BIOINFORMATICS APPLICATIONS NOTE Vol. 31 no. 1 2015, pages 123–125
doi:10.1093/bioinformatics/btu594

*Structural bioinformatics*

Advance Access publication September 2, 2014

# CCharPPI web server: computational characterization of protein–protein interactions from structure

Iain H. Moal†, Brian Jiménez-García† and Juan Fernández-Recio*

Joint BSC-IRB Research Programme in Computational Biology, Department of Life Sciences, Barcelona Supercomputing Center, C/Jordi Girona 29, 08034 Barcelona, Spain

Associate Editor: Alfonso Valencia

**ABSTRACT**

**Summary:** The atomic structures of protein–protein interactions are central to understanding their role in biological systems, and a wide variety of biophysical functions and potentials have been developed for their characterization and the construction of predictive models. These tools are scattered across a multitude of stand-alone programs, and are often available only as model parameters requiring reimplementation. This acts as a significant barrier to their widespread adoption. CCharPPI integrates many of these tools into a single web server. It calculates up to 108 parameters, including models of electrostatics, desolvation and hydrogen bonding, as well as interface packing and complementarity scores, empirical potentials at various resolutions, docking potentials and composite scoring functions.

**Availability and implementation:** The server does not require registration by the user and is freely available for non-commercial academic use at http://life.bsc.es/pid/ccharppi

**Contact:** juanf@bsc.com

Received on June 5, 2014; revised on August 3, 2014; accepted on August 28, 2014

## 1 INTRODUCTION

Protein–protein interactions are involved in most cell processes, and their structural and functional annotation is essential to understand biological and pathological phenomena and to develop new therapeutic approaches. The increasing volume of experimental data on protein–protein interactions at the molecular level offers many opportunities for functional characterization and the construction of predictive models based on properties arising from structure, such as interface geometry, hydrogen bonding, electrostatics and desolvation energy, which act as an intermediate layer between structure and function (Chothia and Janin, 1975; Jones and Thornton, 1997). Indeed, the selection and combination of structure-based potentials within a learning framework have been used for many tasks, often beyond their original development purpose, such as the prediction of binding affinity (Moal *et al*., 2011), kinetics (Moal and Bates, 2012), mutational effects (Agius *et al*., 2013; Moretti *et al*., 2013; Pallara *et al*., 2013), interface design (Fleishman *et al*., 2011; Yu *et al*., 2012), protein–protein docking (Moal *et al*., 2013b) and the detection of hotspots (Lise *et al*., 2009; Zhu and Mitchell 2011), with many further possibilities remaining to be explored (Moal *et al*., 2013a). Although many tools have been developed to calculate structural properties, some of which are available online (Tuncbag *et al*., 2009), their availability and ease of use are an impediment, often requiring the installation of stand-alone programs with different library dependencies, reimplementation of models for which only parameters are given and reformatting of pdb files. Thus, there is a need to consolidate these methods into a single implementation. Here, we present CCharPPI, a web server, which gathers together a large number of these functions, including those on which many of our previous models were based, into a single easy-to-use interface.

## 2 THE WEB SERVER

CCharPPI incorporates many parameter calculation tools into a single web application, which is freely available for academic non-commercial use. Up to 108 intermolecular parameters are calculated for the input protein–protein interface/s, including 43 potential functions, which have been reimplemented (Chuang *et al*., 2008; Feng *et al*., 2010; Lu *et al*., 2003; Liu and Vakser, 2011; Liu *et al*., 2004; Mintseris *et al*., 2007; Moal and Fernandez-Recio, 2013; Pokarowski *et al*., 2005; Rajgaria *et al*., 2008, 2006; Shen and Sali, 2006; Tobi, 2010; Tobi and Bahar, 2006), as well as terms calculated with 11 stand-alone programs (Feliu *et al*., 2011; Li and Liang, unpublished; Lu *et al*., 2008; Mitra and Pal, 2010; Pierce and Weng, 2007, 2008; Ravikant and Elber, 2010; Viswanath *et al*., 2013; Yang and Zhou, 2008a,b; Zhang and Zhang, 2010; Zhou and Skolnick, 2011) and 4 packages: FireDock (Andrusier *et al*., 2007), PyRosetta (Chaudhury *et al*., 2010), SIPPER (Pons *et al*., 2011) and PyDock (Cheng *et al*., 2007). A detailed list of individual parameters is given online (http://life.bsc.es/pid/ccharppi/info/faq_and_help#descriptors). Users can easily calculate descriptors of interest using a clear workflow. There are three different input sources: a protein databank ID code for automatic retrieval, an uploaded complex in PDB format, or a compressed batch job file for analysing multiple interfaces, for instance, those derived from docking predictions. The web front end acts as user input source and makes results available for display and download. The back end polls for queued projects and schedules jobs for parallel execution. The distribution of descriptor values can be visualized by clicking the descriptor name on the results page. For comparison, values are shown against a background distribution pre-calculated using a set of diverse non-redundant

*To whom correspondence should be addressed.
†The authors wish it be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

complexes of known affinity (Kastritis *et al*., 2011), with the relationship between affinity values and the pre-calculated descriptor values indicated by a scatter plot. Two pre-calculated datasets are available from the website. The first consists of the structural affinity benchmark (Kastritis *et al*., 2011), a set of 144 complexes with experimentally determined affinity. The second consists of 157 wild-type complexes and 2731 unique mutations in the SKEMPI database (Moal and Fernandez-Recio, 2012), as modelled using FoldX (Guerois *et al*., 2002). Computational time for calculating all descriptors is typically <5 min and took <15 min for the largest complex tested (the FAB/influenza haemagglutinin, PDBid 2VIS). Calculations are quicker when executed in parallel using the batch mode, with the 157 wild-type and 2731 unique mutants in the SKEMPI set taking 18 h to complete, and the 144 complexes in the structural affinity benchmark completing in 1 h 20 min. The server has been tested on major browser for MacOS, Ubuntu 12.4, Windows 7 and Windows 8.

## 3 CONCLUSIONS

In conclusion, we have brought together many different methods for characterizing protein–protein interactions, and provide pre-calculated descriptors for two datasets, one of which is also used to provide a visual comparison of uploaded complexes with complexes of known affinity. The ease with which these descriptors can be calculated can accelerate the prototyping of reproducible predictive models, allow users to mix and match different functional forms to model physical phenomena, find new terms for their scoring functions and characterize their complexes of interest. For researchers interested in local execution or incorporation into their own software, all scripts and code are available on request. We intent to expand the pre-calculated datasets, as well as the features as new methods become available.

## ACKNOWLEDGEMENTS

The methods implemented have come from many laboratories, and the authors would like to thank all those who have made their parameters, code and software available, and for clarifying questions regarding licencing.

*Funding*: The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme (FP7/2007-2013) under REA grant agreement PIEF-GA-2012-327899 and grant BIO2013-48213-R from Spanish Ministry of Economy and Competitiveness.

*Conflict of interest*: none declared.

## REFERENCES

Agius,R. *et al*. (2013) Characterizing changes in the rate of protein-protein dissociation upon interface mutation using hotspot energy and organization. *PLoS Comput. Biol*., **9**, e1003216.

Andrusier,N. *et al*. (2007) FireDock: fast interaction refinement in molecular docking. *Proteins*, **69**, 139–159.

Chaudhury,S. *et al*. (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics*, **26**, 689–691.

Cheng,T.M. *et al*. (2007) pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins*, **68**, 503–515.

Chothia,C. and Janin,J. (1975) Principles of protein-protein recognition. *Nature*, **256**, 705–708.

Chuang,G.Y. *et al*. (2008) DARS (Decoys As the Reference State) potentials for protein-protein docking. *Biophys. J*., **95**, 4217–4227.

Feliu,E. *et al*. (2011) On the analysis of protein-protein interactions via knowledge-based potentials for the prediction of protein-protein docking. *Protein Sci*., **20**, 529–541.

Feng,Y. *et al*. (2010) Potentials'R' Us web-server for protein energy estimations with coarse-grained knowledge-based potentials. *BMC Bioinformatics*, **11**, 92.

Fleishman,S.J. *et al*. (2011) Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J. Mol. Biol*., **414**, 289–302.

Guerois,R. *et al*. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol*., **320**, 369–387.

Jones,S. and Thornton,J.M. (1997) Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol*., **272**, 121–132.

Kastritis,P. *et al*. (2011) A structure-based benchmark for protein-protein binding affinity. *Protein Sci*., **20**, 482–491.

Lise,S. *et al*. (2009) Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. *BMC Bioinformatics*, **10**, 365.

Liu,S. and Vakser,I.A. (2011) DECK: Distance and environment-dependent, coarse-grained, knowledge-based potentials for protein-protein docking. *BMC Bioinformatics*, **12**, 280.

Liu,S. *et al*. (2004) A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins*, **56**, 93–101.

Lu,H. *et al*. (2003) Development of unified statistical potentials describing protein-protein interactions. *Biophys. J*., **84**, 1895–1901.

Lu,M. *et al*. (2008) OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing. *J. Mol. Biol*., **376**, 288–301.

Mintseris,J. *et al*. (2007) Integrating statistical pair potentials into protein complex prediction. *Proteins*, **69**, 511–520.

Mitra,P. and Pal,D. (2010) New measures for estimating surface complementarity and packing at protein-protein interfaces. *FEBS Lett*., **584**, 1163–1168.

Moal,I.H. and Bates,P.A. (2012) Kinetic rate constant prediction supports the conformational selection mechanism of protein binding. *PLoS Comput. Biol*., **8**, e1002351.

Moal,I.H. and Fernandez-Recio,J. (2012) SKEMPI: a Structural kinetic and energetic database of mutant protein interactions and its use in empirical models. *Bioinformatics*, **28**, 2600–2607.

Moal,I.H. and Fernandez-Recio,J. (2013) Intermolecular contact potentials for protein-protein interactions extracted from binding free energy changes upon mutation. *J. Chem. Theory Comput*., **9**, 3715–3727.

Moal,I.H. *et al*. (2011) Protein-protein binding affinity prediction on a diverse set of structures. *Bioinformatics*, **27**, 3002–3009.

Moal,I.H. *et al*. (2013a) Scoring functions for protein-protein interactions. *Curr. Opin. Struct. Biol*., **23**, 862–867.

Moal,I.H. *et al*. (2013b) The scoring of poses in protein-protein docking: current capabilities and future directions. *BMC Bioinformatics*, **14**, 286.

Moretti,R. *et al*. (2013) Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. *Proteins*, **81**, 1980–1987.

Pallara,C. *et al*. (2013) Expanding the frontiers of protein-protein modeling: from docking and scoring to binding affinity predictions and other challenges. *Proteins*, **81**, 2192–2200.

Pierce,B. and Weng,Z. (2007) ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins*, **67**, 1078–1086.

Pierce,B. and Weng,Z. (2008) A combination of rescoring and refinement significantly improves protein docking performance. *Proteins*, **72**, 270–279.

Pokarowski,P. *et al*. (2005) Inferring ideal amino acid interaction forms from statistical protein contact potentials. *Proteins*, **59**, 49–57.

Pons,C. *et al*. (2011) Scoring by intermolecular pairwise propensities of exposed residues (SIPPER): a new efficient potential for protein-protein docking. *J. Chem. Inf. Model*., **51**, 370–377.

Rajgaria,R. *et al*. (2006) A novel high resolution Calpha–Calpha distance dependent force field based on a high quality decoy set. *Proteins*, **65**, 726–741.

Rajgaria,R. *et al*. (2008) Distance dependent centroid to centroid force fields using high resolution decoys. *Proteins*, **70**, 950–970.

Ravikant,D.V. and Elber,R. (2010) PIE-efficient filters and coarse grained potentials for unbound protein-protein docking. *Proteins*, **78**, 400–419.

Shen,M.Y. and Sali,A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci*., **15**, 2507–2524.

Tobi,D. (2010) Designing coarse grained-and atom based-potentials for protein-protein docking. *BMC Struct. Biol.*, **10**, 40.

Tobi,D. and Bahar,I. (2006) Optimal design of protein docking potentials: efficiency and limitations. *Proteins*, **62**, 970–981.

Tuncbag,N. *et al.* (2009) A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Brief. Bioinformatics*, **10**, 217–232.

Viswanath,S. *et al.* (2013) Improving ranking of models for protein complexes with side chain modeling and atomic potentials. *Proteins*, **81**, 592–606.

Yang,Y. and Zhou,Y. (2008a) *Ab initio* folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. *Protein Sci.*, **17**, 1212–1219.

Yang,Y. and Zhou,Y. (2008b) Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins*, **72**, 793–803.

Yu,C.M. *et al.* (2012) Rationalization and design of the complementarity determining region sequences in an antibody-antigen recognition interface. *PLoS One*, **7**, e33340.

Zhang,J. and Zhang,Y. (2010) A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction. *PLoS One*, **5**, e15386.

Zhou,H. and Skolnick,J. (2011) GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. *Biophys. J.*, **101**, 2043–2052.

Zhu,X. and Mitchell,J.C. (2011) KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins*, **79**, 2671–2683.