Bioinformatics, 32(10), 2016, 1589–1591
doi: 10.1093/bioinformatics/btw031
Advance Access Publication Date: 21 January 2016
Applications Note

Systems biology

ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity

Daniel K. Witvliet[1,2,†], Alexey Strokach[1,3,†], Andrés Felipe Giraldo-Forero[4], Joan Teyra[1], Recep Colak[1,3] and Philip M. Kim[1,2,3,*]

[1]Donnelly Center for Cellular and Biomolecular Research, University of Toronto, [2]Department of Molecular Genetics, University of Toronto, [3]Department of Computer Science, University of Toronto, Toronto, ON, Canada and [4]Department of Electrical and Electronic Engineering, Universidad Nacional De Colombia, Manizales, Colombia

*To whom correspondence should be addressed.
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Associate Editor: Jonathan Wren

Received on 19 August 2015; revised on 31 December 2015; accepted on 17 January 2016

Abstract

Summary: ELASPIC is a novel ensemble machine-learning approach that predicts the effects of mutations on protein folding and protein–protein interactions. Here, we present the ELASPIC web-server, which makes the ELASPIC pipeline available through a fast and intuitive interface. The web-server can be used to evaluate the effect of mutations on any protein in the Uniprot database, and allows all predicted results, including modeled wild-type and mutated structures, to be managed and viewed online and downloaded if needed. It is backed by a database which contains improved structural domain definitions, and a list of curated domain–domain interactions for all known proteins, as well as homology models of domains and domain–domain interactions for the human proteome. Homology models for proteins of other organisms are calculated on the fly, and mutations are evaluated within minutes once the homology model is available.
Availability and implementation: The ELASPIC webserver is available online at http://elaspic.kimlab.org.
Contact: pm.kim@utoronto.ca or pi@kimlab.org
Supplementary data: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Sequence variation in the human genome is a crucial aspect in biomedical research. Many variants can directly cause Mendelian diseases and can confer susceptibility or protection to more multifactorial diseases, such as cancer (Risch and Merikangas, 1996). To unravel the mechanism behind those phenotypic effects we first need to understand the biochemical impact of mutations. A number of computational tools have been developed to predict the effect of mutations on protein stability or binding (Dehouck *et al.*, 2011, 2013; Tian *et al.*, 2010). However, those tools require a resolved crystal structure of the protein, precluding their application on a proteome-wide scale. There are webservers which use structural information to predict the effect of mutations (De Baets *et al.*, 2012; Venselaar *et al.*, 2010), and which map mutations to proteins in a 3D protein interaction network (Mosca et al., 2015). However, none of those resources predict the effect of mutations on protein–protein interaction affinity. Here, we present a webserver that can be used to predict the effect of mutations on protein stability and protein interaction affinity, for any protein for which a homology model can be constructed. The webserver is based on the ELASPIC pipeline (Berliner *et al.*, 2014), but drastically simplifies the experience for the end user. It includes improved domain definitions for all proteins in the UniProt database, homology models of domains and domain–domain interactions for all human proteins, and mutation

impact results for all human mutations implicated in hereditary diseases and cancers. New mutations are evaluated within minutes if homology models are available, and under an hour if those models are to be generated. The webserver offers a streamlined and intuitive interface for analyzing the results, and makes all predictions and structures available for download.

2 Methods

2.1 Domain boundary definition

In order to generate structural models, structurally sound domain boundary definitions are needed. For this, we generated the Profs pipeline that unifies domain definitions obtained from CATH and Pfam. Profs domain boundaries are optimized to be both structurally and evolutionarily consistent. Unlike Gene3D (Lees *et al.*, 2012), which uses hidden Markov models (HMMs) that are based on CATH, Profs uses HMMs that are based on Pfam. Therefore, all functional annotations that have been assigned to Pfam domains can easily be transferred to Profs domains. For more information, see the Supplementary material.

2.2 ELASPIC pipeline

ELASPIC constructs homology models of domains and domain–domain interactions, and it uses those models, together with sequential and other features, to predict the energetic impact of a mutation on the stability of a single domain or the affinity between two domains. A more detailed description of the ELASPIC method, including cross-validation and comparison with existing approaches, can be found in the Supplementary material and the original publication.

2.3 Webserver implementation

The ELASPIC webserver was implemented using the Django framework. Its database contains over 800 million protein cross-references, imported from Uniprot, allowing it to recognize protein input in different formats.

2.4 Overview

Unlike previous methods, ELASPIC takes advantage of homology modeling to create protein structures. Therefore, the webserver only requires a protein or gene identifier and a point mutation. An alternative batch input page is available, which takes a list of point mutations with the standard form <protein>.<mutation> either by direct input or as a file upload. The users can still submit their own PDB structure if required.

The main output of the webserver is the predicted change in the Gibbs free energy ($\Delta\Delta G$) of folding and/or binding for every domain and interface affected by the mutation. A specific domain or interface can be selected to display more detailed information, including a graphic showing the location of the domain in the protein, a Jmol applet showing superimposed wild-type and mutant structures (e.g. Fig. 1) and, for interface mutations, an interactive connectivity network showing the affected interactions. The structures can be downloaded for further inspection or calculations.

2.5 Case study

We applied our approach to a list of somatic mutations found in cancer and catalogued in the COSMIC database. EP300, a transcriptional coactivator, is mutated with a high frequency in different cancer types. ELASPIC predicts those mutations to decrease the stability of EP300 and decrease its affinity to downstream hypoxia-regulators HIF1A and EPAS1, and the SRC/p160 nuclear receptor coactivator family proteins NCOA1, NCOA2 and NCOA3. Interaction models created by the ELASPIC webserver show the structural changes that lead to the loss in affinity (Fig. 1). The p160 family proteins are thought to promote EP300-mediated HIF1A and EPAS1 activation (Carrero *et al.*, 2000). ELASPIC predicts that the

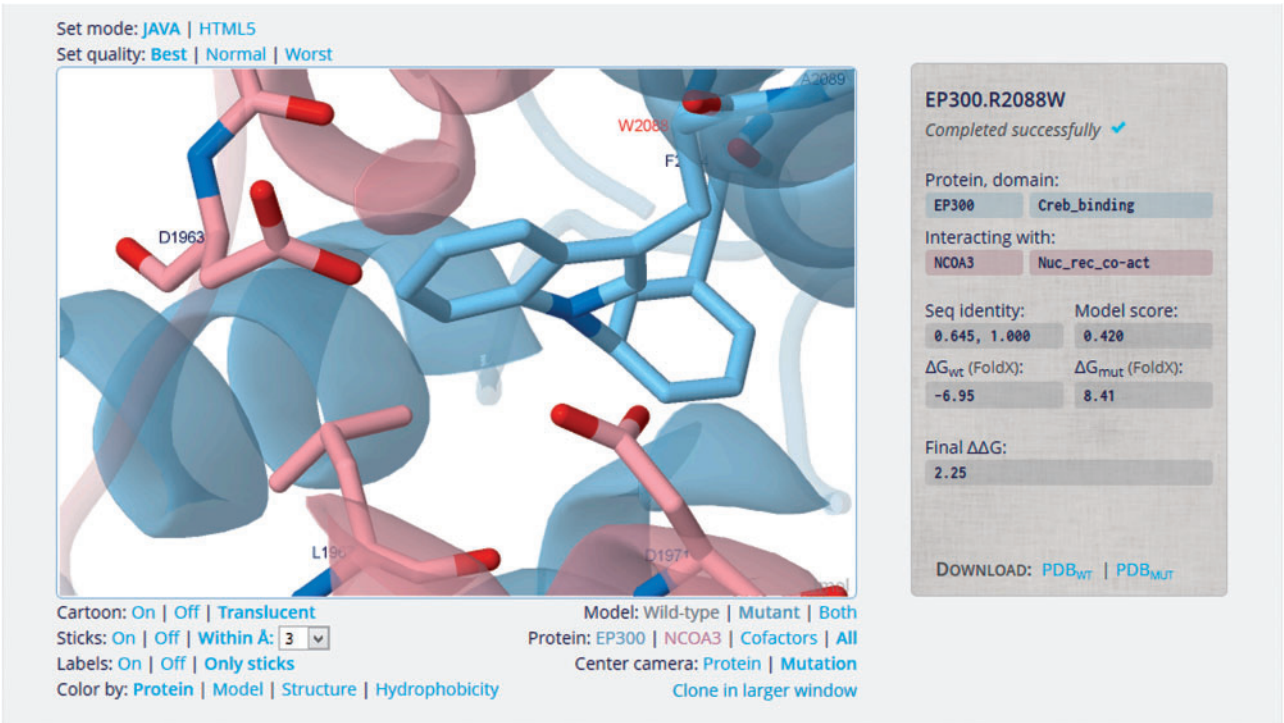

Fig. 1. Structural model of mutated interaction between EP300 (blue) and NCOA3 (red), as presented by the ELASPIC webserver. Mutation R2088W destroys an important salt bridge to aspartic acid residues of NCOA3 in the binding interface, resulting in a severely decreased binding affinity

increase in cancer susceptibility caused by mutations in EP300 is due to those interactions being hindered. Indeed, it has been shown that disrupting the interaction between EP300 and HIF1A inhibits apoptosis and promotes renal cancer cells survival (Khan *et al.*, 2011). A more in-depth analysis is available in the Supplementary material.

Acknowledgements

We thank Niklas Berliner for providing and helping to debug the original ELASPIC pipeline scripts. We thank the SciNet HPC consortium for computing time.

Funding

P.M.K. acknowledges support from a NSERC Discovery Grant (#386671).

Conflict of Interest: none declared.

References

Berliner,N. *et al.* (2014) Combining structural modeling with ensemble machine learning to accurately predict protein fold stability and binding affinity effects upon mutation. *PLoS One*, **9**, e107353.

Carrero,P. *et al.* (2000) Redox-regulated recruitment of the transcriptional coactivators CREB-binding protein and SRC-1 to hypoxia-inducible factor 1α. *Mol. Cell. Biol.*, **20**, 402–415.

De Baets,G. *et al.* (2012) SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Res.*, **40**, D935–D939.

Dehouck,Y. *et al.* (2011) PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, **12**, 151.

Dehouck,Y. *et al.* (2013) BeAtMuSiC: prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res.*, **41**, W333–W339.

Khan,M.N. *et al.* (2011) Factor inhibiting HIF (FIH-1) promotes renal cancer cell survival by protecting cells from HIF-1α-mediated apoptosis. *Br. J. Cancer*, **104**, 1151–1159.

Lees,J. *et al.* (2012) Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis. *Nucleic Acids Res.*, **40**, D465–D471.

Mosca,R. *et al.* (2015) dSysMap: exploring the edgetic role of disease mutations. *Nat. Methods*, **12**, 167–168.

Risch,N. and Merikangas,K. (1996) The future of genetic studies of complex human diseases. *Science*, **273**, 1516–1517.

Tian,J. *et al.* (2010) Predicting changes in protein thermostability brought about by single-or multi-site mutations. *BMC Bioinformatics*, **11**, 370.

Venselaar,H. *et al.* (2010) Protein structure analysis of mutations causing inheritable diseases. An e-science approach with life scientist friendly interfaces. *BMC Bioinformatics*, **11**, 548.