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 reformating 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; Raigaria 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
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 intend 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