Bioinformatics, 33(24), 2017, 3996–3998
doi: 10.1093/bioinformatics/btx430
Advance Access Publication Date: 6 July 2017
Applications Note

OXFORD

Structural bioinformatics

AlloSigMA: allosteric signaling and mutation analysis server

Enrico Guarnera[1,†], Zhen Wah Tan[1,†], Zejun Zheng[1,†] and Igor N. Berezovsky[1,2,*]

[1]Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore 138671, Singapore and [2]Department of Biological Sciences (DBS), National University of Singapore (NUS), Singapore 117597, Singapore

*To whom correspondence should be addressed.
†The authors wish it to be known that these authors contributed equally to the work.
Associate Editor: Alfonso Valencia

Received on March 1, 2017; revised on June 5, 2017; editorial decision on June 29, 2017; accepted on July 3, 2017

Abstract

Motivation: Allostery is an omnipresent mechanism of the function modulation in proteins via either effector binding or mutations in the exosites. Despite the growing number of online servers and databases devoted to prediction/classification of allosteric sites and their characteristics, there is a lack of resources for an efficient and quick estimation of the causality and energetics of allosteric communication.
Results: The AlloSigMA server implements a unique approach on the basis of the recently introduced structure-based statistical mechanical models of allosteric signaling. It provides an interactive framework for estimating the allosteric free energy as a result of the ligand(s) binding, mutation(s) and their combinations. Latent regulatory exosites and allosteric effect of mutations can be detected and explored, facilitating the research efforts in protein engineering and allosteric drug design.
Availability and implementation: The AlloSigMA server is freely available at http://allosigma.bii.a-star.edu.sg/home/.
Contact: igorb@bii.a-star.edu.sg

1 Introduction

One of the consequences of the pervasive presence of the allosteric signaling phenomena in the wide spectrum protein of types (Berezovsky *et al.*, 2017; Guarnera and Berezovsky, 2016; Gunasekaran *et al.*, 2004) and molecular machines (Cui and Karplus, 2008; Guarnera and Berezovsky, 2016; Mitternacht and Berezovsky, 2011) is the development of many web-based resources dedicated to the detection/listing of allosteric sites (Goncearenco *et al.*, 2013; Guarnera and Berezovsky, 2016; Shen *et al.*, 2016). However, efficient online applications for the physics-based (Guarnera and Berezovsky, 2016; Rodgers *et al.*, 2013) analysis of allosteric signaling, which would allow one to quickly estimate the causality and energetics of the process are still lacking. Additionally, recently reported enrichment of allosteric sites with deleterious mutations (Shen *et al.*, 2017) shows that the analysis of allosteric effects of mutations is an important component in the understanding of the mechanisms of cancerogenesis, calling for the development of relevant computational approaches and their web implementations. AlloSigMA server is aimed at providing a quantitative tool for the analysis of the energetics of allosteric communication, allowing users to quickly estimate in energy terms the allosteric effects of ligand binding, mutations, and their combinations. The quantification of allosteric effects offers a rational guide to the experimental researcher in the selection of allosterically relevant binding sites and/or mutations, which can facilitate the design of experimental efforts towards modulation of the protein activity.

2 Methods

2.1 Theoretical background

We use here the structure-based statistical mechanical model of allostery (SBSMMA), which allows one to explore the causality and energetics of allosteric signaling in the general case of a protein perturbed by the allosteric ligand(s) (Guarnera and Berezovsky, 2016) and mutation(s) (Kurochkin *et al.*, 2017). The resulting per-residue allosteric free energy is obtained by solving the statistical mechanical problem for the ensemble of all possible protein local configurations in the unbound/wild-type (0), bound (B), mutated (M) and bound/mutated (BM) states, respectively, leading to the relations

$$\Delta g_i^{(B)} = \frac{1}{2} k_B T \sum_\mu \ln \frac{\varepsilon_{\mu,i}^{(B)}}{\varepsilon_{\mu,i}^{(0)}}, \quad \Delta g_i^{(M)} = \frac{1}{2} k_B T \sum_\mu \ln \frac{\varepsilon_{\mu,i}^{(M)}}{\varepsilon_{\mu,i}^{(0)}}, \qquad (1)$$
$$\Delta g_i^{(BM)} = \Delta g_i^{(B)} + \Delta g_i^{(M)}$$

where i is the residue index. The $\varepsilon_{\mu,\,i}^{(P)}$ are parameters associated to the normal modes $\boldsymbol{e}_\mu^{(P)}$ of the protein in a state (P), and they are components of the allosteric potential:

$$U_i^{(P)}(\sigma) = \frac{1}{2} \sum_\mu \varepsilon_{\mu,\,i}^{(P)} \sigma_\mu^2, \qquad (2)$$

where $\sigma = (\sigma_1, \ldots, \sigma_\mu, \ldots)$ is a vector of Gaussian variables with variance $1/\varepsilon_{\mu,\,i}^{(P)}$, each of which is associated with a corresponding normal mode. The allosteric free energies are thus obtained by integrating over all the C_α residue displacements identified by the vector σ.

The parameters $\varepsilon_{\mu,\,i}^{(P)} = \sum_j \left| \boldsymbol{e}_{\mu,\,i}^{(P)} - \boldsymbol{e}_{\mu,\,j}^{(P)} \right|^2$ are calculated from the modes $\boldsymbol{e}_\mu^{(P)}$ that characterize the dynamics of a protein in either one of considered states: unbound/wild-type (0), bound (B) or mutated (M). They are obtained as the orthonormal modes of the Hessian matrices $K^{(P)} = \partial^2 E^{(P)} / \partial \boldsymbol{r}_i \partial \boldsymbol{r}_j$, with $E^{(P)}(\boldsymbol{r})$ the harmonic energies associated with the corresponding protein state (P).

The energy function associated with C_α harmonic model of the protein in the unbound/wild-type (0) is

$$E^{(0)}(\boldsymbol{r} - \boldsymbol{r}^0) = \sum\nolimits_{\langle i,\,j \rangle} k_{i,\,j} \left(d_{i,\,j} - d_{i,\,j}^0 \right)^2, \qquad (3)$$

where $d_{i,\,j}$ and $d_{i,\,j}^0$ are the interatomic distances between C_α atoms in the generic and reference structures, respectively, and $k_{i,\,j}$ is a distance-dependent force constant. The energy function of the protein bound state (B) for a particular site S is

$$E^{(B)}(\boldsymbol{r} - \boldsymbol{r}^0, S) = \sum\nolimits_{\langle i,\,j \rangle \notin S} k_{i,\,j} \left(d_{i,\,j} - d_{i,\,j}^0 \right)^2 + \alpha \sum\nolimits_{\langle i,\,j \rangle} k_{i,\,j} \left(d_{i,\,j} - d_{i,\,j}^0 \right)^2 \qquad (4)$$

where the second term defines binding as a harmonic restraint with α being the corresponding stiffening parameter ($\alpha = 100$, see Guarnera and Berezovsky, 2016). The energy function associated with a mutated protein state (M), with point mutation on residue m is

$$E^{(M)}(\boldsymbol{r} - \boldsymbol{r}^0, m) = \sum\nolimits_{\langle i,\,j \rangle : i \neq m} k_{i,\,j} \left(d_{i,\,j} - d_{i,\,j}^0 \right)^2 + \theta \sum\nolimits_{\langle m,\,j \rangle} k_{i,\,j} \left(d_{i,\,j} - d_{i,\,j}^0 \right)^2 \qquad (5)$$

where θ determines the type of mutation. Two types of mutations are defined: UP-mutation ($\uparrow M,\ \theta = 100$), which models the situation of an actual mutation to a bulky residue with over-stabilizing effects on the local contact network; conversely, DOWN-mutation ($\downarrow M,\ \theta = 10^{-2}$) models the destabilization of residue's contact network similarly to Ala/Gly-like mutations.

2.2 Input, preprocessing and processing

For the server input the user can provide either the PDB ID of an existing protein X-ray structure or upload an individual file with protein coordinates in the standard PDB format.

The preprocessing starts from the ordered list of biological assemblies in the PISA database, according to the solvation free energy gain upon assembly formation (Krissinel and Henrick, 2007). If no assembly is found, the structure will be fetched from the Protein Databank as is. Next, a comprehensive list of binding sites is compiled and mapped to the correct chains of the considered protein structure on the basis of the first 10 best matching homologs (99% sequence identity) as generated in the VAST server (Madej *et al.*, 2014).

The processing requires the user to select binding sites of interest and make the choice of the UP-mutations (stabilizing) and DOWN-mutations (destabilizing). It is also possible to remove some of the protein chains in case if it is only required to analyze a part of the protein complex/oligomer (see the online 'Tutorial' for details).

2.3 Implementation

AlloSigMA server is written in Python using the Flask framework (http://flask.pocoo.org/). The calculation of the allosteric free energy is implemented in Python. The interactive web interface is powered by the JavaScript libraries jQuery (http://www.jquery.com/) and D3.js (http://d3js.org/), and by JSmol (http://wiki.jmol.org/index.php/JSmol). The normal mode analysis is done using the C_α harmonic model implemented in the Molecular Modeling Toolkit (MMTK, Hinsen, 2000). Ten lowest frequency normal modes are used in calculations (Guarnera and Berezovsky, 2016). Finally, the server is interfaced with the Protein Databank (Berman *et al.*, 2000) and PDBePISA (Krissinel and Henrick, 2007). The system can process structures with up to 7000 amino acid residues, and up to 40 protein chains.

3 Results

AlloSigMA is designed for the interactive exploratory analysis of the allosteric signaling in proteins given a single crystal structure. Any combination of binding site(s) obtained in preprocessing along with site(s) of interest and mutation(s) manually designated by the user will be processed. The server provides a graphical output, including two panels showing the protein chains and structure, respectively, with residues colored according to the corresponding per-residues allosteric free energy values. All the results are also available in a downloadable zip archive with files ready for further analysis and machine processing. The archive contains files that allow a user to restore previously obtained sessions.

To illustrate the potential and the major options provided by the server, we use the tetrameric enzyme phosphofructokinase (PFK), which displays a classical example of allostery (see online Tutorial for details). The PFK's substrate is fructose-6-phosphate (site F6P), and the protein is allosterically activated by adenosine-diphosphate and inhibited by phosphoenolpyruvate (PEP). Figure 1 shows the result of the allosteric signaling caused by the binding of phosphoglycorate (PGA), an analog of the PEP inhibitor. The color gradient in both sequence and structure representations indicates negative (red, decreased residue dynamics) and positive (blue, increased dynamics) allosteric free energies, respectively, as a result of the PGA binding. The allosteric effect can also be monitored at the level of sites by considering the mean allosteric free energy in the sites of interest. That is especially important when one is interested to check how the dynamics of a catalytic site is allosterically affected by either binding or mutations, or both combined.

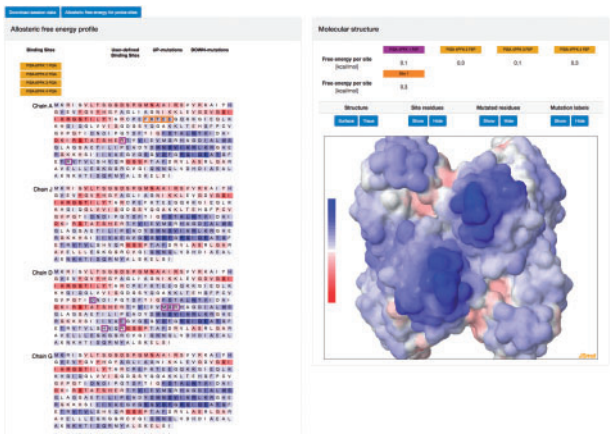

Fig. 1. Output of the AlloSigMA server for the analysis of allosteric communication in Phosphofructokinase (3PFK) originated by the binding of PGA (phosphoglycerate—analog of the inhibitor PEP) ligands. The panels contain protein chains and protein surface with residues colored according to the per-residues allosteric free energy values observed upon the PGA binding. The color gradient shows increased (blue) and decreased (red) dynamics expressed in per-residue allosteric free energy (see online Tutorial for further details)

4 Conclusions

AlloSigMA server is an interactive exploratory framework for the analysis of allosteric signaling, which implements the SBSMMA and provides a quick estimate of the allosteric free energy. It is applicable to a wide range of proteins—from small monomeric structures to large protein complexes (Guarnera and Berezovsky, 2016). The allosteric signaling caused by the ligand binding to known allosteric sites, to sites of interest designated by the user, by the over-stabilizing and destabilizing mutations, and combinations thereof can be explored. We hope, therefore, that AlloSigMA will become a valuable tool in the investigation of allosteric effects, search for latent regulatory exosites, analysis of clinical high-throughput data on mutations (Shen *et al.*, 2017), and in the emerging field of allosteric mutagenesis (Kurochkin *et al.*, 2017; Shen *et al.*, 2017).

Funding

Biomedical Research Council, Agency for Science, Technology, and Research (A*STAR), Singapore.

Conflict of interest: none declared.

References

Berezovsky,I.N. *et al.* (2017) Protein function machinery: from basic structural units to modulation of activity. *Curr. Opin. Struct. Biol.*, **42**, 67–74.

Berman,H.M. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

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

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

Guarnera,E. and Berezovsky,I.N. (2016) Allosteric sites: remote control in regulation of protein activity. *Curr. Opin. Struct. Biol.*, **37**, 1–8.

Guarnera,E. and Berezovsky,I.N. (2016) Structure-based statistical mechanical model accounts for the causality and energetics of allosteric communication. *PLoS Comput. Biol.*, **12**, e1004678.

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

Hinsen,K. (2000) The molecular modeling toolkit: a new approach to molecular simulations. *J. Comput. Chem.*, **21**, 79–85.

Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.

Kurochkin,I.V. *et al.* (2017) Toward allosterically increased catalytic activity of insulin-degrading enzyme against amyloid peptides. *Biochemistry*, **56**, 228–239.

Madej,T. *et al.* (2014) MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res.*, **42(Database issue)**, D297–D303.

Mitternacht,S. and Berezovsky,I.N. (2011) Coherent conformational degrees of freedom as a structural basis for allosteric communication. *PLoS Comput. Biol.*, **7**, e1002301.

Rodgers,T.L. *et al.* (2013) Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors. *PLoS Biol.*, **11**, e1001651.

Shen,Q. *et al.* (2017) Proteome-scale investigation of protein allosteric regulation perturbed by somatic mutations in 7,000 cancer genomes. *Am. J. Hum. Genet.*, **100**, 5–20.

Shen,Q. *et al.* (2016) ASD v3.0: unraveling allosteric regulation with structural mechanisms and biological networks. *Nucleic Acids Res.*, **44**, D527–D535.