Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Databases and ontologies

PROXiMATE: A database of mutant protein-protein complex thermodynamics and kinetics

Sherlyn Jemimah, K. Yugandhar, M. Michael Gromiha*

Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai - 600036, India

*To whom correspondence should be addressed.

Associate Editor: Prof. Alfonso Valencia

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: We have developed PROXiMATE, a database of thermodynamic data for more than 6000 missense mutations in 174 heterodimeric protein-protein complexes, supplemented with interaction network data from STRING database, solvent accessibility, sequence, structural and functional information, experimental conditions and literature information. Additional features include complex structure visualization, search and display options, download options and a provision for users to upload their data.

Availability and Implementation: The database is freely available at http://www.iitm.ac.in/bioinfo/PROXiMATE/. The website is implemented in Python, and supports recent versions of major browsers such as IE10, Firefox, Chrome and Opera.

Contact: gromiha@iitm.ac.in

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Protein-protein interactions mediate a range of cellular functions, including metabolism, signaling and ubiquitination (Keskin*et. al.*, 2008). Missense mutations tend to alter the stability and specificity of interacting proteins as well as disrupt crucial functions, and some of them lead to diseases (Yip *et. al.*, 2004). The effect of these mutations can be assessedand quantified by their impact on the thermodynamics and kinetics of protein-protein interactions. Experimentally determined thermodynamic data is a prerequisite for designing effective algorithms for predicting the binding affinity change upon mutation in protein-protein complexes. In order to facilitate this, we present PROXiMATE (PROtein-protein compleX MutAtion ThErmodynamics), a thermodynamic database of missense mutations in heterodimeric protein-protein complexes, with structural and functional information, and links to major databases for protein sequence, structure and interaction networks. PROXiMATE can be accessed at http://www.iitm.ac.in/bioinfo/PROXiMATE/.

2 Contents

We have collected experimental thermodynamic data from literature and integrated with information available in other resources. The oligomeric state of the complexes used in experiments is not explicitly stated in the literature. However, most of the interacting proteins are expressed separately for in vitro experiments and hence the data may be considered for heterodimeric complexes. A detailed illustration of the contents is provided in **Figure 1**. We have included experimental conditions, functional classification (**Supplementary Table S1**) and binding affinity values from the primaryliterature source. The interacting proteins are linked to their respective accession numbers from UniProt (UniProt Consortium, 2015) as well as the experimentally determined complex crystal structure from PDB (Rose *et. al.*, 2017), if available. The JSmol applet (Hanson*et. al.*, 2013) provides visuals of the protein-protein complex. DSSP v2.0.4 (Kabsch and Sander, 1983) has been used to calculate relative accessibility and assign secondary structure for the wild-type residue at mutant positions. We have provided the solvent accessibility for both interacting dimers and the whole complex. The number of chains in PDB is also mentioned in the database. The interface information for each mutant is provided using a distance-based cutoff of 3.5Å (Gromiha and Yugandhar, 2017). Network data from STRING (Szklarczyk*et. al.*, 2015) assign the complexes in the context of their immediate protein-protein interaction network.We have also highlighted the specific nodes in the annotation table at the website. An example is shown in **Supplementary Figure S1**.

The database contains 6296 mutations in 174 complexes, collected from literature published in the years 1988-2016. Hotspots (i.e. those with a free energy change (ΔΔG) of more than 1.5 kcal/mol; Gao et al. 2004)

represent approximately one-third of the entries. While a majority of the entries represent single mutations and/or Ala mutations, efforts have been made to include multiple and non-Ala mutations. Also, the database includes functional classification of the complexes and data collected from diverse experimental techniques.

3 Unique Features

While there have been previous efforts to compile thermodynamic data for mutant protein-protein complexes (Thorn and Bogan, 2001; Kumar and Gromiha, 2006; Moal and Fernández-Recio, 2012; Sirin*et. al.*, 2016), these databases have several limitations. Most of them are not maintained currently (last checked on 27January 2017) and provide no search options. Some databases target only a particular class of complexes or mutations. PROXiMATE is larger than previous databases and provides structural and network information, complex structure visualization, and user-friendly search, display, download and upload options. We have also provided the exact location in a research article, where the data can be found and a mapping between Pubmed ID and entry numbers along with the name of the complex. A detailed comparison of the databases is provided in **Supplementary Table S2**.

4 Applications

PROXiMATE can provide unbiased training and validation datasets for the development of algorithms to predict binding affinity changes due to missense mutations (Dehouck*et. al.*, 2013; Pires*et. al.*, 2014; Brender and Zhang, 2015; Petukh*et. al.*, 2016; Li *et. al.*, 2016; Gromiha *et al.* 2016) in protein-protein interactions. The database will also aid the study of disease-causing mutations in the progression, diagnosis and treatment of various diseases, and provide possible drug targets and novel therapy options(Watkins and Arora, 2015). Further, PROXiMATE can provide experimental data for the identification of mutants that exhibit increased affinity to their interacting partners.The applications are described further in Supplementary Information.

Acknowledgements

We thank the members of the Protein Bioinformatics Lab for their valuable feedback. We appreciate the efforts of Ambuj Srivastava and P. Prabhakaran in uploading and testing the website. We acknowledge the Indian Institute of Technology Madras for computational facilities. We also thank the reviewers for their useful comments and suggestions.

Funding

This work has been partially supported by theMinistry of Human Resource and Development, India to SJ, and Department of Science and Technology, Government of India to MMG (EMR/2016/001476).

Conflict of Interest: none declared.

References

Brender,J.R. and Zhang,Y. (2015) Predicting the Effect of Mutations on Protein-Protein BindingInteractions through Structure-Based Interface Profiles. *PLoSComput.Biol.*, **11**(10), e1004494.

Dehouck,Y.,*et. al.* (2013) BeAtMuSiC: prediction of changes in protein–protein binding affinity on mutations.*NucleicAcids Res.*,**41**(W1), W333–W339.

Gao, Y., et al. (2004) Structure-based method for analyzing protein-protein interfaces. *J. Mol. Model.***10**(1), 44-54.

Gromiha,M.M. *et al.* (2016) Protein-protein interactions: scoring schemes and binding affinity. *Curr.Opin. Str. Biol.*,**44**, 31-38.

Gromiha,M.M. and Yugandhar,K. (2017) Integrating computational methods and experimental data for understanding the recognition mechanism and binding affinity of protein-protein complexes. *Prog.Biophys.Mol. Biol.* DOI: 10.1016/j.pbiomolbio.2017.01.001

Hanson,R.M.*et. al.* (2013) JSmol and the next-generation web-based representation of 3D molecular structure as applied to Proteopedia. *Isr. J. Chem.*, **53**, 207-16.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-637.

Keskin,O. *et. al.* (2008) Principles of Protein-Protein Interactions: What are the Preferred Ways forProteins to Interact?*Chem. Rev.*, **108**,1225-44.

Kumar,M.D. and Gromiha,M.M. (2006) PINT: Protein-protein Interactions Thermodynamic Database. *Nucleic AcidsRes.*,**34**, D195-8.

Li,M., *et. al.* (2016)MutaBind estimates and interprets the effects of sequence variants on protein–protein interactions. *NucleicAcids Res.*,**44**(W1), W494-501.

Moal,I.H. and Fernández-Recio,J. (2012) SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. *Bioinformatics*, **28**, 2600-7.

Petukh,M.,*et. al.* (2016) SAAMBE: Webserver to Predict the Charge of Binding Free Energy Caused by Amino Acids Mutations. *Int. J. Mol. Sci.*, **17**, E547.

Pires,D.E., *et. al.* (2014)mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, **30**, 335-42.

Rose,P.W., *et. al.* (2017) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.***45**(D1):D271-D281.

Sirin,S. *et. al.* (2016) AB-Bind: Antibody binding mutational database for computational affinity predictions. *Protein Sci.*, **25**, 393-409.

Szklarczyk,D. *et. al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *NucleicAcidsRes.*, **43**(Database issue), D447-52.

Thorn,K.S. and Bogan,A.A. (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, **17**, 284-5.

UniProt Consortium. (2015) UniProt: a hub for protein information. *NucleicAcidsRes.*, **43**(Database issue), D204-12.

Yip,Y.L. *et. al.* (2004) The Swiss-Prot variant page and the ModSNP database: a resource for sequence and structure information on human protein variants. *Hum. Mutat.*,**23**, 464-70.

Watkins,A.M. and Arora,P.S. (2015) Structure-based inhibition of protein-protein interactions. *Eur. J. Med. Chem.*, **94**, 480-488.

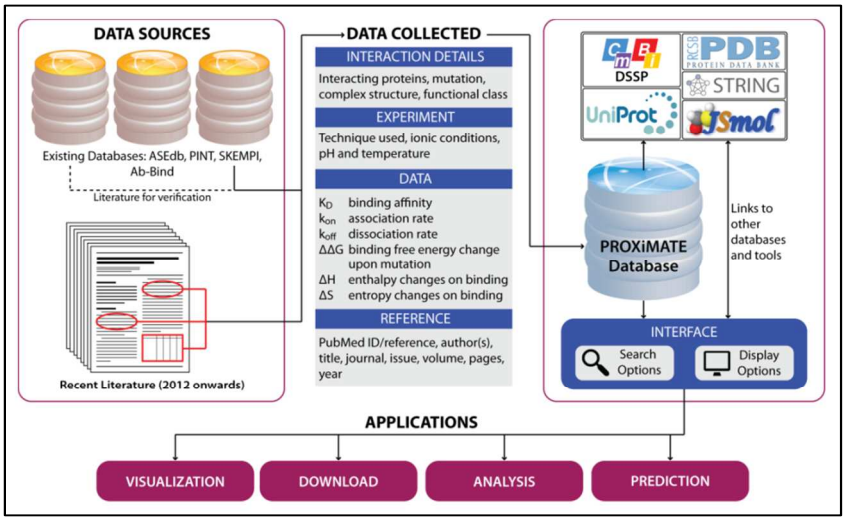

Fig. 1. Schematic diagram describing data collection, workflow and applications of PROXiMATE database