


Big data and artificial intelligence discover novel drugs targeting proteins without 3D structure and overcome the undruggable targets

Huiqin He,¹ Benquan Liu,¹ Hongyi Luo,¹ Tingting Zhang,¹ Jingwei Jiang ²

To cite: He H, Liu B, Luo H, et al. Big data and artificial intelligence discover novel drugs targeting proteins without 3D structure and overcome the undruggable targets. *Stroke & Vascular Neurology* 2020;5:e000323. doi:10.1136/svn-2019-000323

Received 28 December 2019
Revised 29 February 2020
Accepted 3 March 2020
Published Online First
29 March 2020

ABSTRACT

The discovery of targeted drugs heavily relies on three-dimensional (3D) structures of target proteins. When the 3D structure of a protein target is unknown, it is very difficult to design its corresponding targeted drugs. Although the 3D structures of some proteins (the so-called undruggable targets) are known, their targeted drugs are still absent. As increasing crystal/cryogenic electron microscopy structures are deposited in Protein Data Bank, it is much more possible to discover the targeted drugs. Moreover, it is also highly probable to turn previous undruggable targets into druggable ones when we identify their hidden allosteric sites. In this review, we focus on the currently available advanced methods for the discovery of novel compounds targeting proteins without 3D structure and how to turn undruggable targets into druggable ones.

INTRODUCTION

As we all know, there are shortcomings in the traditional drug research and development field: (1) research and development cycle is too long; a new drug can take 14–15 years or more to develop¹; (2) research and development cost is expensive (US\$2.6 billion per new drug); (3) the success rate of drug research and development is very low. Although conventional methods such as combinatorial chemistry (CC) and high-throughput screening (HTS) can be used for effective drug development, they are cumbersome processes and require huge investment. In the past decade, as traditional approaches to drug development have become increasingly difficult, computer-aided drug design (CADD) had been used to improve the success rate of drug discovery. And it has become a powerful tool for the development of small molecule therapies with higher hit rates than HTS and CC.² When both HTS and CADD methods were used to screen a novel inhibitor of protein tyrosine phosphatase-1B, a key target for type II diabetes, CADD showed higher hit rates.³

CADD is a method based on computer chemistry to predict and calculate the

relationship between ligand and receptor through computer simulation, to carry out the design and optimisation of lead compounds. After years of efforts and exploration, the CADD method has gradually become mature and played an important role in drug design and development. These include proteomics, genomics, structural biology, bioinformatics, chemoinformatics and HTS.⁴ CADD usually has two different calculation methods for drug design: ligand-based drug design (ie, indirect drug design, LBDD) and structure-based drug design (ie, direct drug design, SBDD).⁵ LBDD is generally preferred only when compounds with activity data are available; SBDD relies on information about the real three-dimensional (3D) structure of proteins or 3D models reconstructed by homologous modeling.⁶ This article mainly introduces SBDD.

Protein Data Bank (PDB) is one of the most commonly used repositories of biomolecular structure information, most of which is determined by X-ray crystallography, while the smaller group is determined by cryogenic electron microscopy (Cryo-EM) and nuclear magnetic resonance techniques (NMR) spectrum⁷; this provides support for SBDD. However, it is difficult to obtain the actual structure of many specific proteins which cannot be expressed or purified at a certain scale. When the target protein does not have a well-resolved 3D structure, the homologous modelling method can be used to predict its structure and guide further experimental work, which is also a relatively reliable method.^{8,9} When the hidden allosteric sites located on proteins with 3D structure are identified, it is possible to turn previously undruggable targets into druggable ones. Based on our previous research experience in artificial intelligence (AI)-assisted drug discovery, we review the above methods and successful cases to provide a theoretical basis for subsequent research work.



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Jiangsu Key Lab of Drug Screening, China Pharmaceutical University, Nanjing, China

²Institute of Pharmacologic Science, China Pharmaceutical University, Nanjing, China

Correspondence to

Dr Jingwei Jiang;
jiangjingwei@cpu.edu.cn

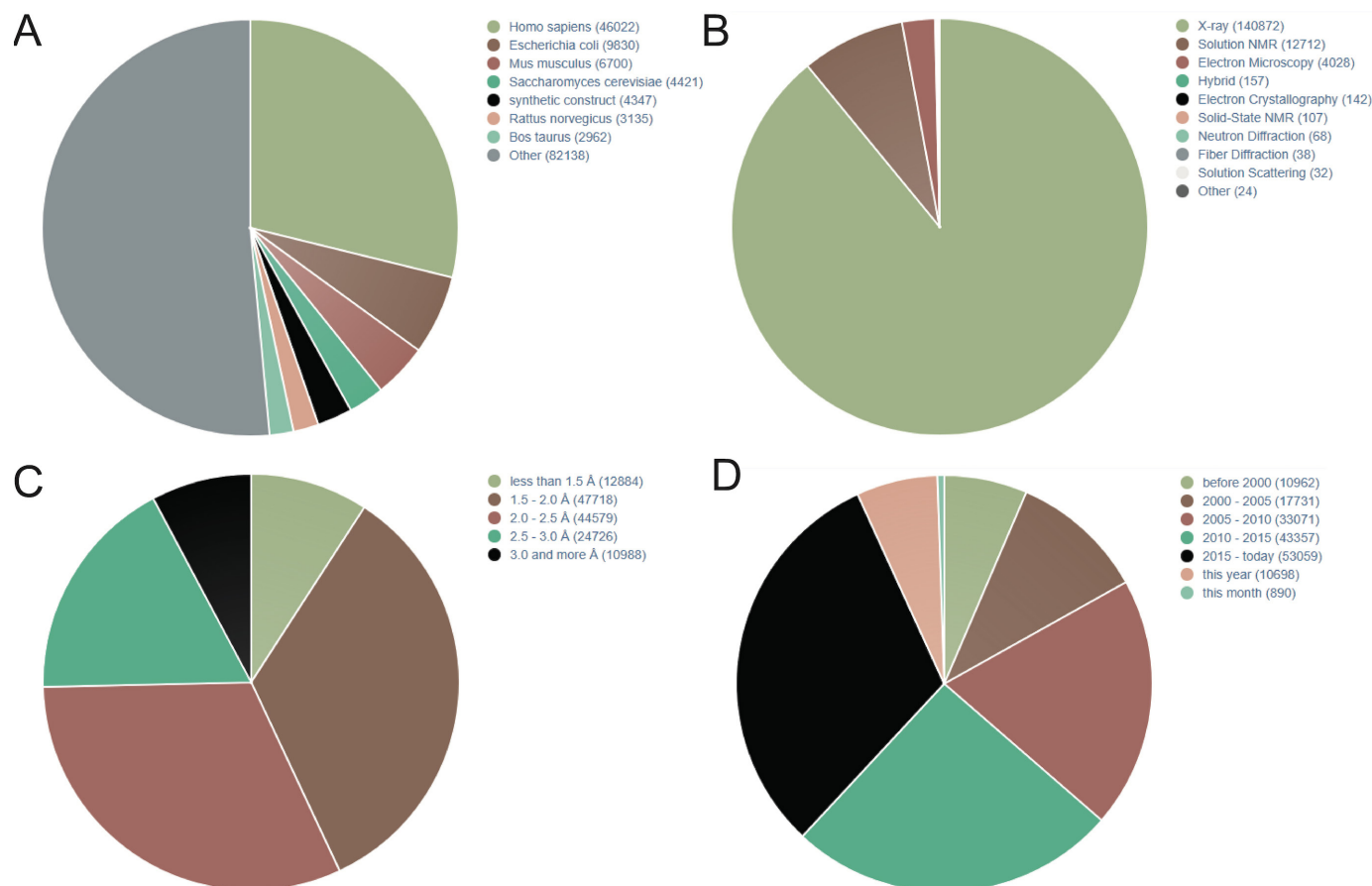


Figure 1 Statistics of three-dimensional (3D) protein structures in Protein Data Bank (PDB; December, 2019). (A) 3D protein structures distributed in different species. (B) 3D protein structures parsed by different methods. (C) The resolution of 3D protein structures in PDB. (D) Increasing trend of published 3D protein structures during last two decades. NMR, nuclear magnetic resonance.

TARGET PROTEIN DATABASE

The structure and function of protein are an important research content in life science. The 3D folding structure of protein natural folding determines its function. And it is very important for drug research and development. At present, the methods to obtain the 3D structure model of real protein include: X-ray crystal diffraction techniques, NMR and Cryo-EM techniques. With rapid development of these technologies, more and more protein structures are being analysed. So, researchers established many protein 3D structure information databases.

PDB is a commonly used protein database, which was founded in 1971. It is a single global repository, which stores the 3D structures of biological macromolecules and their complexes determined by experiments. It is also the first open-access digital resource in the biological field.¹⁰ In addition to the 3D coordinates, the PDB file also contains the general information needed for all sedimentary structures and the information unique to the structure determination method.¹¹ Figure 1 shows that according to PDB statistics, 3D protein structures are rapidly increasing. In addition, according to different research purposes, the researchers further constructed a

more detailed database of protein receptors, as shown in the following table. (table 1)¹²

PREDICTION OF PROTEIN 3D STRUCTURES

In recent years, with the development of genomics/transcriptomics/proteomics, more and more protein sequences have been identified. For proteins with no available 3D structure, the above method cannot be used for drug design. Therefore, researchers have designed a series of new methods to predict and construct protein 3D structure model based on bioinformatics. There are three common methods: (1) homologous modelling: the 3D structure of the target sequence is deduced by looking for its homologous protein. The commonly used software is SWISS-MODEL, which requires that a known structure with a consistency of the target sequence $\geq 30\%$ as the template. For example, Dhanavade *et al* predicted the structure of cysteine protease, an important target of Alzheimer's disease, by homologous modelling.¹³ (2) Folding recognition: also known as threading method, it is used to find proteins that have no significant homologous relationship with the target sequence but have the same structural folding type, and finally establish the structural

Table 1 Protein structure databases

Database	URL	Description
Therapeutic Target Database (TTD)	https://db.idrblab.org/ttd/	Known potential therapeutic targets, including protein and nucleic acid targets, were collected from the existing literature. However, the 3D structures of many of these targets have not yet been resolved.
Potential Drug-Target Database (PDTD)	http://www.dddc.ac.cn/pdtd/	Similar to TTD, PDTD contains only protein receptors and is a protein receptor with a known 3D structure.
Drug Adverse Reaction Database (DART)	http://bidd.nus.edu.sg/group/drt	This is for protein receptors that have adverse drug reactions
Small Molecule Transcription Factor Database (SM-TF)	http://zoulab.dalton.missouri.edu/SM-TF/	The structures of transcription factors that bind to small molecules are collected from the Protein Data Bank (PDB). The dysfunction of transcription factors is directly related to many diseases and is the target protein of many drugs.
PDBbind (a comprehensive collection of binding affinities for the protein-ligand complexes in the PDB)	WWW.pdbbind.org.cn	All protein-ligand complex structures with experimentally measured binding affinity data in the PDB database were collected. In addition to experimental binding affinity data (Kd, Ki or IC50), the database also provides the cleaned protein 3D structure and corresponding ligand structure.

3D, three-dimensional; IC50, half maximal (50%) inhibitory concentration of a substance.

model. The commonly used software is I-TASSER. It can start with amino acid sequences and construct a 3D structure model by recombining the fragments cut from the thread template, and then match the structure model with the known proteins in the functional database to predict the biological function of the target proteins.¹⁴ Spinocerebellar ataxia type 2 and 3 are two common autosomal dominant ataxia syndromes. Polyglutamine (poly-Q) is considered to be an important factor in the pathogenesis of the disease. Wen *et al* used I-TASSER to predict the structure of poly-Q.¹⁵ (3) Ab initio prediction: the 3D structure of the protein is predicted from the sequence itself. One can use AlphaFold,¹⁶ C-I-TASSER¹⁷ and so on. Both are a mixture of deep learning and traditional

algorithms. Because of the large amount of calculation, this method is not commonly used.

In practice, we conducted many protein models by homologous modelling with I-TASSER. Based on these models, we have successfully screened agonists/antagonists for given proteins without any published 3D structure (unpublished data provided by Jingwei Jiang). Figure 2 shows the recently published crystal structure of extracellular domain of VISTA (PDB code: 6o1l) and a corresponding VISTA homologous reconstructed model we constructed 2 years ago. We had obtained a very precise virtual VISTA 3D model (root mean SD=3.6 Å, structurally aligned to 6o1l) for virtual screening of drug discovery. As a result, several agonists/antagonists lead compounds

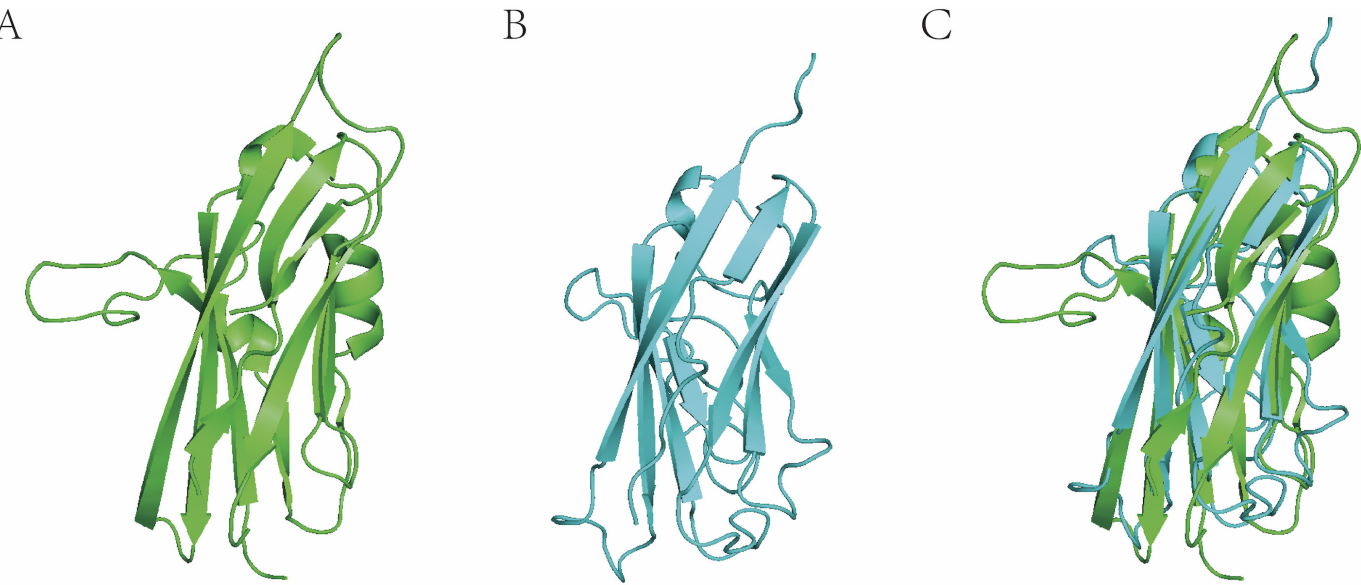


Figure 2 Three-dimensional (3D) protein structural alignment of VISTA (crystal structure) and homologous reconstructed model. (A) VISTA crystal structure (6o1l, resolution=1.85 Å). (B) VISTA model. (C) 3D structural alignment between real 3D structure and 3D model (root mean SD=3.6 Å).

effective at nanomolar concentration have been discovered (unpublished data provided by Jingwei Jiang).

Homologous modelling is the most successful method in protein structure prediction at present. The boundary between folding recognition and homologous is becoming more and more blurred. Homologous modelling plays an important role in finding templates, while folding recognition uses the details of sequences to improve the accuracy of sequence alignment. Ab initio prediction of protein structure is an immature research field, but the development potential is very great, because this method is to predict the high-level structure of protein from its primary structure. In practical applications, homologous modelling is generally considered first. If the consistency between the template sequence and the target sequence is less than 30%, then the folding recognition method is recommended. If the result of structural evaluation is still not ideal, or the target sequence is less than 100 amino acids, ab initio prediction can be used. However, its performance is not good if the sequence is longer than 100 amino acids.

PREDICTION OF LIGAND-BINDING SITES ON PROTEIN 3D MODELS

The biochemical function of protein plays an important role in organism. However, the physical properties and physiological functions of many proteins have not been discovered. Proteins perform biological functions mainly by binding to their ligands. Identification of binding sites in proteins is essential for understanding their interactions with ligands as well as other proteins, and is very helpful for the discovery and rational design of therapeutic compounds.¹⁸ Although many methods have been developed to screen ligand-binding sites based on highly similar protein structures, those included in PDB are very limited. In fact, many proteins do not have a resolved crystal structure, using predicted low-resolution protein models to find and verify ligand binding sites has become a challenging problem. At present, with the continuous development of computer technology, we can make full use of the methods and advantages of AI to predict the binding sites of ligands on protein 3D models.

For some given amino acid sequences of benchmark proteins, the protein structure can be predicted by a variety of servers such as I-TASSER¹⁹ and PredictProtein.²⁰ Subsequently, a variety of developed programs can be used to predict protein–ligand binding sites. For example, BSP-SLIM,²¹ as a method of ligand–protein blind docking using low-resolution protein structures, can use a given sequence to predict protein structure models through I-TASSER. The assumed ligand-binding sites were transferred from the holographic template structure of the model and the ligand–protein docking conformation was constructed by using the shape and chemical matching of the ligand and the negative image of the binding pocket. In addition, according to the Structural Classification of Proteins database, some similar or very distant

homologous proteins can have common binding sites.²² According to this, FINDFITE²³ can identify the ligand-bound structural template by algorithm and superimpose it on the structural model of distant or near homologous proteins to identify the hypothetical binding sites. On the other hand, the improved version of FINDSITE^{fit} of FINDSITE can elevate the accuracy and accuracy by filtering out the false-positive ligands in the thread recognition template.²⁴ In this way, the ligand-binding sites on the 3D model of proteins without crystal structure can be predicted through a variety of available servers.

In practice, we have successfully applied large-scale structural alignment to predict a lot of ligand-binding sites and discover corresponding novel agonists/antagonists (unpublished data provided by Jingwei Jiang). Figure 3 shows the 3D structure of gasdermin D (GSDMD). We structurally aligned the human GSDMD (6n9o) to mouse GSDMD (6n9n) and predicted the polymerisation protein-binding pocket located on human GSDMD. Based on the predicted pocket, we have successfully discovered several novel antagonists inhibiting the polymerisation of human GSDMD both in vitro and in vivo within 1 month (unpublished data provided by Jingwei Jiang).

UNDRUGGABLE TARGETS AND OVERCOMING UNDRUGGABLE TARGETS

The term ‘Undruggable’ is used to describe a protein that is not pharmacologically capable of being targeted; however, efforts are being made to turn these proteins into ‘druggable’, so it is more appropriate to describe it as ‘difficult to administer’ or ‘currently unavailable’.²⁵ Over the past few decades, many pathways of tumourigenesis have been discovered with the efforts of scientists. Many proteins involved in cancer development, particularly kinases, provide drug targets. However, many proteins, such as RAS, MYC and p53, are considered as ‘Undruggable targets’. RAS mutations are early events in tumour progression, and there is substantial experimental evidence that sustained expression of RAS mutations is necessary for tumour maintenance. Despite more than 30 years of hard work, effective pharmacological inhibitors of RAS oncoprotein have not been approved by FDA (Food and Drug Administration) yet.²⁶ MYC is a transcription factor that is involved in a variety of cancer-promoting programs and is often overexpressed in advanced stages of cancer, and various attempts have been made to overcome its lack of medicinally binding pockets.²⁵ p53 is the most frequently altered gene in human cancer, with p53 mutations present in about 50% of all invasive tumours. It is clear that the p53 mutant protein is an important target, but it has traditionally been considered undruggable.²⁷ These targets are due to their large or flat protein–protein interaction interfaces or lack deep protein-binding pockets.²⁸ Therefore, overcoming these ‘so-called undruggable targets’ is one of the main challenges of drug discovery.

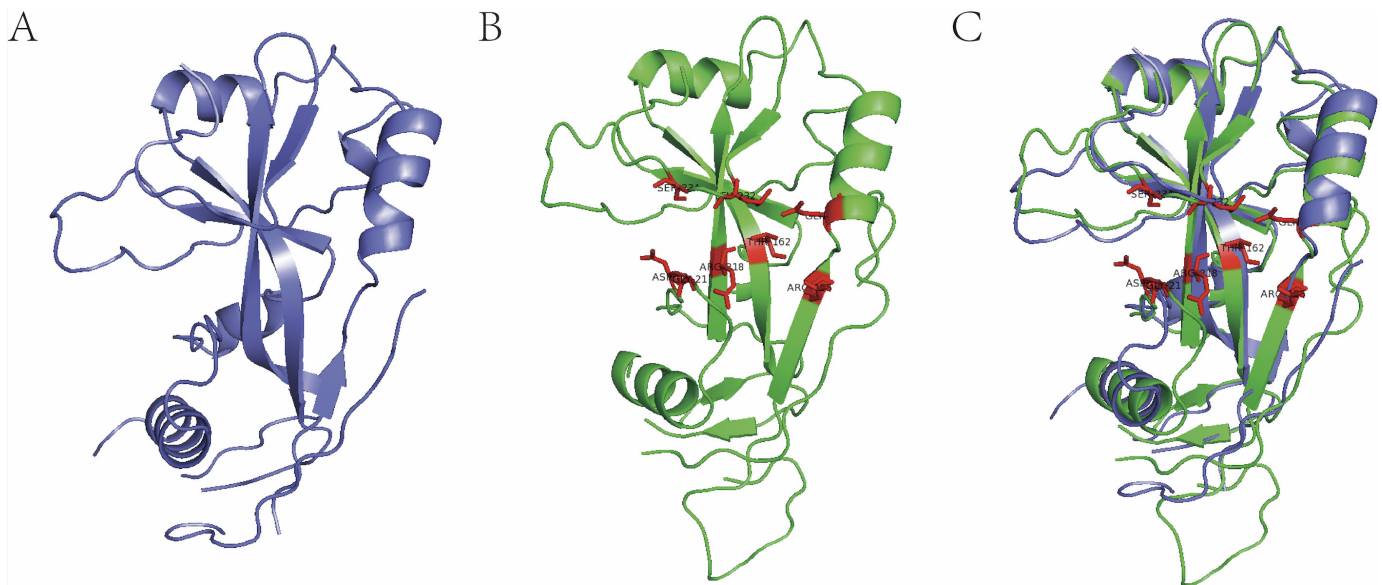


Figure 3 Identification of polymerisation pocket on gasdermin D (GSDMD) by 3D protein structural alignment. (A) Human GSDMD (6n9o). (B) Mouse GSDMD (6n9n), red amino acid residues are verified polymerisation-associated residues. (C) 3D alignment for the identification of corresponding polymerisation pocket residues on human GSDMD.

In recent years, researchers have proposed many new methods to solve this problem: (1) induce target protein degradation²⁹; (2) block pathways downstream of the target³⁰; (3) discover hidden allosteric sites, and so on. The discovery of hidden allosteric sites in these three approaches will greatly expand the types of existing drug targets and provide opportunities for the discovery of allosteric drugs. Therefore, the intervention of hidden allosteric sites is an effective method.

HIDDEN ALLOSTERIC SITES

In the design of targeted drugs, the combination of drugs and biological macromolecules is like the relationship between key and keyhole. An effective drug is like a key to the configuration of targeted macromolecules, so the question becomes to find the 'keyhole' and quickly match the 'key'. Hidden allosteric sites are more challenging than allosteric sites because these sites are not visible in proteins even in the well-resolved crystal structure. Since the biological macromolecular protein in solution is not static, but dynamic, there is a combination of multiple conformations coexisting with different energies. The hidden allosteric site is a binding pocket that does not exist in the protein crystal structure, but it becomes available as the protein fluctuates. Such sites may have unknown biological functions and can be used as important targets for drug design.³¹

There are several methods to identify hidden allosteric sites³²: (1) large-scale unbiased molecular dynamics (MD) simulations: in theory, a variety of protein conformations can be obtained by large-scale unbiased MD simulation of proteins. Several low-filled conformations obtained by MD simulation cannot be captured by experimental methods. These small conformations may have hidden allosteric

sites, which can be used in the design of allosteric regulators. For instance, the discovery of allosteric sites for the b2-adrenalin receptors³³; (2) combined ensemble-based docking and MD simulations, for example, andrographolide derivatives can inhibit the conformational transformation of RAS³⁴; (3) accelerated MD simulations: compared with traditional MD simulation, accelerated MD by introducing a non-negative press or potential into the potential energy surface can promote the conformational transition between low energy states and effectively enhance the conformational sampling space of proteins, especially for proteins with slow time scale dynamics. Yang used this method to identify potential allosteric sites of interleukin-1 receptor³⁵; (4) MD-based Markov state analysis: the Markov state model is based on a large number of MD simulations. The advantage of these models is that they can identify hidden allosteric sites that are usually located in small conformations, which can be captured from multiple simulations rather than a single simulation, for example, the discovery of the allosteric site of TEM-1 β -lactamase.³¹ In conclusion, the discovery of hidden allosteric sites has greatly expanded the range of available drug targets, and these calculations have enabled us to predict hidden allosteric sites with greater accuracy, providing an opportunity for the discovery of allosteric drugs.

In addition, some tool used online can also be used to identify allosteric sites. (1) Allosite (<http://mdl.shsmu.edu.cn/AST>): this is the first publicly available allosteric site recognition method established by Professor Zhang Jian's research group of Shanghai Jiaotong University in 2013.³⁶ This method discovers and validates new targeted allosteric on many proteins such as CDK2 α .³⁷ (2) GO model (<http://www.ligbuilder.org/cavity/home.php>):

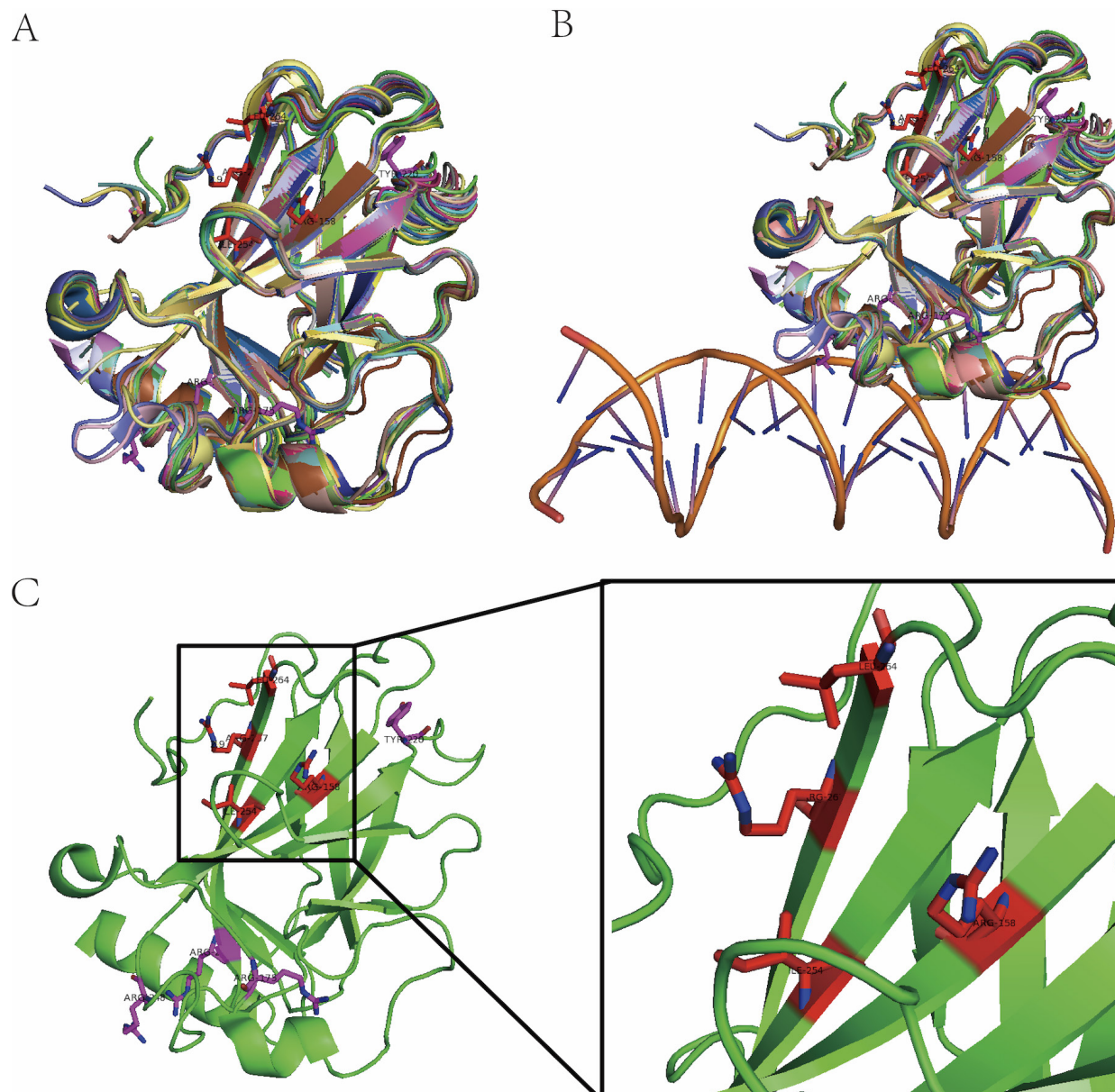


Figure 4 Identification of allosteric pocket on mutant p53. (A) Three-dimensional (3D) structural alignment of 43 p53 DNA binding domain. (B) DNA binding pocket is composed of fluctuating coiled structures (double helix structure represents DNA). (C) Allosteric pocket (red amino acid residues) on p53(4lof) predicted by molecular dynamic simulation and hot spot mutant residues (purple).

Professor Lai Luhua of Peking University established a coarse-grained two-state Go model to identify allosteric sites. This method predicted two potential allosteric sites in *Escherichia coli* phosphoglycerate lactate dehydrogenase.³⁸ (3) PARS (<http://bioinf.uab.cat/pars>): Panjkovich and Daura developed a method to identify the location of allosteric sites by normal mode analysis. PARS has successfully identified a number of previously discovered allosteric sites.³⁹ (4) SPACER (<http://allostery.bii.a-star.edu.sg>): Goncarencu *et al* developed a method of allosteric site recognition based on the combination of Monte Carlo simulation and NMR.⁴⁰

In the past few years, we have successfully used MD simulation to identify a potential allosteric pocket on

DNA binding domain of p53 (figure 4). Based on this pocket, we have screened an allosteric lead compound both effective efficient in vitro and in vivo (unpublished data provided by Jingwei Jiang).

CONCLUSION REMARK

In recent years, with the development of computational advancements and the unremitting efforts of scientists, they have improved innovative methods to overcome targets that are previously considered undruggable. At present, some lead compounds related to the undruggable targets have been developed. Some small molecules are already in clinical trials such as AMG-510 (a KRAS

G12C inhibitor, Clinical Trial: NCT03600883). Soon, AI technology has also developed by leaps and bounds. Scientists have used computer technology to assist in the discovery of new drug targets and in search for hidden targets as well as undruggable targets, and have achieved certain positive results. Although AI cannot guarantee 100% accuracy just by computation, it still has many advantages compared with traditional methods in terms of cost and efficiency. This will also be the main trend in the future to discover potential allosteric sites on undruggable targets.

Contributors All authors wrote the manuscript. JJ provided guidance and modifications.

Funding This work is supported by NSFC (No. 81872892 and No. 2018ZX09735001-004), "Double First-Class" University project (No. CPU2018GY20 and No. CPU2018GY38).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Jingwei Jiang <http://orcid.org/0000-0002-9163-3633>

REFERENCES

- Dar KB, Bhat AH, Amin S, et al. Modern computational strategies for designing drugs to curb human diseases: a prospect. *Curr Top Med Chem* 2018;18:2702–19.
- Katsila T, Spyroulias GA, Patrinos GP, et al. Computational approaches in target identification and drug discovery. *Comput Struct Biotechnol J* 2016;14:177–84.
- Doman TN, McGovern SL, Witherbee BJ, et al. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* 2002;45:2213–21.
- Cheng F, Liang H, Butte AJ, et al. Personal Mutanomes meet modern oncology drug discovery and precision health. *Pharmacol Rev* 2019;71:1–19.
- Sliwoski G, Kothiwale S, Meiler J, et al. Computational methods in drug discovery. *Pharmacol Rev* 2014;66:334–95.
- Kalyanamoorthy S, Chen Y-PP. Structure-based drug design to augment hit discovery. *Drug Discov Today* 2011;16:831–9.
- Rose PW, Prić A, Altunkaya A, et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res* 2017;45:D271–81.
- Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. *Brief Bioinform* 2009;10:579–91.
- Ul-Haq Z, Saeed M, Halim SA, et al. 3D structure prediction of human β 1-adrenergic receptor via threading-based homology modeling for implications in structure-based drug designing. *PLoS One* 2015;10:e0122223.
- Burley SK, Berman HM, Kleywegt GJ, et al. Protein data bank (PDB): the single global macromolecular structure Archive. *Methods Mol Biol* 2017;1607:627–41.
- Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res* 2000;28:235–42.
- Xu X, Huang M, Zou X. Docking-based inverse virtual screening: methods, applications, and challenges. *Biophys Rep* 2018;4:1–16.
- Dhanavade MJ, Jalkute CB, Barage SH, et al. Homology modeling, molecular docking and MD simulation studies to investigate role of cysteine protease from *Xanthomonas campestris* in degradation of A β peptide. *Comput Biol Med* 2013;43:2063–70.
- Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 2015;43:W174–81.
- Wen J, Scoles DR, Facelli JC. Effects of the enlargement of polyglutamine segments on the structure and folding of ataxin-2 and ataxin-3 proteins. *J Biomol Struct Dyn* 2017;35:504–19.
- AlQuraishi M. AlphaFold at CASP13. *Bioinformatics* 2019;35:4862–5.
- Zheng W, Li Y, Zhang C, et al. Deep-learning contact-map guided protein structure prediction in CASP13. *Proteins* 2019;87:1149–64.
- Konc J, Janezic D. ProBIS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* 2010;26:1160–8.
- Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010;5:725–38.
- Rost B, Yachdav G, Liu J. The PredictProtein server. *Nucleic Acids Res* 2004;32:W321–6.
- Lee HS, Zhang Y. BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins* 2012;80:93–110.
- Russell RB, Sasieni PD, Sternberg MJ. Supersites within superfolds. Binding site similarity in the absence of homology. *J Mol Biol* 1998;282:903–18.
- Brylinski M, Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc Natl Acad Sci U S A* 2008;105:129–34.
- Zhou H, Skolnick J. FINDSITE(comb): a threading/structure-based, proteomic-scale virtual ligand screening approach. *J Chem Inf Model* 2013;53:230–40.
- Dang CV, Reddy EP, Shokat KM, et al. Drugging the 'undruggable' cancer targets. *Nat Rev Cancer* 2017;17:502–8.
- Cox AD, Fesik SW, Kimmelman AC, et al. Drugging the undruggable RAS: mission possible? *Nat Rev Drug Discov* 2014;13:828–51.
- Duffy MJ, Synnott NC, Crown J. Mutant p53 as a target for cancer treatment. *Eur J Cancer* 2017;83:258–65.
- McCormick F. KRAS as a therapeutic target. *Clin Cancer Res* 2015;21:1797–801.
- Ray D, Cuneo KC, Rehemtulla A, et al. Inducing oncoprotein degradation to improve targeted cancer therapy. *Neoplasia* 2015;17:697–703.
- Hallin J, Engstrom LD, Hargis L, et al. The KRAS^{G12C} Inhibitor MRTX849 Provides Insight toward Therapeutic Susceptibility of KRAS-Mutant Cancers in Mouse Models and Patients. *Cancer Discov* 2020;10:54–71.
- Bowman GR, Bolin ER, Hart KM, et al. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc Natl Acad Sci U S A* 2015;112:2734–9.
- Lu S, Ji M, Ni D, et al. Discovery of hidden allosteric sites as novel targets for allosteric drug design. *Drug Discov Today* 2018;23:359–65.
- Dror RO, Pan AC, Arlow DH, et al. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci U S A* 2011;108:13118–23.
- Hocker HJ, Cho K-J, Chen C-YK, et al. Andrographolide derivatives inhibit guanine nucleotide exchange and abrogate oncogenic Ras function. *Proc Natl Acad Sci U S A* 2013;110:10201–6.
- Yang C-Y. Identification of potential small molecule allosteric modulator sites on IL-1R1 ectodomain using accelerated conformational sampling method. *PLoS One* 2015;10:e0118671.
- Huang W, Lu S, Huang Z, et al. Allosteric: a method for predicting allosteric sites. *Bioinformatics* 2013;29:2357–9.
- Jiang H-M, Dong J-K, Song K, et al. A novel allosteric site in casein kinase 2 α discovered using combining bioinformatics and biochemistry methods. *Acta Pharmacol Sin* 2017;38:1691–8.
- Qi Y, Wang Q, Tang B, et al. Identifying allosteric binding sites in proteins with a two-state Go model for novel allosteric effector discovery. *J Chem Theory Comput* 2012;8:2962–71.
- Panjikovich A, Daura X. PARS: a web server for the prediction of protein allosteric and regulatory sites. *Bioinformatics* 2014;30:1314–5.
- Goncearenco A, Mitternacht S, Yong T, et al. SPACER: server for predicting allosteric communication and effects of regulation. *Nucleic Acids Res* 2013;41:W266–72.