Send Orders for Reprints to reprints@benthamscience.net

Current Medicinal Chemistry, XXXX, *XX*, 1-00 1

REVIEW ARTICLE

Binding Specificity and Local Frustration in Structure-based Drug Discovery

Zhiqiang Yan[1,2,*], Yuqing Li[2,3], Ying Cao[1,2], Xuetao Tao[4], Jin Wang[1,2,5,*] and Yongsheng Jiang[1,*]

[1]*Joint Research Centre on Medicine, the Affiliated Xiangshan Hospital of Wenzhou Medical University, Ningbo, Zhejiang, China;* [2]*Center for Theoretical Interdisciplinary Sciences, Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, Zhejiang, China;* [3]*Zhejiang Key Laboratory of Petrochemical Environmental Pollution Control, National Engineering Research Center for Marine Aquaculture, Marine Science and Technology College, Zhejiang Ocean University, Zhoushan, Zhejiang, China;* [4]*The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China;* [5]*Department of Chemistry and Physics, State University of New York at Stony Brook, Stony Brook, NY, USA*

ARTICLE HISTORY

Received: December 12, 2024
Revised: February 15, 2025
Accepted: February 28, 2025

DOI:
10.2174/0109298673376099250428054846

Abstract: Evolution has optimized proteins to balance stability and function by reducing unfavorable energy states, leading to regions of flexibility and frustration on protein surfaces. These locally frustrated regions correspond to functionally important areas, such as active sites and regions for ligand binding and conformational plasticity. Typical strategies of structure-based drug discovery primarily concentrate on enhancing the binding affinity during compound screening and target identification. However, this often overlooks the binding specificity, which is critical for distinguishing specific binding partners from competing ones and avoiding off-target effects. According to the energy landscape theory, optimization of the intrinsic binding specificity involves globally minimizing the frustrations existing in the biomolecular interactions. Recent studies have demonstrated that identifying local frustrations provides a promising approach for screening more specific compounds binding with targets, and quantifying binding specificity complements typical strategies that focus on binding affinity only. This review explores the principles and strategies of computationally quantifying the binding specificity and local frustrations and discusses their applications in structure-based drug discovery. Moreover, given the advancements of artificial intelligence in protein science, this review aims to motivate the integration of AI and available approaches in quantifying the binding specificity and local frustration. We expect that an AI-powered prediction model will accelerate the drug discovery process and improve the success rate of hit compounds.

Keywords: Energy landscape, biomolecular recognition, binding specificity, local frustration, drug screening, binding site, cryptic site.

1. INTRODUCTION

Drug discovery involves multiple stages, from target and lead compound identifications to clinical trials. The process is generally time-consuming and expensive and encounters a high risk of failure during the clinical phase [1, 2]. Therefore, preclinical development is critical to reduce the cost and improve the success rate of a lead compound [3, 4]. Structure-based drug discovery (SBDD) plays a crucial role throughout the drug development process by providing a rational basis to increase the overall success rate [4, 5]. The main concept of SBDD is the rational design and optimization of drug candidates based on the three-dimensional structure of the biological target (often a protein) [6, 7]. It directly leverages the detailed structural information of the target to guide the design of drug candidates that can effectively bind to and modulate the target's activity. With the breakthrough of artificial intelligence, like Alphafold2, in predicting protein

*Address correspondence to these authors at the Joint Research Centre on Medicine, the Affiliated Xiangshan Hospital of Wenzhou Medical University, Ningbo, Zhejiang, China;
E-mails: zqyan@ucas.ac.cn; Jin.wang.1@stonybrook.edu; shenren34127@163.com

structures, high-quality structure information readily available can accelerate the process of SBDD and significantly enhance its capability and efficiency [8-11].

The key step of SBDD is computational docking and scoring. Traditionally, the goal of this step is to identify the binding site and maximize binding affinity, which is employed and ranked in the virtual screening [12, 13]. However, this usually neglects the importance of the binding specificity of the drug-target system [14-18]. For instance, a drug is intended to selectively bind to its biological target (on-target effect) and avoid side effects by binding to unintended targets (off-target effect). In principle, binding specificity refers to how selectively a drug candidate binds to its intended target compared to all other potential targets. In practice, the implementation is impossible both computationally and experimentally because the targets are not all known, and the task is huge even if all the targets are available. This demands the development of strategies to address the quantification of binding specificity. The issue is illustrated through a thought experiment in which shuffling the target sequences is analogous to randomizing the binding sites across the entire surface of the target protein (Fig. **1**) [14, 15, 19, 20]. This shifts the concept of binding specificity from the conventional approach. In the conventional approach, a compound identifies a specific protein among all possible protein targets. Instead, the intrinsic binding specificity refers to a compound identifying a specific binding site (or pocket) on the target protein. In other words, intrinsic binding specificity focuses solely on the interaction between the compound and the known protein target rather than the protein universe. Therefore, intrinsic binding specificity merely involves the optimization of the interactions between the compound and the known protein target.

Naturally occurring proteins have been evolved to minimize energetic conflicts in balancing structural stability and functional binding. Besides being optimized for folding, proteins have evolved primarily for functional binding, often leading to a trade-off where stability and function may conflict [21-25]. Even if the protein is completely folded, not all conflicting interactions have been minimized [26, 27]. This results in native proteins being marginally stable, with regions of flexibility and frustration that are crucial for their biological functions [28-31]. It has been demonstrated that these locally frustrated regions generally correspond to functionally important areas, such as active sites and regions involved in conformational changes. Identifying these frustrated regions cannot only help in understanding how a protein performs its biological function and how it interacts with other molecules but also provide potential binding sites for drugs [32-34]. By recognizing these regions, researchers can design molecules that alleviate frustration by binding to these patches, thereby stabilizing the protein-ligand complex and potentially increasing the binding specificity. This review attempts to introduce the principles and strategies for computationally quantifying the binding specificity and local frustrations, as well as their applications in structure-based drug discovery. Furthermore, this review outlines potential future developments in utilizing binding specificity and local frustration in drug discovery.

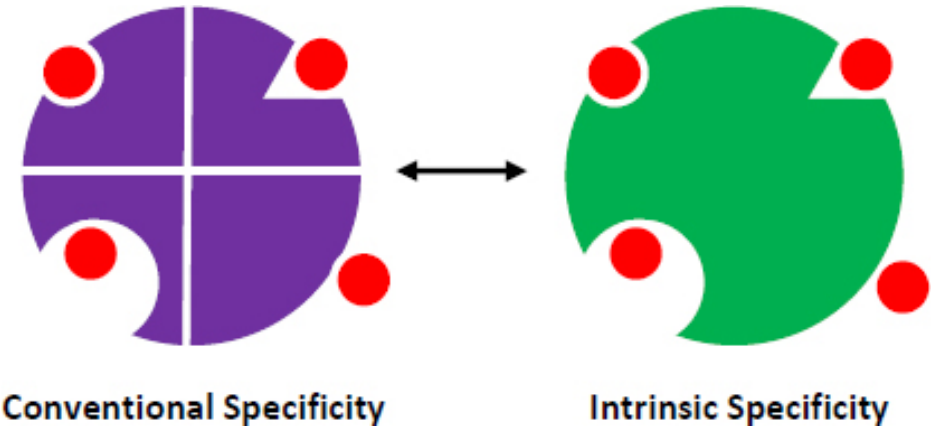

Fig. (1). Thought experiment on the equivalence of conventional and intrinsic binding specificity for protein-ligand binding. Conventional binding specificity refers to the ligand (red) binding onto the intended protein target rather than all other possible proteins (purple), while intrinsic binding specificity refers to the ligand binding onto the intended binding site rather than all other binding sites (green); they are equivalent if the protein is large enough compared to the ligand. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

2. METHODOLOGY

The reviewed literature was searched from Google Scholar, ScienceDirect, and Web of Science. The keywords used for searching were "biomolecular recognition", "binding specificity", "local frustration", "drug screening", "binding site", "cryptic site", "energy function", "multi-state conformations," and the combination of them. The searched literature was included if it met the following criteria. First, the published or released year should be between 2000 and 2025, except for the original literature on the funneled energy landscape of protein folding. Second, the written language is English only. Third, the literature irrelevant to the topics of binding specificity and local frustration, as well as drug discovery, was manually disregarded.

In addition, the web server for computing local frustrations (Frustrometer server) was searched on Google by using the keyword "frustration calculation". The data for drawing the figures were obtained from the corresponding literature reviewed.

3. INTRINSIC SPECIFICITY OF BIOMOLECULAR BINDING

3.1. Quantification of Intrinsic Binding Specificity for Protein-ligand Interactions

The quantification of the intrinsic binding specificity was carried out by applying the well-established energy landscape theory of protein-ligand binding [14, 15, 19, 20]. According to the theory, specific protein-ligand binding interactions adhere to the principle of minimal frustration, which leads to funnelled energy landscape (Fig. **2**). According to the principle, the conflicted interactions in the specific protein-ligand binding should be minimally frustrated to enable high affinity and specificity, *i.e.*, the energetic conflicts or unfavorable interactions are minimized once the protein and ligand bind together. A dimensionless parameter was derived to quantify the funnelness of the energy landscape of protein-ligand binding [14, 15, 19, 20], which is computed using Eq. **1**:

$$\Lambda = \sqrt{\frac{K_B}{2S}}\frac{\delta E}{\Delta E} \quad (1)$$

Λ can be readily quantified by computationally generating an ensemble of binding decoys. δE is the energy gap between the energy of the native binding conformation and the average energy of the decoy conformation ensemble, and ΔE is the energy roughness or the width of the Gaussian-like energy distribution of the conformation ensemble (Fig. **2**), K_B is the Boltzmann constant, and S is the conformational entropy of the system. Larger Λ means higher binding specificity. Λ can be used as a quantitative global specificity measure for biomolecular binding.

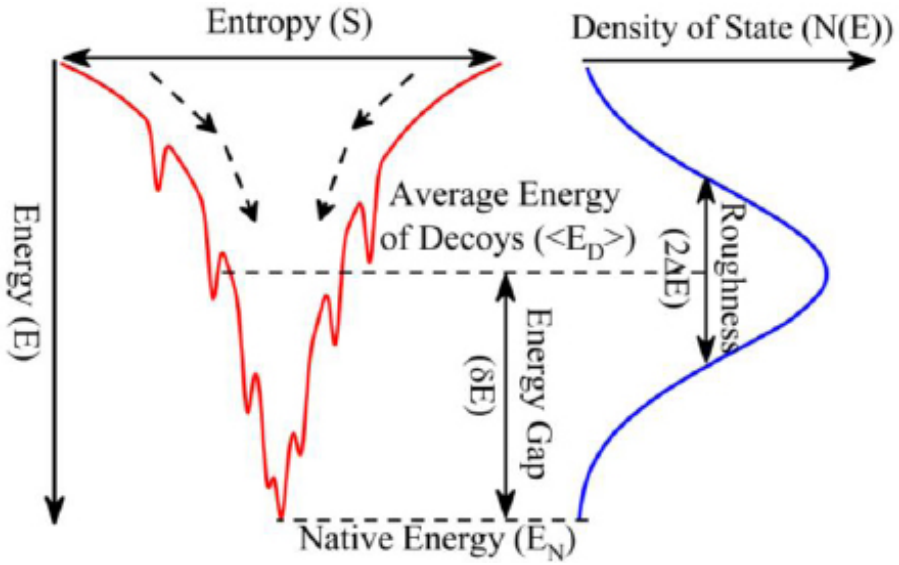

Fig. (2). Quantification of binding specificity. Binding energy landscape with a funneled shape towards the "native" state, with the corresponding distribution of binding energies, including the roughness or standard deviation of energy (ΔE), the average energy of decoys (< E_D >), and the energy gap (δE). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

3.2. Two-dimensional Drug Screening with Both Specificity and Affinity

An ideal target-based drug is designed to achieve both high affinity and specificity for its target within the crowded cellular environment. Affinity determines the stability of the drug-target complex, while specificity ensures the drug's ability to selectively bind to its intended target over other competing biomolecules and to differentiate itself from other small molecules [14-19]. Traditional virtual drug screening methods often focus solely on ranking binding affinity, overlooking the crucial role of binding specificity. The ignorance of binding specificity may contribute significantly to the drug's side effects.

To emphasize the critical role of binding specificity, a two-dimensional drug screening strategy has been proposed, which incorporates specificity into the screening process [14, 15, 19]. This approach has been computationally validated in the drug-cyclooxygenase (COX) system. Traditional nonsteroidal anti-inflammatory drugs (NSAIDs) are considered nonselective because they inhibit both COX-1 and COX-2 enzymes. While the inhibition of COX-2 accounts for the anti-inflammatory effects of NSAIDs, the inhibition of COX-1 can lead to toxicity and side effects, such as peptic ulceration. Therefore, COX-1 inhibition is often undesirable, whereas COX-2 inhibition is intended. COX-2 selective drugs represent a newer class of NSAIDs that specifically inhibit COX-2, allowing COX-1 to perform its essential functions. These drugs serve as selective pain relievers and fever reducers without the associated side effects of traditional NSAIDs [35]. COX-2 and COX-1 are isoenzymes that compete for binding with small molecules (Fig. **3a**). For testing the screening performance of a two-dimensional drug screening strategy [15, 19], 37 selective and 20 non-selective drugs were manually collected

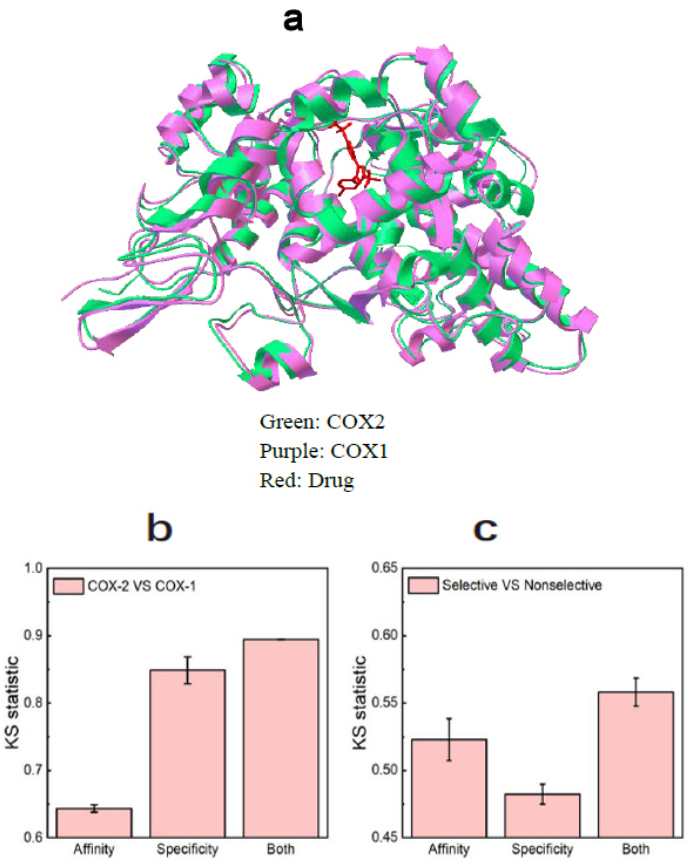

Fig. (3). Two-dimensional drug screening test on the drug-COX system. (**a**) The aligned structures of COX2 (green, PDB ID: 1CX2) and COX-1 (purple, PDB ID: 1Q4G) with a selective drug (SC-558) inside the pocket (red); **(b)** the Kolmogorov-Smirnov statistic (KS statistic) of discrimination between COX-2 and COX-1 when binding with selective drugs of COX-2; (**c**) KS statistic of discrimination between selective drugs and nonselective drugs upon binding with COX-2. The data for drawing is taken from the literature [19]. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

from the published literature. Kolmogorov-Smirnov (K-S) test was carried out to quantify the discrimination of COX-2 (PDB ID: 1CX2) against COX-1 (PDB ID: 1Q4G) when binding with selective drugs and the discrimination of selective drugs against non-selective drugs upon binding with COX-2. It has been demonstrated that the two-dimensional drug screening strategy is more effective than affinity-based methods alone in distinguishing between selective and nonselective COX-2 drugs (Fig. **3c**) [15, 19]. Additionally, it provides a greater ability to differentiate between COX-2 and COX-1 binding of selective drugs (Fig. **3b**). The enhanced discriminatory power of two-dimensional screening would improve the identification of selective drugs and increase the hit rate of lead compounds from compound libraries.

3.3. Scoring Function Development with Both Specificity and Affinity

The scoring function is central to structure-based virtual database screening [36, 37]. Recently, AI-based AlphaFold3 has revolutionized the prediction of biomolecular interactions [38], significantly improving the structure quality of predicted binding complexes and greatly enhancing docking capabilities. However, the higher prediction accuracy of structural models does not necessarily translate to increased scoring or discriminatory power in distinguishing specific binding complexes from competing ones [39, 40]. Previous studies have attempted to incorporate binding specificity into the development of scoring functions, resulting in an approach known as SPA (SPecificity and Affinity) [15, 41-45]. The concept of SPA is that maximizing intrinsic binding specificity involves globally minimizing the frustrations present in biomolecular interactions. SPA has demonstrated superior scoring and discriminatory power compared to widely used academic and industrial scoring functions [15, 41-45]. This development strategy establishes a framework for optimizing scoring functions by integrating ligand binding specificity.

4. LOCAL FRUSTRATION OF BIOMOLECULAR INTERACTIONS

4.1. Quantification of Local Frustration

Strong, energetic conflicts are generally minimized in folded native states, allowing protein sequences to spontaneously fold according to the principle of minimal frustration [46-48]. The principle of minimal frustration guarantees the funneled energy landscape of protein folding. However, local deviations from this principle are necessary to encode the complex energy landscapes required for active biological functions [28-31]. Achieving minimal frustration across the entire structure often results in localized frustration in specific regions of a protein, where the interactions between amino acids are not fully energetically minimized. These frustrated regions are crucial for maintaining functional flexibility and can be compatible with ligand binding. For instance, residues in or near the binding site may experience energetic conflicts that are relieved and minimally frustrated upon ligand recognition.

Quantifying local frustration has proven to be an effective method for identifying these functionally important regions within a protein's structure. Local frustration can be assessed in three ways: configurational frustration and mutational frustration for the contact, and mutational frustration for the position [49]. The frustration index F_i for a residue position or contact i is calculated using Eq. **2**:

$$F_i = \left(E_i^N - < E_i^U >\right) / \sqrt{(1/M)\sum_{k=1}^{M}\left(E_i^N - < E_i^U >\right)^2} \quad (2)$$

Where E_i^N is the "native" energy of residue position or contact i, and E_i^U is the reference energy of the same residue position or contact, obtained by scuffling the amino acids or local environment M times. In this manner, the individual position or contact can be roughly classified as being either minimally frustrated, highly frustrated, or neutrally frustrated with regard to their frustration level. For example, a residue position or contact is defined as minimally frustrated if $F_i < -0.78$, highly frustrated if $F_i > 1.0$, and neutrally frustrated if the values fall between these two limits [49]. The frustration index can be viewed as a localized version of the global specificity measure Λ for biomolecular binding [14, 15, 19, 20]. This energy landscape theory-inspired algorithm has been developed as a web server Frustratometer and a package FrustratometerR, which are both available to quantify the degree of local frustration for proteins [50, 51].

4.2. Identification of Drug-binding Sites with Local Frustration

Studies have reported that enzyme-active sites and regions responsible for conformational changes are often enriched with highly frustrated interactions [29-32, 49]. By analysing the frustration indices with or without the binding of substrate, it is straightforward to determine whether the conflict of frustrations is mini-

mized upon binding. Drug binding specificity is closely linked to minimal frustration, which discriminates frustrated interactions. Selective drugs tend to bind effectively to frustrated pockets in protein targets that become minimally frustrated upon binding [33]. Consequently, local frustration indices are valuable for identifying potential ligand-binding sites and assessing whether a binding site for a specific drug will become minimally frustrated or remain frustrated after binding. If the site still remains frustrated after binding, then the protein may undergo structural rearrangements, or the drug may change its configuration and orientation, which potentially leads to side effects. Frustration analysis thus offers a promising approach for screening more specific compounds in drug discovery. Based on the identification of binding sites with local frustration patterns in the unbound protein structures, a protein-ligand binding site predictor, FrustraPocket, has been recently developed using a machine learning algorithm [34]. Its package link is https://github.com/CamilaClemente/FrustraPocket. An example of ligand binding site prediction with mutational frustration through FrustraPocket is shown in Fig. (**4**).

5. FUTURE OUTLOOK

5.1. Protein Dynamics Enhances Prediction of Binding Specificity

The quantification of binding specificity discussed above is generally based on the static structures. Proteins actually are not static; they exhibit conformational flexibility that allows them to adapt to different substrates [52, 53]. Proteins often interconvert between different conformations that are either in dynamic equilibrium or triggered by ligand binding. For example, conformational changes can help proteins fit precisely around a ligand (induced fit) or select the optimal conformation for binding (conformation selection), thereby affecting the specificity of the interaction [54, 55]. The dynamic conformational change ultimately allows proteins to perform their biological functions by adapting the ligands. Advances in molecular dynamics simulation algorithms and experimental techniques, as well as the integration of AI and the energy landscape [55-60], have led to the increasing availability of multi--state structures and their motions, providing more opportunities to describe the mechanisms of specific binding and develop specific inhibitors.

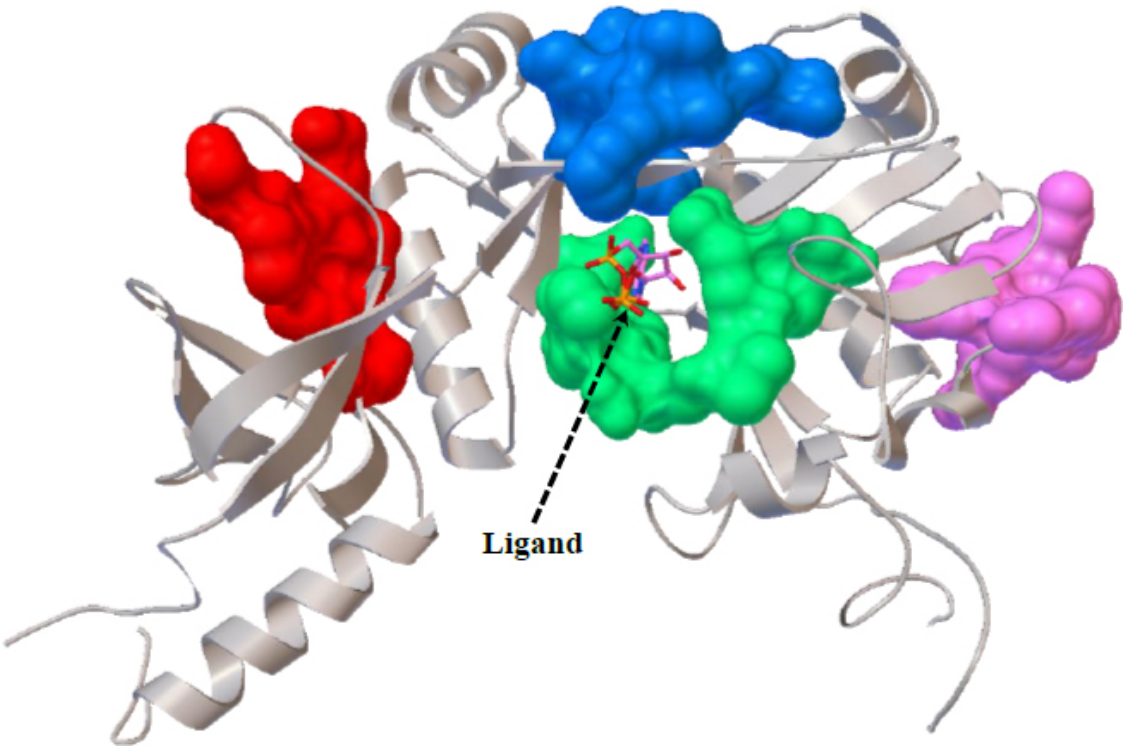

Fig. (4). Example of ligand binding site prediction by FrustraPocket. The protein is ATP-dependent DNA ligase (PDB ID: 1A0I). Four predicted binding sites are shown in different colors, and the ligand ATP is shown using the sticks. The data for drawing is obtained from the link: https://github.com/CamilaClemente/FrustraPocket/ [34]. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

5.2. Detecting Cryptic Binding Site of Undruggable Target with Local Frustration

The number of biologically validated drug targets for complex diseases has considerably increased due to the rise in genomics and proteomics [61, 62]. However, undruggable protein targets are also discovered with few characteristics of conventional druggable targets. The term undruggable refers to target proteins whose functional interfaces are flat and lack defined pockets for ligand interaction, making classic drug discovery strategies facing a challenge [63, 64]. For this reason, significant efforts were put toward alternative strategies, including identifying allosteric ligands and characterizing hidden (cryptic) allosteric pockets [65-67]. A typical example of an undruggable target is KRAS, one of the most frequently mutated oncogenes, for which an allosteric inhibitor has recently been approved by the FDA [68, 69].

Conventional approaches to predict ligand binding sites are generally based on the geometric definition of the protein's pocket [70, 71]. However, the pockets identified with geometric means are often large, which could neglect the cryptic sites. This challenge could be alleviated by detecting the sites with local frustration, which combines geometric and energetic characteristics [34]. Studies have reported that protein structural fluctuations often lead to the formation of cryptic pockets, which could present druggable sites for undruggable targets [72, 73]. This matches the predictive power of local frustration since the regions with highly frustrated interactions tend to be flexible and change conformation [32]. Thus, targeting those sites detected by local frustration could provide a number of compelling opportunities for drug discovery.

5.3. AI-Powered Drug Discovery with Binding Specificity and Local Frustration

As AI continues to evolve rapidly, it holds great promise for advancing the prediction of binding specificity and ligand binding sites. Traditional methods for predicting binding specificity often rely on static structures, which ignores the dynamic nature of protein-ligand interactions. However, AI-driven models can integrate vast datasets, including structural and sequence information as well as biophysical principles, to predict binding specificity with higher accuracy. For example, by feeding Alphafold2 with similarity-clustered sequences or highly frustrated sequences, researchers can predict multi-state structures and dynamic pathways [59, 74]. This provides opportunities for refining the binding specificity of protein-ligand complex with only the static structure. Also, by learning features of local frustration from known binding sites, FrustraPocket can identify potential protein-ligand binding sites only from the unbound forms. Moreover, the calculations of binding specificity and local frustration are contingent upon the precision of the interaction potentials between atoms or residues. AI-based potentials have not only demonstrated greater accuracy but also exhibited a significantly faster computing speed [75-77]. Given the advancements in the integration of AI and traditional approaches in quantifying the binding specificity and local frustration, it is reasonable to expect that an AI-powered prediction model will accelerate the drug discovery process and improve the success rate.

CONCLUSION

Funneled energy landscape theory has proven successful in explaining the thermodynamics and kinetics of biomolecules, such as the processes of protein folding and binding. In light of the minimal frustration principle of the theory, the global binding specificity of protein-ligand interactions and local frustration index across the whole protein structure can be quantified computationally. This provides potential promises in developing algorithms for computer-aided drug discovery. Traditional strategies of drug discovery, both *in silico* and *in vitro*, mainly concentrated on the binding affinities as the screening criteria. In this review, we have introduced the development of another screening criteria *in silico*, *i.e.*, quantified binding specificity, for structure-based drug screening. With the incorporation of binding specificity, two-dimensional drug screening strategies were proposed and validated. Further, we reviewed algorithms for the identification of binding sites with quantified local frustrations. The local frustration indexes were found to be effective in discovering the drug-binding sites on the protein surface.

This review proposes three potential directions to further improve the accuracy of these algorithms and enhance their applications in structure-based drug discovery. Firstly, the growing accessibility of multi-state structures and their dynamic motions offers greater opportunities to elucidate the mechanisms of specific binding. This, in turn, provides more avenues for the development of highly specific inhibitors with binding specificity analysis. Secondly, targeting binding sites identified through local frustration analysis presents a wealth of promising opportunities for uncovering cryptic ligand-binding sites. These cryptic sites could serve as druggable sites for undruggable targets. Thirdly, AI algorithms can analyze vast amounts of data quickly and identify patterns that might be overlooked by traditional methods. In addition, AI-based potentials have

demonstrated advantages in accuracy and computing speed. We expect that this review will motivate more integrations of these algorithms with the latest AI advances, thereby improving the success rate of hit compounds.

AUTHORS' CONTRIBUTIONS

The authors confirm their contribution to the paper as follows: study conception and design: Z.Y., J.W.; data collection, analysis and interpretation of results: Z.Y.; Writing-draft manuscript: Z.Y., Y.L, Y.C., X.T.; Writing-Review and Editing: Z.Y., J.W., Y.J; All authors reviewed the results and approved the final version of the manuscript.

LIST OF ABBREVIATIONS

SBDD = Structure-based Drug Discovery

NSAIDs = Nonsteroidal Anti-inflammatory Drugs

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

This study is financially supported by the National Natural Science Foundation of China (No. 12474192 and 12234019), Zhejiang Provincial Natural Science Foundation of China (No. LZ24A040002), Beijing National Laboratory for Condensed Matter Physics (No. 2023BNLCMPKF009), and Wenzhou Institute, University of Chinese Academy of Sciences (No. WIUCASQD2022019).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the High--Performance Computing Center of Wenzhou Institute, University of Chinese Academy of Sciences, for providing computational resources and the Affiliated Xiangshan Hospital of Wenzhou Medical University for the grant.

REFERENCES

[1] Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B,* **2022**, *12*(7), 3049-3062. http://dx.doi.org/10.1016/j.apsb.2022.02.002 PMID: 35865092

[2] Berdigaliyev, N.; Aljofan, M. An overview of drug discovery and development. *Future Med. Chem.,* **2020**, *12*(10), 939-947. http://dx.doi.org/10.4155/fmc-2019-0307 PMID: 32270704

[3] Wu, F.; Zhou, Y.; Li, L.; Shen, X.; Chen, G.; Wang, X.; Liang, X.; Tan, M.; Huang, Z. Computational approaches in preclinical studies on drug discovery and development. *Front Chem.,* **2020**, *8*, 726. http://dx.doi.org/10.3389/fchem.2020.00726 PMID: 33062633

[4] Sadybekov, A.V.; Katritch, V. Computational approaches streamlining drug discovery. *Nature,* **2023**, *616*(7958), 673-685. http://dx.doi.org/10.1038/s41586-023-05905-z PMID: 37100941

[5] Özçelik, R.; van Tilborg, D.; Jiménez-Luna, J.; Grisoni, F. Structure-based drug discovery with deep learning. *ChemBioChem,* **2023**, *24*(13), e202200776. http://dx.doi.org/10.1002/cbic.202200776 PMID: 37014633

[6] Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med. (Berl.),* **2000**, *78*(5), 269-281. http://dx.doi.org/10.1007/s001090000084 PMID: 10954199

[7] Merz, K.M.; Ringe, D.; Reynolds, C.H. Pharmacophore methods. In: *Drug Design: Structure- and Ligand-Based Approaches*; Cambridge University Press, **2010**; pp. 137-150. http://dx.doi.org/10.1017/CBO9780511730412

[8] Kim, G.; Lee, S.; Levy Karin, E.; Kim, H.; Moriwaki, Y.; Ovchinnikov, S.; Steinegger, M.; Mirdita, M. Easy and accurate protein structure prediction using ColabFold. *Nat. Protoc.,* **2024**, *20*(3), 620-642. http://dx.doi.org/10.1038/s41596-024-01060-5 PMID: 39402428

[9] Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science,* **2023**, *379*(6637), 1123-1130. http://dx.doi.org/10.1126/science.ade2574 PMID: 36927031

[10] Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; Millán, C.; Park, H.; Adams, C.; Glassman, C.R.; DeGiovanni, A.; Pereira, J.H.; Rodrigues, A.V.; van Dijk, A.A.; Ebrecht, A.C.; Opperman, D.J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M.K.; Dalwadi, U.; Yip, C.K.; Burke, J.E.; Garcia, K.C.; Grishin, N.V.; Adams, P.D.; Read, R.J.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science,* **2021**, *373*(6557), 871-876. http://dx.doi.org/10.1126/science.abj8754 PMID: 34282049

[11] Binder, J.L.; Berendzen, J.; Stevens, A.O.; He, Y.; Wang, J.; Dokholyan, N.V.; Oprea, T.I. AlphaFold illuminates half of the dark human proteins. *Curr. Opin. Struct. Biol.,* **2022**, *74*, 102372. http://dx.doi.org/10.1016/j.sbi.2022.102372 PMID: 35439658

[12] Maia, E.H.B.; Assis, L.C.; de Oliveira, T.A.; da Silva, A.M.; Taranto, A.G. Structure-based virtual screening: From classical to artificial intelligence. *Front Chem.,* **2020**, *8*, 343.

http://dx.doi.org/10.3389/fchem.2020.00343 PMID: 32411671

[13] Rester, U. From virtuality to reality - Virtual screening in lead discovery and lead optimization: A medicinal chemistry perspective. *Curr. Opin. Drug Discov. Devel.,* **2008**, *11*(4), 559-568. PMID: 18600572

[14] Wang, J.; Zheng, X.; Yang, Y.; Drueckhammer, D.; Yang, W.; Verkhivker, G.; Wang, E. Quantifying intrinsic specificity: A potential complement to affinity in drug screening. *Phys. Rev. Lett.,* **2007**, *99*(19), 198101. http://dx.doi.org/10.1103/PhysRevLett.99.198101 PMID: 18233118

[15] Yan, Z.; Wang, J. Specificity quantification of biomolecular recognition and its implication for drug discovery. *Sci. Rep.,* **2012**, *2*(1), 309. http://dx.doi.org/10.1038/srep00309 PMID: 22413060

[16] Huggins, D.J.; Sherman, W.; Tidor, B. Rational approaches to improving selectivity in drug design. *J. Med. Chem.,* **2012**, *55*(4), 1424-1444. http://dx.doi.org/10.1021/jm2010332 PMID: 22239221

[17] Kawasaki, Y.; Freire, E. Finding a better path to drug selectivity. *Drug Discov. Today,* **2011**, *16*(21-22), 985-990. http://dx.doi.org/10.1016/j.drudis.2011.07.010 PMID: 21839183

[18] Schmidt, F.; Matter, H.; Hessler, G.; Czich, A. Predictive *in silico* off-target profiling in drug discovery. *Future Med. Chem.,* **2014**, *6*(3), 295-317. http://dx.doi.org/10.4155/fmc.13.202 PMID: 24575966

[19] Yan, Z.; Zheng, X.; Wang, E.; Wang, J. Thermodynamic and kinetic specificities of ligand binding. *Chem. Sci. (Camb.),* **2013**, *4*(6), 2387-2395. http://dx.doi.org/10.1039/c3sc50478f

[20] Wang, J.; Verkhivker, G.M. Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys. Rev. Lett.,* **2003**, *90*(18), 188101. http://dx.doi.org/10.1103/PhysRevLett.90.188101 PMID: 12786043

[21] Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D.S. How protein stability and new functions trade off. *PLOS Comput. Biol.,* **2008**, *4*(2), e1000002. http://dx.doi.org/10.1371/journal.pcbi.1000002 PMID: 18463696

[22] Luque, I.; Leavitt, S.A.; Freire, E. The linkage between protein folding and functional cooperativity: Two sides of the same coin? *Annu. Rev. Biophys. Biomol. Struct.,* **2002**, *31*(1), 235-256. http://dx.doi.org/10.1146/annurev.biophys.31.082901.134215 PMID: 11988469

[23] Yan, Z.; Wang, J. Superfunneled energy landscape of protein evolution unifies the principles of protein evolution, folding, and design. *Phys. Rev. Lett.,* **2019**, *122*(1), 018103. http://dx.doi.org/10.1103/PhysRevLett.122.018103 PMID: 31012725

[24] Yan, Z.; Wang, J. Funneled energy landscape unifies principles of protein binding and evolution. *Proc. Natl. Acad. Sci. USA,* **2020**, *117*(44), 27218-27223. http://dx.doi.org/10.1073/pnas.2013822117 PMID: 33067388

[25] Yan, Z.; Wang, J. Evolution shapes interaction patterns for epistasis and specific protein binding in a two-component signaling system. *Commun. Chem.,* **2024**, *7*(1), 13. http://dx.doi.org/10.1038/s42004-024-01098-2 PMID: 38233668

[26] Nassar, R.; Dignon, G.L.; Razban, R.M.; Dill, K.A. The protein folding problem: The role of theory. *J. Mol. Biol.,* **2021**, *433*(20), 167126. http://dx.doi.org/10.1016/j.jmb.2021.167126 PMID: 34224747

[27] Onuchic, J.N.; Wolynes, P.G. Theory of protein folding. *Curr. Opin. Struct. Biol.,* **2004**, *14*(1), 70-75. http://dx.doi.org/10.1016/j.sbi.2004.01.009 PMID: 15102452

[28] Gonzalo Parra, R.; Komives, E.A.; Wolynes, P.G.; Ferreiro, D.U. Frustration in physiology and molecular medicine. *arXiv,* **2025**. http://dx.doi.org/10.48550/arXiv.2502.03851

[29] Ferreiro, D.U.; Komives, E.A.; Wolynes, P.G. Frustration in biomolecules. *Q. Rev. Biophys.,* **2014**, *47*(4), 285-363. http://dx.doi.org/10.1017/S0033583514000092 PMID: 25225856

[30] Ferreiro, D.U.; Komives, E.A.; Wolynes, P.G. Frustration, function and folding. *Curr. Opin. Struct. Biol.,* **2018**, *48*, 68-73. http://dx.doi.org/10.1016/j.sbi.2017.09.006 PMID: 29101782

[31] Freiberger, M.I.; Guzovsky, A.B.; Wolynes, P.G.; Parra, R.G.; Ferreiro, D.U. Local frustration around enzyme active sites. *Proc. Natl. Acad. Sci. USA,* **2019**, *116*(10), 4037-4043. http://dx.doi.org/10.1073/pnas.1819859116 PMID: 30765513

[32] Froes, T.Q.; Castilho, M.S. The interplay between protein frustration and hotspot formation. *J. Braz. Chem. Soc.,* **2024**, *35*(10), e–20240168. http://dx.doi.org/10.21577/0103-5053.20240168

[33] Chen, M.; Chen, X.; Schafer, N.P.; Clementi, C.; Komives, E.A.; Ferreiro, D.U.; Wolynes, P.G. Surveying biomolecular frustration at atomic resolution. *Nat. Commun.,* **2020**, *11*(1), 5944. http://dx.doi.org/10.1038/s41467-020-19560-9 PMID: 33230150

[34] Freiberger, M.I.; Clemente, C.M.; Valero, E.; Pombo, J.G.; Leonetti, C.O.; Ravetti, S. FrustraPocket: A protein–ligand binding site predictor using energetic local frustration. *bioRxiv,* **2022**.

[35] Chakraborti, A.; Garg, S.; Kumar, R.; Motiwala, H.; Jadhavar, P. Progress in COX-2 inhibitors: A journey so far. *Curr. Med. Chem.,* **2010**, *17*(15), 1563-1593. http://dx.doi.org/10.2174/092986710790979980 PMID: 20166930

[36] Carlsson, J.; Luttens, A. Structure-based virtual screening of vast chemical space as a starting point for drug discovery. *Curr. Opin. Struct. Biol.,* **2024**, *87*, 102829. http://dx.doi.org/10.1016/j.sbi.2024.102829 PMID: 38848655

[37] Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discov.,* **2004**, *3*(11), 935-949. http://dx.doi.org/10.1038/nrd1549 PMID: 15520816

[38] Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature,* **2024**, *630*(8061), 493-500. http://dx.doi.org/10.1038/s41586-024-07487-w PMID: 38718835

[39] Scardino, V.; Di Filippo, J.I.; Cavasotto, C.N. How good are AlphaFold models for docking-based virtual screen-

ing? *iScience,* **2023**, *26*(1), 105920. http://dx.doi.org/10.1016/j.isci.2022.105920 PMID: 36686396

[40] Liu, Z.; Su, M.; Han, L.; Liu, J.; Yang, Q.; Li, Y.; Wang, R. Forging the basis for developing protein–ligand interaction scoring functions. *Acc. Chem. Res.,* **2017**, *50*(2), 302-309. http://dx.doi.org/10.1021/acs.accounts.6b00491 PMID: 28182403

[41] Yan, Z.; Wang, J. Optimizing the affinity and specificity of ligand binding with the inclusion of solvation effect. *Proteins,* **2015**, *83*(9), 1632-1642. http://dx.doi.org/10.1002/prot.24848 PMID: 26111900

[42] Yan, Z.; Guo, L.; Hu, L.; Wang, J. Specificity and affinity quantification of protein–protein interactions. *Bioinformatics,* **2013**, *29*(9), 1127-1133. http://dx.doi.org/10.1093/bioinformatics/btt121 PMID: 23476023

[43] Yan, Z.; Wang, J. Optimizing scoring function of protein-nucleic acid interactions with both affinity and specificity. *PLoS One,* **2013**, *8*(9), e74443. http://dx.doi.org/10.1371/journal.pone.0074443 PMID: 24098651

[44] Yan, Z.; Wang, J. Incorporating specificity into optimization: Evaluation of SPA using CSAR 2014 and CASF 2013 benchmarks. *J. Comput. Aided Mol. Des.,* **2016**, *30*(3), 219-227. http://dx.doi.org/10.1007/s10822-016-9897-0 PMID: 26879323

[45] Yan, Z.; Wang, J. SPA-LN: A scoring function of ligand–nucleic acid interactions *via* optimizing both specificity and affinity. *Nucleic Acids Res.,* **2017**, *45*(12), e110-e110. http://dx.doi.org/10.1093/nar/gkx255 PMID: 28431169

[46] Bryngelson, J.D.; Wolynes, P.G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA,* **1987**, *84*(21), 7524-7528. http://dx.doi.org/10.1073/pnas.84.21.7524 PMID: 3478708

[47] Wolynes, P.G. Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie,* **2015**, *119*, 218-230. http://dx.doi.org/10.1016/j.biochi.2014.12.007 PMID: 25530262

[48] Gianni, S.; Freiberger, M.I.; Jemth, P.; Ferreiro, D.U.; Wolynes, P.G.; Fuxreiter, M. Fuzziness and frustration in the energy landscape of protein folding, function, and assembly. *Acc. Chem. Res.,* **2021**, *54*(5), 1251-1259. http://dx.doi.org/10.1021/acs.accounts.0c00813 PMID: 33550810

[49] Ferreiro, D.U.; Hegler, J.A.; Komives, E.A.; Wolynes, P.G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA,* **2007**, *104*(50), 19819-19824. http://dx.doi.org/10.1073/pnas.0709915104 PMID: 18077414

[50] Parra, R.G.; Schafer, N.P.; Radusky, L.G.; Tsai, M.Y.; Guzovsky, A.B.; Wolynes, P.G.; Ferreiro, D.U. Protein Frustratometer 2: A tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.,* **2016**, *44*(W1), W356-W360. http://dx.doi.org/10.1093/nar/gkw304 PMID: 27131359

[51] Rausch, A.O.; Freiberger, M.I.; Leonetti, C.O.; Luna, D.M.; Radusky, L.G.; Wolynes, P.G.; Ferreiro, D.U.; Parra, R.G. FrustratometeR: An R-package to compute local frustration in protein structures, point mutants and MD simulations. *Bioinformatics,* **2021**, *37*(18), 3038-3040. http://dx.doi.org/10.1093/bioinformatics/btab176 PMID: 33720293

[52] Lill, M.A. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry,* **2011**, *50*(28), 6157-6169. http://dx.doi.org/10.1021/bi2004558 PMID: 21678954

[53] Tokuriki, N.; Tawfik, D.S. Protein dynamism and evolvability. *Science,* **2009**, *324*(5924), 203-207. http://dx.doi.org/10.1126/science.1169375 PMID: 19359577

[54] Gianni, S.; Dogan, J.; Jemth, P. Distinguishing induced fit from conformational selection. *Biophys. Chem.,* **2014**, *189*, 33-39. http://dx.doi.org/10.1016/j.bpc.2014.03.003 PMID: 24747333

[55] Lu, W.; Zhang, J.; Huang, W.; Zhang, Z.; Jia, X.; Wang, Z.; Shi, L.; Li, C.; Wolynes, P.G.; Zheng, S. DynamicBind: Predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nat. Commun.,* **2024**, *15*(1), 1071. http://dx.doi.org/10.1038/s41467-024-45461-2 PMID: 38316797

[56] Monteiro da Silva, G.; Cui, J.Y.; Dalgarno, D.C.; Lisi, G.P.; Rubenstein, B.M. High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nat. Commun.,* **2024**, *15*(1), 2464. http://dx.doi.org/10.1038/s41467-024-46715-9 PMID: 38538622

[57] Bryant, P.; Noé, F. Structure prediction of alternative protein conformations. *Nat. Commun.,* **2024**, *15*(1), 7328. http://dx.doi.org/10.1038/s41467-024-51507-2 PMID: 39187507

[58] Sala, D.; Engelberger, F.; Mchaourab, H.S.; Meiler, J. Modeling conformational states of proteins with AlphaFold. *Curr. Opin. Struct. Biol.,* **2023**, *81*, 102645. http://dx.doi.org/10.1016/j.sbi.2023.102645 PMID: 37392556

[59] Guan, X.; Tang, Q.Y.; Ren, W.; Chen, M.; Wang, W.; Wolynes, P.G.; Li, W. Predicting protein conformational motions using energetic frustration analysis and AlphaFold2. *Proc. Natl. Acad. Sci. USA,* **2024**, *121*(35), e2410662121. http://dx.doi.org/10.1073/pnas.2410662121 PMID: 39163334

[60] Song, X.; Bao, L.; Feng, C.; Huang, Q.; Zhang, F.; Gao, X.; Han, R. Accurate prediction of protein structural flexibility by deep learning integrating intricate atomic structures and Cryo-EM density information. *Nat. Commun.,* **2024**, *15*(1), 5538. http://dx.doi.org/10.1038/s41467-024-49858-x PMID: 38956032

[61] Ahn, N.G.; Wang, A.H.J. Proteomics and genomics: Perspectives on drug and target discovery. *Curr. Opin. Chem. Biol.,* **2008**, *12*(1), 1-3. http://dx.doi.org/10.1016/j.cbpa.2008.02.016 PMID: 18302945

[62] Rask-Andersen, M.; Almén, M.S.; Schiöth, H.B. Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.,* **2011**, *10*(8), 579-590. http://dx.doi.org/10.1038/nrd3478 PMID: 21804595

[63] Dang, C.V.; Reddy, E.P.; Shokat, K.M.; Soucek, L. Drugging the 'undruggable' cancer targets. *Nat. Rev. Cancer,* **2017**, *17*(8), 502-508. http://dx.doi.org/10.1038/nrc.2017.36 PMID: 28643779

[64] Xie, X.; Yu, T.; Li, X.; Zhang, N.; Foster, L.J.; Peng, C.; Huang, W.; He, G. Recent advances in targeting the "un-

druggable" proteins: From drug discovery to clinical trials. *Signal Transduct. Target. Ther.,* **2023**, *8*(1), 335. http://dx.doi.org/10.1038/s41392-023-01589-z PMID: 37669923

[65] Colombo, G. Computing allostery: From the understanding of biomolecular regulation and the discovery of cryptic sites to molecular design. *Curr. Opin. Struct. Biol.,* **2023**, *83*, 102702. http://dx.doi.org/10.1016/j.sbi.2023.102702 PMID: 37716095

[66] Zhang, G.; Zhang, J.; Gao, Y.; Li, Y.; Li, Y. Strategies for targeting undruggable targets. *Expert Opin. Drug Discov.,* **2022**, *17*(1), 55-69. http://dx.doi.org/10.1080/17460441.2021.1969359 PMID: 34455870

[67] Rehman, A.U.; Lu, S.; Khan, A.A.; Khurshid, B.; Rasheed, S.; Wadood, A.; Zhang, J. Hidden allosteric sites and De-Novo drug design. *Expert Opin. Drug Discov.,* **2022**, *17*(3), 283-295. http://dx.doi.org/10.1080/17460441.2022.2017876 PMID: 34933653

[68] Ostrem, J.M.; Peters, U.; Sos, M.L.; Wells, J.A.; Shokat, K.M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature,* **2013**, *503*(7477), 548-551. http://dx.doi.org/10.1038/nature12796 PMID: 24256730

[69] Blair, H.A. Sotorasib: First approval. *Drugs,* **2021**, *81*(13), 1573-1579. http://dx.doi.org/10.1007/s40265-021-01574-2 PMID: 34357500

[70] Xia, Y.; Pan, X.; Shen, H.B. A comprehensive survey on protein-ligand binding site prediction. *Curr. Opin. Struct. Biol.,* **2024**, *86*, 102793. http://dx.doi.org/10.1016/j.sbi.2024.102793 PMID: 38447285

[71] Henrich, S.; Salo-Ahen, O.M.H.; Huang, B.; Rippmann, F.F.; Cruciani, G.; Wade, R.C. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.,* **2010**, *23*(2), 209-219. http://dx.doi.org/10.1002/jmr.984 PMID: 19746440

[72] Amaro, R.E. Will the real cryptic pocket please stand out? *Biophys. J.,* **2019**, *116*(5), 753-754. http://dx.doi.org/10.1016/j.bpj.2019.01.018 PMID: 30739726

[73] Meller, A.; Ward, M.; Borowsky, J.; Kshirsagar, M.; Lotthammer, J.M.; Oviedo, F.; Ferres, J.L.; Bowman, G.R. Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nat. Commun.,* **2023**, *14*(1), 1177. http://dx.doi.org/10.1038/s41467-023-36699-3 PMID: 36859488

[74] Wayment-Steele, H.K.; Ojoawo, A.; Otten, R.; Apitz, J.M.; Pitsawong, W.; Hömberger, M.; Ovchinnikov, S.; Colwell, L.; Kern, D. Predicting multiple conformations *via* sequence clustering and AlphaFold2. *Nature,* **2024**, *625*(7996), 832-839. http://dx.doi.org/10.1038/s41586-023-06832-9 PMID: 37956700

[75] Wang, T.; He, X.; Li, M.; Li, Y.; Bi, R.; Wang, Y.; Cheng, C.; Shen, X.; Meng, J.; Zhang, H.; Liu, H.; Wang, Z.; Li, S.; Shao, B.; Liu, T.Y. Ab initio characterization of protein molecular dynamics with AI^{2}BMD. *Nature,* **2024**, *635*(8040), 1019-1027. http://dx.doi.org/10.1038/s41586-024-08127-z PMID: 39506110

[76] Zhang, L.; Han, J.; Wang, H.; Car, R.; e, W. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.,* **2018**, *120*(14), 143001. http://dx.doi.org/10.1103/PhysRevLett.120.143001 PMID: 29694129

[77] Yang, H.; Xiong, Z.; Zonta, F. Construction of a deep neural network energy function for protein physics. *J. Chem. Theory Comput.,* **2022**, *18*(9), 5649-5658. http://dx.doi.org/10.1021/acs.jctc.2c00069 PMID: 35939398