Article

Towards a General Intermolecular Binding Affinity Calculator

Wei Li [1]

[1] Contrebola Institute of Computational Interstructural Biophysics, No. 88, Fuxing East Road, Nantong City, 226000, Jiangsu Province, P. R. China; wli148@aucklanduni.ac.nz

Abstract: Thanks to the continued development of experimental structural biology and the half-a-century old Protein Data Bank, 2021 saw a big step forward in the development of protein structure prediction with deep learning algorithms. Recently, DeepMind's AlphaFold has determined the structures of $\sim$ 200 million proteins from 1 million species. The speed of this progress raise the question of what becomes possible for computational drug discovery and design when we have a systems-wide account of the structures and motions of most proteins. Therefore, this article puts forward the concept of a general intermolecular binding affinity calculator (GIBAC): $K_{\mathrm{d}} = f(molA, molB, envPara)$, towards the acceleration of traditional computer-aided drug design (CADD) and artificial intelligence-integrated drug discovery (AIDD), for both small molecules and biologics such as therapeutic proteins.

Keywords: intermolecular binding affinity; drug target binding affinity; computer-aided drug design (CADD); artificial intelligence-integrated drug discovery (AIDD); machine learning

Intermolecular binding affinity: a definition and brief introduction

Intermolecular binding affinity is the strength of the binding between a single molecule to its ligand/binding partner, and is typically measured and expressed as the equilibrium dissociation constant (K_d), to evaluate and rank order strength of intermolecular binding [1,2]. The smaller the K_d value, the greater the binding affinity; The larger the K_d value, the weaker the binding affinity [3,4]. In theory, K_d is influenced by a range of factors, including non-covalent intermolecular interactions such as hydrogen bonding, electrostatics, hydrophobics and Van der Waals forces between the two binding molecules [5–7], and also environmental parameters (*envPara*) such as pH, ionic strength and temperature [8–12]. Furthermore, K_d may also be affected by the presence of other molecules, where there are multiple interacting partners, in cases such as IL-2 [13,14] and $Ca_V1.2$ [15–17].

From drug target K_d calculation to intermolecular K_d calculation: a generalization

Whether it is mall molecule or therapeutic protein, drugs exert their desired pharmacological effects through binding to and interacting with specific disease-related target(s) [18–20]. Hence, drug target binding affinity (DT K_d) is used to describe the strength of binding between a drug molecule and its target. Be it obtained through experimental measurement or computational prediction, adequate and accurate knowledge of DT K_d is crucial both in early drug discovery and screening, during drug repurposing, and in avoiding undue risk of toxicity mediated by drug-drug interactions [21–23]. Towards this end, this article proposes a general intermolecular binding affinity calculator (GIBAC), as in Equation 1 below,

$$K_{\mathrm{d}} = f(molA, molB, envPara) \tag{1}$$

given that DT K_d can easily be generalized as intermolecular binding affinity, with *drug* as molecule A (*molA* in Equation 1) and *target* as molecule B (*molB* in Equation 1). Moreover, this article puts forward a road map for the construction of a GIBAC, and a set of key ingredients to ensure its accuracy and efficiency.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2022 doi:10.20944/preprints202208.0213.v1

GIBAC: an integrative definition for both mall molecules and therapeutic proteins

As discussed above, DT K_d is an essential parameter in drug discovery and design, both computationally and experimentally [24–26]. A set of computational tools have therefore already been developed to calculate DT K_d with various algorithms, including molecular mechanics-based calculations [27–30] and machine-learning based predictions [31–33].

$$K_d = f(molAstruc, molBstruc, ABinterface, envPara) \quad (2)$$

Take Prodigy [2,34] for instance, protein-protein or protein-ligand K_d is calculated using the binary interfacial features, i.e., interfacial contacts between the two interacting partners (Equation 2), including interstructural hydrogen bonding, electrostatics, hydrophobics and Van der Waals forces between the two binding molecules [5,6,35,36].

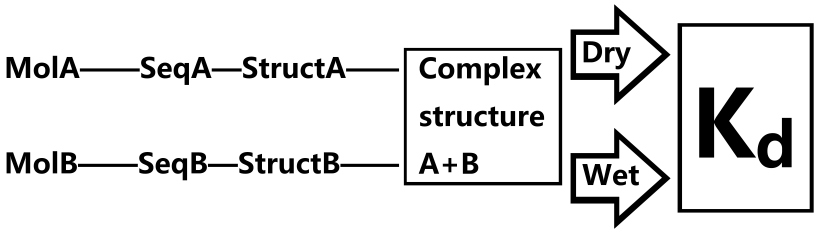

Figure 1. A flow chart of experimental (wet-lab approach) determination and computational (dry-lab approach) prediction of K_d (Equation 2). In the wet-lab approach, K_d is measured using tools such as isothermal titration calorimetry on the experimental sample of $molA+molB$ complex, while with the dry-lab approach, K_d is calculated by tools such as Prodigy with a structural (experimental or computational) model of $molA+molB$ complex structure.

Overall, Equation 2 for K_d calculation requires sufficient structural information, including a complex structure of molecule A ($molAstruc$ in Equation 2) bound to molecule B ($molBstruc$ in Equation 2), and its binary interfacial structural features ($ABinterface$ in Equation 2) [5,24,35–37]. For further generalization and simplification, this article for the first time puts forward a GIBAC in Equation 3 as below,

$$K_d = f(molAstring, molBstring, envPara) \quad (3)$$

In Equation 3, $molAstring$ or $molBstring$ represents a sequence of amino acids, i.e., a string of letters for protein A or B, or a string of SMILES (Simplified Molecular Input Line Entry System) characters to represent the chemical structure of small molecule A or B, while $envPara$ represents environmental parameters such as pH, ionic strength and temperature [8–10,12].

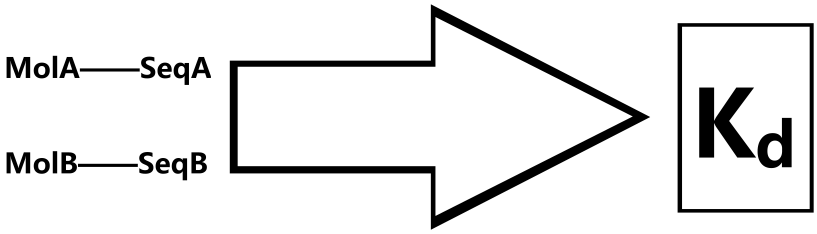

Figure 2. A flow chart of GIBAC for K_d calculation (Equation 3).

From Figures 1 and 2, it is quite clear that Equation 3 (GIBAC) does not require as input any structural information for molecule A or B, as required by computational tools such as Prodigy (Equation 2) [2,34]. While the generalization from Equation 2 to Equation 3 aims to include DT K_d calculation even for interacting partners whose complex structural information (experimental or computational) is not available yet [38], it is by no means

 doi:10.20944/preprints202208.0213.v1

indicating that structural information is negligible or dispensable. Instead, a vast amount of reliable structural information [38], experimental complex structural information [39–43] in particular, is a prerequisite for the construction of a GIBAC with reasonable accuracy for early stage drug discovery and design [44,45].

Experimental structural information is essential to build an accurate GIBAC

By definition, intermolecular binding affinity or DT K_d calculation will not be accurate unless experimental three-dimensional structural data (intermolecular binding mode, structural features such as electrostatics, hydrophobics, Van der Waals forces, etc.) are experimentally characterized and measured or computationally predicted with acceptable accuracy, highlighting the need for structure-based K_d calculation [46], instead of K_d calculation based on sequence information alone.

For instance, BindProfX is a computational tool to calculate protein-protein binding free-energy changes ($\Delta\Delta G$) induced by single- and multiple-mutations, and is particularly useful for designing and engineering protein-protein interactions with desired (enhanced [24] or suppressed [47]) binding affinities [42]. By definition, $\Delta\Delta G$ is the same as K_d, in that both parameters boil down to the strength of the binding between two interacting molecules, i.e., the energy required to seperate them [48–51]. When BindProfX predicts $\Delta\Delta G$ of interface residue mutations, it uses *iAlign* to align the protein-protein interaction (PPI) interface structure of target protein to a set of PPI interface structures from PIFACE database [42], highlighting the essential role of experimental structural data in DT K_d calculation.

While resolving the three dimensional structure of a protein is a critical step in modern drug discovery today, experimental three-dimensional structural data, however, is harder to obtain than sequential data, because experimental methods for determining the structure of protein-ligand complexes are still quite expensive and time-consuming [8]. This is where computational methods come in for protein structural prediction, i.e., predicting protein structure from its amino acid sequence, i.e., the holy grail of structural biology [52–54].

Protein structure prediction: a brief retrospective

This year (2022) marks the 51st year since the beginning of the Worldwide PDB Consortium (wwPDB), which has acted as a catalyst for developments in both experimental and, more importantly, computational methods for integrative structural biology [41,55–57].

Take protein structure prediction (PSP) for instance, where advances driven by machine learning techniques have occured beyond our wildest expectations [58]. To date, $\sim$ 193,455 (as of July 29, 2022) experimental structures represents only a small fraction of the billions of known protein sequences [59]. As a result, accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics [59]. While the progresses in PSP have ebbed and flowed historically, the past two years saw dramatic advances driven by the increasing neuralization [60–63] of PSP algorithms, such as AlphaFold and RoseTTAFold [59,63–66], whereby computations previously based on energy models and sampling procedures are replaced by neural networks [67–73].

Recently, DeepMind's AlphaFold has determined the structures of $\sim$ 200 million proteins from 1 million species [38], and $\sim$35% of them are deemed highly accurate, which means they are as good as experimentally determined structures [38]. Another 45% were deemed confident enough to rely on for many applications [38]. The data for the 200 million structures will be freely available on a DeepMind database [38].

Road map and key ingredients of an accurate and efficient GIBAC

Artificial intelligence algorithm is a broad field consisting of machine learning algorithms and deep learning algorithms with four key steps: build, train, test, evaluate. Similar to the development of any artificial intelligence-based tool, building an accurate GIBAC requires both **data** and **algorithm** (Figure 3), and in particular, machine-learning/deep-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2022 doi:10.20944/preprints202208.0213.v1

learning algorithms (as the backbone of artificial intelligence) need to be tailored to the situation where a GIBAC is applicable.

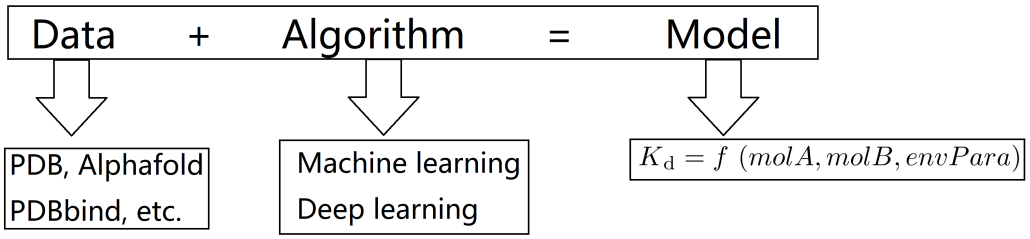

Figure 3. GIBAC's accuracy requires a $data + algorithm = model$ approach with algorithms tailored to GIBAC applications, and a hybrid (experimental & synthetic) approach for datasets (as the lifeblood of artificial intelligence) with reasonable size, compared to the size of the universe of small molecules and therapeutic proteins.

With respect to the $data + algorithm = model$ approach (Figure 3), a varity of artificial intelligence algorithms have already been developed and applied in drug discovery and design and DT K_d calculation [74–79]. Thus, another key aspect to build an accurate GIBAC is a hybrid (experimental & synthetic) approach for data sets used to train machine-learning/deep-learning algorithms. For example, deep learning is a subdiscipline of artificial intelligence that uses neural networks to extract patterns and make predictions from large data sets. As a result, building a deep-learning-based GIBAC would require a vast amount of structural and K_d data, including but not limited to

1. experimental structural data from PDB [40,56].
2. computational structural data from AlphaFold database [38], and synthetic structural data obtained from homology structural modeling tools [53,80,81].
3. experimental complex ($molA$+$molB$, Figure 1) structural data from PDB [40,56].
4. computational complex structural data from AlphaFold-Multimer [82] and synthetic complex structural data from homology structural modeling tools [53,80,81] with experimentally measured structures as modeling templates, such as antigen-antibody, ligand-receptor, small molecule-target complexes.
5. experimental K_d data from PDBbind, a comprehensive collection of experimentally measured binding affinity data for the protein-ligand complexes deposited in the PDB [39,40,83–86].
6. synthetic K_d data from computational tools such as Prodigy [2,34] with complex structural data as input.

Since collecting experimental structural and K_d data from the real world is complicated, expensive and time-consuming, this is where synthetic data comes in. In the history of artificial intelligence, synthetic data is not a new idea, but it is now approaching a critical inflection point in terms of real-world impact. According to a widely referenced Gartner study, 60% of all data used in the development of artificial intelligence will be synthetic rather than real by 2024. In the case here for the building of an accurate GIBAC, a huge amount of synthetic structural and K_d data is needed, because currently available experimental structural and K_d data is far from enough, compared with the size of the universe of small molecules and therapeutic proteins [38].

Take small molecule for example, the number of possible compounds up to molecular weight 500 Dalton is $\sim 10^{60}$ [87], a huge number that the human mind is not well equipped to handle. Yet, this number (10^{60}) is able to describe the degree of complexity of only a peptide consisting of $\sim$ 47 amino acid residues, that is, excluding the complexity due to installations of nonnatural amino acids, antibody- and polyethylene glycol (PEG)-coupling, and with modified glycosylation(s), methylation(s) or phosphorylation(s), etc. However, the entire protein universe, as we currently know it, possesses a complexity much more larger than that of a 47 amino-acid peptide. To this end, an example of $Ca_V1.2$ (consisting

of more than 2000 AA) [15–17] should suffice to support this hypothesis and make a solid call for: (A) the applications of AI algorithms to ensure the feasibility of an accurate GIBAC; and (B) vast amount of computational power to ensure the efficiency of GIBAC.

Conclusion and Discussion

Towards the acceleration of traditional CADD and AIDD [88,89], this article puts forward the concept of a GIBAC, and argues that the time is now ripe to build an accurate and efficient GIBAC for the prediction of novel interactions with desired affinities [24,47,90,91] on the genome scale, better characterization of signaling networks and design of novel binding partners, either small molecules or therapeutic proteins, for various disease-related targets [43,92–95].

With regard to the entire protein universe, however, further challenges still remain for GIBAC, when it comes to K_d calculations involving intrinsically disordered proteins, because the flexibility involved makes structural modeling and structure-based K_d prediction a rather difficult (if possible) problem, to which an experimental approach is perhaps the only feasible solution in practice [54,96–99].

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

1. Kastritis, P.L.; Rodrigues, J.P.; Folkers, G.E.; Boelens, R.; Bonvin, A.M. Proteins Feel More Than They See: Fine-Tuning of Binding Affinity by Properties of the Non-Interacting Surface. *Journal of Molecular Biology* **2014**, *426*, 2632–2652.
2. Vangone, A.; Bonvin, A.M. Contacts-based prediction of binding affinity in protein–protein complexes. *eLife* **2015**, *4*.
3. Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* **2018**, *34*, i821–i829.
4. Wang, K.; Zhou, R.; Li, Y.; Li, M. DeepDTAF: a deep learning method to predict protein–ligand binding affinity. *Briefings in Bioinformatics* **2021**, *22*.
5. Bitencourt-Ferreira, G.; de Azevedo Junior, W.F. Electrostatic Potential Energy in Protein-Drug Complexes. *Current Medicinal Chemistry* **2021**, *28*, 4954–4971.
6. Fahmy, A.; Wagner, G. Optimization of van der Waals Energy for Protein Side-Chain Placement and Design. *Biophysical Journal* **2011**, *101*, 1690–1698.
7. Umeyama, H.; Morokuma, K. The origin of hydrogen bonding. An energy decomposition study. *J. Am. Chem. Soc.* **1977**, *99*, 1316–1332.
8. Li, W. Gravity-driven pH adjustment for site-specific protein pKa measurement by solution-state NMR. *Measurement Science and Technology* **2017**, *28*, 127002.
9. Webb, H.; Tynan-Connolly, B.M.; Lee, G.M.; Farrell, D.; O'Meara, F.; Søndergaard, C.R.; Teilum, K.; Hewage, C.; McIntosh, L.P.; Nielsen, J.E. Remeasuring HEWL pKa values by NMR spectroscopy: Methods, analysis, accuracy, and implications for theoretical pKa calculations. *Proteins: Structure, Function, and Bioinformatics* **2010**, *79*, 685–702.
10. Hansen, A.L.; Kay, L.E. Measurement of histidine pK_a values and tautomer populations in invisible protein states. *Proceedings of the National Academy of Sciences* **2014**, *111*.
11. Søndergaard, C.R.; Olsson, M.H.M.; Rostkowski, M.; Jensen, J.H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values. *Journal of Chemical Theory and Computation* **2011**, *7*, 2284–2295.
12. Olsson, M.H.M.; Søndergaard, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *Journal of Chemical Theory and Computation* **2011**, *7*, 525–537.
13. Votavova, P.; Tomala, J.; Kovar, M. Increasing the biological activity of IL-2 and IL-15 through complexing with anti-IL-2 mAbs and IL-15Rα-Fc chimera. *Immunology Letters* **2014**, *159*, 1–10.
14. Shevach, E.M. Application of IL-2 therapy to target T regulatory cell function. *Trends in Immunology* **2012**, *33*, 626–632.
15. Catterall, W.A.; Few, A.P. Calcium Channel Regulation and Presynaptic Plasticity. *Neuron* **2008**, *59*, 882–901.
16. Catterall, W.A. Structure and Regulation of Voltage-Gated Ca^{2+} Channels. *Annual Review of Cell and Developmental Biology* **2000**, *16*, 521–555.

17. Li, W.; Shi, G. How $Ca_V1.2$-bound verapamil blocks Ca^{2+} influx into cardiomyocyte: Atomic level views. *Pharmacological Research* **2019**, *139*, 153–157.
18. Ding, X.; Zhang, B. DeepBAR: A Fast and Exact Method for Binding Free Energy Computation. *The Journal of Physical Chemistry Letters* **2021**, *12*, 2509–2515.
19. Trosset, J.Y.; Cavé, C. In Silico Drug–Target Profiling. In *Target Identification and Validation in Drug Discovery*; Springer New York, 2019; pp. 89–103.
20. Agrawal, P.; Raghav, P.K.; Bhalla, S.; Sharma, N.; Raghava, G.P.S. Overview of free software developed for designing drugs based on protein-small molecules interaction. *Curr. Top. Med. Chem.* **2018**, *18*, 1146–1167.
21. Zeng, Y.; Chen, X.; Luo, Y.; Li, X.; Peng, D. Deep drug-target binding affinity prediction with multiple attention blocks. *Briefings in Bioinformatics* **2021**, *22*.
22. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery* **2018**, *18*, 41–58.
23. Han, K.; Cao, P.; Wang, Y.; Xie, F.; Ma, J.; Yu, M.; Wang, J.; Xu, Y.; Zhang, Y.; Wan, J. A Review of Approaches for Predicting Drug–Drug Interactions Based on Machine Learning. *Frontiers in Pharmacology* **2022**, *12*.
24. Li, W. Strengthening Semaglutide-GLP-1R Binding Affinity via a Val27-Arg28 Exchange in the Peptide Backbone of Semaglutide: A Computational Structural Approach. *Journal of Computational Biophysics and Chemistry* **2021**, *20*, 495–499.
25. Rifaioglu, A.S.; Atalay, R.C.; Kahraman, D.C.; Doğan, T.; Martin, M.; Atalay, V. MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery. *Bioinformatics* **2020**, *37*, 693–704.
26. D'Souza, S.; Prema, K.; Balaji, S. Machine learning models for drug–target interactions: current knowledge and future directions. *Drug Discovery Today* **2020**, *25*, 748–756.
27. Malone, F.D.; Parrish, R.M.; Welden, A.R.; Fox, T.; Degroote, M.; Kyoseva, E.; Moll, N.; Santagati, R.; Streif, M. Towards the simulation of large scale protein–ligand interactions on NISQ-era quantum computers. *Chemical Science* **2022**, *13*, 3094–3108.
28. Fuji, H.; Qi, F.; Qu, L.; Takaesu, Y.; Hoshino, T. Prediction of Ligand Binding Affinity to Target Proteins by Molecular Mechanics Theoretical Calculation. *Chemical and Pharmaceutical Bulletin* **2017**, *65*, 461–468.
29. Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M.P. Molecular mechanics methods for predicting protein–ligand binding. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5166–5177.
30. Cavasotto, C.N. Binding Free Energy Calculation Using Quantum Mechanics Aimed for Drug Lead Optimization. In *Methods in Molecular Biology*; Springer US, 2020; pp. 257–268.
31. Soni, A.; Bhat, R.; Jayaram, B. Improving the binding affinity estimations of protein–ligand complexes using machine-learning facilitated force field method. *Journal of Computer-Aided Molecular Design* **2020**, *34*, 817–830.
32. Ballester, P.J.; Mitchell, J.B.O. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* **2010**, *26*, 1169–1175.
33. Bitencourt-Ferreira, G.; de Azevedo, W.F. Machine Learning to Predict Binding Affinity. In *Methods in Molecular Biology*; Springer New York, 2019; pp. 251–273.
34. Xue, L.C.; Rodrigues, J.P.; Kastritis, P.L.; Bonvin, A.M.; Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **2016**, p. btw514.
35. Li, W. Delving deep into the structural aspects of a furin cleavage site inserted into the spike protein of SARS-CoV-2: A structural biophysical perspective. *Biophysical Chemistry* **2020**, *264*, 106420.
36. Li, W. Extracting the Interfacial Electrostatic Features from Experimentally Determined Antigen and/or Antibody-Related Structures inside Protein Data Bank for Machine Learning-Based Antibody Design **2020**.
37. Li, W. How do SMA-linked mutations of *SMN1* lead to structural/functional deficiency of the SMA protein? *PLOS ONE* **2017**, *12*, e0178519.
38. Callaway, E. The entire protein universe: AI predicts shape of nearly every known protein. *Nature* **2022**.
39. Wang, R.; Fang, X.; Lu, Y.; Yang, C.Y.; Wang, S. The PDBbind Database: Methodologies and Updates. *Journal of Medicinal Chemistry* **2005**, *48*, 4111–4119.
40. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research* **2018**, *47*, D520–D528.
41. Bonvin, A.M.J.J. 50 years of PDB: a catalyst in structural biology. *Nature Methods* **2021**, *18*, 448–449.

42. Xiong, P.; Zhang, C.; Zheng, W.; Zhang, Y. BindProfX: Assessing Mutation-Induced Binding Affinity Change by Protein Interface Profiles with Pseudo-Counts. *Journal of Molecular Biology* **2017**, *429*, 426–434.
43. Erijman, A.; Rosenthal, E.; Shifman, J.M. How Structure Defines Affinity in Protein-Protein Interactions. *PLoS ONE* **2014**, *9*, e110085.
44. Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R.K. Artificial intelligence in drug discovery and development. *Drug Discovery Today* **2021**, *26*, 80–93.
45. Bao, J.; He, X.; Zhang, J.Z.H. DeepBSP—a Machine Learning Method for Accurate Prediction of Protein–Ligand Docking Structures. *Journal of Chemical Information and Modeling* **2021**, *61*, 2231–2240.
46. Gilson, M.K.; Zhou, H.X. Calculation of Protein-Ligand Binding Affinities. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36*, 21–42.
47. Li, W. Designing rt-PA Analogs to Release its Trapped Thrombolytic Activity. *Journal of Computational Biophysics and Chemistry* **2021**, *20*, 719–727.
48. Boyles, F.; Deane, C.M.; Morris, G.M. Learning from the ligand: using ligand-based features to improve binding affinity prediction. *Bioinformatics* **2019**.
49. Karimi, M.; Wu, D.; Wang, Z.; Shen, Y. DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics* **2019**, *35*, 3329–3338.
50. Hu, F.; Jiang, J.; Wang, D.; Zhu, M.; Yin, P. Multi-PLI: interpretable multi-task deep learning model for unifying protein–ligand interaction datasets. *Journal of Cheminformatics* **2021**, *13*.
51. Jensen, O.; Brockmöller, J.; Dücker, C. Identification of Novel High-Affinity Substrates of OCT1 Using Machine Learning-Guided Virtual Screening and Experimental Validation. *Journal of Medicinal Chemistry* **2021**, *64*, 2762–2776.
52. Anfinsen, C.B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230.
53. Hasani, H.J.; Barakat, K. Homology Modeling: an Overview of Fundamentals and Tools. *International Review on Modelling and Simulations (IREMOS)* **2017**, *10*, 129.
54. Antunes, D.A.; Abella, J.R.; Devaurs, D.; Rigo, M.M.; Kavraki, L.E. Structure-based Methods for Binding Mode and Binding Affinity Prediction for Peptide-MHC Complexes. *Current Topics in Medicinal Chemistry* **2019**, *18*, 2239–2255.
55. Crystallography: Protein Data Bank. *Nature New Biology* **1971**, *233*, 223–223.
56. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **2003**, *10*, 980–980.
57. Li, W. Half-a-century Burial of ρ, and φ in PDB **2021**.
58. Lutomski, C.A.; El-Baba, T.J.; Robinson, C.V.; Riek, R.; Scheres, S.H.; Yan, N.; AlQuraishi, M.; Gan, L. The next decade of protein structure. *Cell* **2022**, *185*, 2617–2620.
59. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
60. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710.
61. Hey, T.; Butler, K.; Jackson, S.; Thiyagalingam, J. Machine learning and big scientific data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2020**, *378*, 20190054.
62. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins: Structure, Function, and Bioinformatics* **2019**, *87*, 1141–1148.
63. AlQuraishi, M. Machine learning in protein structure prediction. *Current Opinion in Chemical Biology* **2021**, *65*, 1–8.
64. AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **2019**, *35*, 4862–4865.
65. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876.
66. Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences* **2020**, *117*, 1496–1503.

67. Pinheiro, F.; Santos, J.; Ventura, S. AlphaFold and the amyloid landscape. *Journal of Molecular Biology* **2021**, p. 167059.
68. Ruff, K.M.; Pappu, R.V. AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of Molecular Biology* **2021**, p. 167208.
69. Higgins, M.K. Can We AlphaFold Our Way Out of the Next Pandemic? *Journal of Molecular Biology* **2021**, p. 167093.
70. Strodel, B. Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins. *Journal of Molecular Biology* **2021**, p. 167182.
71. Ayoub, R.; Lee, Y. Protein structure search to support the development of protein structure prediction methods. *Proteins: Structure, Function, and Bioinformatics* **2021**, *89*, 648–658.
72. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596.
73. Layer, G.; Reichelt, J.; Jahn, D.; Heinz, D.W. Structure and function of enzymes in heme biosynthesis. *Protein Science* **2010**, *19*, 1137–1161.
74. Gupta, R.; Srivastava, D.; Sahu, M.; Tiwari, S.; Ambasta, R.K.; Kumar, P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular Diversity* **2021**, *25*, 1315–1360.
75. Hessler, G.; Baringhaus, K.H. Artificial Intelligence in Drug Design. *Molecules* **2018**, *23*, 2520.
76. Deciphering critical amino acid residues to modify and enhance the binding affinity of ankyrin scaffold specific to capsid protein of human immunodeficiency virus type 1. *Asian Pacific Journal of Allergy and Immunology* **2017**.
77. Olabi, M.; Stein, M.; Wätzig, H. Affinity capillary electrophoresis for studying interactions in life sciences. *Methods* **2018**, *146*, 76–92.
78. Veit-Acosta, M.; de Azevedo Junior, W.F. The Impact of Crystallographic Data for the Development of Machine Learning Models to Predict Protein-Ligand Binding Affinity. *Current Medicinal Chemistry* **2021**, *28*, 7006–7022.
79. Meng, Z.; Xia, K. Persistent spectral–based machine learning (PerSpect ML) for protein-ligand binding affinity prediction. *Science Advances* **2021**, *7*.
80. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018**, *46*, W296–W303.
81. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, *25*, 1605–1612.
82. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; et al. Protein complex prediction with AlphaFold-Multimer **2021**.
83. Greenidge, P.A.; Kramer, C.; Mozziconacci, J.C.; Wolf, R.M. MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement. *Journal of Chemical Information and Modeling* **2012**, *53*, 201–209.
84. Liu, Z.; Li, J.; Liu, J.; Liu, Y.; Nie, W.; Han, L.; Li, Y.; Wang, R. Cross-Mapping of Protein-Ligand Binding Data Between ChEMBL and PDBbind. *Molecular Informatics* **2015**, *34*, 568–576.
85. Liu, Z.; Li, Y.; Han, L.; Li, J.; Liu, J.; Zhao, Z.; Nie, W.; Liu, Y.; Wang, R. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* **2014**, *31*, 405–412.
86. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. In *Methods in Molecular Biology*; Springer New York, 2017; pp. 627–641.
87. Ruddigkeit, L.; van Deursen, R.; Blum, L.C.; Reymond, J.L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *Journal of Chemical Information and Modeling* **2012**, *52*, 2864–2875.
88. Rezaei, M.A.; Li, Y.; Wu, D.; Li, X.; Li, C. Deep Learning in Drug Design: Protein-Ligand Binding Affinity Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2022**, *19*, 407–417.
89. Tanoori, B.; Jahromi, M.Z.; Mansoori, E.G. Binding affinity prediction for binary drug–target interactions using semi-supervised transfer learning. *Journal of Computer-Aided Molecular Design* **2021**, *35*, 883–900.
90. Quijano-Rubio, A.; Ulge, U.Y.; Walkey, C.D.; Silva, D.A. The advent of de novo proteins for cancer immunotherapy. *Current Opinion in Chemical Biology* **2020**, *56*, 119–128.

91. Silva, D.A.; Yu, S.; Ulge, U.Y.; Spangler, J.B.; Jude, K.M.; Labão-Almeida, C.; Ali, L.R.; Quijano-Rubio, A.; Ruterbusch, M.; Leung, I.; et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **2019**, *565*, 186–191.
92. Mason, D.M.; Friedensohn, S.; Weber, C.R.; Jordi, C.; Wagner, B.; Meng, S.M.; Ehling, R.A.; Bonati, L.; Dahinden, J.; Gainza, P.; et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nature Biomedical Engineering* **2021**, *5*, 600–612.
93. Yan, J.; Qiu, Y.; dos Santos, A.M.R.; Yin, Y.; Li, Y.E.; Vinckier, N.; Nariai, N.; Benaglio, P.; Raman, A.; Li, X.; et al. Systematic analysis of binding of transcription factors to noncoding variants. *Nature* **2021**, *591*, 147–151.
94. Li, H.; Lu, G.; Sze, K.H.; Su, X.; Chan, W.Y.; Leung, K.S. Machine-learning scoring functions trained on complexes dissimilar to the test set already outperform classical counterparts on a blind benchmark. *Briefings in Bioinformatics* **2021**, *22*.
95. Ortet, P.C.; Muellers, S.N.; Viarengo-Baker, L.A.; Streu, K.; Szymczyna, B.R.; Beeler, A.B.; Allen, K.N.; Whitty, A. Recapitulating the Binding Affinity of Nrf2 for KEAP1 in a Cyclic Heptapeptide, Guided by NMR, X-ray Crystallography, and Machine Learning. *Journal of the American Chemical Society* **2021**, *143*, 3779–3793.
96. Marasco, D.; Scognamiglio, P.L. Identification of inhibitors of biological interactions involving intrinsically disordered proteins. *Int. J. Mol. Sci.* **2015**, *16*, 7394–7412.
97. Villoutreix, B.O.; Kuenemann, M.A.; Poyet, J.L.; Bruzzoni-Giovanelli, H.; Labbe, C.; Lagorce, D.; Sperandio, O.; Miteva, M.A. Drug-Like ProteinProtein Interaction Modulators: Challenges and Opportunities for Drug Discovery and Chemical Biology. *Molecular Informatics* **2014**, *33*, 414–437.
98. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction **2022**.
99. Antunes, D.A.; Devaurs, D.; Moll, M.; Lizée, G.; Kavraki, L.E. General Prediction of Peptide-MHC Binding Modes Using Incremental Docking: A Proof of Concept. *Scientific Reports* **2018**, *8*.