BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 9 2014, pages 1314–1315
doi:10.1093/bioinformatics/btu002

Structural bioinformatics Advance Access publication January 9, 2014

PARS: a web server for the prediction of Protein Allosteric and Regulatory Sites

Alejandro Panjkovich[1] and Xavier Daura[1,2,*]

[1]Institute of Biotechnology and Biomedicine (IBB), Universitat Autònoma de Barcelona (UAB), Bellaterra 08193, Spain and [2]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona 08010, Spain

Associate Editor: Alfonso Valencia

ABSTRACT

Summary: The regulation of protein activity is a key aspect of life at the molecular level. Unveiling its details is thus crucial to understanding signalling and metabolic pathways. The most common and powerful mechanism of protein-function regulation is allostery, which has been increasingly calling the attention of medicinal chemists due to its potential for the discovery of novel therapeutics. In this context, PARS is a simple and fast method that queries protein dynamics and structural conservation to identify pockets on a protein structure that may exert a regulatory effect on the binding of a small-molecule ligand.

Availability: PARS is freely available as a web server at http://bioinf.uab.cat/pars.

Contact: xavier.daura@uab.cat

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

Received on June 5, 2013; revised on November 30, 2013; accepted on December 31, 2013

1 INTRODUCTION

Tight regulation of protein function is fundamental to life. Proteins involved in metabolic pathways, signalling cascades and genomic transcription, among other processes within the living cell, are commonly under allosteric regulation, i.e. their activity is modified through the binding of a ligand molecule to the protein in a site different from the active site. In fact, allostery is one of the most powerful protein-function regulation mechanisms, as it allows proteins to sense and immediately respond to changes in their environment (Fenton, 2008). Traditional drug-design approaches focusing on active or primary binding sites can be, therefore, extended by exploiting allosteric sites, as shown by current efforts on GPCRs (Melancon *et al*., 2012) or farnesyl pyrophosphate synthase (Jahnke *et al*., 2010). An advantage of targeting allosteric sites therapeutically is a reduced risk of secondary adverse effects. This is because allosteric sites appear to be significantly less conserved than active sites across homolog proteins (Waelbroeck, 2003), enabling the design of allosteric drugs with high specificity for a single protein within a family. This observation has motivated the development of allosteric drugs for the regulation of phosphodiesterase 4D, for which active site inhibitors cause emesis, a dose-limiting side effect (Burgin *et al*., 2010). Moreover, a drug-discovery approach based on allosteric sites may result in the development of not only novel drug-like inhibitors but activators as well (Peracchi and Mozzarelli, 2011). Other concepts such as the notion of serendipitous allosteric sites, which have no known ligand in nature but can become functional in the presence of an 'opportunistic' ligand (Hardy and Wells, 2004) or the idea that almost any protein may be regulated allosterically (Gunasekaran *et al*., 2004), contribute to the current interest in the pharmacological exploitation of allosteric sites. In this context, it is expected that a deeper understanding of the properties of allosteric sites and their identification would help streamlining the design and discovery of novel therapeutic drugs (Nussinov and Tsai, 2013).

Despite growing efforts, the atomic-level details that explain the functional relationship between distant sites in the same protein molecule have not been elucidated for most of the known cases of allostery (Peracchi and Mozzarelli, 2011). This has motivated a growth in the number of computational studies on protein allosteric sites, including the recent publication of three web servers: SPACER (Goncearenco *et al*., 2013) and MCPath (Kaya *et al*., 2013), which study the allosteric communication across residues in a protein structure, and Allosite (Huang *et al*., 2013), which predicts allosteric sites on protein structures using a machine-learning approach. Here, we present PARS, which queries protein flexibility and structural conservation to predict the location of allosteric sites. The implementation emphasizes ease-of-use and speed, whereas maintaining high accuracy as benchmarked against a large set of known allosteric proteins. Even though our method (as the servers mentioned earlier in the text) requires a protein structure to work, we have added an initial homology modelling step that allows the user to run predictions starting from protein sequence in cases where a known structure is not available.

2 DESCRIPTION

The current perspective on allosteric transitions and associated regulatory events relies on the 'population shift' concept (Cui and Karplus, 2008). Briefly, the protein or protein complex explores different conformations (both active and inactive) in solution, and the allosteric ligand 'shifts' the population or ensemble of conformations on binding, effectively modulating the protein's activity rate (Kar *et al*., 2010). The conformational space explored by the protein can be sampled using computational methods such as molecular dynamics (Chiappori *et al*., 2012) or normal mode analysis (NMA) (Dykeman and Twarock, 2010). In our approach, protein dynamics are queried

*To whom correspondence should be addressed.

 © The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

through NMA, allowing fast large-scale (potentially genome-wide) analyses. The method developed previously (Panjkovich and Daura, 2012) is now implemented as a web server working as follows: (i) initially, the user uploads a protein-structure file (PDB format) and selects which chains and ligands should be considered for the calculations. Alternatively, the user may submit a protein sequence and the server will attempt to generate a structural model by homology to known structural templates (Eswar *et al*., 2008). (ii) Once the job is submitted, the LIGSITE^{cs} program (Huang and Schroeder, 2006) is used to predict putative ligand-binding sites, where a simplified representation of a small molecule (i.e. a set of dummy atoms forming an octahedron) will be placed to simulate the presence of a ligand. (iii) NMA is carried out for the apo (unbound) structure. (iv) For each potential ligand-binding site, an NMA is executed for the protein–ligand complex using a simplified ligand representation; if a cocrystallized ligand is selected, the calculation will be performed for the corresponding binding site. (v) If a significant difference is found between the NMA-derived B-factors of the apo and ligand-bound states of the protein (Wilcoxon–Mann–Whitney's $P \leq 0.05$), the binding site is marked as potentially allosteric.

Complementarily, if enough structural data are available for the protein family, the structural conservation of each pocket is also measured. We have previously shown that allosteric sites tend to be structurally conserved within a protein family (Panjkovich and Daura, 2010) and that the incorporation of this measure improves the capacity of the method based on dynamics to correctly identify allosteric sites (Panjkovich and Daura, 2012). Immediately after submission the user is provided with a link where results can be accessed once the calculation has finished. Optionally, the user can choose to be notified by email when results are available.

A standard run on a 500-residue protein will take $\sim$3 min. Even though calculation times increase exponentially with the number of residues, larger systems can be processed within reasonable time (e.g. 1000 residues $\approx$ 12 min). The results are delivered as a ranking and, if the browser allows it, the binding sites can be explored online on the 3D protein structure by means of the Jmol package (http://www.jmol.org/). To complement the analysis, active site residues are predicted and reported on the structure using a previously described methodology (Mistry *et al*., 2007). Binding sites and ligands are color coded according to given thresholds for the level of structural conservation ($\geq$ 50%) and the predicted effect on protein dynamics ($P \leq 0.05$). The protein structure file including the position of identified sites, a visualization script and a table with calculated values can be downloaded as well for further processing. Normal modes and derived B-factors can be explored in detail using specifically designed interfaces that are available at the results web page. A more detailed description of the server and its parameters is provided as Supplementary Material and available as online documentation on the website.

Predicting allosteric sites computationally is not trivial, and to illustrate the success rate that the user can expect, we have benchmarked PARS using 102 allosteric proteins. Briefly, the results show that if we consider all predicted cavities matching an allosteric site in the benchmark set, 44% appear in the first position and 18% on the second position of the PARS ranking, whereas a total of 73% are found in the top three positions. Full benchmark details and performance statistics are available as Supplementary Material.

We expect that this easy-to-use and relatively fast web server will prove useful for medicinal chemists and other researchers studying the regulation of protein function for biochemical characterization and other applied tasks such as binding-site prioritization for virtual drug screening.

Funding: Seventh Research Framework Programme of the European Union (ref. HEALTH-F3-2009-223101). FPU Scholarship from MICINN, Spanish Government (to A.P.).

Conflict of Interest: none declared.

REFERENCES

Burgin,A.B. *et al*. (2010) Design of phosphodiesterase 4d (PDE4D) allosteric modulators for enhancing cognition with improved safety. *Nat. Biotechnol*., **28**, 63–70.

Chiappori,F. *et al*. (2012) Molecular mechanism of allosteric communication in hsp70 revealed by molecular dynamics simulations. *PLoS Comput. Biol*., **8**, e1002844.

Cui,Q. and Karplus,M. (2008) Allostery and cooperativity revisited. *Protein Sci*., **17**, 1295–1307.

Dykeman,E.C. and Twarock,R. (2010) All-atom normal-mode analysis reveals an RNA-induced allostery in a bacteriophage coat protein. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*., **81** (Pt. 1), 031908.

Eswar,N. *et al*. (2008) Protein structure modeling with modeller. *Methods Mol. Biol*., **426**, 145–159.

Fenton,A.W. (2008) Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem. Sci*., **33**, 420–425.

Goncearenco,A. *et al*. (2013) Spacer: server for predicting allosteric communication and effects of regulation. *Nucleic Acids Res*., **41**, W266–W272.

Gunasekaran,K. *et al*. (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins*, **57**, 433–443.

Hardy,J.A. and Wells,J.A. (2004) Searching for new allosteric sites in enzymes. *Curr. Opin. Struct. Biol*., **14**, 706–715.

Huang,B. and Schroeder,M. (2006) Ligsitecsc: predicting ligand binding sites using the connolly surface and degree of conservation. *BMC Struct. Biol*., **6**, 19.

Huang,W. *et al*. (2013) Allosite: a method for predicting allosteric sites. *Bioinformatics*, **29**, 2357–2359.

Jahnke,W. *et al*. (2010) Allosteric non-bisphosphonate fpps inhibitors identified by fragment-based discovery. *Nat. Chem. Biol*., **6**, 660–666.

Kar,G. *et al*. (2010) Allostery and population shift in drug discovery. *Curr. Opin. Pharmacol*., **10**, 715–722.

Kaya,C. *et al*. (2013) Mcpath: Monte Carlo path generation approach to predict likely allosteric pathways and functional residues. *Nucleic Acids Res*., **41**, W249–W255.

Melancon,B.J. *et al*. (2012) Allosteric modulation of seven transmembrane spanning receptors: theory, practice, and opportunities for central nervous system drug discovery. *J. Med. Chem*., **55**, 1445–1464.

Mistry,J. *et al*. (2007) Predicting active site residue annotations in the pfam database. *BMC Bioinformatics*, **8**, 298.

Nussinov,R. and Tsai,C.-J. (2013) Allostery in disease and in drug discovery. *Cell*, **153**, 293–305.

Panjkovich,A. and Daura,X. (2010) Assessing the structural conservation of protein pockets to study functional and allosteric sites: implications for drug discovery. *BMC Struct. Biol*., **10**, 9.

Panjkovich,A. and Daura,X. (2012) Exploiting protein flexibility to predict the location of allosteric sites. *BMC Bioinformatics*, **13**, 273.

Peracchi,A. and Mozzarelli,A. (2011) Exploring and exploiting allostery: Models, evolution, and drug targeting. *Biochim. Biophys. Acta*, **1814**, 922–933.

Waelbroeck,M. (2003) Allosteric drugs acting at muscarinic acetylcholine receptors. *Neurochem. Res*., **28**, 419–422.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016