

Databases and ontologies

ASBench: benchmarking sets for allosteric discovery

Wenkang Huang[†], Guanqiao Wang[†], Qiancheng Shen[†], Xinyi Liu, Shaoyong Lu, Lv Geng, Zhimin Huang and Jian Zhang*

Department of Pathophysiology, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

Associate Editor: Janet Kelso

Received on November 11, 2014; revised on March 16, 2015; accepted on March 17, 2015

Abstract

Summary: Allostery allows for the fine-tuning of protein function. Targeting allosteric sites is gaining increasing recognition as a novel strategy in drug design. The key challenge in the discovery of allosteric sites has strongly motivated the development of computational methods and thus high-quality, publicly accessible standard data have become indispensable. Here, we report benchmarking data for experimentally determined allosteric sites through a complex process, including a 'Core set' with 235 unique allosteric sites and a 'Core-Diversity set' with 147 structurally diverse allosteric sites. These benchmarking sets can be exploited to develop efficient computational methods to predict unknown allosteric sites in proteins and reveal unique allosteric ligand–protein interactions to guide allosteric drug design.

Availability and implementation: The benchmarking sets are freely available at <http://mdl.shsmu.edu.cn/asbench>.

Contact: jian.zhang@sjtu.edu.cn

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online

1 Introduction

Allostery is a major regulatory mechanism that is of fundamental importance in a wide variety of biological phenomena. The propagation of allosteric signals from allosteric sites induced by effectors binding to different, often distant, functional sites allows for exquisite control of protein functional activity (Nussinov and Tsai, 2013). The structural diversity of allosteric sites endows allosteric effectors with the benefits of higher selectivity and lower toxicity. Targeting allosteric sites, as a novel tactic in drug design, has therefore prompted intense interest in the development of allosteric drugs (Lu *et al.*, 2014a).

Despite the enticing advantages of allosteric regulation, modern allosteric drug discovery faces considerable challenges. In particular, the vast majority of allosteric sites in proteins are as-yet undiscovered as a result of the difficulty in identifying such sites experimentally. This difficulty has led to a paucity of structural and mechanistic insights into the characterization of allosteric sites, and

it has also impeded computational development for the identification of allosteric sites (Lu *et al.*, 2014b). To address this issue, we recently constructed the Allosteric Database (ASD; Huang *et al.*, 2014), which is a collection of experimentally determined allosteric proteins, and modulators to provide a platform for the application and discovery of allosteric sites. Nevertheless, the collection of raw allosteric sites suffers from redundancy and the inclusion of low quality sites. It is therefore imperative to offer a more stringent set of allosteric sites for the development of viable computational approaches for allosteric site prediction and to reveal the unique characteristics of allosteric ligand–protein interactions for drug design.

To this end, two benchmarking sets of allosteric sites were compiled through a complex process, designated as a 'Core set' and a 'Core-Diversity set'. The former totally contains 235 unique allosteric site, and the latter encompasses 147 structurally diverse allosteric sites. These high-quality benchmarking sets can be

exploited to develop computational approaches to predict the location of as-yet-unknown allosteric sites in proteins (Huang *et al.* 2013; Panjkovich and Daura, 2014), and they contain representative specific allosteric ligand–protein interactions that can guide structure-based allosteric drug discovery.

2 Methods

The latest version of ASD (v2.0, July 2014), containing 1743 experimentally determined allosteric complexes, was used to develop the benchmarking sets of allosteric sites. A set of rules was applied to select the qualified allosteric sites and remove site redundancy, yielding the ‘Core set’ composed of 235 representative complexes for allosteric sites (Supplementary Table S1). Subsequently, the structural similarity of the allosteric sites in the ‘Core set’ was assessed, and 147 structural complexes with diverse allosteric sites were further extracted to constitute the ‘Core-Diversity set’ (Supplementary Table S2). Both datasets are deposited in ASBench. Detailed information about the process is provided in the ‘Materials and Methods’ of Supplementary Information.

3 Results

3.1 Benchmarking sets

The ‘Core set’ in ASBench includes 235 unique allosteric sites, which are most frequently found in bacteria and humans (Supplementary Fig. S1) and are primarily distributed in several classes of proteins,

such as transferases (35%), hydrolases (15%), oxidoreductases (9%) and transcription factors (8%) (Fig. 1A). The polar solvent-accessible surface area and the pocket volume of these allosteric sites are mainly within the range of 10–1500 and 150–6500 Å³, respectively (Fig. 1B). An analysis of pocket similarity, measured by the Pocket Similarity score (PS-score), revealed the existence of structural similarity between several allosteric sites, with the largest PS-score of 0.976 (Fig. 1C, left). Eliminating the structurally redundant allosteric sites in the ‘Core set’ yielded 147 diverse allosteric sites that constitute the ‘Core-Diversity set’, with the largest PS-score of 0.491 (Fig. 1C, right). Interestingly, multiple allosteric sites within an individual protein are observed in the benchmarking sets (Supplementary Table S3). For example, four allosteric sites are currently known in the human glycogen phosphorylase (Fig. 1D). Using the ‘Core-Diversity set’ of allosteric sites, it is possible to develop efficient computational models to detect undiscovered allosteric sites in proteins and subsequently screen or design compounds that target these pockets. In addition, the pharmacophoric and structural properties of allosteric sites, coupled with the paradigms of allosteric ligand–protein interactions, can be extracted from the ‘Core set’ for use in pharmaceutical R&D.

3.2 Usage

ASBench provides a variety of interfaces and graphical visualizations to facilitate the viewing and analysis of allosteric sites in the benchmarking sets, including structural and pharmacophoric properties. This tool allows for browsing the sets and provides a search filter

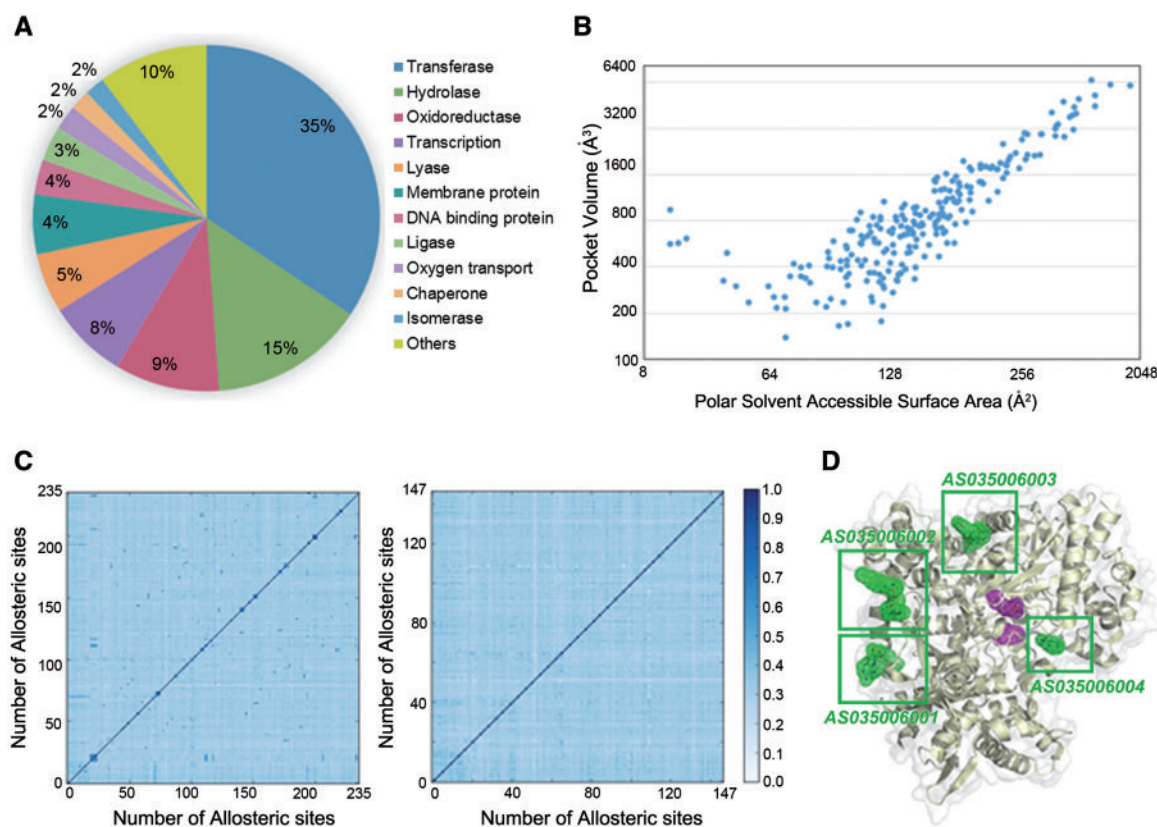


Fig. 1. Data features in the benchmarking sets. (A) Class distribution of proteins in the ‘Core set’. (B) Distribution of allosteric sites based on the properties of ‘Pocket volume’ and ‘Polar solvent accessible surface area’ in the ‘Core set’. (C) Structural similarity between allosteric sites in the ‘Core set’ (left) and the ‘Core-Diversity set’ (right). (D) Diverse functional ligands bound to human glycogen phosphorylase. Allosteric modulators are framed and colored in green, and the substrate is colored in purple

for flexible query (Supplementary Fig. S2). A full (or query) list of allosteric sites in the benchmarking sets integrated within a panel is displayed by clicking from the homepage of entries. Then, checking the selected site in the panel opens a new browser window with a detailed view of the representative allosteric complex containing the site. In addition, other complex structures containing the same site and different allosteric sites within the same protein are hyperlinked under 'Related Structures'. Finally, all data in the benchmarking sets can be downloaded using the 'Download' menu on the homepage.

4 Conclusion

The benchmarking sets in ASBench provide the high-quality data of allosteric sites for the development of efficient computational methods on the prediction of unknown allosteric sites and the exploration of unique allosteric ligand–protein interactions. It is thus expected to provide valuable avenues for allosteric drug discovery.

Funding

This work was supported by National Natural Science Foundation of China (81322046, 81302698, 81473137) and Shanghai Rising-Star Program (13QA1402300).

Conflict of Interest: none declared.

References

- Huang,W. *et al.* (2013) Allosite: a method for predicting allosteric sites. *Bioinformatics*, **29**, 2357–2359.
- Huang,Z. *et al.* (2014) ASD v2.0: updated context and novel features focusing on allosteric regulation. *Nucleic Acids Res.*, **42**, D510–D516.
- Lu,S. *et al.* (2014a) Harnessing allostery: a novel approach to drug discovery. *Med. Res. Rev.*, **34**, 1242–1285.
- Lu,S. *et al.* (2014b) Recent computational advances in the identification of allosteric sites in proteins. *Drug Discov. Today*, **19**, 1595–1560.
- Nussinov,R. and Tsai,C.-J. (2013) Allostery in disease and in drug discovery. *Cell*, **153**, 293–305.
- Panjikovich,A. and Daura,X. (2014) PARS: a web server for the prediction of protein allosteric and regulatory sites. *Bioinformatics*, **30**, 1314–1315.