Published online 21 May 2018 *Nucleic Acids Research, 2018, Vol. 46, Web Server issue* ***W127–W132***
doi: 10.1093/nar/gky375

Kinact: a computational approach for predicting activating missense mutations in protein kinases

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

[1]Department of Biochemistry and Molecular Biology, Bio21 Institute, University of Melbourne, [2]Department of Biochemistry, University of Cambridge and [3]Instituto René Rachou, Fundação Oswaldo Cruz

Received January 31, 2018; Revised April 15, 2018; Editorial Decision April 26, 2018; Accepted April 28, 2018

ABSTRACT

Protein phosphorylation is tightly regulated due to its vital role in many cellular processes. While gain of function mutations leading to constitutive activation of protein kinases are known to be driver events of many cancers, the identification of these mutations has proven challenging. Here we present Kinact, a novel machine learning approach for predicting kinase activating missense mutations using information from sequence and structure. By adapting our graph-based signatures, Kinact represents both structural and sequence information, which are used as evidence to train predictive models. We show the combination of structural and sequence features significantly improved the overall accuracy compared to considering either primary or tertiary structure alone, highlighting their complementarity. Kinact achieved a precision of 87% and 94% and Area Under ROC Curve of 0.89 and 0.92 on 10-fold cross-validation, and on blind tests, respectively, outperforming well established tools ($P < 0.01$). We further show that Kinact performs equally well on homology models built using templates with sequence identity as low as 33%. Kinact is freely available as a user-friendly web server at http://biosig.unimelb.edu.au/kinact/.

INTRODUCTION

The ability of cells to recognize and correctly respond to their microenvironment is crucial for survival. In order to dynamically respond to cellular signals, fast dynamic switches are required. Protein phosphorylation is the most widespread type of post-translational modification, with over one-third of the proteins in the human proteome phosphorylated (1). The dynamic equilibrium between phosphorylation and dephosphorylation is stringently regulated, and provides a rapid mechanism to modulate protein behaviour and activity across most signalling pathways (2). Loss of control over this regulation process, through the introduction of dominant activating mutations in kinases and the consequent hyperphosphorylation of their targets can have many phenotypic consequences, including the development and metastasis of many cancers (3–7), and the development of other metabolic disorders (8).

Advances in next generation sequencing techniques are leading to the identification of a range of novel mutations, including in kinases. In the absence of experimental information, it is currently challenging to identify mutations that are likely to lead to constitutive activation of kinases. While many computational approaches have been proposed for predicting the effects of mutations that disrupt activity, these approaches have been shown to be of limited success to predict gain of function mutations, as also shown on this work, despite the important roles they play in many diseases, particularly in cancer.

To fill this gap, here we present Kinact, a machine learning-based predictive model and web server. Using our graph-based signatures, the method was tailored to accurately identify kinase activating mutations from a combination of sequence and structural information.

MATERIALS AND METHODS

Data sets

Mutations were derived from three mutational databases with experimental evidence of their functional consequence: Kindriver (9); ClinVar (10); and Ensembl (11). Kinase mutations were divided into two groups based upon the available experimental evidence: activating and non-activating mutations. The non-activating group is represented by variations that either disrupt activity (inactivating) or have no significant biological effect (neutral). The activating mutations were defined by a significant experimentally measured increase in kinase activity.

The complete data set contained 384 mutations (260 activating and 124 non-activating) distributed across 42 proteins, of which 256 (186 activating and 70 non-activating) could be mapped onto experimentally solved 3D structures. Supplementary Figures S1 and S2 of Supplementary Mate-

*To whom correspondence should be addressed. Email: douglas.pires@minas.fiocruz.br
Correspondence may also be addressed to David B. Ascher. Tel: +61 90354794; Email: david.ascher@unimelb.edu.au

rials summarises the composition and the class distribution of mutations over the data set.

The dataset of mutations with experimental structures available, which account for 256 mutations, was randomly split into training and blind test sets. The proportion of activating and non-activating mutations on training and blind test sets is similar to observed on the original dataset as an attempt to prevent bias on the final method. The training set is comprised of 179 mutations (130 activating and 49 non-activating) that were used to train Kinact under 10-fold cross validation. The remaining 77 (56 activating and 21 non-activating) were used as blind test for validating the predictive model, minimizing the risk of overfitting. In order to assess the quality of the sub sets selected for training and blind test we repeated this process 20 times and the final version of the web server was built using the predictive model with best performance. Average and standard deviation values are reported on Supplementary Materials.

In addition, 41 mutations (24 activating and 17 non-activating in 14 kinases) that did not have experimentally solved structures available, therefore were not part of the original 256 mutations, had their structure modelled using homology modelling for further evaluation of Kinact predictive performance as a blind test.

Feature engineering

The task of predicting and understanding the effects of mutations in proteins at a molecular level has been tackled by approaches using different biological features, each with their own assumptions and limitations. Protein structural and sequence features have been the two most popular categories of attributes used by these computational methods. Sequence-based features have focussed predominantly on the analysis of sequence residue conservation throughout a protein family and homologs (12) and sequence composition (13). By contrast, previous studies have used a wide range of structural features, including secondary structure, solvent accessibility and dihedral angles (14,15). Significant effort has also been employed on more computationally intensive approaches to model mutation effects from the use of force fields and energy terms, to molecular dynamics simulations (16,17).

As an alternative, the use of graph-based structural signatures have been shown to be a scalable and effective approach for modeling the residue environment, which was successfully employed to train machine learning-based methods to predict and elucidate effects of mutations on protein stability and interactions with their partner (18–26). Moreover, these have also been used to provide insights into the molecular mechanisms of mutations and how they lead to disease and disease predisposition (27–33) and drug resistance (34–41). These graph-based signatures are predominantly composed of distance patterns extracted from the wildtype residue environment, which together with a pharmacophore modelling of its components, has been shown to be an effective way to model both geometry and physicochemical composition of protein regions.

Despite these diverse range of approaches, a combination of sequence and structural information has also been proven to be valuable when predicting damaging mutations (42,43). Based on these assumptions, graph-based signatures together with complementary sequence and structural information were used to build a predictive model. This complementary information included: (a) wild-type residue environment descriptors, (b) wild-type residue interactions, (c) predicted stability changes upon mutation, (d) sequence-based predicted effects on protein function and (e) the mCSM mutation pharmacophore modelling. A total of 82 different attributes (72 structural and 10 sequence-based) were calculated for each mutation in our dataset. These were then provided as evidence to train and test supervised learning algorithms using the Weka Tool Kit (44). The attributes used on this work were categorised into six different groups and summarised in Supplementary Table S1 of Supplementary Materials.

WEB SERVER

We have implemented Kinact as a user-friendly, freely available web server (http://biosig.unimelb.edu.au/kinact/). The server front end was built using Bootstrap framework version 3.3.7, while the back-end was built in Python via the Flask framework (Version 0.12.2). It is hosted on a Linux server running Apache.

Input

The server provides two different input options for the user (Supplementary Figure S4). The 'Single mutation' option allows users to predict whether a given mutation will lead to protein kinase activation or not. This option requires the user to provide a PDB (45) file or PDB accession code of the kinase, the point mutation specified as a string containing the wild-type residue one-letter code, its corresponding residue number and the mutant residue one-letter code, and the chain identifier of the wild-type residue. The primary sequence of the kinase of interest in fasta format is also required. The 'Mutation list' option allows users to upload a list of mutations in a file for batch processing. In order to aid users to submit their jobs, sample submission entries are available on the submission page and a help page is available via the top navigation bar.

Output

For the 'Single mutation' option, as shown in Figure 1, the web server displays in the output page the prediction outcome of Kinact, the details of the user input data, such as structure of wild-type and mutant residues, and also information on the kinase group in which the submitted structure was assigned to, based on sequence similarity according to the Standard Kinase Classification Scheme (46).

In addition, Kinact provides a set of analyses to help users investigate in greater detail the impact of the mutation. All resources displayed within the analysis section, including Pymol Sessions and the Multiple Sequence Alignment in fasta format, are made available for download.

The first item in the analysis section (Supplementary Figure S5) allows users to explore the 3D structure and the inter-residue interactions established by the wild-type residue, calculated by Arpeggio (47). Below this, users can

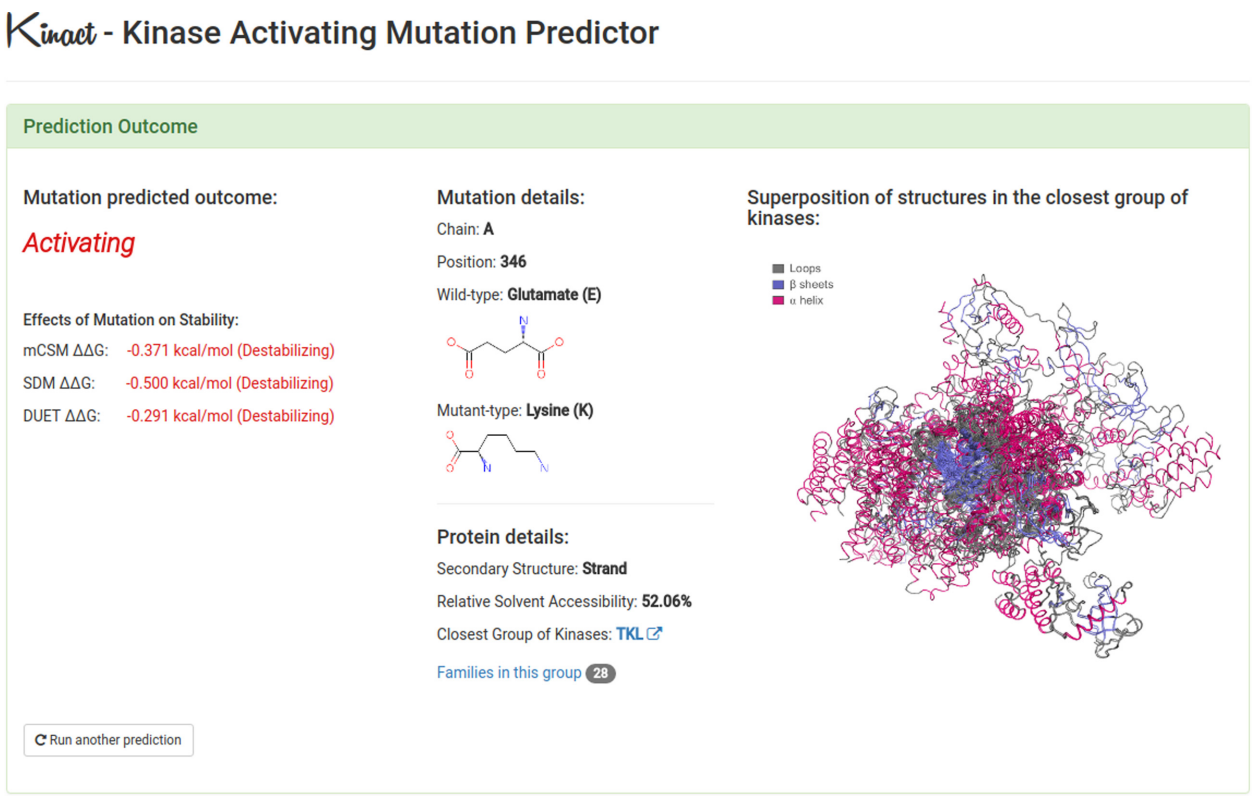

Figure 1. Web server results page for a single mutation prediction. The predicted outcome is shown alongside with complementary information on the submitted protein and the details of the mutation being evaluated. In addition, Kinact displays information on the group of homologue protein kinases according to the Standard Kinase Classification Scheme. The effects of mutation on protein stability calculated by mCSM (21), SDM (43) and DUET (25) are also shown.

also explore the conservation of residues with the structure of the wild-type kinase (Supplementary Figure S6). The 3D structure of the kinase of interest is displayed and colored according to conservation within the kinase sub-group, from red (not conserved) to blue (conserved). The structures are displayed in an interactive viewer implemented with 3Dmol.js (48).

Finally, users can also explore, within the analysis section, a multiple sequence alignment of the sequence of the provided structure and those from the closest kinase group according to the Standard Kinase Classification Scheme, assigned by similarity (Supplementary Figure S7). Previously experimentally characterised point mutations within any kinases of the group are highlighted, enabling users to rapidly identify through homology the effect of mutations at the corresponding residue position.

For the 'Mutation list' option, the server output is shown as a downloadable table (Supplementary Figure S8) and users also have the option to analyse each mutation separately, similarly to what was described for the 'Single mutation' option.

VALIDATION

In order to evaluate the quality of the training and blind test sets used we performed a resampling of these subsets 20 times and evaluated the performance of the predictive model on each split using AUC and precision. All values for the blind tests are reported on Supplementary Materials for each sample. Average and standard deviation are also shown and no bias was identified. Here we compare the performance of the best predictive model of Kinact with widely used tools to study the effects of mutations in proteins functions PolyPhen2 (42), SIFT (12) and wKinMut2 (49), a tool to identify and interpret pathogenic variants in human protein kinases.

Performance on cross validation

In order to better evaluate the contribution of structure and sequence-based attributes on the performance of supervised learning algorithms, three different predictive models were generated. The first model uses only attributes that rely on protein sequence information, which include mutation tolerance predictions (12,42), as well as a pharmacophore difference vector between wild-type and mutant residues, as proposed by the mCSM signatures (21), for this model we used the complete original dataset of 384 mutations. The second model uses only structural attributes calculated using the experimental structural data from the PDB. These include the graph-based structural signatures and complementary descriptors described in Supplementary Table S1 of Supplementary materials. Finally, the third model was constructed based on a combination of all attributes, using both sequence and structural data. For the models that used structural data on their predictions we used only the

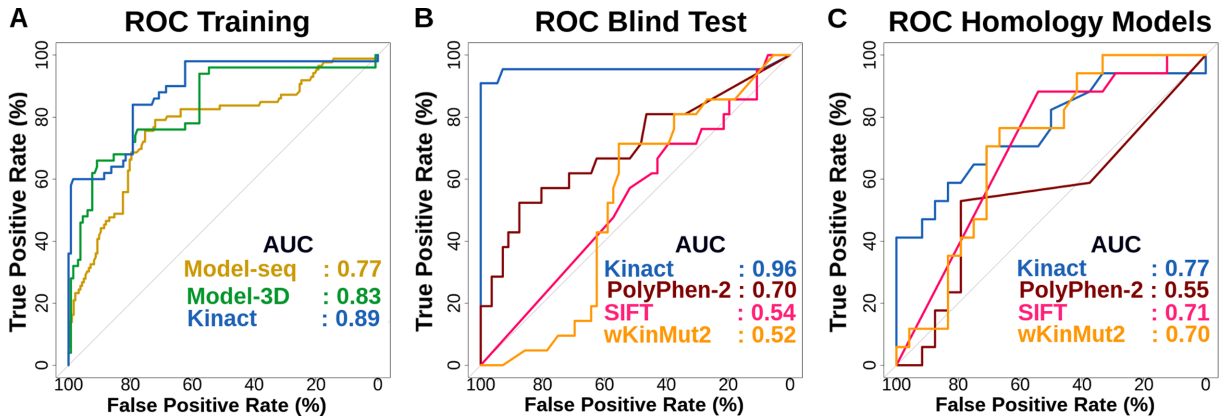

Figure 2. Comparative performance of Kinact. The ROC curves obtained for the training data set for models using sequence information alone, structural information alone, and the Kinact combined model is shown in (**A**). Kinact (AUC of 0.89), performs significantly better (P-value < 0.01) than the models using either just sequence or structural data (AUC of 0.77 and 0.83, respectively). In order to compare the performance of Kinact against the widely used tools SIFT, PolyPhen-2 and wKinMut2, a blind test (**B**) over a non-redundant test was evaluated and Kinact (AUC of 0.96) significantly (P-value < 0.01) outperformed all three methods (AUC of 0.54, 0.70 and 0.52, respectively). Using homology models (**C**), Kinact was also able to accurately identify activating mutations (AUC of 0.77), and again outperformed the other methods.

dataset of mutations with experimental structure available, which accounts for 256 mutations.

In order to run and assess the performance of the machine learning algorithms, we split each dataset into 70% of the mutations for training and 30% for blind test. In that sense, for the model that uses only sequence-based data we used 268 mutations for training (182 activating and 86 non-activating) and 116 for blind test (77 activating and 39 non-activating). For the other two models that used structure-based features 179 mutations were used for training and 77 mutations for blind test as previously described. All models were trained under 10-fold cross validation. Supplementary Figure S3 of Supplementary Materials summarises the distribution of activating and non-activating mutations in training and blind test sets for all models. Machine learning methods, evaluation procedures and performance metrics used are described in Supplementary Data.

A series of experiments were carried out to assess the performance of Kinact to predict whether a given mutation was likely to lead to constitutive activation of a kinase. The ROC curves across the training data set for models using sequence information alone, structure-based features alone, and the Kinact model that combines both attribute classes are shown in Figure 2. Details on the evaluation metrics for each algorithm are summarised on Supplementary Tables S2-S4 in Supplementary materials. Across the complete training set, Kinact achieved a Precision of 87% and Area Under ROC Curve of 0.89, significantly higher than the models using either just sequence or structural data (AUC of 0.77 and 0.83, and Precision of 0.78 and 0.81, respectively, $P < 0.01$). The final predictive models were trained using the full training set and all the performance evaluation metrics were calculated considering the average values for all 10 folds from cross validation.

Blind test

In order to properly evaluate the method's predictive performance and generalization, Kinact was initially evaluated against a separate, independent, non-redundant blind test set comprised of 77 missense mutations in protein kinases with available experimental structures, achieving a precision of 97% and Area Under ROC Curve of 0.96. When comparing with other methods, Kinact significantly outperformed (Figure 2B) all three methods (P-value < 0.01). Looking specifically at the activating mutations, SIFT predicted 55% of mutations as deleterious (score < 0.05), PolyPhen-2 classified 84% as probably damaging (score > 0.85), and wKinMut2 predicted 62% of mutations as disease related (score > 0), while Kinact correctly classified 99% of them. Comparisons of Kinact with tools that assess the effects of mutations on protein stability are described on Supplementary Materials.

Homology models

The performance of the web server to accurately classifying mutations using homology models was evaluated using a set of 41 mutations in kinases without experimentally resolved structures. Homology models of the kinases were generated by Modeller (50) using experimentally resolved structures down to 33% sequence identity. Using the homology models, Kinact was able to accurately identify activating mutations (AUC of 0.77 and precision of 0.78), providing confidence and robustness in the applicability of this approach beyond experimental structures to those that are computationally modelled. This was also significantly better than PolyPhen-2, SIFT and wKinMut2 (Figure 2C). When comparing the performance of the methods specifically at the activating mutations, Kinact was able to classify correctly 100% of mutations, while SIFT predicted 75% as deleterious (score < 0.05), PolyPhen-2 classified 83% as probably damaging (score > 0.85), and wKinMut2 predicted 77% as disease related (score > 0).

CONCLUSION

We present here, Kinact, a predictive model and web server tailored for identifying kinase activating mutations using

graph-based signatures, sequence and structural data. Kinact conveniently combines high-performance, open access, web visualization tools to assist research on how mutations affect protein kinases activity as well as prioritise mutations for further investigation. Given the importance of these variants in the context of many diseases, especially on the development of many types of cancer, and also that widely used tools have not been able to successfully predict gain of function mutations, we believe Kinact will be a useful tool to help identify and understand the role of these mutations. The method is freely available as a user friendly and easy to use web server at http://biosig.unimelb.edu.au/kinact/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Australian Government Research Training Program Scholarship [to C.H.M.R]; Jack Brockhoff Foundation [JBF 4186, 2016 to D.B.A.]; Newton Fund RCUK-CONFAP Grant awarded by the Medical Research Council (MRC) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) [MR/M026302/1 to D.B.A. and D.E.V.P.]; National Health and Medical Research Council of Australia [APP1072476 to D.B.A.]; Victorian Life Sciences Computation Initiative (VLSCI), an initiative of the Victorian Government, Australia, on its Facility hosted at the University of Melbourne [UOM0017]; Instituto René Rachou (IRR/FIOCRUZ Minas), Brazil and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [to D.E.V.P.]; Department of Biochemistry and Molecular Biology, University of Melbourne [to D.B.A.]. Funding for open access charge: Instituto René Rachou (IRR/FIOCRUZ Minas).
Conflict of interest statement. None declared.

REFERENCES

1. Cohen,P. (2002) The origins of protein phosphorylation. *Nat. Cell Biol.*, **4**, E127–E130.
2. Salazar,C. and Hofer,T. (2009) Multisite protein phosphorylation–from molecular mechanisms to kinetic models. *FEBS J.*, **276**, 3177–3198.
3. Bose,R., Kavuri,S.M., Searleman,A.C., Shen,W., Shen,D., Koboldt,D.C., Monsey,J., Goel,N., Aronson,A.B., Li,S. *et al.* (2013) Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov.*, **3**, 224–237.
4. Chirgadze,D.Y., Ascher,D.B., Blundell,T.L. and Sibanda,B.L. (2017) DNA-PKcs, allostery, and DNA double-strand break repair: defining the structure and setting the stage. *Methods Enzymol.*, **592**, 145–157.
5. Grabiner,B.C., Nardi,V., Birsoy,K., Possemato,R., Shen,K., Sinha,S., Jordan,A., Beck,A.H. and Sabatini,D.M. (2014) A diverse array of cancer-associated MTOR mutations are hyperactivating and can predict rapamycin sensitivity. *Cancer Discov.*, **4**, 554–563.
6. Sibanda,B.L., Chirgadze,D.Y., Ascher,D.B. and Blundell,T.L. (2017) DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. *Science*, **355**, 520–524.
7. Tiacci,E., Pettirossi,V., Schiavoni,G. and Falini,B. (2017) Genomics of hairy cell leukemia. *J. Clin. Oncol.*, **35**, 1002–1010.
8. Lahiry,P., Torkamani,A., Schork,N.J. and Hegele,R.A. (2010) Kinase mutations in human disease: interpreting genotype-phenotype relationships. *Nat. Rev. Genet.*, **11**, 60–74.
9. Simonetti,F.L., Tornador,C., Nabau-Moreto,N., Molina-Vila,M.A. and Marino-Buslje,C. (2014) Kin-Driver: a database of driver mutations in protein kinases. *Database (Oxford)*, **2014**, bau104.
10. Landrum,M.J., Lee,J.M., Riley,G.R., Jang,W., Rubinstein,W.S., Church,D.M. and Maglott,D.R. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.*, **42**, D980–D985.
11. Hubbard,T., Barker,D., Birney,E., Cameron,G., Chen,Y., Clark,L., Cox,T., Cuff,J., Curwen,V., Down,T. *et al.* (2002) The Ensembl genome database project. *Nucleic Acids Res.*, **30**, 38–41.
12. Ng,P.C. and Henikoff,S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, **31**, 3812–3814.
13. Capriotti,E., Fariselli,P. and Casadio,R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, **33**, W306–W310.
14. Chasman,D. and Adams,R.M. (2001) Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J. Mol. Biol.*, **307**, 683–706.
15. Guerois,R., Nielsen,J.E. and Serrano,L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, **320**, 369–387.
16. Schymkowitz,J., Borg,J., Stricher,F., Nys,R., Rousseau,F. and Serrano,L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.
17. Zimmermann,M.T., Urrutia,R., Oliver,G.R., Blackburn,P.R., Cousin,M.A., Bozeck,N.J. and Klee,E.W. (2017) Molecular modeling and molecular dynamic simulation of the effects of variants in the TGFBR2 kinase domain as a paradigm for interpretation of variants obtained by next generation sequencing. *PLoS One*, **12**, e0170822.
18. Pires,D.E. and Ascher,D.B. (2016) mCSM-AB: a web server for predicting antibody-antigen affinity changes upon mutation with graph-based signatures. *Nucleic Acids Res.*, **44**, W469–W473.
19. Pires,D.E. and Ascher,D.B. (2016) CSM-lig: a web server for assessing and comparing protein-small molecule affinities. *Nucleic Acids Res.*, **44**, W557–W561.
20. Pires,D.E. and Ascher,D.B. (2017) mCSM-NA: predicting the effects of mutations on protein-nucleic acids interactions. *Nucleic Acids Res.*, **45**, W241–W246.
21. Pires,D.E., Ascher,D.B. and Blundell,T.L. (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, **30**, 335–342.
22. Pires,D.E., Blundell,T.L. and Ascher,D.B. (2015) Platinum: a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. *Nucleic Acids Res.*, **43**, D387–D391.
23. Pires,D.E., Blundell,T.L. and Ascher,D.B. (2016) mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Sci. Rep.*, **6**, 29575.
24. Pires,D.E., Chen,J., Blundell,T.L. and Ascher,D.B. (2016) In silico functional dissection of saturation mutagenesis: Interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. *Sci. Rep.*, **6**, 19848.
25. Pires,D.E., Ascher,D.B. and Blundell,T.L. (2014) DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.*, **42**, W314–W319.
26. Rodrigues,C.H., Pires,D.E. and Ascher,D.B. (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.*, doi:10.1093/nar/gky300.
27. Casey,R.T., Ascher,D.B., Rattenberry,E., Izatt,L., Andrews,K.A., Simpson,H.L., Challis,B., Park,S.M., Bulusu,V.R., Lalloo,F. *et al.* (2017) SDHA related tumorigenesis: a new case series and literature review for variant interpretation and pathogenicity. *Mol. Genet. Genomic Med.*, **5**, 237–250.
28. Jafri,M., Wake,N.C., Ascher,D.B., Pires,D.E., Gentle,D., Morris,M.R., Rattenberry,E., Simpson,M.A., Trembath,R.C., Weber,A. *et al.* (2015) Germline mutations in the CDKN2B tumor suppressor gene predispose to renal cell carcinoma. *Cancer Discov.*, **5**, 723–729.
29. Nemethova,M., Radvanszky,J., Kadasi,L., Ascher,D.B., Pires,D.E., Blundell,T.L., Porfirio,B., Mannoni,A., Santucci,A., Milucci,L. *et al.* (2016) Twelve novel HGD gene variants identified in 99 alkaptonuria patients: focus on 'black bone disease' in Italy. *Eur. J. Hum. Genet.*, **24**, 66–72.
30. Soardi,F.C., Machado-Silva,A., Linhares,N.D., Zheng,G., Qu,Q., Pena,H.B., Martins,T.M.M., Vieira,H.G.S., Pereira,N.B.,

Melo-Minardi,R.C. *et al.* (2017) Familial STAG2 germline mutation defines a new human cohesinopathy. *NPJ Genom. Med.*, **2**, 7.
31. Trezza,A., Bernini,A., Langella,A., Ascher,D.B., Pires,D.E.V., Sodi,A., Passerini,I., Pelo,E., Rizzo,S., Niccolai,N. *et al.* (2017) A computational approach from gene to structure analysis of the human ABCA4 transporter involved in genetic retinal diseases. *Invest. Ophthalmol. Vis. Sci.*, **58**, 5320–5328.
32. Usher,J.L., Ascher,D.B., Pires,D.E., Milan,A.M., Blundell,T.L. and Ranganath,L.R. (2015) Analysis of HGD gene mutations in patients with alkaptonuria from the United Kingdom: Identification of novel mutations. *JIMD Rep.*, **24**, 3–11.
33. Hnizda,A., Fabry,M., Moriyama,T., Pachl,P., Kugler,M., Brinsa,V., Ascher,D.B., Carroll,W.L., Novak,P., Zaliova,M. *et al.* (2018) Relapsed acute lymphoblastic leukemia-specific mutations in NT5C2 cluster into hotspots driving intersubunit stimulation. *Leukemia*, **In Press**.
34. Albanaz,A.T.S., Rodrigues,C.H.M., Pires,D.E.V. and Ascher,D.B. (2017) Combating mutations in genetic disease and drug resistance: understanding molecular mechanisms to guide drug design. *Expert Opin. Drug Discov.*, **12**, 553–563.
35. Pandurangan,A.P., Ascher,D.B., Thomas,S.E. and Blundell,T.L. (2017) Genomes, structural biology and drug discovery: combating the impacts of mutations in genetic disease and antibiotic resistance. *Biochem. Soc. Trans.*, **45**, 303–311.
36. Phelan,J., Coll,F., McNerney,R., Ascher,D.B., Pires,D.E., Furnham,N., Coeck,N., Hill-Cawthorne,G.A., Nair,M.B., Mallard,K. *et al.* (2016) Mycobacterium tuberculosis whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. *BMC Med.*, **14**, 31.
37. Hawkey,J., Ascher,D.B., Judd,L.M., Wick,R.R., Kostoulias,X., Cleland,H., Spelman,D.W., Padiglione,A., Peleg,A.Y. and Holt,K.E. (2018) Evolution of carbapenem resistance in Acinetobacter baumannii during a prolonged infection. *Microb.Genomics*, **4**, e000165.
38. Vedithi,S.C., Malhotra,S., Das,M., Daniel,S., Kishore,N., George,A., Arumugam,S., Rajan,L., Ebenezer,M., Ascher,D.B. *et al.* (2018) Structural implications of mutations conferring rifampin resistance in mycobacterium leprae. *Sci. Rep.*, **8**, 5016.
39. Karmakar,M., Globan,M., Fyfe,J.A.M., Stinear,T.P., Johnson,P.D.R., Holmes,N.E., Denholm,J.T. and Ascher,D.B. (2018) Analysis of a novel pncA mutation for susceptibility to Pyrazinamide therapy. *Am. J. Respir. Crit. Care Med.*, **In Press**.
40. Singh,V., Donini,S., Pacitto,A., Sala,C., Hartkoorn,R.C., Dhar,N., Keri,G., Ascher,D.B., Mondesert,G., Vocat,A. *et al.* (2017) The inosine monophosphate dehydrogenase, GuaB2, is a vulnerable new bactericidal drug target for tuberculosis. *ACS Infect. Dis.*, **3**, 5–17.
41. Holt,K.E., McAdam,P., Thai,P.V.K., Thuong,N.T.T., Ha,D.T.M.H., Lan,N.N., Lan,N.H., Nhu,N.T.Q., Hai,H.T., Ha,V.T.N. *et al.* (2018) Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for EsxW Beijing variant in Vietnam. *Nat. Genet.*, **In Press**.
42. Adzhubei,I.A., Schmidt,S., Peshkin,L., Ramensky,V.E., Gerasimova,A., Bork,P., Kondrashov,A.S. and Sunyaev,S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
43. Pandurangan,A.P., Ochoa-Montano,B., Ascher,D.B. and Blundell,T.L. (2017) SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.*, **45**, W229–W235.
44. Hall,M., Frank,E., Holmes,G., Pfahringer,B., Reutemann,P. and Witten,I.H. (2009) The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, **11**, 10–18.
45. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
46. Manning,G., Whyte,D.B., Martinez,R., Hunter,T. and Sudarsanam,S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–1934.
47. Jubb,H.C., Higueruelo,A.P., Ochoa-Montano,B., Pitt,W.R., Ascher,D.B. and Blundell,T.L. (2017) Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J Mol Biol*, **429**, 365–371.
48. Rego,N. and Koes,D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, **31**, 1322–1324.
49. Vazquez,M., Pons,T., Brunak,S., Valencia,A. and Izarzugaza,J.M. (2016) wKinMut-2: identification and Interpretation of Pathogenic Variants in Human Protein Kinases. *Hum. Mutat.*, **37**, 36–42.
50. Webb,B. and Sali,A. (2014) Protein structure modeling with MODELLER. *Methods Mol. Biol.*, **1137**, 1–15.