Structural bioinformatics

Rhapsody: predicting the pathogenicity of human missense variants

Luca Ponzoni 1,*, Daniel A. Peñaherrera 1, Zoltán N. Oltvai 1,2,3 and Ivet Bahar 1,1*  

1Department of Computational and Systems Biology, 2Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15261, USA and 3Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA

*To whom correspondence should be addressed.

Received on August 6, 2019; revised on December 27, 2019; editorial decision on February 19, 2020; accepted on February 21, 2020

Abstract

Motivation: The biological effects of human missense variants have been studied experimentally for decades but predicting their effects in clinical molecular diagnostics remains challenging. Available computational tools are usually based on the analysis of sequence conservation and structural properties of the mutant protein. We recently introduced a new machine learning method that demonstrated for the first time the significance of protein dynamics in determining the pathogenicity of missense variants.

Results: Here, we present a new interface (Rhapsody) that enables fully automated assessment of pathogenicity, incorporating both sequence coevolution data and structure- and dynamics-based features. Benchmarked against a dataset of about 20 000 annotated variants, the methodology is shown to outperform well-established and/or advanced prediction tools. We illustrate the utility of Rhapsody by in silico saturation mutagenesis studies of human H-Ras, phosphatase and tensin homolog and thiopurine S-methyltransferase.

Availability and implementation: The new tool is available both as an online webserver at http://rhapsody.csb.pitt.edu and as an open-source Python package (GitHub repository: https://github.com/prody/rhapsody; PyPI package installation: pip install prody-rhapsody). Links to additional resources, tutorials and package documentation are provided in the ‘Python package’ section of the website.

Contact: bahar@pitt.edu or lponzoni@pitt.edu or lponzoni@keiserlab.org

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Single-nucleotide polymorphisms (SNPs) are single DNA base pair changes that are inherited (germline variants) or occur during the organism’s lifetime (somatic variants). A SNP located in a coding region of the DNA may lead to the translation of the gene codon into a different amino acid than the wild-type (non-synonymous SNPs), giving rise to a single amino acid variant (SAV or missense variant). Both synonymous and non-synonymous SNPs can perturb the normal activity of a cell. For example, synonymous SNPs can affect splicing, regulatory mechanisms and gene and/or protein expression levels although they do not affect the encoded protein’s sequence. SAVs can additionally have molecular effects, e.g. by altering a protein’s orthosteric or allosteric sites, its interaction with substrates or its stability.

More than half of the mutations implicated in human inherited diseases are estimated to be associated with SAVs (Stenson et al., 2017). As a result, devising analytical and computational approaches for predicting their effect has been of broad interest, but equally challenging due to complex effects in the cell. In recent years, it became evident that comprehensive approaches integrating multiple perspectives are the only viable solutions to achieve higher accuracy in pathogenicity predictions and to interpret experimental data at the molecular level. In the case of SAVs, this means understanding not only the significance of the mutated amino acid vis-à-vis the biological function of the protein, often captured by sequence-based conservation models, but also its importance for the fold stability and conformational mechanics and interactions, both intra- and intermolecular (Ancien et al., 2018).

Significant progress has been made in tools that focus on protein sequence conservation and residue coevolution, such as context-dependent modeling of sequence evolution (Feinauer and Weigt, 2017; Hopf et al., 2017) in recent years. In contrast, structure-based modeling approaches have been lagging behind compared to sequence-based approaches in evaluating the effect of SAVs, even though the first-generation classifiers that take account of 3D structures have shown considerable success (Adzhubei et al., 2010; Ancien et al., 2018; Capriotti and Altman, 2011). The importance of considering structure, or solvent accessibility, especially when relatively few homologs are available, has been pointed out in early studies (Saunders and Baker, 2002) and in more recent works based on residue network analysis (Brown et al., 2017; Brown and Tastan...
We illustrate the utility of Rhapsody by way of applications to human H-Ras, a highly conserved G-protein belonging to Ras subfamily of small GTases for which deep mutational scanning data have been recently reported (Bandaru et al., 2017), and to two human proteins featured in a recent Critical Assessment of Genome Interpretation (CAGI) competition (Andreotti et al., 2019): PIP3 phosphatase, also called phosphatase and tensin homolog (PTEN) and thiopurine S-methyltransferase (TPMT). The new tool provides not only an efficient independent assessment of potential pathogenic effect of mutations, but also mechanistic insights into the molecular basis of the observed and/or predicted effects.

2 Materials and methods

2.1 Development of an upgraded dynamics-based pathogenicity predictor

Three groups of features, sequence-based (SEQ), structure-based (STR) and dynamics-based (DYN), computed for each position along the sequence and/or specific amino acid substitution (e.g. '2011 T 10 G A'), in UniProt coordinate, a indexing variant GI (other GTase H-Ras), are used for training a random forest classifier, following the approach described in our earlier work (Ponzoni and Bahar, 2018). In the original version of the algorithm, SEQ features were computed by the PolyPhen-2 server (Adzhubei et al., 2010), STR features by using structural data from the PDB and DYN features by the PolyDy (Bakan et al., 2011). This classifier proved to achieve accuracy levels comparable to, if not better, than 11 existing tools (Ponzoni and Bahar, 2018).

In this study, we introduce two upgraded versions, referred to as 'reduced' and 'full' Rhapsody classifiers. Supplemental Table S1 provides a detailed list of the features used in both versions along with their definition and interpretation. The reduced version includes BLOSUM62 amino acid substitution scores (Henikoff and Henikoff, 1992) as an additional feature and upgraded DYN features calculations (Fig. 1A and B). The full Rhapsody classifier uses as additional features the mutation site entropy and coevolution properties deduced from Pfam domains (El-Gebali et al., 2019).

We also designed a new interface (http://rhapsody.csb.pitt.edu) that enables efficient use of the algorithm and visualization of its output. A detailed description of random forest features and hyper-parameter optimization, Python package implementation and interface design is presented in Supplementary Materials and Methods.

2.2 Construction of an integrated dataset of annotated human variants

The dataset for training the algorithm has been generated by combining five publicly available datasets [HumVar (Adzhubei et al., 2010), ExoVar (Li et al., 2013), PredictSNP (Bendil et al., 2014), VariBench (Thuesberg et al., 2011) and SwissVar (Mottaz et al., 2010)] with the Humsavar DB of all human missense variants annotated in the UniProtKB/SwissProt DB and the ClinVar archive of reports on the level of concordance between human variations and phenotypes (Landrum et al., 2016). Supplementary Table S2 provides information on the content of these datasets and their level of agreement. After filtering out discordant labels, we obtained an 'Integrated Dataset' (IDS) of 87,726 SAVs, of which 27,655 could be mapped onto PDB structures, a prerequisite for computing STR/DYN features, and 23,085 had PDB structures with at least 150 residues.

The ClinVar DB provides a reliability level for each variant, with the help of zero (weak) to four (best) 'review stars' assigned to each SAV, based on the number of, and consensus between, various sources. Variants with 'no assertion' or 'no assertion criteria provided' are assigned 0-star; those characterized by 'single submitter' or 'conflicting interpretations' are assigned 1-star; a 2-star assignment refers to 'no conflicts and multiple submitters'; 3-star, to 'reviewed by experts' and 4-star, to 'practice guideline'. As will be shown in Section 3, removal of the 0-star cases led to improved prediction accuracy. The final, optimized integrated dataset (OPTIDS) after
eliminating these low-confidence cases contains 20,361 SAVs with at least 1 ClinVar review star, mapped onto 2828 unique chains in the PDB, each containing at least 150 residues.

3 Results

3.1 Cross-validation and comparison with other tools

In a preliminary analysis (Fig. 1C), we monitored the average area under the ROC curve (AUROC) attained by the full classifier in a 10-fold cross-validation procedure while gradually excluding from the IDS those SAVs with lower ClinVar rating. The exclusion of SAVs with 0-stars helped improve the accuracy (blue curve in Fig. 1C). This was followed by a plateau or minimal decrease in accuracy when further excluding 1-, 2-, 3- and 4-star SAVs. These additional changes were within the error bars computed from 10 cross-validation iterations, so we opted to exclude SAVs with 0-stars only, which accounted for \(\sim12\%\) of cases, from our training dataset in all subsequent analyses.

This OPTIDS was used for evaluating the accuracy of the classifier through cross-validation. In Figure 1B, we compare the performances of three variants of Rhapsody against PolyPhen-2 (Adzhubei et al., 2010) and EVmutation (Hopf et al., 2017). The colored bars represent accuracy measurements for each method’s predictions. For the three Rhapsody variants on the left, we calculated the average AUROC and associated SDs from a 10-fold cross-validation on OPTIDS, while for PolyPhen-2 and EVmutation, we plotted the AUROC values over the same dataset of variants. The light green bars in the background indicate the actual number of SAVs that could be evaluated by each approach. The cross-validation for Rhapsody classifiers has been carried out through random partitioning of OPTIDS, stratified by mutation classes to ensure equivalent bias in each fold (gray bars in Fig. 1B). Additional low-redundancy measurements have been performed as more stringent tests, by removing the variants of the same residue (‘residue-stratification’, orange bars) or within the same protein (‘protein-stratification’, blue bars) from the training subsets. Each of these steps resulted in lower estimates of accuracy, by up to \(\sim0.03\).

We notice that the full Rhapsody classifier outperforms both PolyPhen-2 and EVmutation, based on AUROC values. Similar conclusions could be drawn by evaluating the performance of these methods with other metrics, such as the Matthews correlation coefficient (MCC) and F1-score, as presented in Supplementary Figure S1. The latter are known to be less affected by high class imbalance (bias toward deleterious mutations in OPTIDS), and therefore may provide a better estimate of accuracy. Note that about 70% of our training dataset consists of deleterious variants while an opposite composition bias is observed in naturally-occurring human variants (Lek et al., 2016). To mitigate the effect of such imbalances, the random forest models have been trained by assigning to training examples weights inversely proportional to class frequency.

The full Rhapsody classifier is also seen to outperform the reduced version, although within the error margins defined by the metrics’ SD. Further comparison with the original version introduced in 2018 (Ponzoni and Bahar, 2018), presented in Figure 2C, shows the statistically significant improvement achieved in the full version, using two different ENMs, the Gaussian Network Model (GNM) (Li et al., 2016) and the Anisotropic Network Model (ANM) (Eyal et al., 2015), for evaluating DYN properties. However, the introduction of Pfam-derived features in the full classifier comes at the cost of a slight decrease in coverage, since Pfam domains often do not encompass the full span of a protein sequence, but only those portions that are preserved across species. In this regard, PolyPhen-2 has the widest coverage, being able to return a prediction even for variants without a PDB structure.

In addition to the full and reduced versions of Rhapsody, we also considered a third option, designated as ‘Rhapsody + EVmut’, which incorporated the EVmutation ‘epistatic’ score AE within the feature set. This variant slightly improved upon the full classifier, but it also further reduced the coverage. Of note, the integration of EVmutation and Rhapsody leads to significantly more accurate predictions than EVmutation used alone.

In the above comparative evaluations, we note that PolyPhen-2’s training dataset partially overlaps with OPTIDS, as discussed earlier (Ponzoni and Bahar, 2018), which may lead to an overestimation of the accuracy of PolyPhen-2 (Grimm et al., 2015). More generally, it is not always possible nor feasible to account for such ‘training biases’, unless a completely novel and independent testing dataset is designed. In order to facilitate future assessments, the output from our algorithm explicitly acknowledges whenever a tested variant is also listed in the training dataset. We presented in Supplementary Figure S2 an additional comparison of the outputs from Rhapsody with those from 27 other tools currently compiled in dbNSFP, a DB of functional predictions and annotations for all potential non-synonymous single-nucleotide variants in the human genome (Liu et al., 2011, 2016). Yet, the same type of training bias may also hold...
for the precomputed outputs in dbNSFP which may preclude an objective assessment, even though an exhaustive list of metrics has been considered therein. The large discrepancies in the accuracy levels for individual classes [neutral and deleterious SAVs, indicated by suffixes ‘(0)’ and ‘(1)’, respectively] observed for all methods reflects the imbalance of the dataset and the challenges associated with it.

Finally, we carried out an additional benchmarking study against predictions from SNPs3D (Yue et al., 2006). The latter is notable among pathogenicity prediction tools because it evaluates the functional consequences of a SAV by assessing its impact on structural stability, in addition to identifying candidate genes for specific diseases and providing information on the relationships between these candidates. For this comparison, a new classifier was trained. A relatively small subset of variants in our OPTIDS was chosen as a test set, given the availability of precomputed predictions from SNPs3D, and the proteins containing those variants were excluded from the training set. The results presented in Supplementary Figure S3 show equal or better performance of Rhapsody in general over SNPs3D using a broad range of metrics, even on this particularly challenging (imbalanced) test set that included a small proportion of deleterious SAVs, strongly departing from the composition of OPTIDS.

Overall, these results confirm the usefulness of including intrinsic dynamics features in the context of functional assessment of variants, and further demonstrate the power of adopting an integrative approach that incorporates coevolution analysis into supervised learning approaches, thus taking advantage of its superior predictive power compared to single amino acid conservation properties.

### 3.3 Higher accuracy achieved with larger structures

Figure 2C illustrates the dependency of pathogenicity prediction accuracy on the minimum size of the PDB structure included in the evaluation of the STR and DYN features. More detailed results with different metrics are presented in the Supplementary Figure S5. A slight improvement in accuracy is observed when excluding structures with fewer than \(N = 150\) residues, and again when limiting the analysis to structures with at least 500 residues. Examination of the dependency of feature weights on protein size illustrated in Supplementary Figure S7 indicated that the observed pattern did not originate from differences in feature weights which remained relatively constant in the range \(N < 300\). The increased accuracy upon exclusion of small (\(N < 150\)) structures could be attributed to the fact that sequence/structure data in this range might be incomplete and not representative of the intact protein. Conversely, the relatively high accuracy in the range \(N > 500\) could reflect the more complete inclusion of physical and evolutionary interactions between sequentially distal but spatially close neighbors in the multi-domain or multi-subunit proteins.
The existence of a direct correlation between prediction accuracy and size of PDB structures, if any, is blurred by the concurrent changes in the training dataset size and composition (blue and red curves, respectively, in Figure 2D). The non-monotonic behavior of the AUROC plot in Figure 2C could thus be attributed to the changing imbalance between deleterious and neutral variants in the training dataset at different PDB size cutoffs. Such non-uniform distributions are also viewed in the breakdown of the IDS population and imbalance at various PDB chain length intervals in Supplementary Figure S6. However, the pattern observed in Figure 2C is robustly displayed by other metrics that are less susceptible to dataset imbalance, namely MCC and F1-score (Supplementary Fig. S5). Thus, we deemed it safe to use the SAVs with \( N > 150 \) for training purposes.

3.4 Application to H-Ras

3.4.1 Saturation mutagenesis analysis of human H-Ras protein

Kuriyan and coworkers recently presented results from deep mutational scanning of human H-Ras (Bandaru et al., 2017), a highly conserved signaling protein which transduces signals through a nucleotide-dependent switch between active (GTP-bound) and inactive (GDP-bound) conformations. The impact of a single mutation on the protein’s normal activity was experimentally linked to the survival of the hosting bacterial system and quantified by a ‘fitness score’ (\( \Delta E \)), under different contexts. Here, we focus on the complete ‘regulated Ras’ experimental setup, designed to include regulatory factors that might constrain Ras sequence variability and that are necessary to obtain a realistic assessment of mutants’ fitness.

Figure 3 presents the results from our so-called ‘in silico saturation mutagenesis’ analysis. The results are presented in a 20 x \( N \) heat map (Figure 3A) where the entries are color-coded by pathogenicity probability (Supplementary Materials and Methods) predicted for all 19 possible substitutions at each of the \( N = 171 \) structurally resolved sequence positions of H-Ras (UniProt sequence ID: P01112). The entries corresponding to the wild-type amino acids are in white. The map structure mirrors that of analogous maps of experimental fitness measurements (Bandaru et al., 2017).

The structure-dependent (STR) and dynamics-based (DYN) features required by Rhapsody were computed on the active, GTP-bound conformation of H-Ras (PDB ID: 6Q21, chain A). Computations repeated for the inactive state (PDB ID: 4Q21, chain A) showed that the predictions were very similar (Supplementary Figs. S8 and S9), with the main differences localized at the switches I and II (Fig. 3B). These results are consistent with the robustness of ENM results to structural details, i.e. H-Ras structural dynamics is predominantly defined by its 3D fold, which defines its inter-residue contact topology. The contact topology, in turn, determines the intrinsically accessible spectrum of motions. The impact of SAVs on collective mechanics can thus be inferred from either active or inactive state, provided that the overall fold remains unchanged.

At first glance, the heat maps in Figure 3A show an alternating pattern of blue (neutral) and red (pathogenic) vertical bands that loosely correlate with either secondary structure or surface exposure...
of residues (top strips). Such a pattern can also be discerned in the bottom panels of Figure 3A. The red curve therein shows the residue-based pathogenicity profile predicted by Rhapsody upon averaging the entries in the corresponding column of the map. Analogous profiles obtained using PolyPhen-2 (blue), EVmutation (green) and experimental fitness scores for ‘regulated-Ras’ (Bandaru et al., 2017) (AE, gray) reveal an overall agreement between computations and experiments.

Rhapsody performs better than EVmutation and PolyPhen-2 when comparing the predicted residue-averaged pathogenicities with experimental data, as can be seen in Supplementary Table S3. The table lists the Spearman’s rank-order correlations, $\rho$, between experimental and (different types of) computational data. For the ‘regulated’ case (Fig. 4A and Supplementary Fig. S10), $\rho = 0.60$ and 0.57 for Rhapsody predictions based on the inactive and active states, respectively, as opposed to $\rho = 0.52$ and 0.51 for EVmutation and PolyPhen-2. Both Rhapsody and EVmutation outperform PolyPhen-2 in predicting individual fitness scores ($|\rho| \approx 0.42$ versus 0.36). We also estimated the prediction accuracies using AUROC and AUPRC as metrics. These required a binary labeling of variants (neutral/pathogenic) that cannot be readily deduced from the distribution of experimental $\Delta E$ values, see Supplementary Figure S11. We arbitrarily set the median of the distribution as a cutoff to assign binary labels to variants (Supplementary Fig. S11). The 40th and 60th percentiles have also been considered and used to compute uncertainty bands, represented in figure by semi-transparent blue/red shades. See also Supplementary Figure S12.

A visualization of Rhapsody incorrect predictions on Ras 3D structure (Fig. 5B and C) reveals that most False Negatives are localized on the protein’s surface, while False Positives are generally found in less exposed positions. A possible explanation is that the method is inherently biased toward the identification of residues important for the fold stability or internal dynamics, while locations subjected to other kinds of constraints, e.g. allostery and interactions with other proteins and small molecules, are more difficult to evaluate with the current set of features.

3.4.2 Analysis of H-Ras variants in gnomAD

We tested our predictions on a set of human variants found in healthy individuals, as collected by the gnomAD DB (Karczewski et al., 2019). The assumption is that those substitutions seen in the 140 000 people tested (mostly normal population) are somewhat permissive. We therefore compared the distribution of predictions obtained by Rhapsody on this set of gnomAD SAVs with the corresponding fitness scores from the experimental study considered above (Bandaru et al., 2017).

The results, illustrated in Figure 6, show that the predictions for the gnomAD SAVs are skewed toward ‘neutral’ classification in both distributions, with 49 out of 82 total variants classified as ‘neutral’ or ‘probably neutral’ by our algorithm. Of note, 3 out of 4 ‘high count’ SAVs (i.e. seen in 10 or more people) are interpreted as non-pathogenic by Rhapsody, while 2 out of 4 SAVs have a fitness score $\Delta E$, as measured in the saturation mutagenesis study, significantly lower than the wild-type amino acid (when choosing the median of all values as cutoff).

3.5 Application to PTEN and TPMT variants from CAGI competition

As an additional test, we considered a dataset of over 7000 SAVs for the tumor suppressor protein PTEN and the enzyme TPMT. The pathogenicity of these proteins’ variants has been recently investigated by massively parallel sequencing (VAMP-seq), a functional assay that measures the steady-state abundance of variants in cultured human cells (Matreyek et al., 2018). The results for PTEN/TPMT datasets were featured in the fifth edition of CAGI, a series
of competitions that aim to objectively assess computational methods on blind prediction tasks (Andreoletti et al., 2019). We performed a direct comparison of our predictor with other computational methods that, although adapted for the specific challenges proposed in the competition, were tasked with providing blind predictions, without having access to the experimental results (Pejaver et al., 2019). To ensure as much as possible a similar unbiased evaluation, new Rhapsody classifiers were trained by excluding SAVs of PTEN (56 deleterious, 1 neutral) and TPMT (3 deleterious, 4 neutral) from our training dataset, as previously done for H-Ras.

We first evaluated the predictions from Rhapsody, PolyPhen-2 and EVmutation by computing their Spearman’s correlation with experimental ‘protein-abundance’ scores from VAMP-seq data (Supplementary Fig. S16). Low-abundance variants were found to be enriched in pathogenic variants and they correlated with low protein thermodynamic stability (Matreyek et al., 2018), thus abundance score has been used as a proxy for variant impact on proteins (Pejaver et al., 2019). A classification of variants into ‘abundance’ classes (‘low-abundance’, ‘possibly low-abundance’, ‘possibly WT-like’ and ‘WT-like’) was also provided (Matreyek et al., 2018), thus allowing the use of other class-based accuracy metrics, such as AUROC, MCC and F1 score. Based on these metrics, we see in Figure 7 that Rhapsody and EVmutation are distinguished by their respective higher accuracy levels on TPMT and PTEN variants, and both consistently outperform PolyPhen-2. EVmutation, however, could only provide predictions for a small fraction (≈13%) of PTEN variants.

Prediction accuracies from participants to the CAGI5 challenge, described in (Pejaver et al., 2019) (data available from CAGI website to registered users only), are also shown in aggregated form as violin plots in Figure 7. In both cases, Rhapsody, EVmutation and Polyphen-2 all fall within the range of prediction accuracies measured for CAGI predictors. In the case of TPMT, we notice that Rhapsody consistently ranks between the median of the CAGI methods and the best-performing one.

These results demonstrate the validity of Rhapsody predictions in tasks specifically designed for testing computational methods, and against tools specifically adapted for these tasks. The modest performances demonstrated by all methods, on the other hand, also highlight the need for more effective computational approaches. Systematic assessment campaigns such as CAGI constitute an invaluable platform for evaluating the progress in the field.

4 Discussion

In the present study, we presented a novel machine learning approach for evaluating the functional impact of human SAVs, and illustrated its application to H-Ras, PTEN and TPMT. In a strict sense, Rhapsody, like many other tools in the field, predicts whether a given mutation is neutral or deleterious to protein activity, whereas pathogenicity entails many other factors, including inheritance pattern, penetrance, expressivity and environment. Thus, the outcome from the tool rather indicates a potential to be pathogenic. The newly introduced interface, Rhapsody, integrates dynamical features computed from the ENM-based analyses of protein structures and attains a state-of-the-art accuracy for predicting such a potential with a relatively simple design. We also highlighted how the method can be used not only for hypothesis generation (predictions for variants of unknown significance) but also for hypothesis testing, by providing a unified framework for comparing the predictive power of new as well as more established features. For instance, we demonstrated the utility of including in our machine-learning algorithm the ENM-derived dynamics-based features, in addition to more traditional features such as sequence conservation and structural accessibility, and emphasized the need for a better integration with coevolution analysis that recently showed significant success in evaluating the effect of SAVs.

Through the analysis of saturation mutagenesis studies and other experimental and clinical data, we identified the strengths and limitations of our approach and compared it against other prediction tools. We observed a general robustness of computational predictions, especially in the identification of residue sites that are sensitive to any mutation, regardless of the specific amino acid substitution. This information can be invaluable for the study of the functional
mechanisms of proteins, especially when projected on the 3D structures. The use of structure-based properties, in combination with sequence conservation properties (reduced classifier), can be used as an alternative approach to the more sophisticated coevolutionary analysis, whenever the latter cannot be applied due to lack of suitable multiple sequence alignments. The current algorithm has been designed to be easily expandable with new features and functionalities. Structural features such as those used in the FEATURE framework for protein annotation (Halperin et al., 2008) could be incorporated in future versions for possibly enhancing the utility of Rhapsody.

The comparison with clinical and experimental data also revealed a few issues that need to be resolved in order to advance the field. Apart from the obvious shortcomings such as the imbalance of available datasets toward pathogenic variants and the often-contradictory clinical interpretations in different databases, we reported our difficulties in interpreting data from large-scale experimental studies. These studies provide unique opportunities for dramatically increasing the size of training datasets. However, there is a need for a systematic definition of what is considered as a 'pathogenic' variant, that would account for both loss-of-function and gain-of-function effects in relation to the biological role of the affected protein.

We expect future improvements to our method to address some of these shortcomings. A recent ENM study has demonstrated how the consideration of the intact structures of multimers, complexes or assemblies improves the accuracy of predicted fluctuation spectrum of residues, and predictions from that server (DynOmicS) (Li et al., 2017) could be used for evaluating context-dependent structural and dynamic properties. For example, a region that is deemed to be tolerant to mutations by virtue of its solvent-exposure in the PDB resolved structure, may become a buried site in a complex/assembly, and a substitution at that region could alter its binding properties. A recent study has demonstrated how disease-associated SAVs are likely to be located at singlet hot spots at protein–protein interfaces (Ozdemir et al., 2018). Consideration of the involvement of residues in interfacial interactions is expected to improve the prediction accuracy of current algorithms.

Another possible improvement would be the consideration of the signature dynamics of the protein family to which the investigated protein belongs, as opposed to the dynamics of the protein alone (Zhang et al., 2019). In the same way as variations in sequence among family members point to sites that can, or cannot, tolerate mutations, family-based analyses can provide deeper insights into sites whose mechanistic properties are indispensable for function or for differentiation among subfamily members. Finally, a decomposition of the mode spectrum could help extract information on high-energy localization (hot) spots emerging as peaks in high frequency modes, as well as the hinge regions between domains, where substitutions may be detrimental (Dorantes-Gilardi et al., 2018; Rodrigues et al., 2018; Sayilgan et al., 2019).

The Rhapsody algorithm is provided both as an open-source Python package (pip install prody-rhapsody) and a web tool (http://rhapsody.csb.pitt.edu). The latter has been designed as a user-friendly service that requires minimal user input or computing skills, but also allows for some customization, such as selecting or uploading a specific PDB structure. The Rhapsody webserver can be used for both obtaining predictions on a list of human SAVs (batch query) and for visualizing a complete in silico saturation mutagenesis analysis of a human sequence, akin to those presented in Figure 3 for H-Ras. Finally, the site offers tutorials, training data (OPTIDS) and precomputed features needed for reproducing all results presented here, or for analyzing new variants. The documentation also explains how to train a model on a completely different set of features and using a different training dataset, thus providing researchers with a flexible tool for analyzing personalized datasets and testing new predictors with the help of all the functionalities implemented in Rhapsody.

Acknowledgements

The authors are grateful to Dr. Reece Hart for constructive comments.

Author contributions

The project was designed by L.P. and I.B. The in silico models were generated by L.P. with input from I.B. The Rhapsody Python package, documentation, tutorials and webserver were implemented by L.P. Clinical interpretations of the variants were contributed by Z.N.O. D.A.P contributed to data retrieval and processing. The manuscript was written by L.P., Z.N.O. and I.B. All authors approved the manuscript.

Funding

This work was supported by National Institutes of Health [P41 GM103712 and P30 DA035778 to I.B.].

Conflict of Interest: none declared.

References

Abdul Samad,F. et al. (2016) A comprehensive in silico analysis on the structural and functional impact of SNPs in the congenital heart defects associated with NKX2-5 gene—a molecular dynamic simulation approach. PLoS One, 11, e0153999.

Adzhubei,I.A. et al. (2010) A method and server for predicting damaging missense mutations. Nat. Methods, 7, 248–249.
Andreolietti,G. et al. (2019) Reports from the fifth edition of CAGI: the critical assessment of genome interpretation. *Hum. Mutat.*, 40, 1197–1201.

Bahar,I. et al. (2010) Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.*, 39, 23–42.

Bakan,A. et al. (2011) ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*, 27, 1575–1577.

Bandaru,P. et al. (2017) Deconstruction of the Ras switching cycle through saturation mutagenesis. *Elife*, 6, e27810.

Bend,J. et al. (2014) PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput. Biol.*, 10, e1003440.

Brown,D.K. et al. (2017) Structure-based analysis of single nucleotide variants in the renin-angiotensinogen complex. *Glob. Heart*, 12, 121–132.

Brown,D.K. and Tastan Bishop,O. (2017) Role of structural bioinformatics in drug discovery by computational SNP analysis: analyzing variation at the protein level. *Glob. Heart*, 12, 151–161.

Capriotti,E. and Altman,R.B. (2011) Improving the prediction of Eyal,E. et al. (2015) The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics*, 31, 1487–1489.

Hu,Z. et al. (2014) A general framework for estimating the relative functional significance of missense variants. *Hum. Mutat.*, 35, 128–135.

Kumar,A. and Purohit,R. (2014) Use of long-term molecular dynamics simulation in predicting cancer associated SNPs. *PLoS Comput. Biol.*, 10, e1003318.

Landrum,M.J. et al. (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, 44, D862–D868.

LaRusch,J. et al. (2014) Mechanisms of CFTR functional variants that impair regulated bicarbonate permeation and increase risk for pancreatitis but not for cystic fibrosis. *PLoS Genet.*, 10, e1004376.

Lek,M. et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536, 285–291.

Li,H. et al. (2016) iGNM 2.0: the Gaussian network model database for biomolecular structural dynamics. *Nucleic Acids Res.*, 44, D415–D422.

Li,H. et al. (2017) DynOmics: dynamics of structural proteome and beyond. *Nucleic Acids Res.*, 45, W374–W380.

Li,M.-X. et al. (2013) Predicting Mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet.*, 9, e1003143.

Liu,X. et al. (2011) dbNSFP: a lightweight database of human non-synonymous SNPs and their functional predictions. *Hum. Mutat.*, 32, 894–899.

Liu,X. et al. (2016) dbNSFP v3.0: a one-stop database of functional predictions and annotations for human non-synonymous and splice-site SNVs. *Hum. Mutat.*, 37, 235–241.

Matreyek,K.A. et al. (2018) Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.*, 50, 874–882.

Mottaz,A. et al. (2010) Easy retrieval of single amino acid polymorphisms and phenotype information using SwissVar. *Bioinformatics*, 26, 851–852.

Ng,P.C. and Henikoff,S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, 31, 3812–3814.

Ozdemir,E.S. et al. (2018) Analysis of single amino acid variations in single hot spots of protein–protein interfaces. *Bioinformatics*, 34, 7495–8001.

Ponzoni,L. and Bahar,I. (2018) Structural dynamics is a determinant of the functional significance of missense variants. *Proc. Natl. Acad. Sci. USA*, 115, 4164–4169.

Priya Doss,C.G. et al. (2014) Integrating *in silico* prediction methods, molecular docking, and molecular dynamics simulation to predict the impact of ALK missense mutations in structural perspective. *Biomed. Res. Int.*, 2014, 1–14.

Rodrigues,C.H. et al. (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.*, 46, W350–W355.

Saunders,C.T. and Baker,D. (2002) Evaluation of structural and evolutionary contributions to deleterious mutation prediction. *J. Mol. Biol.*, 322, 891–901.

Sayilgan,J.F. et al. (2019) Protein dynamics analysis reveals that missense mutations in cancer-related genes appear frequently on hinge-neighboring residues, *Prot. Struct. Funct. Bioinform.*, 87, 512–519.

Stenson,P.D. et al. (2017) The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.*, 136, 665–677.

Thuk Berg,J. et al. (2011) Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.*, 32, 338–368.

Yue,P. et al. (2006) SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics*, 7, 166.

Zhang,S. et al. (2019) Shared signature dynamics tempered by local fluctuations enables fold adaptability and specificity. *Mol. Biol. Evol.*, 36, 2053–2068.