Allosite: a method for predicting allosteric sites

Wenkang Huang[1,†], Shaoyong Lu[1,†], Zhimin Huang[1], Xinyi Liu[1], Linkai Mou[1], Yu Luo[1], Yanlong Zhao[1], Yaqin Liu[1], Zhongjie Chen[1], Tingjun Hou[2] and Jian Zhang[1,*]

[1]Department of Pathophysiology, Shanghai Jiao-Tong University School of Medicine, Shanghai 200025, China and [2]Functional Nano & Soft Materials Laboratory, Soochow University, Suzhou, Jiangsu 215123, China

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: The use of allosteric modulators as preferred therapeutic agents against classic orthosteric ligands has colossal advantages, including higher specificity, fewer side effects and lower toxicity. Therefore, the computational prediction of allosteric sites in proteins is receiving increased attention in the field of drug discovery. Allosite is a newly developed automatic tool for the prediction of allosteric sites in proteins of interest and is now available through a web server.

Availability: The Allosite server and tutorials are freely available at http://mdl.shsmu.edu.cn/AST

Contact: jian.zhang@sjtu.edu.cn

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

1 INTRODUCTION

Allostery is a fundamental process that regulates a protein's functional activity by means of changes in the conformation and dynamics induced by the action of an effector at a site distinct from the active site, also termed as the allosteric site. From the perspective of drug discovery, because allosteric sites evolved under lower sequence-conservation pressure compared with the evolutionarily conserved active sites, targeting allosteric sites can provide unprecedented advantages in terms of higher specificity, fewer side effects and lower toxicity (Nussinov *et al.*, 2011). Therefore, identification of allosteric sites in proteins of interest is a prerequisite for the development of allosteric drugs with a wide range of chemotypes.

Thus far, several methods have been tried to detect allosteric sites in proteins, including the geometry-based generic predictor, 'binding leverage', and normal mode analysis (Mitternacht *et al.*, 2011a, b; Panjkovich and Daura, 2012). Owing to the limited size of the datasets in these studies, as well as the lack of knowledge about common characteristics of allosteric sites, such methods often suffer from an overly specialized range of applications (Fukunaga, 1990), and therefore it is imperative to develop novel and general methods to precisely locate allosteric sites in proteins.

Currently, a vast array of known allosteric sites has been discovered by high-throughput screening followed by X-ray crystallography, phage display and tethering, yet most of these discoveries are serendipitous. This reliance on serendipity can be ascribed to the pronounced deficiency of understanding of the structural characteristics of allosteric sites. With our continuous endeavors invested in recruiting the allosteric proteins and allosteric sites deposited in the allosteric database (ASD) (Huang *et al.*, 2011), together with recent preliminary understanding of the characteristics of allosteric proteins and modulators (Li *et al.*, 2013; Wang *et al.*, 2012), we hope to lay the foundation for exploiting a high-efficiency model to detect allosteric sites.

In this study, we present for the first time a web service, Allosite, which uses elegant algorithms such as pocket-based analysis and support vector machine (SVM) classifier to predict the location of allosteric sites in proteins. Allosite's model uses a rigorous selection of high-quality datasets for training and exhibits a high accuracy of ∼95% on the test set. More importantly, the prediction of novel allosteric sites for several proteins using Allosite was experimentally supported by mutagenesis. Thus, the prediction capability and the user-friendly interface of this web server can be of great benefit to biologists and medicinal chemists by serving as a starting point for identifying the location of allosteric sites and following allosteric drug design.

2 METHODS

Ninety crystal structures of non-redundant allosteric proteins with a resolution better than 3 Å were carefully extracted from ASD. By means of a discriminated feature selection method, a subset of 21 site descriptors best suited to delineate the characteristics of allosteric sites was selected and the parameters in SVM were optimized. Then, the SVM model for allosteric site identification was trained and tested, and a final model was deployed in the web server of Allosite. Detailed information about the process of model construction is provided in the Supplementary information.

3 RESULTS

We have extensively tested our algorithm, and the results are summarized in the Supplementary Table S1. The accuracies of the 5-fold cross-validation for the three training sets are all >92%. Using the models derived from training data, the prediction accuracies on test sets are all >95%, and both the sensitivities and specificities are >83%, indicating the satisfactory prediction capability of our model in the finding of allosteric sites. To further validate the reliability of the Allosite server in practice, additional allosteric proteins with unknown allosteric sites were collected from ASD (Supplementary Table S2).

---

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
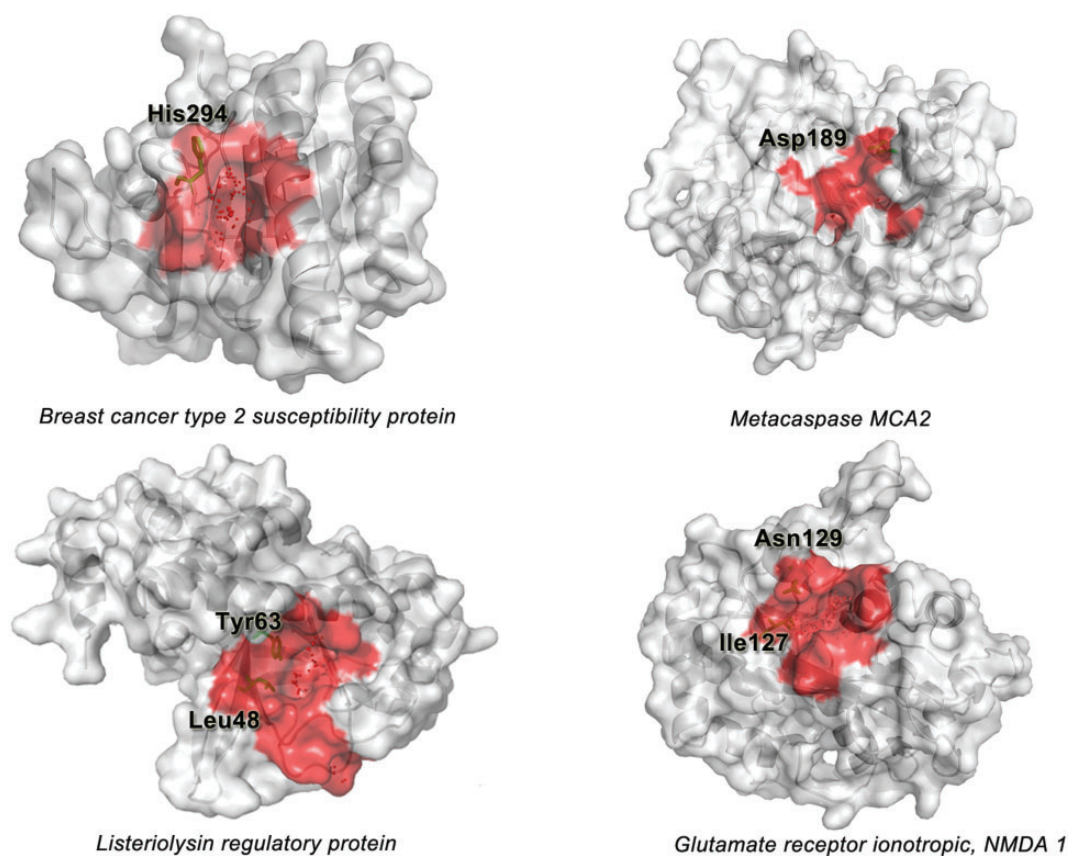
Fig. 1. View of the potential allosteric sites of four proteins predicted by Allosite. Regions of allosteric sites are highlighted in red on the surface. Side chains of crucial residues supported by experimental mutagenesis in the allosteric sites are shown as stick and labeled. Detailed information is described in the Supplementary Table S2

Allosite is capable of identifying potential allosteric sites in all systems. In the four of five cases, we have found mutations that affect the orthostatic functions of the proteins in the area of the predicted allosteric sites by literature survey (Fig. 1), which give support to the predictive results of Allosite in real applications. In addition, the predicted allosteric site in coagulation factor IX also could provide a useful hint to guide experimental validation (Supplementary Table S2). Perhaps more importantly, the findings can be used to design compounds against these pockets.

3.1 Usage and output

For each job, the user can specify the proteins of interest either with a PDB ID or by uploading PDB files under 'Base Structure'. After defining a query protein, a 'Job Name' must be set before submitting, which allows the users to locate their queries in the 'Job Queue'. Once the run is submitted, a unique Job ID appears on the submission page. The users can apply this Job ID or Job Name to track the job's status or to access results in the 'Job Queue' page of Allosite. On completion of a job, a button labeled 'Finished' emerges in the 'Job Queue'. A representative run of an Allosite job takes 15–30 s, depending on the size and the complexity of the input protein.

Clicking the 'Finished' button links to the result, which contains the interactive 3D representation of the protein structure and the pocket properties of the predicted allosteric site (Supplementary Fig. S2). The residues of the predicted allosteric site can be downloaded for offline analysis under the 'Download Results' panel. Further, the predicted allosteric site and residues forming the allosteric site can be accessed in a separate figure in a Jmol applet for visual analysis by clicking the 'Show Pocket' button.

The Allosite Web site includes a step-by-step tutorial in the 'Help' page. The server requires Java and Javascript to be enabled, and has been tested on all major web browsers.

4 CONCLUSION

The Allosite web server conveniently provides a user-friendly interface to predict the allosteric sites of a protein structure of interest. Furthermore, critical properties that are pertinent to characterize the allosteric site are offered. To the best of our knowledge, this web server is the first of its kind and will be of considerable value to scientists interested in allosteric drug design.

Conflict on Interest: none declared.

REFERENCES

Fukunaga,K. (1990) *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, NY, Boston.

Huang,Z. *et al.* (2011) ASD: a comprehensive database of allosteric proteins and modulators. *Nucleic Acids Res.*, **39**, D663–D669.

Li,X. *et al.* (2013) Toward an understanding of the sequence and structural basis of allosteric proteins. *J. Mol. Graph. Model.*, **40**, 30–39.

Mitternacht,S. and Berezovsky,I.N. (2011a) A geometry-based generic predictor for catalytic and allosteric sites. *Protein Eng. Des. Sel.*, **24**, 405–409.

Mitternacht,S. and Berezovsky,I.N. (2011b) Binding leverage as a molecular basis for allosteric regulation. *PLoS Comput. Biol.*, **7**, e1002148.

Nussinov,R. *et al.* (2011) Allo-network drugs: harnessing allostery in cellular networks. *Trends Pharmacol. Sci.*, **32**, 686–693.

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

Wang,Q. *et al.* (2012) Toward understanding the molecular basis for chemical allosteric modulator design. *J. Mol. Graph. Model.*, **38**, 324–333.