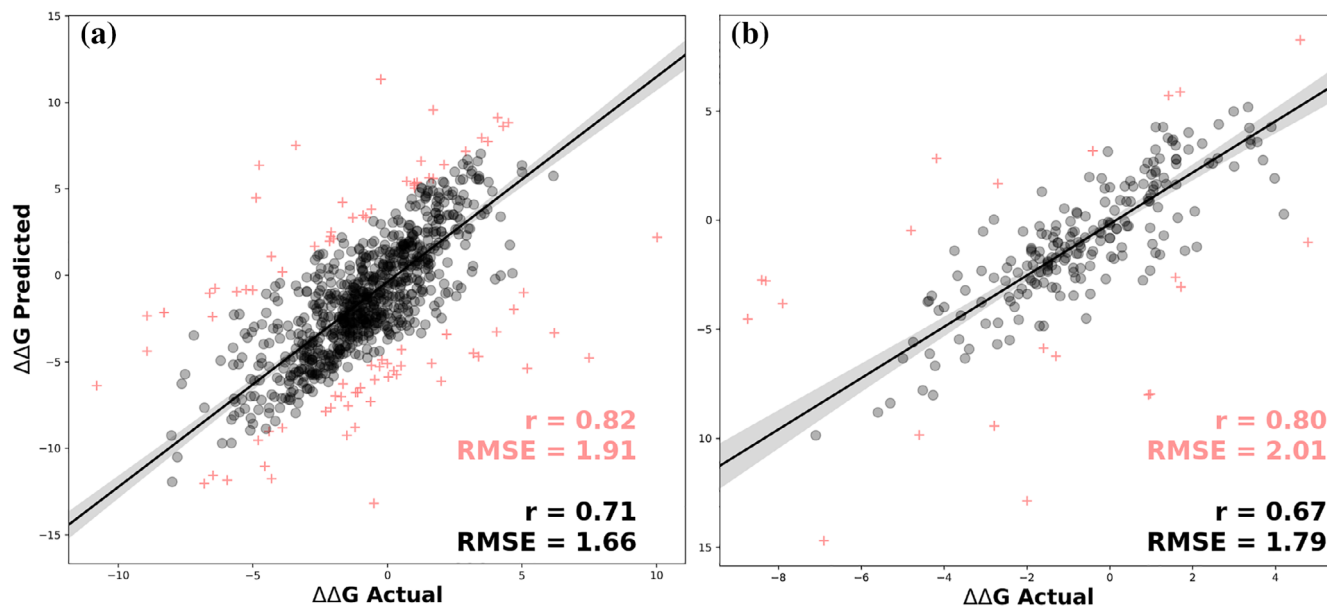
2.2 | Predicting the effects of multiple point mutations

The performance of our approach to predict the effects of multiple point mutations on protein stability was then assessed. DynaMut2 achieved a Pearson's correlation of $r = 0.71$ (RMSE = 1.66 kcal/mol) under 10-fold cross-validation and $r = 0.67$ (RMSE = 1.79 kcal/mol) on our non-redundant test set. The comparable performance between cross-validation and blind test set again gave confidence in the generalizability of the approach. This significantly outperformed the previously reported performances of DDGun, DDGun3D, Maestro, and FoldX, whose correlations ranged from 0.37 to 0.55 on the experimental multiple point mutations in ProTherm.[52] Performances were consistent when considering only 90% of the data, with DynaMut2 achieving $r = 0.82$ (RMSE = 1.91) and $r = 0.80$ (RMSE = 2.01) on 10-fold cross-validation and blind-test, respectively (Figure 3). This indicates that outlier predictions were not having a significant effect on the correlations.

The unbalanced nature of the training dataset was evident when we analyzed the performance of our final model on stabilizing and destabilizing multiple mutations separately (Table 3). Overall, DynaMut2 was able to correctly classify 80% of multiple missense mutations (AUC 0.84) in the blind test set, including 93% of the destabilizing mutations, providing confidence in the ranking ability of the approach. As expected, however, across our non-redundant test set DynaMut2 shows a better performance toward predicting multiple mutations with a destabilizing effect, achieving a Pearson's correlation of $r = 0.56$ (RMSE = 2.66), while for stabilizing entries the performance drops to $r = 0.42$ (RMSE = 2.94). These results indicate a need for new experimental data on multiple point mutations, especially those with stabilizing effects, since the use of hypothetical reverse mutations is likely to add more uncertainty to the model.

2.3 | Web server

We have implemented DynaMut2 as a freely available and user-friendly web server (http://biosig.unimelb.edu.au/

TABLE 2 Comparative performance across the S276 blind test of experimental $\Delta\Delta G$

Method	R	MAE (kcal/Mol)
DynaMut2	0.52	0.88
DeepDDG	0.55	0.86
SDM	0.48	1.02
mCSM	0.46	0.90
I-mutant	0.45	0.91
STRUM	0.44	0.88
MUpro	0.19*	1.06

*p Value < .05 compared with DynaMut2 using z test.

FIGURE 3 Predictive performance of DynaMut2 on 10-fold cross-validation (a) and non-redundant test sets (b) for multiple point mutations. 10% of outliers are shown as pink crosses

Methods	Overall			Stabilizing			Destabilizing		
	r_p	Tau	r_s	r_p	Tau	r_s	r_p	Tau	r_s
DynaMut2	**0.71**	**0.58**	**0.75**	**0.42**	**0.38**	**0.53**	**0.56**	**0.47**	**0.63**
MAESTRO	0.19*	0.13+	0.19#	0.12*	0.07+	0.08#	0.21*	0.14+	0.21#
FoldX	0.33*	0.21+	0.31#	0.04*	0.06+	0.09#	0.30*	0.19+	0.27#

TABLE 3 Comparative performance on multiple mutations prediction across different correlation coefficients

*p Value < .05 compared with DynaMut2 using Fisher r-to-z transformation.
+p Value <.05 by transforming tau-to-r followed by Fisher r-to-z transformation.
#p Value <.05 by transforming rho-to-r followed by Fisher r-to-z transformation.

dynamut2/). The frontend was developed using Materializecss version 1.0.0 and the backend uses the Flask module (1.0.2) from the Python programming language. The web server is hosted on a Linux machine running Apache.

2.4 | Input

DynaMut2 can be used in three different ways[1]: predicting $\Delta\Delta G$ for single point mutations,[2] predicting $\Delta\Delta G$ for multiple point mutations (up to three), and also analysis of protein dynamics based on NMA. For predicting single point mutations, similarly to our previous implementation of DynaMut, two different inputs are available: "Single mutation" and "List of Mutations". For the Single Mutation option, users are required to provide a protein structure on PDB format or provide a four-digit code of an entry on the PDB, the chain identifier where the mutation occurs and the point mutation defined as string comprising wild-type residue one-letter code, residue position, and mutant residue one-letter code. For the List of Mutations option, users must provide the structure of the protein, similarly to the Single Mutation option, and also upload a file with the list of variants (one per line), following the same mutation code previously defined.

For predicting the effects of multiple mutations, users are required to provide the structure of the protein, as previously described, and also the multiple mutations separated by a comma. DynaMut2 also allows for submitting a list of multiple point mutations to be analyzed in batch. These can be input by uploading a file with one entry of multiple mutations separated by comma per line.

Alternatively, for protein dynamic analysis, users are required to input the protein structure by uploading a file using the PDB format or provide a valid four-digit code for a PDB entry, and also select one of the force fields available to guide structural interactions for NMA. All force field options available are detailed in Table S12.

2.5 | Output

For single point mutations on the "Single Mutation" option, predicted $\Delta\Delta G$ is shown at the top with details of users' input and also the wild-type residue environment (Figure 4). All interatomic contacts calculated with Arpeggio are also displayed as an interactive viewer using

NGL viewer.[53] On the "Mutation List" option, the results are displayed as a downloadable table with options to view details of each variation separately, similarly to the analysis provided by the "Single Mutation", option.

The results for multiple mutations, predictions are displayed at the top of the page with detailed information for each mutation, if a list is provided these results are
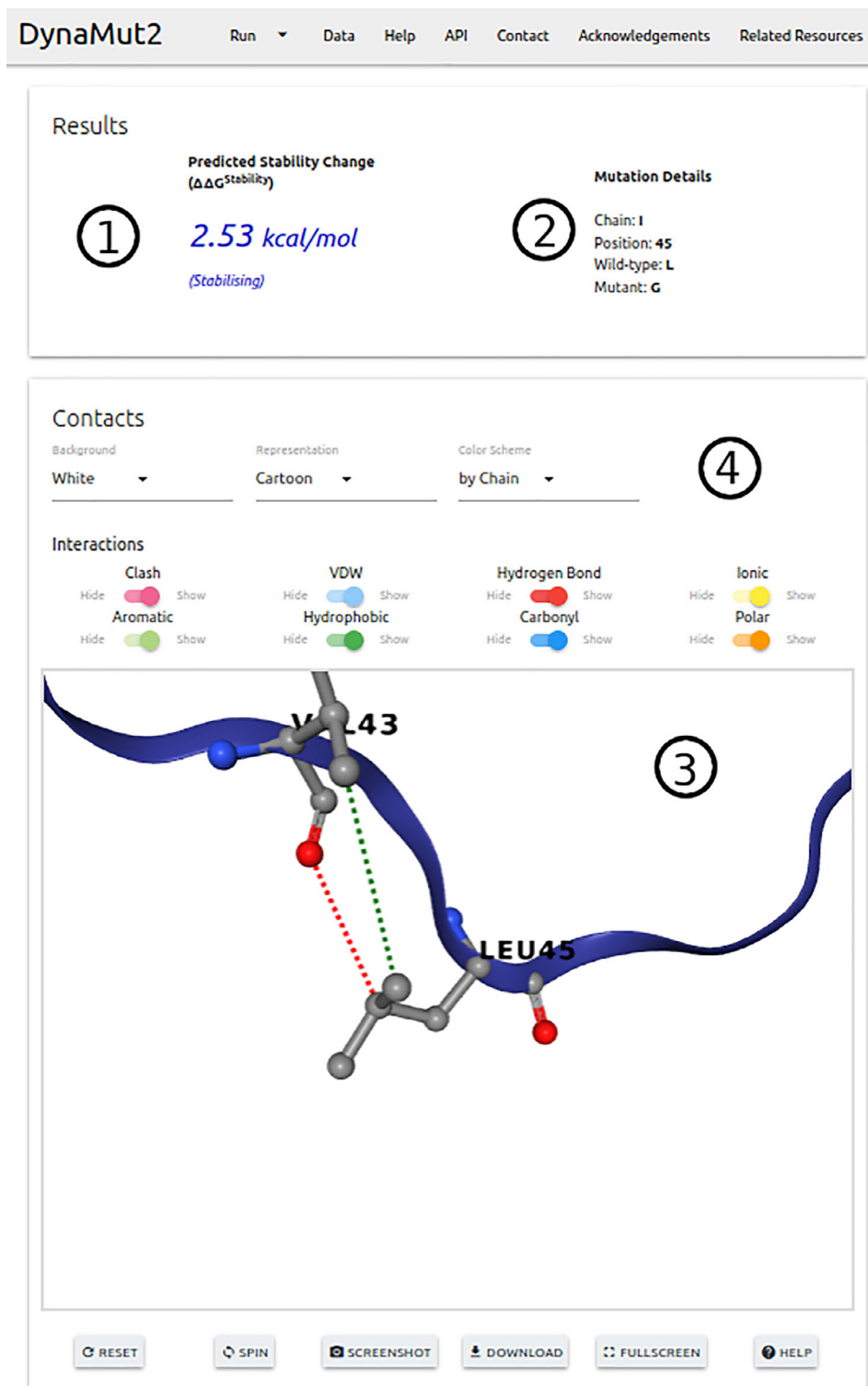
FIGURE 4 DynaMut2 results page. The figure depicts the prediction results page for single-point mutations. The predicted effect of a mutation in stability and dynamics is given as the variation in Gibbs Free Energy (in Kcal/mol) (1), together with complementary information about the mutation provided (2). Users can inspect the wild-type residue environment via an interactive viewer (3), which also allows the visualization of non-covalent interactions established by the mutated residue

shown as a table. An interactive viewer allowing for the analysis of residue contacts is also available.

For NMA submissions, the results are displayed on three panels. The first two provide information on trajectory representation of the molecule motion and porcupine plots, summarizing vector field representation, for the first non-trivial modes. Finally, the last panel displays a residue correlation matrix and structural representations using all modes.

2.6 | API

DynaMut2 conveniently offers an API (Application Programming Interface) to assist users in integrating our predictive tool into their research pipelines. All jobs submitted to DynaMut2 are labeled with a unique identifier, which is used to query the status of the job. Input fields follow the same rules of our website implementation. A full description of these with examples using curl and Python are available at http://biosig.unimelb.edu.au/dynamut2/api.

2.7 | Processing time

Finally, we compared the performance of DynaMut2 with our previous implementation, DynaMut, in terms of processing time for single-point mutations on six different protein structures. For each structure, we submitted a single-point mutation to each server and computed the processing time in seconds. This procedure was repeated 10 times for each mutation and the results are summarized in Table S13 and Figure S2. Clearly, the greater the number of residues comprising the protein structure the longer it takes for our NMA based methods to generate predictions. However, DynaMut2 runs much faster than DynaMut on all sets of experiments with very little differences in each repetition and an improvement of more than six times in the worst-case scenario.

3 | CONCLUSION

Here we present DynaMut2, a tool that incorporates information on protein dynamics and structural environment properties of wild-type residue with our graph-based signatures approach to provide an accurate prediction of mutation effects on stability and dynamics for single and multiple point mutations. Our updated server has shown to outperform other methods on predicting changes in stability caused by single point mutations and also comparable results for when used for estimating ΔT_{m}. In addition, our new approach was significantly

faster compared to the original DynaMut, which will be of great benefit toward large scale analysis and large structures. Finally, we have extended our method to predict the effects of multiple point mutations (double and triple mutants) and an API, which conveniently enables users to programmatically run predictions and represents a great contribution in terms of a novelty for this type of tool. We believe DynaMut2 represents an invaluable resource for the study of protein dynamics and to help understand the role of mutations in diseases. Web server and API with examples are freely available at http://biosig.unimelb.edu.au/dynamut2.

4 | MATERIALS AND METHODS

4.1 | Data set

We have collected experimental data on 2,648 single point mutations on 125 globular proteins from Protherm.[39] Of these, 2,080 are destabilizing ($\Delta\Delta G < 0.0$ kcal/mol) and 568 stabilizing ($\Delta\Delta G > 0.0$ kcal/mol) (Figure S1). To minimize the imbalanced nature of our dataset (Figure S1) and as a sanity check evidenced by other studies, here we use hypothetical reverse mutations.[36,54] However, differently from our previous implementation of DynaMut, hypothetical reverse mutations with more drastic changes on Gibbs free energy ($\Delta\Delta G < -2.0$ kcal/mol or $\Delta\Delta G > 2.0$ kcal/mol) were left out of our study due to uncertainties about the quality and biological significance of the modeled mutant. Our final dataset comprised 4,633 mutations (2,640 destabilizing and 1,993 stabilizing), which were split into 4,022 entries (S4022) for training our predictive model and a non-redundant test set comprising 611 entries (S611), following the protocol from our initial version of DynaMut.[38] For further performance evaluation and comparison with other methods, here we also consider a test set of 276 mutations (S276) with low sequence identity to proteins in the original ProTherm dataset, and an independent test set comprising 173 variants (S173) in six proteins with experimental melting temperature changes available (ΔT_{m}). The latter includes the structure of guanylate kinase (GK) obtained through homology modeling with Modeller[55] using the mouse GK as a template (PDB: 1LVG), similarly to previous works.[56,57]

For the data on multiple point mutations, we were able to extract 1,323 entries from ProTherm; however, since the majority of entries were double and triple mutants (Figure S3) and for the sake of simplicity, here we only considered those two types. Our final dataset comprised 1,098 entries (710 destabilizing and

388 stabilizing) (Figure S4), which were randomly split into train and test sets comprising 872 and 227 entries, respectively.

In this study, we prioritize the use of biological assembly structures author assigned, if not available, for structures generated using NMR for instance, the asymmetric unit was considered. All data used in this study is freely available for download at http://biosig.unimelb.edu.au/dynamut2/data.

4.2 | Normal model analysis

NMA provides a valuable approach for the study of dynamics and accessible conformations in a system as an alternative to time-consuming and computationally expensive Molecular Dynamics simulations. Similarly with our previous work, here we incorporated dynamics properties extracted from the protein structure generated with the module NMA of the bio3D tool.[58]

4.3 | Graph-based signatures

Our in-house graph-based signatures approach to represent molecular structures[35,59–61] has proven to be successful for a range of applications toward the study of protein structure and changes carried out by missense mutations,[35,37,41–48] including phenotypic changes.[16,62,63] These signatures comprise physicochemical and geometrical properties from the wild-type environment based on distance patterns mined from the 3D structure by representing atoms as nodes and their interactions as edges. Physicochemical properties are then defined based upon the amino acid properties, namely pharmacophore, and distance patterns between atoms are summarized as cumulative distribution functions.

4.4 | Analysis of mutation effects

Changes in Gibbs Free energy of folding can occur due to a myriad of factors related and in order to incorporate these properties, we used Arpeggio[40] to calculate the number of hydrophobic contacts involving the wild-type residue and contact potential scores from AAINDEX database.[64]

4.5 | Machine learning

In this study, we used the implementation of the Random Forest algorithm available on the scikit-learn Python library for both the prediction of $\Delta\Delta G$ for single and multiple mutations. In order to avoid the curse of dimensionality and improve performance, we selected our features using an incremental stepwise greedy approach.

5 | GENERAL STATEMENT

Small changes in proteins can have large phenotypic outcomes. By considering the changes of mutations within the context of the protein 3D structure, we have been able to accurately predict the molecular consequences of single and multiple point mutations on protein folding, stability, and dynamics. We have made this tool available through an easy to use website and API.

AUTHOR CONTRIBUTIONS
Carlos Rodrigues: Data curation; formal analysis; investigation; methodology; validation; visualization; writing-original draft. **Douglas Pires:** Formal analysis; methodology; writing-review and editing. **David Ascher:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; supervision; validation; writing-original draft; writing-review and editing.

CONFLICT OF INTEREST
No conflict of interest declared.

ORCID
David B. Ascher https://orcid.org/0000-0003-2948-2413

REFERENCES
1. Byrne JA, Strautnieks SS, Ihrke G, et al. Missense mutations and single nucleotide polymorphisms in ABCB11 impair bile salt export pump processing and function or disrupt pre-messenger RNA splicing. Hepatology. 2009;49:553–567.
2. Pires DE, Chen J, Blundell TL, Ascher DB. In silico functional dissection of saturation mutagenesis: Interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. Sci Rep. 2016;6:19848.

3. Jafri M, Wake NC, Ascher DB, et al. Germline mutations in the CDKN2B tumor suppressor gene predispose to renal cell carcinoma. Cancer Discov. 2015;5:723–729.

4. Usher JL, Ascher DB, Pires DE, Milan AM, Blundell TL, Ranganath LR. Analysis of HGD gene mutations in patients with alkaptonuria from the United Kingdom: Identification of novel mutations. JIMD Rep. 2015;24:3–11.

5. Nemethova M, Radvanszky J, Kadasi L, et al. Twelve novel HGD gene variants identified in 99 alkaptonuria patients: Focus on 'black bone disease' in Italy. Eur J Hum Genet. 2016;24:66–72.

6. Casey RT, Ascher DB, Rattenberry E, et al. SDHA related tumorigenesis: A new case series and literature review for variant interpretation and pathogenicity. Mol Genet Genomic Med. 2017;5:237–250.

7. Soardi FC, Machado-Silva A, Linhares ND, et al. Familial STAG2 germline mutation defines a new human cohesinopathy. NPJ Genom Med. 2017;2:7.

8. Hawkey J, Ascher DB, Judd LM, et al. Evolution of carbapenem resistance in Acinetobacter baumannii during a prolonged infection. Microb Genom. 2018;4:e000165.

9. Hnizda A, Fabry M, Moriyama T, et al. Relapsed acute lymphoblastic leukemia-specific mutations in NT5C2 cluster into hotspots driving intersubunit stimulation. Leukemia. 2018;32:1393–1403.

10. Holt KE, McAdam P, Thai PVK, et al. Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. Nat Genet. 2018;50:849–856.

11. Karmakar M, Globan M, Fyfe JAM, et al. Analysis of a novel pncA mutation for susceptibility to pyrazinamide yherapy. Am J Respir Crit Care Med. 2018;198:541–544.

12. Portelli S, Phelan JE, Ascher DB, Clark TG, Furnham N. Understanding molecular consequences of putative drug resistant mutations in *Mycobacterium tuberculosis*. Sci Rep. 2018;8:15356.

13. Vedithi SC, Malhotra S, Das M, et al. Structural implications of mutations conferring rifampin resistance in mycobacterium leprae. Sci Rep. 2018;8:5016.

14. Ascher DB, Spiga O, Sekelska M, et al. Homogentisate 1,2-dioxygenase (HGD) gene variants, their analysis and genotype-phenotype correlations in the largest cohort of patients with AKU. Eur J Hum Genet. 2019;27:888–902.

15. Hildebrand JM, Kauppi M, Majewski IJ, et al. A missense mutation in the MLKL brace region promotes lethal neonatal inflammation and hematopoietic dysfunction. Nat Commun. 2020;11:3150.

16. Karmakar M, Rodrigues CHM, Horan K, Denholm JT, Ascher DB. Structure guided prediction of pyrazinamide resistance mutations in pncA. Sci Rep. 2020;10:1875.

17. Vedithi SC, Rodrigues CHM, Portelli S, et al. Computational saturation mutagenesis to predict structural consequences of systematic mutations in the beta subunit of RNA polymerase in mycobacterium leprae. Comput Struct Biotechnol J. 2020;18:271–286.

18. Andrews KA, Ascher DB, Pires DEV, et al. Tumour risks and genotype-phenotype correlations associated with germline variants in succinate dehydrogenase subunit genes SDHB, SDHC and SDHD. J Med Genet. 2018;55:384–394.

19. Trezza A, Bernini A, Langella A, et al. A computational approach from gene to structure analysis of the human ABCA4 transporter involved in genetic retinal diseases. Invest Ophthalmol Vis Sci. 2017;58:5320–5328.

20. Ascher DB, Wielens J, Nero TL, Doughty L, Morton CJ, Parker MW. Potent hepatitis C inhibitors bind directly to NS5A and reduce its affinity for RNA. Sci Rep. 2014;4:4765.

21. Phelan J, Coll F, McNerney R, et al. *Mycobacterium tuberculosis* whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. BMC Med. 2016;14:31.

22. Silvino AC, Costa GL, Araujo FC, et al. Variation in human cytochrome P-450 drug-metabolism genes: A gateway to the understanding of *Plasmodium vivax* relapses. PLoS One. 2016;11:e0160172.

23. Raghunathan G, Sokalingam S, Soundrarajan N, Madan B, Munussami G, Lee SG. Modulation of protein stability and aggregation properties by surface charge engineering. Mol Biosyst. 2013;9:2379–2389.

24. Coelho MB, Ascher DB, Gooding C, et al. Functional interactions between polypyrimidine tract binding protein and PRI peptide ligand containing proteins. Biochem Soc Trans. 2016;44:1058–1065.

25. Karpiyevich M, Adjalley S, Mol M, et al. Nedd8 hydrolysis by UCH proteases in plasmodium parasites. PLoS Pathog. 2019;15:e1008086.

26. Singh V, Donini S, Pacitto A, et al. The inosine monophosphate dehydrogenase, GuaB2, is a vulnerable new bactericidal drug target for tuberculosis. ACS Infect Dis. 2017;3:5–17.

27. Singh V, Pacitto A, Donini S, et al. Synthesis and structure-activity relationship of 1-(5-isoquinolinesulfonyl)piperazine analogues as inhibitors of *Mycobacterium tuberculosis* IMPDH. Eur J Med Chem. 2019;174:309–329.

28. White RR, Ponsford AH, Weekes MP, et al. Ubiquitin-dependent modification of skeletal muscle by the parasitic nematode, Trichinella spiralis. PLoS Pathog. 2016;12:e1005977.

29. Park Y, Pacitto A, Bayliss T, et al. Essential but not vulnerable: Indazole sulfonamides targeting inosine monophosphate dehydrogenase as potential leads against *Mycobacterium tuberculosis*. ACS Infect Dis. 2017;3:18–33.

30. Portelli S, Olshansky M, Rodrigues CH, et al. Exploring the structural distribution of genetic variation in SARS-CoV-2 with the COVID-3D online resource. Nat Genet. 2020; in press.

31. Kim DH, Kim MS. Hydrogenases for biological hydrogen production. Bioresour Technol. 2011;102:8423–8431.

32. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005;33:W306–W310.

33. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812–3814.

34. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–249.

35. Pires DE, Ascher DB, Blundell TL. mCSM: Predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics. 2014;30:335–342.

36. Pandurangan AP, Ochoa-Montano B, Ascher DB, Blundell TL. SDM: A server for predicting effects of mutations on protein stability. Nucleic Acids Res. 2017;45:W229–W235.

37. Pires DE, Ascher DB, Blundell TL. DUET: A server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res. 2014; 42:W314–W319.

38. Rodrigues CH, Pires DE, Ascher DB. DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res. 2018;46:W350–W355.

39. Kumar MD, Bava KA, Gromiha MM, et al. ProTherm and ProNIT: Thermodynamic databases for proteins and protein-nucleic acid interactions. Nucleic Acids Res. 2006;34:D204–D206.

40. Jubb HC, Higueruelo AP, Ochoa-Montano B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: A web server for calculating and visualising interatomic interactions in protein structures. J Mol Biol. 2017;429:365–371.

41. Pires DE, Ascher DB. mCSM-AB: A web server for predicting antibody-antigen affinity changes upon mutation with graph-based signatures. Nucleic Acids Res. 2016;44:W469–W473.

42. Pires DE, Ascher DB. CSM-lig: A web server for assessing and comparing protein-small molecule affinities. Nucleic Acids Res. 2016;44:W557–W561.

43. Pires DE, Blundell TL, Ascher DB. mCSM-lig: Quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. Sci Rep. 2016;6:29575.

44. Pires DEV, Ascher DB. mCSM-NA: Predicting the effects of mutations on protein-nucleic acids interactions. Nucleic Acids Res. 2017;45:W241–W246.

45. Myung Y, Pires DEV, Ascher DB. mmCSM-AB: Guiding rational antibody engineering through multiple point mutations. Nucleic Acids Res. 2020;48:W125–W131.

46. Myung Y, Rodrigues CHM, Ascher DB, Pires DEV. mCSM-AB2: Guiding rational antibody design using graph-based signatures. Bioinformatics. 2020;36:1453–1459.

47. Pires DEV, Rodrigues CHM, Ascher DB. mCSM-membrane: Predicting the effects of mutations on transmembrane proteins. Nucleic Acids Res. 2020;48:W147–W153.

48. Rodrigues CHM, Myung Y, Pires DEV, Ascher DB. mCSM-PPI2: Predicting the effects of mutations on protein-protein interactions. Nucleic Acids Res. 2019;47:W338–W344.

49. Caldararu O, Mehra R, Blundell TL, Kepp KP. Systematic investigation of the data set dependency of protein stability predictors. J Chem Inf Model. 2020. https://doi.org/10.1021/acs.jcim.0c00591.

50. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins. 2006;62:1125–1132.

51. Watson MD, Monroe J, Raleigh DP. Size-dependent relationships between protein stability and thermal unfolding temperature have important implications for analysis of protein energetics and high-throughput assays of protein-ligand interactions. J Phys Chem B. 2018;122:5278–5285.

52. Montanucci L, Capriotti E, Frank Y, Ben-Tal N, Fariselli P. DDGun: An untrained method for the prediction of protein stability changes upon single and multiple point variations. BMC Bioinformatics. 2019;20:335.

53. Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlic A, Rose PW. NGL viewer: Web-based molecular graphics for large complexes. Bioinformatics. 2018;34:3755–3758.

54. Thiltgen G, Goldstein RA. Assessing predictors of changes in protein stability upon mutation using self-consistency. PLoS One. 2012;7:e46084.

55. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234:779–815.

56. Sekulic N, Shuvalova L, Spangenberg O, Konrad M, Lavie A. Structural characterization of the closed conformation of mouse guanylate kinase. J Biol Chem. 2002;277:30236–30243.

57. Cao H, Wang J, He L, Qi Y, Zhang JZ. DeepDDG: Predicting the stability change of protein point mutations using neural networks. J Chem Inf Model. 2019;59:1508–1514.

58. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: An R package for the comparative analysis of protein structures. Bioinformatics. 2006;22:2695–2696.

59. Pires DE, Blundell TL, Ascher DB. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem. 2015;58:4066–4072.

60. Kaminskas LM, Pires DEV, Ascher DB. dendPoint: A web resource for dendrimer pharmacokinetics investigation and prediction. Sci Rep. 2019;9:15465.

61. Pires DEV, Ascher DB. mycoCSM: Using graph-based signatures to identify safe potent hits against mycobacteria. J Chem Inf Model. 2020;60:3450–3456.

62. Rodrigues CH, Ascher DB, Pires DE. Kinact: A computational approach for predicting activating missense mutations in protein kinases. Nucleic Acids Res. 2018;46:W127–W132.

63. Karmakar M, Rodrigues CHM, Holt KE, Dunstan SJ, Denholm J, Ascher DB. Empirical ways to identify novel Bedaquiline resistance mutations in AtpE. PLoS One. 2019;14: e0217169.

64. Kawashima S, Kanehisa M. AAindex: amino acid index database. Nucleic Acids Res. 2000;28:374.

SUPPORTING INFORMATION

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