Review



# Towards a Truly General Intermolecular Binding Affinity Calculator for Drug Discovery & Design

Wei Li * and Gary Vottevor

Posted Date: 19 October 2023

doi: 10.20944/preprints202208.0213.v2

Keywords: GIBAC; Biophysics; Structural Biology; Drug discovery \& design; Artificial intelligence-integrated drug discovery

*Article*

# Towards a Truly General Intermolecular Binding Affinity Calculator for Drug Discovery & Design

**Wei Li [1,*] and Gary G. Vottevor [1]**

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

* Correspondence: wli148@aucklanduni.ac.nz

**Abstract:** Intermolecular interactions are the fabrics underlying almost all processes in living organisms, where two cornerstone concepts, intermolecular binding affinity ($K_d$) and binding energy ($\Delta G$), have long been established to physically describe the strengths of biomolecular interactions, e.g., drug-target $K_d$ and $\Delta G$ to describe the strength of drug-target interaction. The past two-three years saw a big step forward in the use of artificial intelligence (AI) in structural biology (e.g., AlphaFold for protein structure prediction) and drug discovery & design. In light of the roles of $K_d$ and $\Delta G$ in drug discovery & design, the speed of this AI progress raises a question of what's next for its practical application in the pharmaceutical industry, in addition to a system-wide account of biomolecular structures and motions. Last August, the concept of a general intermolecular binding affinity calculator (GIBAC) was for the first time coined and proposed in an MDPI-published preprint. Here, this article puts forward an updated conceptual and practical framework of GIBAC, including its inception, definition, construction, practical applications, technical challenges and limitations, and future directions. Moreover, this article argues that the time is now ripe for the construction of such an accurate, precise and efficient GIBAC to be on the agenda of the entire drug discovery & design community, to ensure its applicability & reliability, and to enhance its value in drug R&D in future.

**Keywords:** GIBAC; biophysics; structural biology; drug discovery & design; artificial intelligence-integrated drug discovery

## Introduction

On August 11, 2022, the concept of a general intermolecular binding affinity calculator (GIBAC) was for the first time coined and proposed in an MDPI-published preprint (https://www.preprints.org/manuscript/202208.0213/v1) [1], of which this article puts forward an update, including GIBAC's inception, definition, construction, practical applications, technical challenges and limitations, and future directions. In addition, this article discusses how we can combine the insights of structural biology, biophysics and artificial intelligence to build an accurate, precise and efficient GIBAC, and to navigate the next frontier for its practical application in drug discovery & design.

## Inception of GIBAC: the origin of the idea of it

### *Intermolecular binding affinity is key to how genetic variation affects drug discover & design*

Genetic variation is a fundamental aspect of biology [2–4], which occurs not just in infectious diseases [5,6], but also in non-communicable diseases, e.g., cancer [7], cardiovascular [8,9] and rare diseases [10–12]. Take SARS-CoV-2 for instance [13,14]. The emergence of SARS-CoV-2 variants has had implications for the discovery and design of drugs (small molecules and antibodies) and vaccines. As revealed by genomic surveillance, SARS-CoV-2 has undergone genetic mutations that alter its spike (S) protein, the primary target for discovery and design of neutralizing antibodies and vaccines against SARS-CoV-2 [15,16]. In the meanwhile, these mutations could lead to changes in the virus's antigenic properties (Figure 1), conferring resistance to specific monoclonal antibodies, and reducing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



the effectiveness of existing antibodies and vaccines [17,18]. In addition to antibodies and vaccines, small molecules inhibiting spike-ACE2 protein-protein interactions (PPI) have also been the focus of a variety of R&D efforts to develop potential blockers of viral attachment and entry for SARS-CoV-2 into host cells [19,20]. With this respect, SARS-CoV-2 variants carry genetic changes in critical viral proteins (e.g., the main protease) [21–23], which alter their binding affinity to small molecule inhibitors.

In infectious diseases like COVID-19, therefore, genetic changes of the pathogen(s) lead to structural and functional consequences [24–27], which in turn affect the efficacy of drugs and vaccines against infectious agents such as SARS-CoV-2 [28], calling for an intermolecular $K_d$ calculator to be able to accommodate genetic variations, as described in Equation 1, where $\underline{molA}$, $\underline{molB}$, and $\underline{molC}$ represent S protein, ACE2 on the host cell's membrane, antibody or small molecule inhibitors against SARS-CoV-2, respectively, while *mutation* represents genetic mutation(s) of SARS-CoV-2's S protein, *envPara* represents $K_d$-related enviromental parameters such as temperature, pH [1].

$$K_d = f(molA, molB, molC, envPara, mutation(s)) \tag{1}$$

Of further interest, an intermolecular $K_d$ calculator as described in Equation 1 can strengthen our structural and functional understanding of those genetic variations, which can in turn aid the design of multiple-sites targeting therapeutics with enhanced binding affinities, towards the discovery and development of broad-spectrum inhibitors against infectious pathogens' (e.g., SARS-CoV-2) variants in future [29–32].

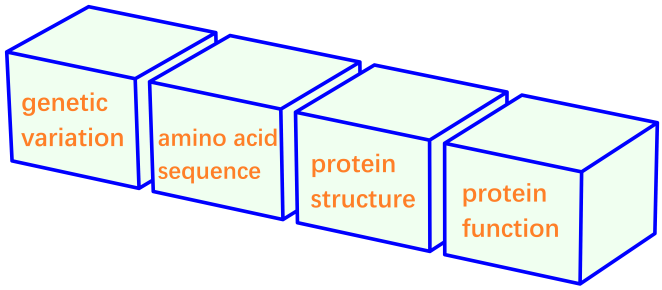


**Figure 1.** Genetic variation-initiated central dogma-like flowchart mechanism through which genetic variation exerts its impact on drug discovery & design, in case `protein function` = `drug target` here.

Similarly, in non-communicable diseases, genetic variations of drug target protein structure(s) are also linked to the discovery & design of therapeutics, including antibodies, small molecule inhibitors and vaccines. For instance, the emergence of resistance to antibody therapies targeting proteins such as EGFR, PD-1 or PD-L1 [33] has been a significant challenge in cancer treatment, as mutations in tumor cells impair the binding of antibodies (or small molecule inhibitors) to these proteins (Figure 1), leading to the development of resistance to therapeutics of a variety of cancers, thereby diminishing their therapeutic efficacy [34,35]. Thus, understanding the mechanisms underlying mutation-induced resistance are crucial for the development of effective strategies to overcome this challenge in cancer immunotherapy [36–38], calling for an intermolecular $K_d$ calculator (as described in Equation 1) to be able to accommodate genetic variations [1], where $\underline{molA}$, $\underline{molB}$, and $\underline{molC}$ represent PD-1, PD-L1 and antibody or small molecule inhibitor against cancer cell's immune escape [39,40], $\underline{mutation}$ represents genetic mutation(s) of PD-1 and/or PD-L1, $\underline{envPara}$ represents $K_d$-related environmental parameters, e.g., temperature, pH [1].

Taken together, therefore, it is necessary to build a general intermolecular binding affinity calculator (GIBAC [1]), which is able to accomodate genetic variation(s), towards efficient design-make-test cycles of therapeutic candidates for the treatments of both infectious and non-communicable diseases [41].

 doi:10.20944/preprints202208.0213.v2



*Towards a truly general intermolecular $K_d$ calculator*

Intermolecular binding affinity, e.g., drug-target binding affinity ($K_d$), is inextricably linked to the mechanism through which genetic variation exerts its impact (Figure 1) on drug discovery & design. Yet, the degree to which genetic variations occur is much lower than that of the complexity of (bio)molecules, which includes a variety of factors such as alternative splicing [10,11], post-translational modifications (PTMs), post-expression modifications (PEMs), chemical and biological space [42–44], three-dimensional structure [45–47], structural folding and conformational dynamics [48,49], intermolecular interactions, induced-fit binding [50,51], conformational transition and dynamics of intrinsically disordered proteins [52–54]. To achieve accurate calculation of intermolecular $K_d$, biomolecular structural information of the interacting partners is indispensable but not always available for physics-based intermolecular $K_d$ calculators such as Prodigy and BindProfX [55–57].

To date, currently available approaches for the calculation of intermolecular $K_d$ include three approaches, i.e., physics-, statistics- and artificial intelligence (AI)-based approaches, all of which are with advantages and disadvantages. For instance, physics-based models [58,59] are as accurate as its physical representation of the underlying principles governing intermolecular interactions, while the complexity of the physical equations make the calculations computationally expensive and time-consuming; Statistics-based models can capture complex relationships between molecular features and binding affinities, and does not require detailed knowledge of the underlying physics. Yet, its performance is highly dependent on the quality and the quantity of available data, with limited accuracy in case statistical approaches do not capture the full complexity of the intermolecular interactions; Lastly, AI-based approaches [60,61] can learn complex patterns and correlations from large datasets, and capture non-linear relationships between molecular features and binding affinities. However, AI approaches require extensive training data, the accuracy of AI models is also highly dependent on the quality and quantity of the training data [46,62,63]. To ensure accuracy and precision of intermolecular $K_d$ calculation, therefore, a combined hybrid approach of AI and physics is herein described and discussed below in the section of **Construction of GIBAC**.

Taken together, none of the three approaches (as of October 19, 2023) is able to collectively meet the standards (listed below) of a GIBAC with adequate accuracy, precision, efficiency and practical applications in drug discovery & design. This article therefore puts forward a conceptual and practical framework to build such a GIBAC, and to ensure its applicability, reliability and efficiency, and to enhance its value in the pharmaceutical industry [41].

1. A truly general intermolecular $K_d$ calculator needs to take into account genetic variations, as described in Equation 1);
2. Structural information of the interacting partners is indispensable but not always available for physics-based calculation of intermolecular $K_d$ such as Prodigy [64,65], making it fall short of a truly general intermolecular $K_d$ calculator;
3. A variety of intermolecular $K_d$-relevant factors need to be taken into account, such as temperature, biomolecular structural dynamics, pH [66–68], site-specific protonation states (e.g., side chain pKa of protein) [56,69–71], PTMs, PEMs [27,72–74], ionic strength, buffer conditions [56,75,76];
4. A truly general intermolecular $K_d$ calculator requires a general forcefield for all types of atoms [77];
5. A truly general intermolecular $K_d$ calculator requires a universal linear string/graph-based notation system for accurate and flexible description and representation of all molecular types and drug modalities [48,78].
6. A truly general intermolecular $K_d$ calculator is able to be used the other way around in drug discovery & design, i.e., to be used as a search engine for drug candidate(s). With the search engine, a list of therapeutic candidates can be retrieved and ranked according to drug-target $K_d$ values, with input parameters including a drug target and a desired drug-target $K_d$ (or a range of $K_d$ values).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



**A biophysical and structural definition of GIBAC**

*The biophysics underlying intermolecular binding affinity*

By definition, intermolecular binding affinity is the physical strength of the binding between a single molecule to its partner, and is typically expressed as the equilibrium dissociation constant ($K_d$) [55,79]:

$$K_d = e^{\frac{-\Delta G}{RT}} \tag{2}$$

where $K_d$ is the binding affinity (Molar, M), $\Delta G$ is the binding free energy (kcal $mol^{-1}$), $R$ is the ideal gas constant ($1.987 \times 10^{-3}$ kcal $Kelvin^{-1}$ $mol^{-1}$), $T$ is the temperature in Kelvin, $ln$ denotes the natural logarithm, $e$ denotes the base of $ln$. For $K_d$ in Equation 2, the smaller the value of $K_d$, the greater the drug-target binding affinity; The larger the value of $K_d$, the weaker the drug-target binding affinity [80,81].

$$\Delta G = -RTln(K_d) \tag{3}$$

For $\Delta G$ in Equation 3 [82], the structural biophysics of a molecular system is governed by a delicate balance between attractive and repulsive forces, where $\Delta G$ is defined as the difference in the free energy between the bound state and the unbound state of the molecules [83]. Specifically, in Equation 3,

1. a negative value of $\Delta G$ ($K_d > 1$ M) indicates energetically favorable binding, i.e., an overall attractive force between the two molecules;
2. a positive value of $\Delta G$ ($K_d < 1$ M) indicates energetically unfavorable binding, i.e., an overall repulsive force between the two molecules;
3. a zero value of $\Delta G$ ($K_d = 1$ M) indicates that no binding or interaction exist between the two molecules, i.e., no energy is required to seperate the two molecules or to keep the two molecules bound.

In practice, $K_d$ and $\Delta G$ are influenced by a variety of factors, including non-covalent intermolecular interactions such as electrostatics (e.g., hydrogen bond or salt bridge) [24–26], hydrophobics and Van der Waals (VdW, Figure 11) forces [84–86], and environmental parameters (abbreviated as *envPara* here) such as pH (e.g., protonation states of ionizable residues' side chains [87–89]), ionic strength and temperature (e.g., in molecular dynamics) [75,76]. Furthermore, $K_d$ and $\Delta G$ may also be affected by the presence of additional molecules, where there are multiple interacting molecules in the system [90].

In light of the fact that drug-target $K_d$ is an essential parameter in drug discovery & design [27,91,92], a series of computational tools have already been developed to calculate drug-target $K_d$, including molecular mechanics-based calculations [93,94] and machine-learning based predictions [60,61,95,96]. Last August, the concept of GIBAC was for the first time coined and proposed in an MDPI preprint [1]. Here, this article puts forward an update of it, i.e., a truly general intermolecular $K_d$ calculator:

$$K_d = f(molecules, envPara) \tag{4}$$

where *molecules* represents the molecular system with the number of interacting partners being $X$ (two, three, or more), and the molecular system is to be described in strings, e.g., amino acid sequences, strings of letters for proteins, or strings of SMILES (Simplified Molecular Input Line Entry System) to represent the chemical structure of small molecules [97–100], or graphs to describe glycosylated proteins for instance, and *envPara* represents environmental parameters [56,69–71].

For instance, in case $X = 3$ (i.e., in Equation 1), one example of GIBAC is a molecular system of PD-1 antibody, PD-1 and PD-L1, where *molA*, *molB* and *molC* represent PD-1, PD-L1 and PD-1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



antibody (Keytruda for instance [101]), respectively. For a GIBAC as described in Equation 1, a traversal intermolecular $K_d$ means the calculations of a list of intermolecular $K_d$ values:

1. $K_{d1}$ ($\Delta G_1$) between $molA$ and $molB$.
2. $K_{d2}$ ($\Delta G_2$) between $molA$ and $molC$.
3. $K_{d3}$ ($\Delta G_3$) between $molC$ and $molB$.
4. $K_{d4}$ ($\Delta G_4$) between $molA + B$ and $molC$.
5. $K_{d5}$ ($\Delta G_5$) between $molA + C$ and $molB$.
6. $K_{d6}$ ($\Delta G_6$) between $molB + C$ and $molA$.

With the discovery & design of PD-1 antibody [33,102] as an example, a set of biophysical principles are defined as below:

1. $K_{d2} < \min([K_{d1},K_{d3},K_{d4},K_{d5},K_{d6}])$ (in Python syntax), to eusure that the complex structure of $molA$ and $molC$ is the most energetically favourable among the six possible situations listed above.
2. the smaller the value of $K_{d2}$, the better from a biophysical (yet not necessarily therapeutic [103,104]) point of view;
3. the smaller the value of $K_{d2}$, the better for the *in vitro* use of the antibody in diagnosis or bioprocessing, e.g., purification [105];
4. $\Delta G_3 = 0$ or $> 0$, to ensure that there is no binding or interaction between $molB$ and $molC$, or, there exists a repulsive force between $molB$ and $molC$;
5. $K_{d2} < K_{d5}$, and $K_{d2} < K_{d4}$, to ensure that the complex structure of $molA$ and $molC$ is much more stable than that of $molA$ and $molB$, such that the antibody is able to disrupt the PD-1-PD-L1 axis;
6. $\Delta G_4 = 0$ or $> 0$, $\Delta G_5 = 0$ or $> 0$, and $\Delta G_6 = 0$ or $> 0$, to ensure the existence of $molC$ (antibody) suppresses the PD-1-PD-L1 axis.

In general, drugs exert their pharmacological effects through binding to and interacting with their target(s) [106,107], be they small molecule inhibitors or biologics, making $K_d$ and $\Delta G$ two cornerstones for drug discovery & design. As a result, a truly general intermolecular $K_d$ calculator (GIBAC [1]) is necessary and useful to ensure adequate, accurate, precise and cost-effective knowledge of the $K_d$ and $\Delta G$, which is pivotal in both early-stage drug discovery & design (drug-target $K_d$), and in drug repurposing (drug-target $K_d$), and also in avoiding undue risk of toxicity mediated by drug-drug interactions (DDI, drug-drug $K_d$) [108–110].

*Structural information is indispensable to GIBAC's accuracy*

In the real world, molecules bind to and interact with each other in their three-dimensional configurations (structures) rather than one-dimensional forms (e.g., protein sequences), making accurate calculation of $K_d$ inextricably linked to accurate and abundant structural information, which provides essential data about the three-dimensional arrangement of molecular system, including residue-specific interactions, structural (e.g., geometric [47]) features at the binding interface, shape complementarity [111,112], electrostatic interactions (hydrogen bonding and salt bridging), hydrophobic interactions, VdW forces (Figure 11), et cetera.

As a matter of fact, extractions of the experimental structural and biophysical features [25,27,113] has long been one critical step in both physics- and AI-based approaches towards intermolecular $K_d$ calculations. Take the physics-based Prodigy [55,56] 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 features, electrostatic and hydrophobic interactions, and VdW forces (Figure 11) between two binding molecules [18,26,84,85]. Prodigy [55,56] as a physics-based intermolecular $K_d$ calculator, therefore, can be described as:

$$K_d = f(ABcomplex, envPara) \tag{5}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



where $ABcomplex$ represents the complex structure of the two interacting partners. It is obvious that Equation 5 requires accurate structural information, which is (to date) not always available for any (bio)molecular system. As a result, intermolecular $K_d$ calculators such as described by Equation 5 needs to be generalized further, where structural information (in Equation 5) is replaced with sequences, strings, or graphs:

$$K_d = f(molAsequence, molBsequence, envPara) \tag{6}$$

or,

$$K_d = f(molAstring, molBstring, envPara) \tag{7}$$

or,

$$K_d = f(molAgraph, molBgraph, envPara) \tag{8}$$

To date, a series of computational methods are currently available that uses various AI algorithms to calculate intermolecular $K_d$ solely from protein sequence information, known as sequence-based or sequence-only approaches [114,115]. As described in Equations 6, 7 and 8, $molAsequence$ ($molAstring$) or $molBsequence$ ($molBstring$) represent sequences of amino acids, i.e., strings of letters for protein $A$ or $B$, or SMILES strings [116–118] characters to represent the chemical structure of small molecules $A$ or $B$, and $molAgraph$ represents biomolecules with PTMs, glycosylated protein for instance.

Towards the construction of a truly general intermolecular $K_d$ calculator, here, this article for the first time calls for the establishment of a universal string-based linear or graph-based notation system for integrated lossless descriptions and representations of molecular fingerprints for all molecule types and drug modalities, with or without PTMs and/or PEMs:

1. PTM: glycosylation [119], phosphorylation [120–122], fatty acid modifications [123–126].
2. PEM: lipidation and fatty acid chain attachment(s) [127], including semaglutide (of Novo Nordisk) [27], insulin icodec (of Novo Nordisk) [128].

With such a notation system, Equation 4 can be rewritten as two sets of input parameters and two outputs, as listed in Table 1.

**Table 1.** A tabular description of Equations 4, 6, 7 and 8.

| Input 1 | Input 2 | Output |
|---|---|---|
| $molAstring$, $molBstring$, ... | $envPara$ | $\Delta G$, $K_d$ |
| $molAgraph$, $molBgraph$, ... | $envPara$ | $\Delta G$, $K_d$ |

In addition to such a linear/graph-based notation system, this article also calls for the development of a general forcefield for all available types of atoms (including unnatural amino acid) [77], and continued validation of it, to ensure accuracy and precision in the calculation of intermolecular $K_d$ for all molecule types and drug modalities with or without PTM or PEM.

## Construction of an accurate and precise GIBAC

### *A key ingredient to build a GIBAC: artificial intelligence*

By definition, calculation of intermolecular $K_d$ is a problem of biophysics and structural biology. In light of the advantages and disadvantages of the currently available approaches (physics, AI and statistics) as discussed above, this article here puts forward a hybrid [129] approach of AI and physics

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



to ensure adequate accuracy, precision and interpretability of the intermolecular $K_d$ calculator, i.e., GIBAC [1].

What's more, the space of molecular types and drug modalities is vast [130], extending beyond proteins and small molecules. This makes a comprehensive physics-based exploration practically impossible (Figure 2) [20,42,131]. Another reason for the hybrid approach is the availability of various AI algorithms, including Graph Convolutional Networks (GCN), Graph Neural Networks (GNN), Graph Transformer Networks (GTN), Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), et cetera. Overall, the ultimate task here is for AI algorithms to accurate and precisely understand the biophysics underlying intermolecular binding & interaction, from a structural and pharmaceutical point of view [132–134].

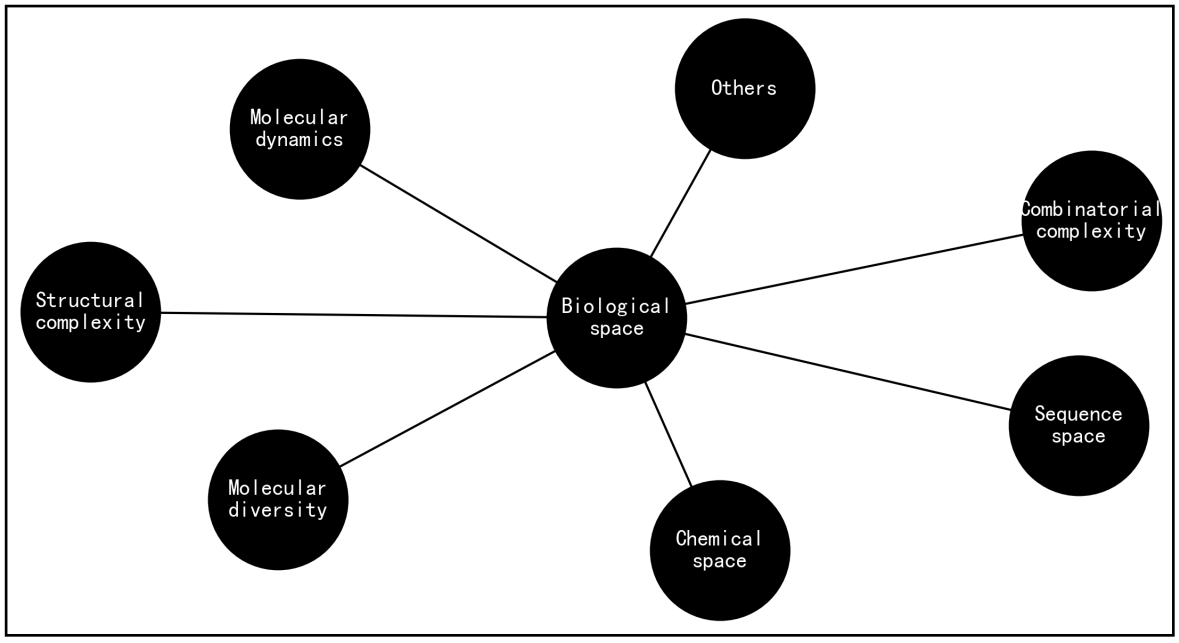


**Figure 2.** Relevant factors of the size of the entire molecular space [42].

*Can AI be a digital crystal ball of drug discovery & design?*

Artificial intelligence (AI) encompasses machine learning (ML) and deep learning (DL), which aim to develop intelligent machines capable of performing tasks that typically require human intelligence, such as understanding natural language, recognizing objects, solving complex problems, and making decisions [5,135,136]. Machine learning focuses on creating algorithms and models that can learn from available data to make predictions or decisions, while deep learning utilizes artificial neural networks with multiple layers to process and learn from data. For AI (including ML and DL), feature extraction is a pivotal process to identify critical features from raw data, capturing important information for the learning process and improving the efficiency and accuracy of AI models [137–139].

As of October 19, 2023, AI is the most hyped technology in 2023 with the advent of generative AI tools such as OpenAI's ChatGPT. In the field of drug discovery & design, however, the current level of enthusiasm surrounding generative AI has elicited diverse reactions in the industry.

1. Dr. Alex Zhavoronkov (co-CEO of Insilico Medicine) thinks that AI can significantly boost the probability of success in drug development, while also agreed that the AI hype has driven valuations for newly-founded companies to levels that likely are not sustainable.
2. 'Nobody in the field is actually using AI,' said Schrödinger CEO Ramy Farid in an interview. 'Describing Schrödinger as a machine-learning company would be like describing Schrödinger as a company that uses Microsoft Office' [140], while Mr. Geoffrey Porges (CFO of Schrödinger)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



considers Schrödinger not as an AI company, but as a proprietary software and drug company [140].

3. 'AI-powered' is tech's meaningless equivalent of all natural', according to Devin Coldewey of TechCrunch, where 'AI-powered' is used to create a perception of advanced technology without providing specific details about how AI is being utilized or what benefits it actually brings to drug discovery & design.

Overall, while the industry itself is concerned about the current AI hype 'inevitably coming back down to earth', there is still optimism about what's next for the field [140], as ultimately AI will prove useful in accelerating this entire process with its ability to analyze and learn from vast amounts of data.

Taken together, drug discovery & design itself is complex, time-consuming, and expensive process, calling for the drug R&D community to get over the current AI hype and back to the basics of a hybrid approach of AI and biophysics [129]. With intermolecular $K_d$ calculator as an example, by leveraging ML and DL techniques, AI can analyze and learn from molecular sequence and structure data, including residue-specific interactions, interstructural features at the binding interface [26,27,47], shape complementarity [111,112], electrostatic interactions (hydrogen bonding and salt bridging), hydrophobic interactions, VdW forces (Figure 11) [141].

Consequently, this article here

1. argues that AI can indeed act as a digital crystal ball of drug discovery & design, provided that it is used in combination with experimental insights from structural biology, biophysics, pharmacology [134], et cetera.
2. puts forward a hybrid approach of AI and biophysics to ensure adequate accuracy, precision and interpretability of a truly general intermolecular $K_d$ calculator, i.e., GIBAC [1], as described in Equation 4 and Table 1.

*Construction of GIBAC: experimental data and tools*

AI algorithms rely on huge amounts of data to learn, train, and improve their performance continuously, where its quantity and quality inextricably linked to the performance of the AI model [46,62,63]. As charted out previously in [1], therefore, the construction of GIBAC requires two key ingredients, i.e., data and algorithm, and is to follow a roadmap as defined by Equation 9 [1].

$$data + algorithm = model \tag{9}$$

To ensure a GIBAC with adequate accuracy and precision, a substantial amount of experimental data with reasonable accuracy is crucial, including Protein Data Bank (PDB) [46,142], PDBbind and BindingDB [143–145], CASF2016 and CASF2013 databases [146,147], DUD-E [148], ChEMBL [149], DrugBank [150], CSAR [151], MUV [152], PDBbind-CN [153], Antigen-Antibody Interaction Database (AgAbDb) [154], NIH molecular libraries initiative database [155], the international ImMunoGeneTics information system (IGMT) [156], et cetera.

Moreover, a wide range of wet-lab tools are also available to ensure continued accumulation of experimental data in structural biology, biophysics, medicinal and computational chemistry, drug discovery & design, including isothermal titration calorimetry (ITC) [157–159], surface plasmon resonance (SPR) [160], nuclear magnetic resonance (NMR) spectroscopy [161], cryogenic electron microscopy (cryo-EM) [162], fluorescence resonance energy transfer (FRET) [163], microscale thermophoresis [164], differential scanning fluorimetry [165], X-ray crystallography [166], mass spectrometry [167], bio-layer interferometry [168], et cetera.

Nonetheless, experimental data and tools alone are useful but insufficient in the construction of a GIBAC with adequate accuracy and precision (Equation 4) [1].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



*Construction of GIBAC: computational data and tools*

As mentioned earlier, exploring the entire molecular space with physical calculation alone is practically impossible due to its size (Figure 2) [20,42,131]. In the construction of GIBAC, a hybrid approach combining AI and physics is therefore described here. However, AI algorithms can still struggle to accurately and precisely calculate intermolecular $K_d$ data that are different from the training data set, which only covers a limited portion of the molecular space (Figure 2) [130]. This is where synthetic data and its generators (i.e., computaional tools) come in [169,170], including:

1. computational structural data from AlphaFold database [171];
2. synthetic (both apo and complex) structural data generators [172–175];
3. molecular docking tools [176].
4. synthetic $K_d$ data generators [55,56,177–180];
5. molecular dynamics simulations tools [70,181,181];
6. side chain placement and energy minimization algorithms [182] to incorporate structural arrangement information of PTMs and PEMs into currently available structural models.

*High-throughput generation of synthetic data for GIBAC*

To illustrate how synthetic structural and biophysical data is generated with reasonable accuracy by currently available computational tools, Modeller [183] and the PD-1/PD-L1 complex structure [184,185] are used as an example to generate synthetic apo and complex structural data.

Specifically, the PD-L1 length is 222 amino acid residues, and the PD-1 length is 134, and their combined length is 356. The experimental PD-1/PD-L1 complex structure (PDB access code: 3BIK) is used as a starting point to generate synthetic data by introducing a limited number ($k$) of site-directed mutations to ensure reasonable accuracy of the synthetic data. With PD-1/PD-L1 complex structure as a template, the size ($s$ in Equation 10) of the synthetic structural data is described as below,

$$s = g(n,k) = \frac{n!}{k!(n-k)!} \times 20^k \tag{10}$$

where $n$ (Equation 10) represents the length of PD-1, PD-L1 or the complex of the two, and $k/n <$ 5% (Equation 10) ensure the overall reasonable accuracy of the synthetic structural data. Thus,

1. for PD-L1 (an apo structure), $s = g(222, 11) = 1.2571 \times 10^{18}$;
2. for PD-1 (an apo structure), $s = g(134, 6) = 7.1779 \times 10^{9}$;
3. for PD-L1/PD-1 (a complex structure), $s = g(356, 17) = 4.5190 \times 10^{28}$.

In short, therefore, for one experimental structure ($n = 134$), a total of $7.17 \times 10^9$ synthetic apo structures are able to be generated based on the experimental structure of PD-1 with reasonable accuracy, i.e., those synthetic apo structures is at least 95% homologous to its experimental template, i.e., the experimental structure of PD-1.

Here, the generation of synthetic structural data is similar to the distribution of the electron cloud model of a hydrogen atom, as shown in Figure 3. A hydrogen atom is composed of a single negatively charged electron, moving around the positively charged proton which is the nucleus of the hydrogen atom, where $R$ (Figure 3) represents the distance between the electron (synthetic data) and the proton (i.e., nucleus, experimental data) of hydrogen, the larger the value of $R$ (Figure 3), the lower the density of the electron cloud of the hydrogen atom, i.e.,, the higher the value of $k/n$ (Equation 10), the lower the homology between the synthetic data and its experimental template, the lower the accuracy of the synthetic structural data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



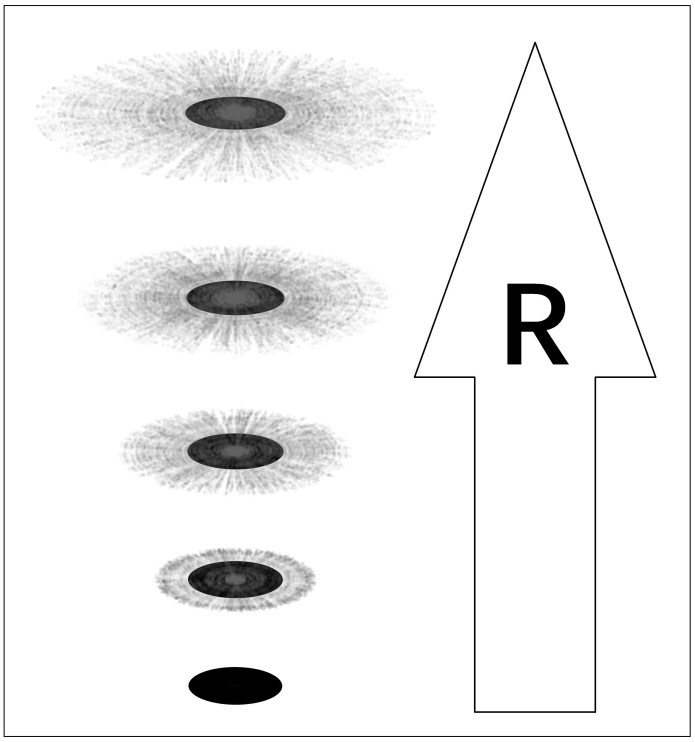


**Figure 3.** An illustration of the generation of synthetic structural data through the electron cloud model of a hydrogen atom. In this figure, the black solid circle represents the hydrogen nuclei, the scattered black dots represent the possible locations of hydrogen's electron, while *R* represents the distance between the electron (synthetic data) and the proton (i.e., nucleus, experimental data) of hydrogen.

Taken together, as shown in Figure 4, the white region represents the experimentally and comutationally uncharted territories (i.e., the ocean, Figure 4), while the rest of Figure 4 consists of islands of synthetic data, with the atomic nuclei representing experimental data, and the electron clouds representing the synthetic data. Thus, Figure 4 is indeed a collection of mini GIBACs (Figure 4) for target-specific calculations and ranking of $K_d$ and $\Delta G$. Specifically, the hydrogen electron cloud models in Figure 4 is able to represent apo and complex structural data (both experimental and synthetic), and also $K_d$ and $\Delta G$ data, both experimental and synthetic, where the white regions of Figure 4 corresponds to uncharted territories of the $K_d$ and $\Delta G$ data. To train an AI- and physics-based intermolecular $K_d$ calculator, therefore, $K_d$ is predefined to be 1 M for white regions of Figure 4, until new experimental $K_d$ data [157–159] is updated in databases as mentioned above [143–145].

Hence, in spite of the size of the entire molecular space (Figure 2) [130], the size of the synthetic data space is also quite considerable, and with a variety of synthetic data generators such as Modeller [183] and Prodigy [55,56,177], it still is conceivable for currently available AI algorithms (or new ones in future) to learn, train, and improve their performance continuously, in the sense that AI algorithms keep learning the biophysics underlying molecular folding [186,187], binding & interaction [132,133], and structural action mechanism [134]. Overall, the situation here is a bit like the generation of a human molecular structural binding and interacting atlas, akin to Human BioMolecular Atlas Program (HuBMAP), a global initiative that aims to assemble spatial maps of biomolecules, including RNA, proteins, and metabolites at single-cell resolution [188,189].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



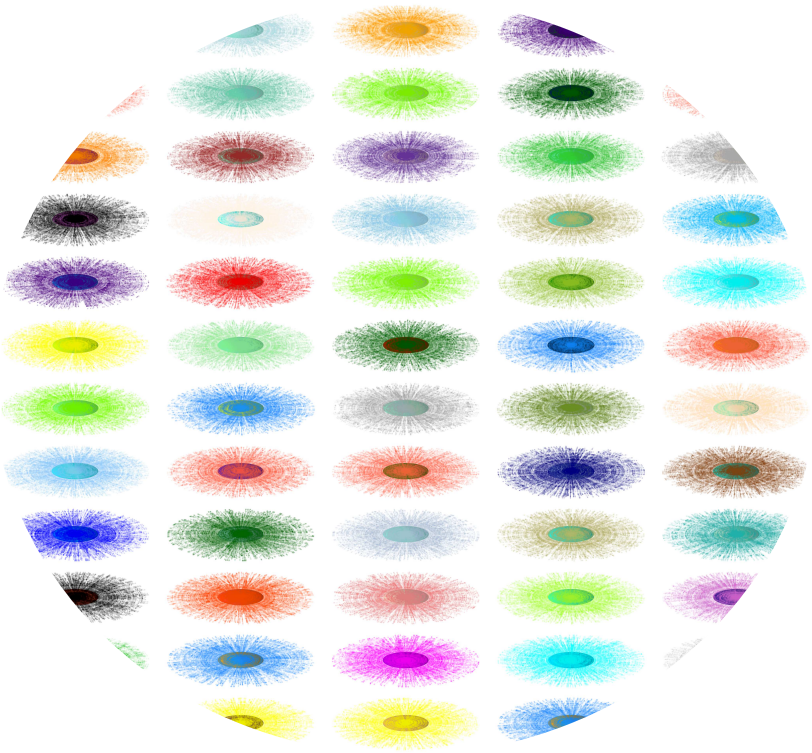

**Figure 4.** A colored depiction of a variety of synthetic data sets, including experimental data (atomic nuclei of hydrogens, Figure 3), synthetic data (atomic electron clouds of hydrogens, Figure 3) and uncharted territories, i.e., the ocean, the white region.

**Application of GIBAC in drug discovery & design**

In recent years, the application of AI (including DL and ML) is becoming increasingly popular in drug discovery & design [2,95,113,190], particularly in lead optimization and ADMET studies, including carcinogenicity, hepatotoxicity, et cetera [113]. For instance, VenomPred is a promising solution for deriving structural toxicophores and assessing the safety profile of compounds.

Here, given the definition of GIBAC as in Equation 4 and Table 1, the discussion of the practical application of GIBAC in drug discovery & design focuses on intermolecular binding and interactions. In biological systems, there are a wide range of intermolecular binding pairs, including including enzyme-substrate [197,198], ligand-receptor [199,200], protein-protein [55,201], ion channel-drug [202, 203], antibody-antigen [18,204,205], DNA-protein [206], RNA-protein [191,207], RNA-RNA [207], hormone-receptor [208], coenzyme-substrate [209], metal ion-protein [134,210], lipid-protein [211], et cetera. By definition, GIBAC can find its use for any binding pair involved in the molecular pathogenesis of human diseases, infectious or non-communicable, including:

1. ligand-receptor binding, e.g., insulin binding to its receptor [100];
2. protein-protein interaction, e.g., TNF-$\alpha$ binding to its receptor;
3. ion channel-drug interaction, e.g., verapamil binding to $Ca_V1.2$ [134];
4. antibody-antigen binding, e.g., Keytruda binding to PD-1 [101];
5. self-association and aggregation, e.g., formation of amyloid-$\beta$ oligomer [212,213].

Overall, GIBAC is useful for drug discovery [217], drug design [218], lead optimization [59], drug repurposing [219], and DDI prediction [195]. Take insulin for instance [100], for which a series of analogues have been designed, synthesized and therapeutically tested towards a better glycaemic control for diabetic patients [220,221]. After injection, insulin exists as multi-forms, including multi-hexamer, di-hexamer, hexamer, dimer or monomer [222], as shown in Figure 6. While insulin mainly binds to its own specific receptor, the insulin receptor (IR, Figure 6), it also binds to IGF-1R [223–225] [197] for an increased mitogenic potential (e.g., continued growth of pre-existing

 doi:10.20944/preprints202208.0213.v2



neoplasms [226,227]), and also binds to IDE for enzyme-mediated degradation of insulin with cells. Moreover, accumulating evidence has led to a concept of the physiological roles of IR isoforms, where predominant IR-A expression may be important for prenatal growth and development, while IR-B expression has a more important role in metabolic insulin action in adults [228].

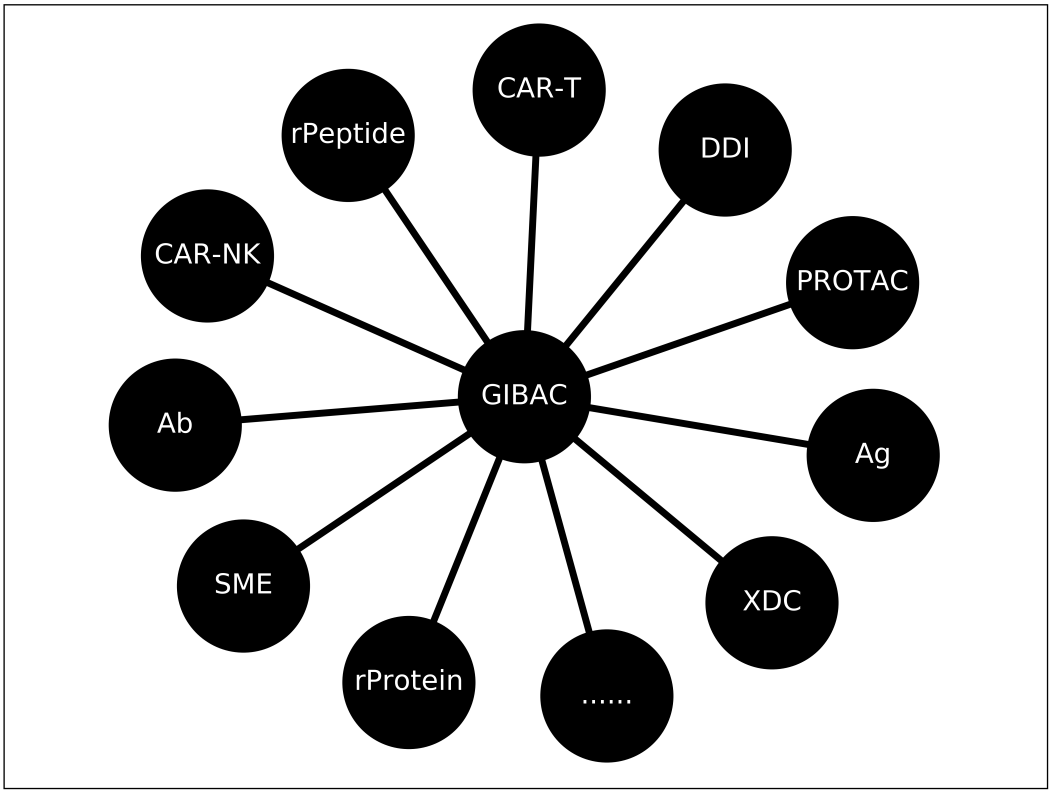


**Figure 5.** A variety of the practical application of GIBAC, including SME (small molecule inhibitor), Ab (antibody), Ag (antigen), XDC (antibody-drug conjugate (ADC), peptide-drug conjugate (PDC), aptamer-drug conjugate (ApDC) [191]), rPeptide (recombinant peptide drug), rProtein (recombinant protein drug), intrabodies [192], proteolysis-targeting chimeric molecules (PROTAC) [193,194], drug-drug interaction (DDI) [195], chimeric antigen receptor T (CAR-T) cell therapy [196].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



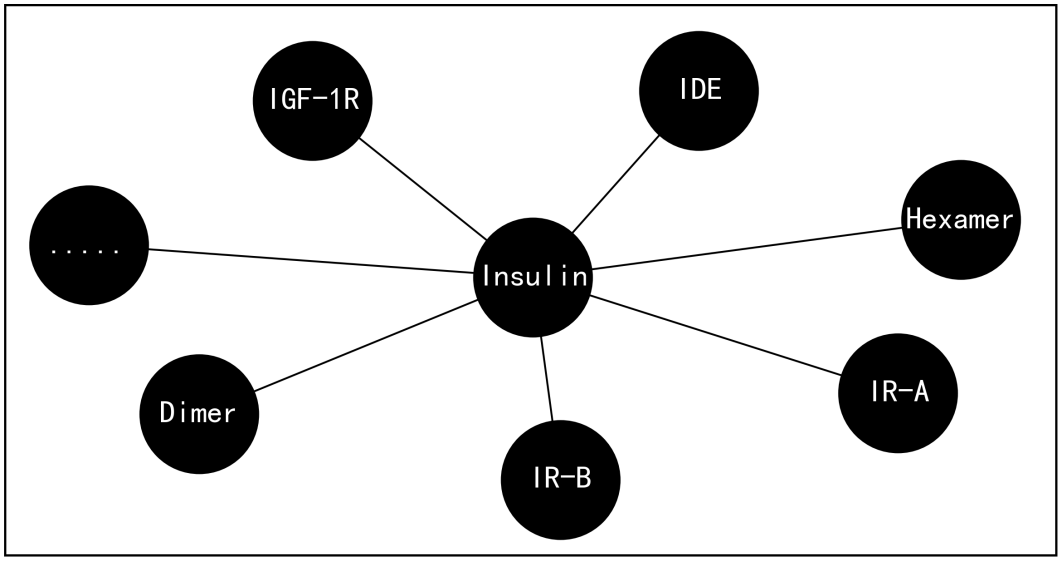


**Figure 6.** An molecular binding and interaction network of insulin. In this figure, IR-A, IR-B, IGF-1R and IDE represent two isoforms of insulin receptor [214], insulin-like factor 1 receptor [215,216] and insulin degrading enzyme (IDE) [197], respectively.

Thus, with GIBAC as a $K_d$-based search engine for therapeutic candidate(s), the problem of drug discovery & design here is to be a matter of plugging into the search engine the name(s) of insulin's binding partner(s) and their associated $K_d$ and $\Delta G$ value(s) or value range(s) specified by the user(s) of GIBAC. Afterwards, the search engine walks traversely through all molecular types and drug modalities, and returns a list of $K_d$-ranked molecular candidates, including but not limited to insulin analogues available to date, for continued improvement of glycaemic control of diabetic patients.

Take ADC for another example, for which internalization is crucial for delivering the cytotoxic drug into the cancer cells [130,229]. To achieve effective internalization and drug release, a high binding affinity between the antigen and antibody is desired. This affinity ensures strong and stable interaction between the ADC and its cancer cell receptor, increasing the likelihood of efficient internalization. Moreover, a high binding affinity enhances ADC retention within the cancer cells, maximizing its cellular exposure to the cytotoxic drug and boosting the therapeutic efficacy of ADC. Thus, GIBAC here is able to act as a search engine for antibodies with high (as high as biophysically possible) affinity and specificity to target antigens, facilitating efficient ADC internalization, minimizing off-target effects, and enhancing therapeutic efficacy [230].

**Technical challenges of GIBAC: openness is the key**

As mentioned above, an exhaustive exploration of the entire molecular space is practically impossible (Figure 2) [20,131], despite the abundance of structural and biophysical data and the tools for synthetic data generation.

Nonetheless, AI algorithms alone are insufficient to build a GIBAC with adequate accuracy and precision, such that it is able to find its use in drug discovery & design. Therefore, this article here proposes an open strategy (Figure 7), because

1. the AI-training and retraining processes require a **huge** amount of data with reasonable accuracy, variety and cost.
2. openness (Figure 7) in data, algorithms, source code, and AI models is essential for promoting transparency, reproducibility, and collaboration within the whole community of drug discovery & design, and facilitates the continued improvement of the performance of GIBAC.
3. the accuracy and the precision of GIBAC is inextricably linked to a variety of interlinked factors, which is to be discussed in the next section.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



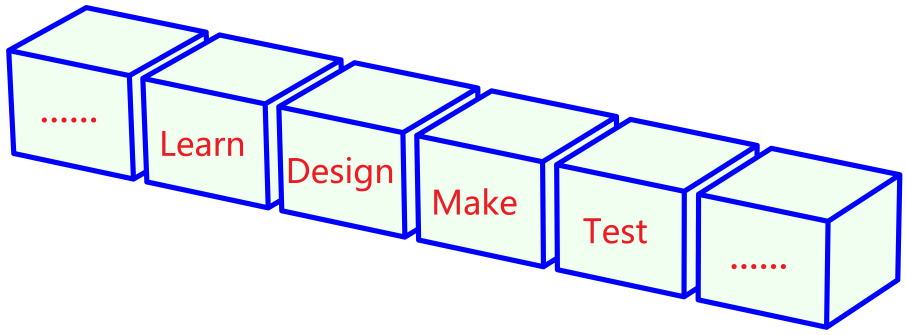


**Figure 7.** An iterative process to build and refine an AI- and physics-based intermolecular $K_d$ calculator (i.e., GIBAC) with adequate accuracy and precision.

**Five interlinked factors linked to the accuracy and the precision of GIBAC**

As two cornerstones for both experimental measurement and computational calculation, **accuracy** and **precision** help ensure the reliability and validity of the results obtained. As an AI- and physics-based intermolecular $K_d$ calculator, GIBAC is no exception here. Take the $K_d$ of antigen-antibody binding for example, where a study of anti-4-1BB monoclonal antibody found that the optimal antibody binding affinity was found to be dependent on its intended application. For blocking antibodies, a high affinity is most effective. However, for agonist antibodies, the results show that an intermediate affinity works best [103,104].

To further highlight the importance of accuracy and precision, this article looked into history and back with experimental measurements of three physical constants: the Planck constant, the Boltzmann constant and the gyromagnetic ratio. Take the Planck constant for the first example, which provides the foundation of quantum physics, and was experimentally measured to be $6.626\ 069\ 57 \times 10^{-34}$ Joule seconds [231]. As the second example, the Boltzmann constant was experimentally measured to be $1.380\ 649 \times 10^{-23}$ Joule $Kelvin^{-1}$ [232]. Lastly, the gyromagnetic ratio of the proton (hydrogen nucleus) in water ($\gamma$) is measured to be $2.67513 \times 10^{8}$ $Weber^{-1}$ $m^2$ $sec^{-1}$ with a total probable error of 5 in $10^6$ [233,234]. Taken together, these three examples highlight the importance of accuracy and precision in both experimental measurement and computational prediction. For drug discovery & design in particular, intermolecular $K_d$ calculations need to be both accurate (close to its true value) and precise (reproducible & consistent, with a total probable error similar to that of $\gamma$ ideally), such that informed decisions could be made during the drug discovery process, to ensure **accurate** and **precise** selection of the most promising drug candidates in the earliest stage of R&D.

Interestingly, the key factors related to GIBAC's accuracy and precision (Figure 8) are interlinked to one another. Starting from experimental structure determination, where experimental techniques such as X-ray crystallography, NMR spectroscopy, cryo-electron microscopy (cryo-EM), or Cryo-electron tomography (cryo-ET) [235,236] are used to record direct experimental data, e.g., X-ray diffraction pattern, chemical shift, cryo-EM images, or cryo-ET tomographies. Afterwards, structure calculation methods are used to refine and interpret the experimental data, which involve mathematical algorithms and computational structural modeling to provides the initial atomic coordinates, with which the forcefield Figure 8 describes the interactions between atoms. Subsequently, energy minimization optimizes the atomic positions to find the lowest energy configuration, and to generate a accurate and complete representation of the (bio)molecule's structure.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



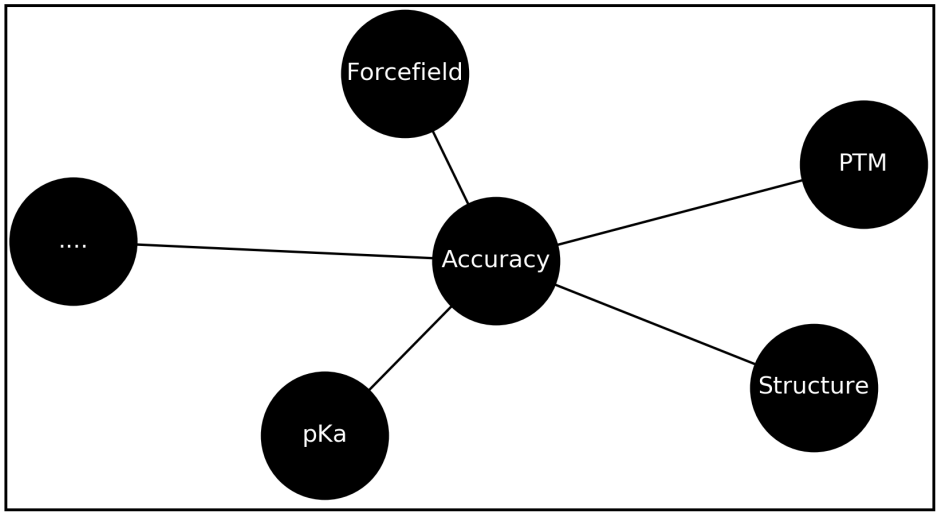


**Figure 8.** Key factors related to the accuracy and the precision of GIBAC.

Specifically, these factors (forcefield, structural information, PTM, PEM, pKa, Figure 8) are interlinked, in the sense that

1. both accurate pKa values and forcefield is required for the energy minimization step in structural calculation for experimental structure determination.
2. for the description of a molecular structure system, a range of factors are required, including forcefield, atomic coordinates, structural information, PTM, PEM, pKa.
3. physics-based pKa calculation requires accurate structural information of protein, PTM, PEM, protein with PEM, or protein with PEM [121,122].
4. accurate structural information and pKa values are useful for continued improvement of the accuracy of forcefield.
5. not only does biomolecules (e.g., protein) have site-specific pKa values, but PTM- or PEM-related chemical groups/moieties (e.g., glycans and lipids, Figure 9) also have site-specific pKa values. For instance, DMPG's and DMPC's head groups have intrinsic pKa values of $\sim 3.5$ and $\sim 1.0$, respectively [70,237].

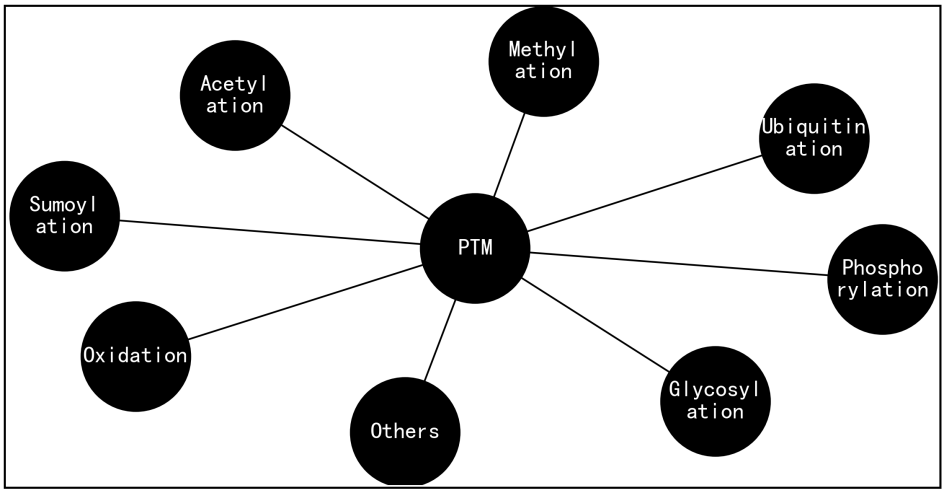


**Figure 9.** A collection of different post-translational modifications (PTMs).

**Towards a truly general site-specific pKa calculator**

To accurately and precisely calculate intermolecular $K_d$, a series of factors need to be considered, including forcefield, atomic coordinates, structural information, PTM, PEM, pKa (Figure 8). Below,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



site-specific pKa is taken as an example to explain why and how it is linked to the accuracy and the precision of GIBAC, and how to tackle this pKa issue.

As is well known, electrostatics (salt bridges, hydrogen bonds, charge-charge attraction or repulsion, et cetera) plays an important role in biomolecular structure and function [67,68,238], in which a specific set of amino acid residues are of particular relevance due to the variable protonation states of their ionizable side chains [66,71]. In an acid-base equilibrium, the dissociation constant is usually written as $K_a = \frac{[A][H]}{[HA]}$, where $[H]$, $[A]$ and $[HA]$ represent the concentrations of protons, the unprotonated and the protonated forms of a titrateable group, respectively. For the titrateable group, $K_a$ is a measure of its acidity, the higher the $K_a$, the higher the acidity. $K_a$ can also be expressed as pKa with the equation $pK_a = -log_{10}K_a$, similar to the way pH is defined [69].

To address the pKa issue, this article here proposes a truly general site-specific protonation constant (pKa) calculator (GSPCC, Equation 11), similar to the way GIBAC is defined in Equation 4:

$$pKa = f(molecules, envPara) \tag{11}$$

Of note, a protonation-deprotonation equilibrium is indeed a ligand-binding reaction with the ligand being a proton ($H^+$). As shown in Equation 11, therefore, GSPCC is essentially a specific form of GIBAC, where proton (a sub-atomic particle) and the ionizable chains constitute the two binding partners. Furthermore, with pH and pKa, protonation state/proton occupancy ($\theta$) can be defined as in Equation 12, to describe the degree to which the entire population of the ionizable side chains are protonated [69]. That is, when $pH = pKa$, the ionizable side chain is half protonated and half deprotonated, i.e., $\theta = 50\%$, assuming that the site-specific protonation/deprotonation follows the classic Henderson-Hasselbalch equation.

$$\theta = \frac{10^{(pK_a - pH)}}{1 + 10^{(pK_a - pH)}} = \frac{10^{pK_a}}{10^{pK_a} + 10^{pH}} \tag{12}$$

**Table 2.** A tabular description of GSPCC (Equation 11).

| Input 1 | Input 2 | Output |
|---|---|---|
| $molAstring, molBstring, ...$ | $envPara$ | pKa |
| $molAgraph, molBgraph, ...$ | $envPara$ | pKa |

Since GSPCC is essentially a specific form of GIBAC, the construction of the two follow the same roadmap as defined by Equation 9 [1]. To date, pKa can be measured by various experimental methods, including UV-spectroscopy [239], a pH-metric approach [240], a capillary electrophoresis-based approach [241] and NMR spectroscopy [69,242–244] In addition to experimental measurements of pKa, computational tools have also been developed, such as protein pKa prediction with ML [245,246] and physics-based pKa calculator, which is able to act as synthetic pKa data generators such as PROPKA [76].

In short, given the importance of electrostatics in intermolecular binding affinity [12,26], it is necessary to first build a GSPCC with adequate accuracy and precision to build a GIBAC with adequate accuracy and precision, to eusure its applicability in drug discovery & design.

### Technical limitations of GIBAC: $K_d$, $K_{on}$, $K_{off}$, ...

While drug-target $K_d$ is an essential parameter for drug discovery & design, it is but one of the many aspects of drug R&D. For instance, $K_d$ and $\Delta G$ has been used to indicate the efficacy of a drug. However, this is not always the case. It has recently been shown that residence time ($RT$) is another better indicator of efficacy than $K_d$ for some systems [247,248].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



In biophysics, the relationship between $K_d$ (dissociation constant), $K_{on}$ (association rate constant), and $K_{off}$ (dissociation rate constant) can be described as $K_d = K_{off}/K_{on}$, while $RT$ refers to the average time a molecule spends bound to its target before dissociation. Thus, The $K_{on}$ represents the rate at which a molecule associates with its target, while the $K_{off}$ represents the rate at which the molecule dissociates from the target.

As a result, while GIBAC is defined as in Equation 4, additional output parameters are necessary for the construction of a general intermolecular biophysics calculator (GIBC) based on the hybrid $AI + physics$ approach, as outlined in Table 3. To this end, a set of biophysical parameters (in addition to the intermolecular $K_d$) allows a further generalization of the GIBAC originally proposed in [1] and defined in Equation 4 and Table 1, leading to the concept of GIBC (Table 3), which is to be one biophysics-based future direction of GIBAC.

**Table 3.** A tabular description of a general intermolecular biophysics calculator.

| Input 1 | Input 2 | Output |
|---|---|---|
| $molAstring, molBstring, ...$ | $envPara$ | $K_d$, $K_{on}$, $K_{off}$, RT |
| $molAgraph, molBgraph, ...$ | $envPara$ | $K_d$, $K_{on}$, $K_{off}$, RT |

**Two future directions of GIBAC: AI + biophysics**

This article puts forward an physics+AI hybrid approach, which is necessary for the construction of a GIBAC with adequate accuracy and precision to be used in drug discovery & design, and which aims at increasing the probability of successfully discovering new drugs, while reducing discovery costs and timelines.

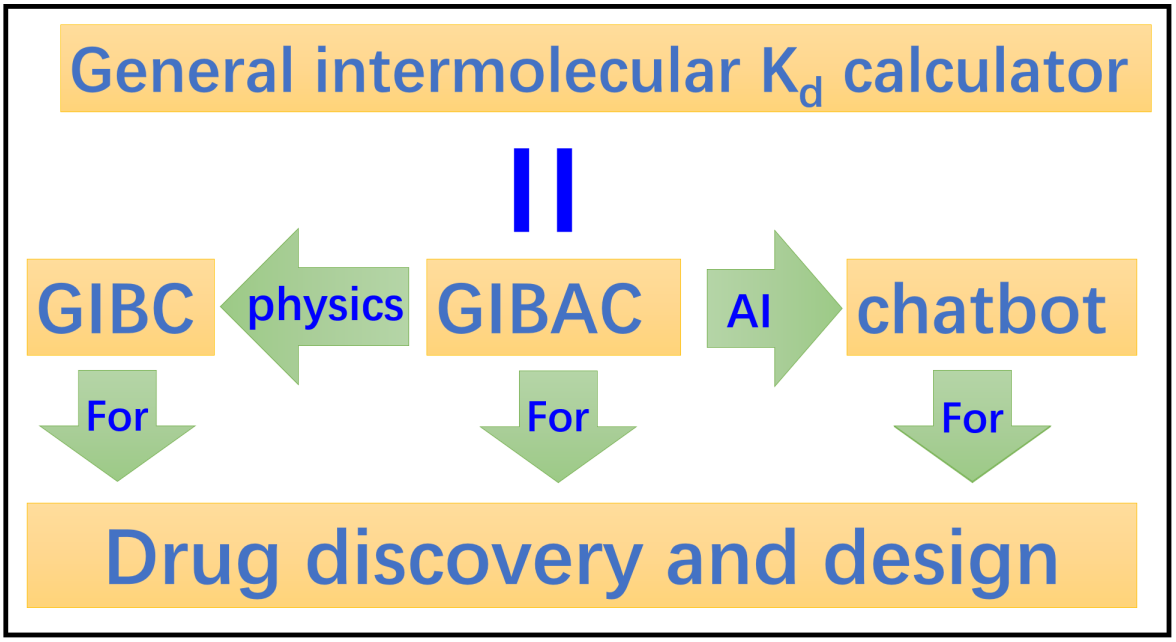


**Figure 10.** Two future directions of GIBAC, i.e., a truly general intermolecular biophysics calculator (GIBC) and a chatbot for drug discovery & design.

Therefore, in addition to an AI- and physics-based GIBC as defined in Tables 3 and 4, this article further discusses the potential of the intermolecular $K_d$ calculator-based search engine (i.e., GIBAC) to act as a ChatGPT-like chatbot for drug discovery & design, which is able to accurately, precisely and efficiently handle questions as below:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



1. for mini GIBAC (Figure 4), can you please generate a $K_d$-ranked list of insulin analogues which binds to IR with a $K_d$ within a desired value range?
2. for mini GIBAC (Figure 4), can you please generate a list of insulin analogues which does not bind to IGF-1R or IDE?
3. for mini GIBAC (Figure 4), can you please generate a list of insulin analogues which does not form dimer or hexamer?
4. for mini GIBAC (Figure 4), can you please generate a $K_d$-ranked list of insulin analogues, which form highly stable dimer or hexamer?
5. for GIBAC (Figure 4), can you generate a $K_d$-ranked list of therapeutic candidates which targets X (i.e., drug target) of different species, e.g., X of human [249], X of cat/dog [250], X of horse [251], et cetera?
6. for GIBAC (Figure 4), can you generate a $K_d$-ranked list of therapeutic candidates which targets X, Y and Z? such as vorolanib [252] or retatrutide [253]?
7. for GIBAC (Figure 4), can you please generate a $K_d$-ranked list of therapeutic small molecule candidates [254,255] which targets $X$ and possesses a $K_d$ within a specified $K_d$ range, and which does not target $Y$ or $Z$?
8. for GIBAC (Figure 4), can you please generate a list of prefusion-stabilizing molecules or introduce a set of mutations into the S protein to stabilize its conformation in the prefusion state? as locking the conformation of the S protein of SARS-CoV-2 into the premembrane-fusion state is essential for subunit vaccine design [26,256,257].

In light of the technical limitations of GIBAC, it is also to be used in combination with other parameters (e.g., LogD [258] or synthesizability [259], et cetera) in drug discovery & design in future. Of one note, GIBAC in itself is no more than a tool, and biology and biophysics are the science underlying all processes in living organisms. As a result, in case GIBAC is actually to be used in future as a chatbot for drug discovery & design, its user may want to follow a science first, technology second principle, and first try to understand the physiology and the pathophysiology for target(s)-specific drug discovery & design.

**Table 4.** A tabular description of a truly general intermolecular biophysics calculator.

| **Input 1** | **Input 2 (*envPara*, et cetera)** | **Output** |
|---|---|---|
| $molAstring$, $molBstring$, ... | pH, pKa, $T$, ionic strength, ... | $\Delta G$, $K_d$, $K_{on}$, $K_{off}$, RT |
| $molAgraph$, $molBgraph$, ... | pH, pKa, $T$, ionic strength, ... | $\Delta G$, $K_d$, $K_{on}$, $K_{off}$, RT |

## Conclusion and discussion

To sum up, this article puts forward a conceptual and practical framework (Figure 11) of a truly general intermolecular binding affinity calculator, including its inception, definition, construction, practical applications, technical challenges and limitations, and future directions (GIBAC [1]).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202208.0213.v2



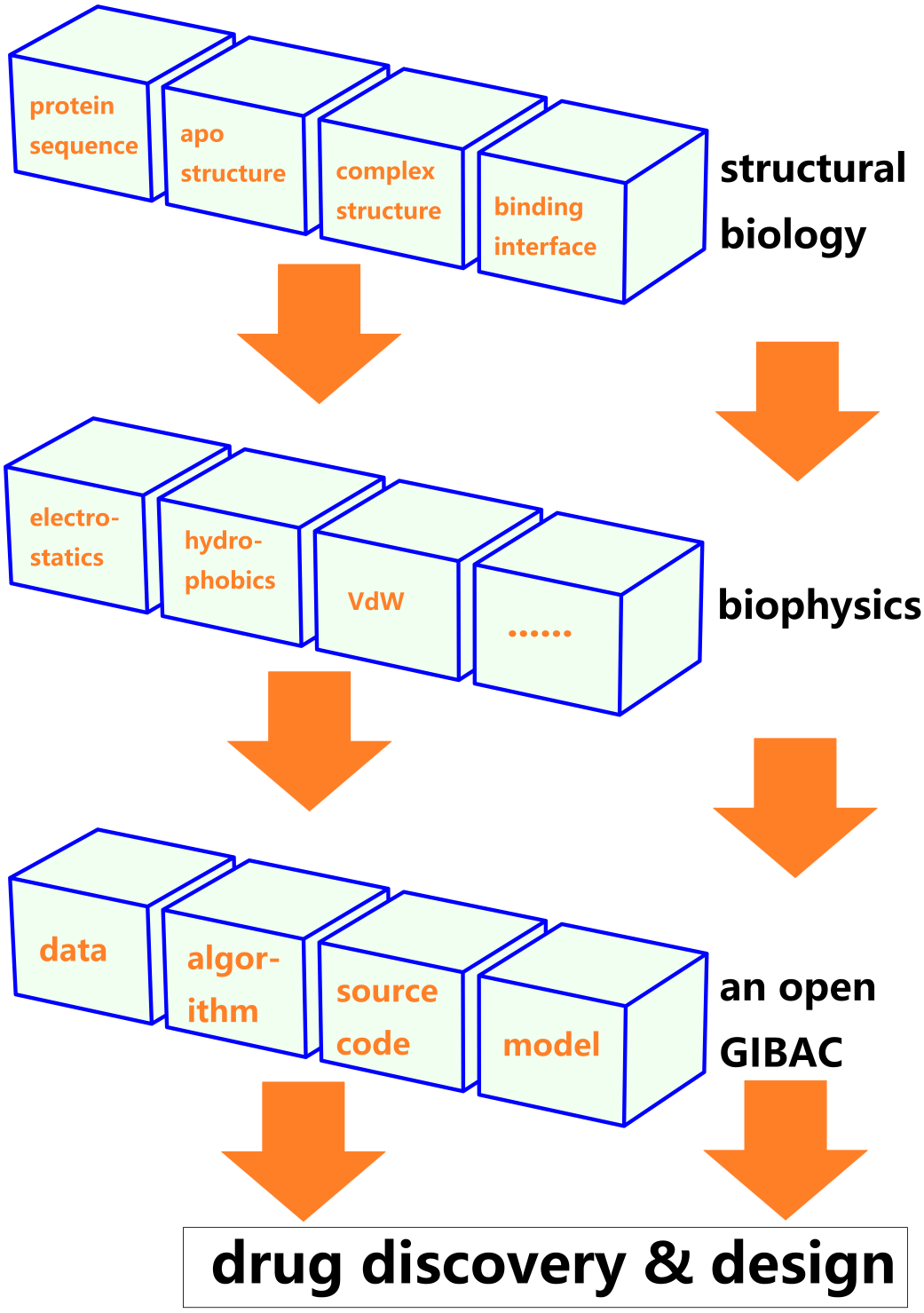


**Figure 11.** GIBAC as a $K_d$-based search engine for drug discovery & design.

As is known throughout the industry, drug R&D is a multi-purpose expensive and arduous task [260–262]. It is also a highly regulated process that can take years and many millions to billions of dollars just to get rejected [41]. As of October 19, 2023, the average time to develop a drug takes between 10 years to 15 years, which includes discovery, lead identification and optimization, preclinical testing, clinical trials, regulatory review, and final regulatory approval.

As the beginning part of the drug R&D, drug discovery & design itself is still a lengthy, costly, difficult, and inefficient yet pivotal process [263]. Given this, only a GIBAC with adequate accuracy,

 doi:10.20944/preprints202208.0213.v2



precision, and efficiency is able to create a paradigm shift [264] from structure-based molecular generation to a search engine for drug discovery & design and find its practical application in the pharmaceutical industry [139,265], aiming at increasing the probability of successfully discovering new drugs, while reducing discovery costs and timelines.

Finally, this article argues that the time is now ripe for the construction of such a GIBAC to be listed on the agenda of the drug R&D community, including in particular structural biologists and biophysicists, medicinal and computational chemists, drug discoverers & designers, and algorithm designers, in light of

1. the crucial roles of $K_d$ and $\Delta G$ in drug discovery & design in continued optimization of drug-target interactions to facilitate the development of therapeutic agents with improved efficacy and safety [108–110].
2. the recent progresses of AI algorithms in drug discovery & design [266,267].
3. a large amount of data [268,269], and a variety of tools [112,175,243,270] in structural biology, bioinformatics, biophysics, drug discovery & design, et cetera.
4. the democratization of high performance computing (HPC) since the beginning of this century and its continued evolution towards scalable quantum computing [271] and perhaps computation beyond silicon [272] in future.
5. the framework (Figure 11) for GIBAC [1] as described here.

**Author Contributions:** Conceptualization, W.L.; methodology, W.L.; software, W.L.; validation, W.L.; formal analysis, W.L.; investigation, W.L.; resources, W.L.; data duration, W.L.; writing–original draft preparation, W.L.; writing–review and editing, W.L.; visualization, W.L.; supervision, W.L.; project administration, W.L.; funding acquisition, not applicable.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** No ethical approval is required.

**Acknowledgments:** The author is grateful to the communities of structural biology, biophysics, medicinal and computational chemistry and algorithm design, for the continued accumulation of knowledge and data for drug discovery & design, and for the continued development of tools (both hardware, software and algorithm) for drug R&D, which contributed immensely to the inception of GIBAC, and which are also to contribute to the construction of an accurate, precise and efficient GIBAC for drug discovery & design.

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

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used OpenAI's ChatGPT in order to improve the readability of the manuscript, and to make it as concise and short as possible. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

## References

1. Li, W. Towards a General Intermolecular Binding Affinity Calculator **2022**.
2. Cheng, J.; Novati, G.; Pan, J.; Bycroft, C.; Žemgulytė, A.; Applebaum, T.; Pritzel, A.; Wong, L.H.; Zielinski, M.; Sargeant, T.; Schneider, R.G.; Senior, A.W.; Jumper, J.; Hassabis, D.; Kohli, P.; Avsec, Ž. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **2023**.
3. Ellegren, H.; Galtier, N. Determinants of genetic diversity. *Nature Reviews Genetics* **2016**, *17*, 422–433.
4. Lynch, M.; Ackerman, M.S.; Gout, J.F.; Long, H.; Sung, W.; Thomas, W.K.; Foster, P.L. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics* **2016**, *17*, 704–714.
5. Wong, F.; de la Fuente-Nunez, C.; Collins, J.J. Leveraging artificial intelligence in the fight against infectious diseases. *Science* **2023**, *381*, 164–170.
6. Murray, C.J.L.; Ikuta, K.S.; Sharara, F.; Swetschinski, L.; Aguilar, G.R.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; Johnson, S.C.; Browne, A.J.; Chipeta, M.G.; Fell, F.; Hackett, S.; Haines-Woodhouse, G.; Hamadani, B.H.K.; Kumaran, E.A.P.; McManigal, B.; Achalapong, S.; Agarwal, R.; Akech, S.; Albertson, S.; Amuasi,

J.; Andrews, J.; Aravkin, A.; Ashley, E.; Babin, F.X.; Bailey, F.; Baker, S.; Basnyat, B.; Bekker, A.; Bender, R.; Berkley, J.A.; Bethou, A.; Bielicki, J.; Boonkasidecha, S.; Bukosia, J.; Carvalheiro, C.; Castañeda-Orjuela, C.; Chansamouth, V.; Chaurasia, S.; Chiurchiù, S.; Chowdhury, F.; Donatien, R.C.; Cook, A.J.; Cooper, B.; Cressey, T.R.; Criollo-Mora, E.; Cunningham, M.; Darboe, S.; Day, N.P.J.; Luca, M.D.; Dokova, K.; Dramowski, A.; Dunachie, S.J.; Bich, T.D.; Eckmanns, T.; Eibach, D.; Emami, A.; Feasey, N.; Fisher-Pearson, N.; Forrest, K.; Garcia, C.; Garrett, D.; Gastmeier, P.; Giref, A.Z.; Greer, R.C.; Gupta, V.; Haller, S.; Haselbeck, A.; Hay, S.I.; Holm, M.; Hopkins, S.; Hsia, Y.; Iregbu, K.C.; Jacobs, J.; Jarovsky, D.; Javanmardi, F.; Jenney, A.W.J.; Khorana, M.; Khusuwan, S.; Kissoon, N.; Kobeissi, E.; Kostyanev, T.; Krapp, F.; Krumkamp, R.; Kumar, A.; Kyu, H.H.; Lim, C.; Lim, K.; Limmathurotsakul, D.; Loftus, M.J.; Lunn, M.; Ma, J.; Manoharan, A.; Marks, F.; May, J.; Mayxay, M.; Mturi, N.; Munera-Huertas, T.; Musicha, P.; Musila, L.A.; Mussi-Pinhata, M.M.; Naidu, R.N.; Nakamura, T.; Nanavati, R.; Nangia, S.; Newton, P.; Ngoun, C.; Novotney, A.; Nwakanma, D.; Obiero, C.W.; Ochoa, T.J.; Olivas-Martinez, A.; Olliaro, P.; Ooko, E.; Ortiz-Brizuela, E.; Ounchanum, P.; Pak, G.D.; Paredes, J.L.; Peleg, A.Y.; Perrone, C.; Phe, T.; Phommasone, K.; Plakkal, N.; de Leon, A.P.; Raad, M.; Ramdin, T.; Rattanavong, S.; Riddell, A.; Roberts, T.; Robotham, J.V.; Roca, A.; Rosenthal, V.D.; Rudd, K.E.; Russell, N.; Sader, H.S.; Saengchan, W.; Schnall, J.; Scott, J.A.G.; Seekaew, S.; Sharland, M.; Shivamallappa, M.; Sifuentes-Osornio, J.; Simpson, A.J.; Steenkeste, N.; Stewardson, A.J.; Stoeva, T.; Tasak, N.; Thaiprakong, A.; Thwaites, G.; Tigoi, C.; Turner, C.; Turner, P.; van Doorn, H.R.; Velaphi, S.; Vongpradith, A.; Vongsouvath, M.; Vu, H.; Walsh, T.; Walson, J.L.; Waner, S.; Wangrangsimakul, T.; Wannapinij, P.; Wozniak, T.; Sharma, T.E.M.W.Y.; Yu, K.C.; Zheng, P.; Sartorius, B.; Lopez, A.D.; Stergachis, A.; Moore, C.; Dolecek, C.; Naghavi, M. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **2022**, *399*, 629–655.

7. Fonseca-Montaño, M.A.; Blancas, S.; Herrera-Montalvo, L.A.; Hidalgo-Miranda, A. Cancer Genomics. *Archives of Medical Research* **2022**, *53*, 723–731.
8. Blaich, A.; Pahlavan, S.; Tian, Q.; Oberhofer, M.; Poomvanicha, M.; Lenhardt, P.; Domes, K.; Wegener, J.W.; Moosmang, S.; Ruppenthal, S.; Scholz, A.; Lipp, P.; Hofmann, F. Mutation of the Calmodulin Binding Motif IQ of the L-type $Ca_V1.2$ $Ca^{2+}$Channel to EQ Induces Dilated Cardiomyopathy and Death. *Journal of Biological Chemistry* **2012**, *287*, 22616–22625.
9. Seisenberger, C.; Specht, V.; Welling, A.; Platzer, J.; Pfeifer, A.; Kühbandner, S.; Striessnig, J.; Klugbauer, N.; Feil, R.; Hofmann, F. Functional Embryonic Cardiomyocytes after Disruption of the L-type $\alpha_{1c}$ ($Ca_V1.2$) Calcium Channel Gene in the Mouse. *Journal of Biological Chemistry* **2000**, *275*, 39193–39199.
10. Hua, Y.; Sahashi, K.; Rigo, F.; Hung, G.; Horev, G.; Bennett, C.F.; Krainer, A.R. Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature* **2011**, *478*, 123–126.
11. Hua, Y.; Liu, Y.H.; Sahashi, K.; Rigo, F.; Bennett, C.F.; Krainer, A.R. Motor neuron cell-nonautonomous rescue of spinal muscular atrophy phenotypes in mild and severe transgenic mouse models. *Genes & Development* **2015**, *29*, 288–297.
12. Li, W. How do SMA-linked mutations of *SMN1* lead to structural/functional deficiency of the SMA protein? *PLOS ONE* **2017**, *12*, e0178519.
13. Hu, B.; Guo, H.; Zhou, P.; Shi, Z.L. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* **2020**, *19*, 141–154.
14. Li, W. Structurally Observed Electrostatic Features of the COVID-19 Coronavirus-Related Experimental Structures inside Protein Data Bank: A Brief Update **2020**.
15. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* **2020**.
16. Lippi, G.; Plebani, M. Laboratory abnormalities in patients with COVID-2019 infection. *Clinical Chemistry and Laboratory Medicine (CCLM)* **2020**, *0*.
17. Chen, K.W.K.; Huang, D.T.N.; Huang, L.M. SARS-CoV-2 variants: Evolution, spike protein, and vaccines. *Biomedical Journal* **2022**, *45*, 573–579.
18. 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**.
19. Bojadzic, D.; Alcazar, O.; Chen, J.; Chuang, S.T.; Capcha, J.M.C.; Shehadeh, L.A.; Buchwald, P. Small-Molecule Inhibitors of the Coronavirus Spike: ACE2 Protein-Protein Interaction as Blockers of Viral Attachment and Entry for SARS-CoV-2. *ACS Infectious Diseases* **2021**, *7*, 1519–1534.

20. Chuang, S.T.; Buchwald, P. Broad-Spectrum Small-Molecule Inhibitors of the SARS-CoV-2 Spike—ACE2 Protein-Protein Interaction from a Chemical Space of Privileged Protein Binders. *Pharmaceuticals* **2022**, *15*, 1084.
21. Singh, D.; Yi, S.V. On the origin and evolution of SARS-CoV-2. *Experimental & Molecular Medicine* **2021**, *53*, 537–547.
22. Liu, L.; Fang, Q.; Deng, F.; Wang, H.; Yi, C.E.; Ba, L.; Yu, W.; Lin, R.D.; Li, T.; Hu, Z.; Ho, D.D.; Zhang, L.; Chen, Z. Natural Mutations in the Receptor Binding Domain of Spike Glycoprotein Determine the Reactivity of Cross Neutralization between Palm Civet Coronavirus and Severe Acute Respiratory Syndrome Coronavirus. *Journal of Virology* **2007**, *81*, 4694–4700.
23. Tao, K.; Tzou, P.L.; Nouhin, J.; Gupta, R.K.; de Oliveira, T.; Pond, S.L.K.; Fera, D.; Shafer, R.W. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nature Reviews Genetics* **2021**, *22*, 757–773.
24. Li, W. Structural and Functional Consequences of the SMA-Linked Missense Mutations of the Survival Motor Neuron Protein: A Brief Update. In *Novel Aspects on Motor Neuron Disease*; IntechOpen, 2019.
25. Li, W. Designing rt-PA Analogs to Release its Trapped Thrombolytic Activity. *Journal of Computational Biophysics and Chemistry* **2021**, *20*, 719–727.
26. 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.
27. 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.
28. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data: from vision to reality. *Eurosurveillance* **2017**, *22*.
29. Abdolmaleki, A.; Shiri, F.; Ghasemi, J.B. Computational Multi-Target Drug Design. In *Methods in Pharmacology and Toxicology*; Springer New York, 2018; pp. 51–90.
30. Burki, T. First shared SARS-CoV-2 genome: GISAID vs virological.org. *The Lancet Microbe* **2023**, *4*, e395.
31. Brito, A.F.; Semenova, E.; Dudas, G.; Hassler, G.W.; Kalinich, C.C.; Kraemer, M.U.G.; Ho, J.; Tegally, H.; Githinji, G.; Agoti, C.N.; Matkin, L.E.; Whittaker, C.; Kantardjiev, T.; Korsun, N.; Stoitsova, S.; Dimitrova, R.; Trifonova, I.; Dobrinov, V.; Grigorova, L.; Stoykov, I.; Grigorova, I.; Gancheva, A.; Jennison, A.; Leong, L.; Speers, D.; Baird, R.; Cooley, L.; Kennedy, K.; de Ligt, J.; Rawlinson, W.; van Hal, S.; Williamson, D.; Singh, R.; Nathaniel-Girdharrie, S.; Edghill, L.; Indar, L.; John, J.S.; Gonzalez-Escobar, G.; Ramkisoon, V.; Brown-Jordan, A.; Ramjag, A.; Mohammed, N.; Foster, J.E.; Potter, I.; Greenaway-Duberry, S.; George, K.; Belmar-George, S.; Lee, J.; Bisasor-McKenzie, J.; Astwood, N.; Sealey-Thomas, R.; Laws, H.; Singh, N.; Oyinloye, A.; McMillan, P.; Hinds, A.; Nandram, N.; Parasram, R.; Khan-Mohammed, Z.; Charles, S.; Andrewin, A.; Johnson, D.; Keizer-Beache, S.; Oura, C.; Pybus, O.G.; Faria, N.R.; Stegger, M.; Albertsen, M.; Fomsgaard, A.; Rasmussen, M.; Khouri, R.; Naveca, F.; Graf, T.; Miyajima, F.; Wallau, G.; Motta, F.; Khare, S.; Freitas, L.; Schiavina, C.; Bach, G.; Schultz, M.B.; Chew, Y.H.; Makheja, M.; Born, P.; Calegario, G.; Romano, S.; Finello, J.; Diallo, A.; Lee, R.T.C.; Xu, Y.N.; Yeo, W.; Tiruvayipati, S.; Yadahalli, S.; Wilkinson, E.; Iranzadeh, A.; Giandhari, J.; Doolabh, D.; Pillay, S.; Rhal, U.; San, J.E.; Msomi, N.; Mlisana, K.; von Gottberg, A.; Walaza, S.; Ismail, A.; Mohale, T.; Engelbrecht, S.; Zyl, G.V.; Preiser, W.; Sigal, A.; Hardie, D.; Marais, G.; Hsiao, M.; Korsman, S.; Davies, M.A.; Tyers, L.; Mudau, I.; York, D.; Maslo, C.; Goedhals, D.; Abrahams, S.; Laguda-Akingba, O.; Alisoltani-Dehkordi, A.; Godzik, A.; Wibmer, C.K.; Martin, D.; Lessells, R.J.; Bhiman, J.N.; Williamson, C.; de Oliveira, T.; Chen, C.; Nadeau, S.; du Plessis, L.; Beckmann, C.; Redondo, M.; Kobel, O.; Noppen, C.; Seidel, S.; de Souza, N.S.; Beerenwinkel, N.; Topolsky, I.; Jablonski, P.; Fuhrmann, L.; Dreifuss, D.; Jahn, K.; Ferreira, P.; Posada-Cespedes, S.; Beisel, C.; Denes, R.; Feldk, M.; Nissen, I.; Santacroce, N.; Burcklen, E.; Aquino, C.; de Gouvea, A.C.; Moccia, M.D.; Grüter, S.; Sykes, T.; Opitz, L.; White, G.; Neff, L.; Popovic, D.; Patrignani, A.; Tracy, J.; Schlapbach, R.; Dermitzakis, E.; Harshman, K.; Xenarios, I.; Pegeot, H.; Cerutti, L.; Penet, D.; Stadler, T.; Howden, B.P.; Sintchenko, V.; Zuckerman, N.S.; Mor, O.; Blankenship, H.M.; de Oliveira, T.; Lin, R.T.P.; Siqueira, M.M.; Resende, P.C.; Vasconcelos, A.T.R.; Spilki, F.R.; Aguiar, R.S.; Alexiev, I.; Ivanov, I.N.; Philipova, I.; Carrington, C.V.F.; Sahadeo, N.S.D.; Branda, B.; Gurry, C.; Maurer-Stroh, S.; Naidoo, D.; von Eije, K.J.; Perkins, M.D.; van Kerkhove, M.; Hill, S.C.; Sabino, E.C.; Pybus, O.G.; Dye, C.; Bhatt, S.; Flaxman, S.; Suchard, M.A.; Grubaugh, N.D.; Baele, G.; Faria, N.R. Global disparities in SARS-CoV-2 genomic surveillance. *Nature Communications* **2022**, *13*.

32. Giovanetti, M.; Benedetti, F.; Campisi, G.; Ciccozzi, A.; Fabris, S.; Ceccarelli, G.; Tambone, V.; Caruso, A.; Angeletti, S.; Zella, D.; Ciccozzi, M. Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. *Biochemical and Biophysical Research Communications* **2021**, *538*, 88–91.
33. Li, Y.; Liang, Z.; Tian, Y.; Cai, W.; Weng, Z.; Chen, L.; Zhang, H.; Bao, Y.; Zheng, H.; Zeng, S.; Bei, C.; Li, Y. High-affinity PD-1 molecules deliver improved interaction with PD-L1 and PD-L2. *Cancer Science* **2018**, *109*, 2435–2445.
34. Sharma, P.; Hu-Lieskovan, S.; Wargo, J.A.; Ribas, A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* **2017**, *168*, 707–723.
35. Banerjee, V.; Oren, O.; Ben-Zeev, E.; Taube, R.; Engel, S.; Papo, N. A computational combinatorial approach identifies a protein inhibitor of superoxide dismutase 1 misfolding, aggregation, and cytotoxicity. *Journal of Biological Chemistry* **2017**, *292*, 15777–15788.
36. Yun, C.H.; Boggon, T.J.; Li, Y.; Woo, M.S.; Greulich, H.; Meyerson, M.; Eck, M.J. Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell* **2007**, *11*, 217–227.
37. Gettinger, S.; Choi, J.; Hastings, K.; Truini, A.; Datar, I.; Sowell, R.; Wurtz, A.; Dong, W.; Cai, G.; Melnick, M.A.; Du, V.Y.; Schlessinger, J.; Goldberg, S.B.; Chiang, A.; Sanmamed, M.F.; Melero, I.; Agorreta, J.; Montuenga, L.M.; Lifton, R.; Ferrone, S.; Kavathas, P.; Rimm, D.L.; Kaech, S.M.; Schalper, K.; Herbst, R.S.; Politi, K. Impaired HLA Class I Antigen Processing and Presentation as a Mechanism of Acquired Resistance to Immune Checkpoint Inhibitors in Lung Cancer. *Cancer Discovery* **2017**, *7*, 1420–1435.
38. Zaretsky, J.M.; Garcia-Diaz, A.; Shin, D.S.; Escuin-Ordinas, H.; Hugo, W.; Hu-Lieskovan, S.; Torrejon, D.Y.; Abril-Rodriguez, G.; Sandoval, S.; Barthly, L.; Saco, J.; Moreno, B.H.; Mezzadra, R.; Chmielowski, B.; Ruchalski, K.; Shintaku, I.P.; Sanchez, P.J.; Puig-Saus, C.; Cherry, G.; Seja, E.; Kong, X.; Pang, J.; Berent-Maoz, B.; Comin-Anduix, B.; Graeber, T.G.; Tumeh, P.C.; Schumacher, T.N.; Lo, R.S.; Ribas, A. Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *New England Journal of Medicine* **2016**, *375*, 819–829.
39. Li, W. Inter-Molecular Electrostatic Interactions Stabilizing the Structure of the PD-1/PD-L1 Axis: A Structural Evolutionary Perspective **2020**.
40. Beatty, G.L.; Gladney, W.L. Immune Escape Mechanisms as a Guide for Cancer Immunotherapy. *Clinical Cancer Research* **2014**, *21*, 687–692.
41. Wong, C.H.; Siah, K.W.; Lo, A.W. Estimation of clinical trial success rates and related parameters. *Biostatistics* **2018**, *20*, 273–286.
42. Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **2004**, *432*, 855–861.
43. Roggia, M.; Natale, B.; Amendola, G.; Maro, S.D.; Cosconati, S. Streamlining Large Chemical Library Docking with Artificial Intelligence: the PyRMD2Dock Approach. *Journal of Chemical Information and Modeling* **2023**.
44. Pharmaceuticals, R. Recursion Bridges the Protein and Chemical Space with Massive Protein-Ligand Interaction Predictions Spanning 36 Billion Compounds, 2023. Accessed: (September 1, 2023).
45. Crystallography: Protein Data Bank. *Nature New Biology* **1971**, *233*, 223–223.
46. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **2003**, *10*, 980–980.
47. Li, W. Half-a-century Burial of ρ, θ and φ in PDB **2021**.
48. Wang, T.; He, X.; Li, M.; Shao, B.; Liu, T.Y. AIMD-Chig: Exploring the conformational space of a 166-atom protein Chignolin with ab initio molecular dynamics. *Scientific Data* **2023**, *10*.
49. Shulman, R.; Wüthrich, K.; Yamane, T.; Patel, D.J.; Blumberg, W. Nuclear magnetic resonance determination of ligand-induced conformational changes in myoglobin. *Journal of Molecular Biology* **1970**, *53*, 143–157.
50. Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **2010**, *35*, 539–546.
51. Teague, S.J. Implications of protein flexibility for drug discovery. *Nature Reviews Drug Discovery* **2003**, *2*, 527–541.
52. Ruan, H.; Yu, C.; Niu, X.; Zhang, W.; Liu, H.; Chen, L.; Xiong, R.; Sun, Q.; Jin, C.; Liu, Y.; Lai, L. Computational strategy for intrinsically disordered protein ligand design leads to the discovery of p53 transactivation domain I binding compounds that activate the p53 pathway. *Chemical Science* **2021**, *12*, 3004–3016.
53. Wang, H.; Xiong, R.; Lai, L. Rational drug design targeting intrinsically disordered proteins. *WIREs Computational Molecular Science* **2023**.

54. Liu, W.; Chen, L.; Yin, D.; Yang, Z.; Feng, J.; Sun, Q.; Lai, L.; Guo, X. Visualizing single-molecule conformational transition and binding dynamics of intrinsically disordered proteins. *Nature Communications* **2023**, *14*.
55. Vangone, A.; Bonvin, A.M. Contacts-based prediction of binding affinity in protein-protein complexes. *eLife* **2015**, *4*.
56. 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.
57. 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.
58. 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.
59. 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.
60. 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.
61. 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.
62. Li, W. Visualising the Experimentally Uncharted Territories of Membrane Protein Structures inside Protein Data Bank **2020**.
63. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein complex prediction with AlphaFold-Multimer **2021**.
64. Cramer, P. AlphaFold2 and the future of structural biology. *Nature Structural & Molecular Biology* **2021**, *28*, 704–705.
65. Read, R.J.; Baker, E.N.; Bond, C.S.; Garman, E.F.; van Raaij, M.J. AlphaFold and the future of structural biology. *Acta Crystallographica Section D Structural Biology* **2023**, *79*, 556–558.
66. Platzer, G.; Okon, M.; McIntosh, L.P. pH-dependent random coil 1H, 13C, and 15N chemical shifts of the ionizable amino acids: a guide for protein pKa measurements. *Journal of Biomolecular NMR* **2014**, *60*, 109–129.
67. Yang, A.S.; Honig, B. On the pH Dependence of Protein Stability. *Journal of Molecular Biology* **1993**, *231*, 459–474.
68. Harris, T.K.; Turner, G.J. Structural Basis of Perturbed pKa Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* **2002**, *53*, 85–98.
69. Li, W. Gravity-driven pH adjustment for site-specific protein pKa measurement by solution-state NMR. *Measurement Science and Technology* **2017**, *28*, 127002.
70. Li, W. Characterising the interaction between caenopore-5 and model membranes by NMR spectroscopy and molecular dynamics simulations. PhD thesis, University of Auckland, 2016.
71. Hansen, A.L.; Kay, L.E. Measurement of histidine pKa values and tautomer populations in invisible protein states. *Proceedings of the National Academy of Sciences* **2014**, *111*, E1705–E1712.
72. Weiss, M. Design of ultra-stable insulin analogues for the developing world. *Journal of Health Specialties* **2013**, *1*, 59.
73. Nuhoho, S.; Gupta, J.; Hansen, B.B.; Fletcher-Louis, M.; Dang-Tan, T.; Paine, A. Orally Administered Semaglutide Versus GLP-1 RAs in Patients with Type 2 Diabetes Previously Receiving 1-2 Oral Antidiabetics: Systematic Review and Network Meta-Analysis. *Diabetes Therapy* **2019**, *10*, 2183–2199.
74. Bucheit, J.D.; Pamulapati, L.G.; Carter, N.; Malloy, K.; Dixon, D.L.; Sisson, E.M. Oral Semaglutide: A Review of the First Oral Glucagon-Like Peptide 1 Receptor Agonist. *Diabetes Technology & Therapeutics* **2020**, *22*, 10–18.
75. 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.
76. 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.

77. Hofmann, D.W.M.; Kuleshova, L.N. A general force field by machine learning on experimental crystal structures. Calculations of intermolecular Gibbs energy with iFlexCryst. *Acta Crystallographica Section A Foundations and Advances* **2023**, *79*, 132–144.
78. Müller, C.E.; Hansen, F.K.; Gütschow, M.; Lindsley, C.W.; Liotta, D. New Drug Modalities in Medicinal Chemistry, Pharmacology, and Translational Science. *ACS Pharmacology & Translational Science* **2021**, *4*, 1712–1713.
79. 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.
80. Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics* **2018**, *34*, i821–i829.
81. Wang, K.; Zhou, R.; Li, Y.; Li, M. DeepDTAF: a deep learning method to predict protein-ligand binding affinity. *Briefings in Bioinformatics* **2021**, *22*.
82. Reif, M.M.; Zacharias, M. Computational Tools for Accurate Binding Free-Energy Prediction. In *Methods in Molecular Biology*; Springer US, 2021; pp. 255–292.
83. Kerdawy, A.E.; Wick, C.R.; Hennemann, M.; Clark, T. Predicting the Sites and Energies of Noncovalent Intermolecular Interactions Using Local Properties. *Journal of Chemical Information and Modeling* **2012**, *52*, 1061–1071.
84. Bitencourt-Ferreira, G.; de Azevedo Junior, W.F. Electrostatic Potential Energy in Protein-Drug Complexes. *Current Medicinal Chemistry* **2021**, *28*, 4954–4971.
85. Fahmy, A.; Wagner, G. Optimization of van der Waals Energy for Protein Side-Chain Placement and Design. *Biophysical Journal* **2011**, *101*, 1690–1698.
86. Umeyama, H.; Morokuma, K. The origin of hydrogen bonding. An energy decomposition study. *J. Am. Chem. Soc.* **1977**, *99*, 1316–1332.
87. Li, W. Gravity-driven pH adjustment for site-specific protein pKa measurement by solution-state NMR. *Measurement Science and Technology* **2017**, *28*, 127002.
88. 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.
89. 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*.
90. 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-15R$\alpha$-Fc chimera. *Immunology Letters* **2014**, *159*, 1–10.
91. 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.
92. 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.
93. 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.
94. 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.
95. Carracedo-Reboredo, P.; Liñares-Blanco, J.; Rodríguez-Fernández, N.; Cedrón, F.; Novoa, F.J.; Carballal, A.; Maojo, V.; Pazos, A.; Fernandez-Lozano, C. A review on machine learning approaches and trends in drug discovery. *Computational and Structural Biotechnology Journal* **2021**, *19*, 4538–4558.
96. 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.
97. Vasselli, J.R.; Frentzas, S.; Weickhardt, A.J.; de Souza, P.L.; Tang, J.; Wyant, T.; Amit, I.; Ofran, Y.; Knickerbocker, A. Trial in progress: A phase 1-2, first-in-human, open label, dose escalation and expansion study of AU-007, a monoclonal antibody that binds to IL-2 and inhibits IL-2R$\alpha$binding, in patients with advanced solid tumors. *Journal of Clinical Oncology* **2022**, *40*, TPS2671–TPS2671.

 doi:10.20944/preprints202208.0213.v2




98. Torres, A.M.; Forbes, B.E.; Aplin, S.E.; Wallace, J.C.; Francise, G.L.; Norton, R.S. Solution structure of human insulin-like growthfactor II. Relationship to receptor and binding protein interactions. *Journal of Molecular Biology* **1995**, *248*, 385–401.
99. Lee, J.; Pilch, P.F. The insulin receptor: structure, function, and signaling. *American Journal of Physiology-Cell Physiology* **1994**, *266*, C319–C334.
100. Rahuel-Clermont, S.; French, C.A.; Kaarsholm, N.C.; Dunn, M.F. Mechanisms of Stabilization of the Insulin Hexamer through Allosteric Ligand Interactions. *Biochemistry* **1997**, *36*, 5837–5845.
101. Robert, C.; Schachter, J.; Long, G.V.; Arance, A.; Grob, J.J.; Mortier, L.; Daud, A.; Carlino, M.S.; McNeil, C.; Lotem, M.; Larkin, J.; Lorigan, P.; Neyns, B.; Blank, C.U.; Hamid, O.; Mateus, C.; Shapira-Frommer, R.; Kosh, M.; Zhou, H.; Ibrahim, N.; Ebbinghaus, S.; Ribas, A. Pembrolizumab versus Ipilimumab in Advanced Melanoma. *New England Journal of Medicine* **2015**, *372*, 2521–2532.
102. Pauken, K.E.; Wherry, E.J. Overcoming T cell exhaustion in infection and cancer. *Trends in Immunology* **2015**, *36*, 265–276.
103. Yu, X.; Orr, C.M.; Chan, H.T.C.; James, S.; Penfold, C.A.; Kim, J.; Inzhelevskaya, T.; Mockridge, C.I.; Cox, K.L.; Essex, J.W.; Tews, I.; Glennie, M.J.; Cragg, M.S. Reducing affinity as a strategy to boost immunomodulatory antibody agonism. *Nature* **2023**, *614*, 539–547.
104. Wülfing, C.; Dovedi, S.J. For optimal antibody effectiveness, sometimes less is more. *Nature* **2023**, *614*, 416–418.
105. Perret, G.; Boschetti, E. Aptamer-Based Affinity Chromatography for Protein Extraction and Purification. In *Aptamers in Biotechnology*; Springer International Publishing, 2019; pp. 93–139.
106. 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.
107. 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.
108. 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*.
109. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery* **2018**, *18*, 41–58.
110. 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*.
111. Agrawal, P.; Singh, H.; Srivastava, H.K.; Singh, S.; Kishore, G.; Raghava, G.P.S. Benchmarking of different molecular docking methods for protein-peptide docking. *BMC Bioinformatics* **2019**, *19*.
112. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research* **2005**, *33*, W363–W367.
113. Stefano, M.D.; Galati, S.; Piazza, L.; Granchi, C.; Mancini, S.; Fratini, F.; Macchia, M.; Poli, G.; Tuccinardi, T. VenomPred 2.0: A Novel In Silico Platform for an Extended and Human Interpretable Toxicological Profiling of Small Molecules. *Journal of Chemical Information and Modeling* **2023**.
114. Yugandhar, K.; Gromiha, M.M. Protein-protein binding affinity prediction from amino acid sequence. *Bioinformatics* **2014**, *30*, 3583–3589.
115. Rube, H.T.; Rastogi, C.; Feng, S.; Kribelbauer, J.F.; Li, A.; Becerra, B.; Melo, L.A.N.; Do, B.V.; Li, X.; Adam, H.H.; Shah, N.H.; Mann, R.S.; Bussemaker, H.J. Prediction of protein-ligand binding affinity from sequencing data with interpretable machine learning. *Nature Biotechnology* **2022**, *40*, 1520–1527.
116. O'Boyle, N.M. Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI. *Journal of Cheminformatics* **2012**, *4*.
117. Hähnke, V.D.; Kim, S.; Bolton, E.E. PubChem chemical structure standardization. *Journal of Cheminformatics* **2018**, *10*.
118. Wiswesser, W.J. 107 Years of Line-Formula Notations (1861-1968). *Journal of Chemical Documentation* **1968**, *8*, 146–150.
119. Herget, S.; Ranzinger, R.; Maass, K.; Lieth, C.W. GlycoCT—a unifying sequence format for carbohydrates. *Carbohydrate Research* **2008**, *343*, 2162–2171.
120. Kawade, R.; Kuroda, D.; Tsumoto, K. How the protonation state of a phosphorylated amino acid governs molecular recognition: insights from classical molecular dynamics simulations. *FEBS Letters* **2019**, *594*, 903–912.

121. Nishi, H.; Shaytan, A.; Panchenko, A.R. Physicochemical mechanisms of protein regulation by phosphorylation. *Frontiers in Genetics* **2014**, *5*.
122. Cohen, P. The role of protein phosphorylation in human health and disease. *European Journal of Biochemistry* **2001**, *268*, 5001–5010.
123. Sud, M.; Fahy, E.; Subramaniam, S. Template-based combinatorial enumeration of virtual compound libraries for lipids. *Journal of Cheminformatics* **2012**, *4*.
124. Foster, J.M.; Moreno, P.; Fabregat, A.; Hermjakob, H.; Steinbeck, C.; Apweiler, R.; Wakelam, M.J.O.; Vizcaíno, J.A. LipidHome: A Database of Theoretical Lipids Optimized for High Throughput Mass Spectrometry Lipidomics. *PLoS ONE* **2013**, *8*, e61951.
125. Capecchi, A.; Probst, D.; Reymond, J.L. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *Journal of Cheminformatics* **2020**, *12*.
126. Krenn, M.; Ai, Q.; Barthel, S.; Carson, N.; Frei, A.; Frey, N.C.; Friederich, P.; Gaudin, T.; Gayle, A.A.; Jablonka, K.M.; Lameiro, R.F.; Lemm, D.; Lo, A.; Moosavi, S.M.; Nápoles-Duarte, J.M.; Nigam, A.; Pollice, R.; Rajan, K.; Schatzschneider, U.; Schwaller, P.; Skreta, M.; Smit, B.; Strieth-Kalthoff, F.; Sun, C.; Tom, G.; von Rudorff, G.F.; Wang, A.; White, A.D.; Young, A.; Yu, R.; Aspuru-Guzik, A. SELFIES and the future of molecular string representations. *Patterns* **2022**, *3*, 100588.
127. Resh, M.D. Fatty acylation of proteins: The long and the short of it. *Progress in Lipid Research* **2016**, *63*, 120–131.
128. Nishimura, E.; Pridal, L.; Glendorf, T.; Hansen, B.F.; Hubálek, F.; Kjeldsen, T.; Kristensen, N.R.; Lützen, A.; Lyby, K.; Madsen, P.; Pedersen, T.Å.; Ribel-Madsen, R.; Stidsen, C.E.; Haahr, H. Molecular and pharmacological characterization of insulin icodec: a new basal insulin analog designed for once-weekly dosing. *BMJ Open Diabetes Research & Care* **2021**, *9*, e002301.
129. Steven, A.C.; Baumeister, W. The future is hybrid. *Journal of Structural Biology* **2008**, *163*, 186–195.
130. Kang, M.; Lu, Y.; Chen, S.; Tian, F. Harnessing the power of an expanded genetic code toward next-generation biopharmaceuticals. *Current Opinion in Chemical Biology* **2018**, *46*, 123–129.
131. Coley, C.W. Defining and Exploring Chemical Spaces. *Trends in Chemistry* **2021**, *3*, 133–145. Special Issue: Machine Learning for Molecules and Materials.
132. Rogers, J.M.; Steward, A.; Clarke, J. Folding and Binding of an Intrinsically Disordered Protein: Fast, but Not 'Diffusion-Limited'. *Journal of the American Chemical Society* **2013**, *135*, 1415–1422.
133. Sugase, K.; Dyson, H.J.; Wright, P.E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007**, *447*, 1021–1025.
134. 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.
135. 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.
136. Bender, A.; Cortés-Ciriano, I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. *Drug Discovery Today* **2021**, *26*, 511–524.
137. Gorai, B.; Vashisth, H. Structural models of viral insulin-like peptides and their analogs. *Proteins: Structure, Function, and Bioinformatics* **2022**, *91*, 62–73.
138. Wu, D.; Sun, J.; Xu, T.; Wang, S.; Li, G.; Li, Y.; Cao, Z. Stacking and energetic contribution of aromatic islands at the binding interface of antibody proteins. *Immunome Research* **2010**, *6*, S1.
139. 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.
140. Dunn, A. 'The more overhype, the harder you fall': Schrödinger CEO warns on AI craze., 2023. Accessed: (September 10, 2023).
141. Smith, G.F. Artificial Intelligence in Drug Safety and Metabolism. In *Artificial Intelligence in Drug Design*; Springer US, 2021; pp. 483–501.
142. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research* **2018**, *47*, D520–D528.
143. 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.
144. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research* **2007**, *35*, D198–D201.

 doi:10.20944/preprints202208.0213.v2




145. 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.
146. Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *Journal of Chemical Information and Modeling* **2018**, *59*, 895–913.
147. Li, Y.; Su, M.; Liu, Z.; Li, J.; Liu, J.; Han, L.; Wang, R. Assessing protein-ligand interaction scoring functions with the CASF-2013 benchmark. *Nature Protocols* **2018**, *13*, 666–680.
148. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry* **2012**, *55*, 6582–6594.
149. 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.
150. Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A.C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z.T.; Han, B.; Zhou, Y.; Wishart, D.S. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research* **2013**, *42*, D1091–D1097.
151. Dunbar, J.B.; Smith, R.D.; Damm-Ganamet, K.L.; Ahmed, A.; Esposito, E.X.; Delproposto, J.; Chinnaswamy, K.; Kang, Y.N.; Kubish, G.; Gestwicki, J.E.; Stuckey, J.A.; Carlson, H.A. CSAR Data Set Release 2012: Ligands, Affinities, Complexes, and Docking Decoys. *Journal of Chemical Information and Modeling* **2013**, *53*, 1842–1852.
152. Rohrer, S.G.; Baumann, K. Maximum Unbiased Validation (MUV) Data Sets for Virtual Screening Based on PubChem Bioactivity Data. *Journal of Chemical Information and Modeling* **2009**, *49*, 169–184.
153. Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. *Journal of Medicinal Chemistry* **2004**, *47*, 2977–2980.
154. Kulkarni-Kale, U.; Raskar-Renuse, S.; Natekar-Kalantre, G.; Saxena, S.A. Antigen-Antibody Interaction Database AgAbDb: A Compendium of Antigen-Antibody Interactions. In *Methods in Molecular Biology*; Springer New York, 2014; pp. 149–164.
155. Austin, C.P.; Brady, L.S.; Insel, T.R.; Collins, F.S. NIH Molecular Libraries Initiative. *Science* **2004**, *306*, 1138–1139.
156. Manso, T.; Folch, G.; Giudicelli, V.; Jabado-Michaloud, J.; Kushwaha, A.; Ngoune, V.N.; Georga, M.; Papadaki, A.; Debbagh, C.; Pégorier, P.; Bertignac, M.; Hadi-Saljoqi, S.; Chentli, I.; Cherouali, K.; Aouinti, S.; Hamwi, A.E.; Albani, A.; Elhassani, M.E.; Viart, B.; Goret, A.; Tran, A.; Sanou, G.; Rollin, M.; Duroux, P.; Kossida, S. IMGT® databases, related tools and web resources through three main axes of research and development. *Nucleic Acids Research* **2021**, *50*, D1262–D1272.
157. Freire, E. Biophysical methods for the determination of protein–ligand binding constants. *Annual review of biophysics* **2009**, *38*, 123–142.
158. Johnson, C.M. Isothermal Calorimetry. In *Protein-Ligand Interactions*; Springer US, 2021; pp. 135–159.
159. Velázquez-Coy, A.; Ohtaka, H.; Nezami, A.; Muzammil, S.; Freire, E. Isothermal Titration Calorimetry. *Current Protocols in Cell Biology* **2004**, *23*.
160. Sauer, U.G. Surface plasmon resonance–a label-free tool for cellular analysis. *Journal of biotechnology* **2008**, *133*, 101–108.
161. Ernst, R.R.; Bodenhausen, G.; Wokaun, A. Principles of nuclear magnetic resonance in one and two dimensions. *Principles of nuclear magnetic resonance in one and two dimensions* **1990**, *14*.
162. de Oliveira, T.M.; van Beek, L.; Shilliday, F.; Debreczeni, J.É.; Phillips, C. Cryo-EM: The Resolution Revolution and Drug Discovery. *SLAS Discovery* **2021**, *26*, 17–31.
163. Haas, E.; Katchalski-Katzir, E.; Steinberg, I.Z. Fluorescence resonance energy transfer. *Journal of biochemical and biophysical methods* **2001**, *49*, 1–16.
164. Seidel, S.A.; Dijkman, P.M.; Lea, W.A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J.S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; others. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **2013**, *59*, 301–315.
165. Pantoliano, M.W.; Petrella, E.C.; Kwasnoski, J.D.; Lobanov, V.S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B.A.; Lane, P.; others. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *Journal of biomolecular screening* **2001**, *6*, 429–440.
166. Drenth, J. Principles of protein X-ray crystallography. *Springer Science & Business Media* **2007**.

 doi:10.20944/preprints202208.0213.v2




167. Smith, R.D.; Anderson, G.A.; Lipton, M.S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T.P.; Veenstra, T.D.; Udseth, H.R. Proteomics, genomics and transcriptomics: the integration of proteomics, genomics and transcriptomics using mass spectrometry. *Nature Reviews Genetics* **2002**, *3*, 601–614.
168. Rich, R.L.; Myszka, D.G. Biolayer interferometry for bioaffinity/biointeraction measurements. *Methods in molecular biology* **2009**, *503*, 287–305.
169. Chen, R.J.; Lu, M.Y.; Chen, T.Y.; Williamson, D.F.K.; Mahmood, F. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering* **2021**, *5*, 493–497.
170. Jadon, A.; Kumar, S. Leveraging Generative AI Models for Synthetic Data Generation in Healthcare: Balancing Research and Privacy. 2023 International Conference on Smart Applications, Communications and Networking (SmartNets). IEEE, 2023.
171. Callaway, E. The entire protein universe: AI predicts shape of nearly every known protein. *Nature* **2022**.
172. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018**, *46*, W296–W303.
173. Hasani, H.J.; Barakat, K. Homology Modeling: an Overview of Fundamentals and Tools. *International Review on Modelling and Simulations (IREMOS)* **2017**, *10*, 129.
174. 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.
175. Tong, A.B.; Burch, J.D.; McKay, D.; Bustamante, C.; Crackower, M.A.; Wu, H. Could AlphaFold revolutionize chemical therapeutics? *Nature Structural & Molecular Biology* **2021**, *28*, 771–772.
176. Agu, P.C.; Afiukwa, C.A.; Orji, O.U.; Ezeh, E.M.; Ofoke, I.H.; Ogbu, C.O.; Ugwuja, E.I.; Aja, P.M. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. *Scientific Reports* **2023**, *13*.
177. Zheng, Z.; Merz, K.M. Calculating protein-ligand binding affinities with MMPBSA: Method and error analysis. *Journal of chemical theory and computation* **2017**, *13*, 4751–4767.
178. Deng, N.J.; Zheng, Q.; Liu, J.; Hao, G.F. Predicting protein–ligand binding affinity with a random matrix framework. *PLoS computational biology* **2012**, *8*, e1002775.
179. Du, X.; Li, Y.; Xia, Y.; Ai, J.; Wu, Y. Molecular dynamics simulations and free energy calculations of protein–ligand interactions: recent advances and future perspectives. *Current pharmaceutical design* **2017**, *23*, 4436–4450.
180. Li, Y.; Han, L.; Liu, Z.; Wang, R. Computational approaches for scoring protein–ligand affinity. *Current topics in medicinal chemistry* **2014**, *14*, 2088–2102.
181. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nature structural biology* **2002**, *9*, 646–652.
182. Canzar, S.; Toussaint, N.C.; Klau, G.W. An exact algorithm for side-chain placement in protein design. *Optimization Letters* **2011**, *5*, 393–406.
183. Webb, B.; Sali, A. Protein Structure Modeling with MODELLER. In *Methods in Molecular Biology*; Springer US, 2020; pp. 239–255.
184. wei Lin, D.Y.; Tanaka, Y.; Iwasaki, M.; Gittis, A.G.; Su, H.P.; Mikami, B.; Okazaki, T.; Honjo, T.; Minato, N.; Garboczi, D.N. The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. *Proceedings of the National Academy of Sciences* **2008**, *105*, 3011–3016.
185. Lin, D.; Tanaka, Y.; Iwasaki, M.; Gittis, A.; Su, H.; Mikami, B.; Okazaki, T.; Honjo, T.; Minato, N.; Garboczi, D. Crystal Structure of the PD-1/PD-L1 Complex, 2008.
186. Li, Z.; Liu, X.; Chen, W.; Shen, F.; Bi, H.; Ke, G.; Zhang, L. Uni-Fold: An Open-Source Platform for Developing Protein Folding Models beyond AlphaFold **2022**.
187. Goldbach, L.; Vermeulen, B.J.A.; Caner, S.; Liu, M.; Tysoe, C.; van Gijzel, L.; Yoshisada, R.; Trellet, M.; van Ingen, H.; Brayer, G.D.; Bonvin, A.M.J.J.; Jongkees, S.A.K. Folding Then Binding vs Folding Through Binding in Macrocyclic Peptide Inhibitors of Human Pancreatic α-Amylase. *ACS Chemical Biology* **2019**, *14*, 1751–1759.
188. Jain, S.; Pei, L.; Spraggins, J.M.; Angelo, M.; Carson, J.P.; Gehlenborg, N.; Ginty, F.; Gonçalves, J.P.; Hagood, J.S.; Hickey, J.W.; Kelleher, N.L.; Laurent, L.C.; Lin, S.; Lin, Y.; Liu, H.; Naba, A.; Nakayasu, E.S.; Qian, W.J.; Radtke, A.; Robson, P.; Stockwell, B.R.; de Plas, R.V.; Vlachos, I.S.; Zhou, M.; Ahn, K.J.; Allen, J.; Anderson,

 doi:10.20944/preprints202208.0213.v2



D.M.; Anderton, C.R.; Curcio, C.; Angelin, A.; Arvanitis, C.; Atta, L.; Awosika-Olumo, D.; Bahmani, A.; Bai, H.; Balderrama, K.; Balzano, L.; Bandyopadhyay, G.; Bandyopadhyay, S.; Bar-Joseph, Z.; Barnhart, K.; Barwinska, D.; Becich, M.; Becker, L.; Becker, W.; Bedi, K.; Bendall, S.; Benninger, K.; Betancur, D.; Bettinger, K.; Billings, S.; Blood, P.; Bolin, D.; Border, S.; Bosse, M.; Bramer, L.; Brewer, M.; Brusko, M.; Bueckle, A.; Burke, K.; Burnum-Johnson, K.; Butcher, E.; Butterworth, E.; Cai, L.; Calandrelli, R.; Caldwell, M.; Campbell-Thompson, M.; Cao, D.; Cao-Berg, I.; Caprioli, R.; Caraccio, C.; Caron, A.; Carroll, M.; Chadwick, C.; Chen, A.; Chen, D.; Chen, F.; Chen, H.; Chen, J.; Chen, L.; Chen, L.; Chiacchia, K.; Cho, S.; Chou, P.; Choy, L.; Cisar, C.; Clair, G.; Clarke, L.; Clouthier, K.A.; Colley, M.E.; Conlon, K.; Conroy, J.; Contrepois, K.; Corbett, A.; Corwin, A.; Cotter, D.; Courtois, E.; Cruz, A.; Csonka, C.; Czupil, K.; Daiya, V.; Dale, K.; Davanagere, S.A.; Dayao, M.; de Caestecker, M.P.; Decker, A.; Deems, S.; Degnan, D.; Desai, T.; Deshpande, V.; Deutsch, G.; Devlin, M.; Diep, D.; Dodd, C.; Donahue, S.; Dong, W.; dos Santos Peixoto, R.; Duffy, M.; Dufresne, M.; Duong, T.E.; Dutra, J.; Eadon, M.T.; El-Achkar, T.M.; Enninful, A.; Eraslan, G.; Eshelman, D.; Espin-Perez, A.; Esplin, E.D.; Esselman, A.; Falo, L.D.; Falo, L.; Fan, J.; Fan, R.; Farrow, M.A.; Farzad, N.; Favaro, P.; Fermin, J.; Filiz, F.; Filus, S.; Fisch, K.; Fisher, E.; Fisher, S.; Flowers, K.; Flynn, W.F.; Fogo, A.B.; Fu, D.; Fulcher, J.; Fung, A.; Furst, D.; Gallant, M.; Gao, F.; Gao, Y.; Gaulton, K.; Gaut, J.P.; Gee, J.; Ghag, R.R.; Ghazanfar, S.; Ghose, S.; Gisch, D.; Gold, I.; Gondalia, A.; Gorman, B.; Greenleaf, W.; Greenwald, N.; Gregory, B.; Guo, R.; Gupta, R.; Hakimian, H.; Haltom, J.; Halushka, M.; Han, K.S.; Hanson, C.; Harbury, P.; Hardi, J.; Harlan, L.; Harris, R.C.; Hartman, A.; Heidari, E.; Helfer, J.; Helminiak, D.; Hemberg, M.; Henning, N.; Herr, B.W.; Ho, J.; Holden-Wiltse, J.; Hong, S.H.; Hong, Y.K.; Honick, B.; Hood, G.; Hu, P.; Hu, Q.; Huang, M.; Huyck, H.; Imtiaz, T.; Isberg, O.G.; Itkin, M.; Jackson, D.; Jacobs, M.; Jain, Y.; Jewell, D.; Jiang, L.; Jiang, Z.G.; Johnston, S.; Joshi, P.; Ju, Y.; Judd, A.; Kagel, A.; Kahn, A.; Kalavros, N.; Kalhor, K.; Karagkouni, D.; Karathanos, T.; Karunamurthy, A.; Katari, S.; Kates, H.; Kaushal, M.; Keener, N.; Keller, M.; Kenney, M.; Kern, C.; Kharchenko, P.; Kim, J.; Kingsford, C.; Kirwan, J.; Kiselev, V.; Kishi, J.; Kitata, R.B.; Knoten, A.; Kollar, C.; Krishnamoorthy, P.; Kruse, A.R.S.; Da, K.; Kundaje, A.; Kutschera, E.; Kwon, Y.; Lake, B.B.; Lancaster, S.; Langlieb, J.; Lardenoije, R.; Laronda, M.; Laskin, J.; Lau, K.; Lee, H.; Lee, M.; Lee, M.; Strekalova, Y.L.; Li, D.; Li, J.; Li, J.; Li, X.; Li, Z.; Liao, Y.C.; Liaw, T.; Lin, P.; Lin, Y.; Lindsay, S.; Liu, C.; Liu, Y.; Liu, Y.; Lott, M.; Lotz, M.; Lowery, L.; Lu, P.; Lu, X.; Lucarelli, N.; Lun, X.; Luo, Z.; Ma, J.; Macosko, E.; Mahajan, M.; Maier, L.; Makowski, D.; Malek, M.; Manthey, D.; Manz, T.; Margulies, K.; Marioni, J.; Martindale, M.; Mason, C.; Mathews, C.; Maye, P.; McCallum, C.; McDonough, E.; McDonough, L.; Mcdowell, H.; Meads, M.; Medina-Serpas, M.; Ferreira, R.M.; Messinger, J.; Metis, K.; Migas, L.G.; Miller, B.; Mimar, S.; Minor, B.; Misra, R.; Missarova, A.; Mistretta, C.; Moens, R.; Moerth, E.; Moffitt, J.; Molla, G.; Monroe, M.; Monte, E.; Morgan, M.; Muraro, D.; Murphy, B.; Murray, E.; Musen, M.A.; Naglah, A.; Nasamran, C.; Neelakantan, T.; Nevins, S.; Nguyen, H.; Nguyen, N.; Nguyen, T.; Nguyen, T.; Nigra, D.; Nofal, M.; Nolan, G.; Nwanne, G.; O'Connor, M.; Okuda, K.; Olmer, M.; O'Neill, K.; Otaluka, N.; Pang, M.; Parast, M.; Pasa-Tolic, L.; Paten, B.; Patterson, N.H.; Peng, T.; Phillips, G.; Pichavant, M.; Piehowski, P.; Pilner, H.; Pingry, E.; Pita-Juarez, Y.; Plevritis, S.; Ploumakis, A.; Pouch, A.; Pryhuber, G.; Puerto, J.; Qaurooni, D.; Qin, L.; Quardokus, E.M.; Rajbhandari, P.; Rakow-Penner, R.; Ramasamy, R.; Read, D.; Record, E.G.; Reeves, D.; Ricarte, A.; Rodríguez-Soto, A.; Ropelewski, A.; Rosario, J.; Roselkis, M.A.; Rowe, D.; Roy, T.K.; Ruffalo, M.; Ruschman, N.; Sabo, A.; Sachdev, N.; Saka, S.; Salamon, D.; Sarder, P.; Sasaki, H.; Satija, R.; Saunders, D.; Sawka, R.; Schey, K.; Schlehlein, H.; Scholten, D.; Schultz, S.; Schwartz, L.; Schwenk, M.; Scibek, R.; Segre, A.; Serrata, M.; Shands, W.; Shen, X.; Shendure, J.; Shephard, H.; Shi, L.; Shi, T.; Shin, D.G.; Shirey, B.; Sibilla, M.; Silber, M.; Silverstein, J.; Simmel, D.; Simmons, A.; Singhal, D.; Sivajothi, S.; Smits, T.; Soncin, F.; Song, Q.; Stanley, V.; Stuart, T.; Su, H.; Su, P.; Sun, X.; Surrette, C.; Swahn, H.; Tan, K.; Teichmann, S.; Tejomay, A.; Tellides, G.; Thomas, K.; Thomas, T.; Thompson, M.; Tian, H.; Tideman, L.; Trapnell, C.; Tsai, A.G.; Tsai, C.F.; Tsai, L.; Tsui, E.; Tsui, T.; Tung, J.; Turner, M.; Uranic, J.; Vaishnav, E.D.; Varra, S.R.; Vaskivskyi, V.; Velickovic, D.; Velickovic, M.; Verheyden, J.; Waldrip, J.; Wallace, D.; Wan, X.; Wang, A.; Wang, F.; Wang, M.; Wang, S.; Wang, X.; Wasserfall, C.; Wayne, L.; Webber, J.; Weber, G.M.; Wei, B.; Wei, J.J.; Weimer, A.; Welling, J.; Wen, X.; Wen, Z.; Williams, M.; Winfree, S.; Winograd, N.; Woodard, A.; Wright, D.; Wu, F.; Wu, P.H.; Wu, Q.; Wu, X.; Xing, Y.; Xu, T.; Yang, M.; Yang, M.; Yap, J.; Ye, D.H.; Yin, P.; Yuan, Z.; Yun, C.; Zahraei, A.; Zemaitis, K.; Zhang, B.; Zhang, C.; Zhang, C.; Zhang, C.; Zhang, K.; Zhang, S.; Zhang, T.; Zhang, Y.; Zhao, B.; Zhao, W.; Zheng, J.W.; Zhong, S.; Zhu, B.; Zhu, C.; Zhu, D.; Zhu, Q.; Zhu, Y.; Börner, K.; and, M.P.S. Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP). *Nature Cell Biology* **2023**.

189. and Michael P. Snyder.; Lin, S.; Posgai, A.; Atkinson, M.; Regev, A.; Rood, J.; Rozenblatt-Rosen, O.; Gaffney, L.; Hupalowska, A.; Satija, R.; Gehlenborg, N.; Shendure, J.; Laskin, J.; Harbury, P.; Nystrom, N.A.; Silverstein, J.C.; Bar-Joseph, Z.; Zhang, K.; Börner, K.; Lin, Y.; Conroy, R.; Procaccini, D.; Roy, A.L.; Pillai, A.; Brown, M.; Galis, Z.S.; Cai, L.; Shendure, J.; Trapnell, C.; Lin, S.; Jackson, D.; Snyder, M.P.; Nolan, G.; Greenleaf, W.J.; Lin, Y.; Plevritis, S.; Ahadi, S.; Nevins, S.A.; Lee, H.; Schuerch, C.M.; Black, S.; Venkataraaman, V.G.; Esplin, E.; Horning, A.; Bahmani, A.; Zhang, K.; Sun, X.; Jain, S.; Hagood, J.; Pryhuber, G.; Kharchenko, P.; Atkinson, M.; Bodenmiller, B.; Brusko, T.; Clare-Salzler, M.; Nick, H.; Otto, K.; Posgai, A.; Wasserfall, C.; Jorgensen, M.; Brusko, M.; Maffioletti, S.; Caprioli, R.M.; Spraggins, J.M.; Gutierrez, D.; Patterson, N.H.; Neumann, E.K.; Harris, R.; deCaestecker, M.; Fogo, A.B.; van de Plas, R.; Lau, K.; Cai, L.; Yuan, G.C.; Zhu, Q.; Dries, R.; Yin, P.; Saka, S.K.; Kishi, J.Y.; Wang, Y.; Goldaracena, I.; Laskin, J.; Ye, D.; Burnum-Johnson, K.E.; Piehowski, P.D.; Ansong, C.; Zhu, Y.; Harbury, P.; Desai, T.; Mulye, J.; Chou, P.; Nagendran, M.; Bar-Joseph, Z.; Teichmann, S.A.; Paten, B.; Murphy, R.F.; Ma, J.; Kiselev, V.Y.; Kingsford, C.; Ricarte, A.; Keays, M.; Akoju, S.A.; Ruffalo, M.; Gehlenborg, N.; Kharchenko, P.; Vella, M.; McCallum, C.; Börner, K.; Cross, L.E.; Friedman, S.H.; Heiland, R.; Herr, B.; Macklin, P.; Quardokus, E.M.; Record, L.; Sluka, J.P.; Weber, G.M.; Nystrom, N.A.; Silverstein, J.C.; Blood, P.D.; Ropelewski, A.J.; Shirey, W.E.; Scibek, R.M.; Mabee, P.; Lenhardt, W.C.; Robasky, K.; Michailidis, S.; Satija, R.; Marioni, J.; Regev, A.; Butler, A.; Stuart, T.; Fisher, E.; Ghazanfar, S.; Rood, J.; Gaffney, L.; Eraslan, G.; Biancalani, T.; Vaishnav, E.D.; Conroy, R.; Procaccini, D.; Roy, A.; Pillai, A.; Brown, M.; Galis, Z.; Srinivas, P.; Pawlyk, A.; Sechi, S.; Wilder, E.; Anderson, J.; .; .; .; .; .; .; .; .; .; .; .; .; .; .; .; and. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **2019**, *574*, 187–192.
190. Sadybekov, A.A.; Sadybekov, A.V.; Liu, Y.; Iliopoulos-Tsoutsouvas, C.; Huang, X.P.; Pickett, J.; Houser, B.; Patel, N.; Tran, N.K.; Tong, F.; Zvonok, N.; Jain, M.K.; Savych, O.; Radchenko, D.S.; Nikas, S.P.; Petasis, N.A.; Moroz, Y.S.; Roth, B.L.; Makriyannis, A.; Katritch, V. Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **2021**, *601*, 452–459.
191. Kinghorn, A.; Fraser, L.; Liang, S.; Shiu, S.; Tanner, J. Aptamer Bioinformatics. *International Journal of Molecular Sciences* **2017**, *18*, 2516.
192. Lobato, M.; Rabbitts, T. Intracellular Antibodies as Specific Reagents for Functional Ablation: Future Therapeutic Molecules. *Current Molecular Medicine* **2004**, *4*, 519–528.
193. Weng, G.; Shen, C.; Cao, D.; Gao, J.; Dong, X.; He, Q.; Yang, B.; Li, D.; Wu, J.; Hou, T. PROTAC-DB: an online database of PROTACs. *Nucleic Acids Research* **2020**, *49*, D1381–D1387.
194. Weng, G.; Cai, X.; Cao, D.; Du, H.; Shen, C.; Deng, Y.; He, Q.; Yang, B.; Li, D.; Hou, T. PROTAC-DB 2.0: an updated database of PROTACs. *Nucleic Acids Research* **2022**, *51*, D1367–D1372.
195. Zhang, W.; Chen, Y.; Liu, F.; Luo, F.; Tian, G.; Li, X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinformatics* **2017**, *18*.
196. Rahnama, R.; Kizerwetter, M.; Yang, H.; Christodoulou, I.; Fearnow, A.; Vorri, S.; Spangler, J.; Bonifant, C. Poster: AML-496 Study of Affinity Tuning in AML-Targeted CAR-NK Cells. *Clinical Lymphoma Myeloma and Leukemia* **2022**, *22*, S136.
197. Fakhrai-Rad, H. Insulin-degrading enzyme identified as a candidate diabetes susceptibility gene in GK rats. *Human Molecular Genetics* **2000**, *9*, 2149–2158.
198. González-Casimiro, C.M.; Merino, B.; Casanueva-Álvarez, E.; Postigo-Casado, T.; Cámara-Torres, P.; Fernández-Díaz, C.M.; Leissring, M.A.; Cózar-Castellano, I.; Perdomo, G. Modulation of Insulin Sensitivity by Insulin-Degrading Enzyme. *Biomedicines* **2021**, *9*, 86.
199. Ferreira, L.G.; Santos, R.N.; Andricopulo, A.D. Computational methods for calculation of protein-ligand binding affinities. *Frontiers in chemistry* **2019**, *7*, 559.
200. Petukh, M.; Stefl, S.; Alexov, E. The Role of Protonation States in Ligand-Receptor Recognition and Binding. *Current Pharmaceutical Design* **2013**, *19*, 4182–4190.
201. Jubb, H.C.; Pandurangan, A.P.; Turner, M.A.; Ochoa-Montaño, B.; Blundell, T.L.; Ascher, D.B. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. *Progress in Biophysics and Molecular Biology* **2017**, *128*, 3–13.
202. Li, W. Calcium Channel Trafficking Blocker Gabapentin Bound to the α-2-δ-1 Subunit of Voltage-Gated Calcium Channel: A Computational Structural Investigation **2020**.
203. Elliott, W.J.; Ram, C.V.S. Calcium Channel Blockers. *The Journal of Clinical Hypertension* **2011**, *13*, 687–689.
204. Wu, D.; Piszczek, G. Rapid Determination of Antibody-Antigen Affinity by Mass Photometry. *Journal of Visualized Experiments* **2021**.

 doi:10.20944/preprints202208.0213.v2




205. Mason, D.M.; Friedensohn, S.; Weber, C.R.; Jordi, C.; Wagner, B.; Meng, S.M.; Ehling, R.A.; Bonati, L.; Dahinden, J.; Gainza, P.; Correia, B.E.; Reddy, S.T. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nature Biomedical Engineering* **2021**, *5*, 600–612.
206. Lane, D.P. p53, guardian of the genome. *Nature* **1992**, *358*, 15–16.
207. Sola, I.; Mateos-Gomez, P.A.; Almazan, F.; Zuñiga, S.; Enjuanes, L. RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biology* **2011**, *8*, 237–248.
208. Iida, K.; Itoh, E.; Kim, D.S.; Rincon, J.P.D.; Coschigano, K.T.; Kopchick, J.J.; Thorner, M.O. Muscle mechano growth factor is preferentially induced by growth hormone in growth hormone-deficientlit/litmice. *The Journal of Physiology* **2004**, *560*, 341–349.
209. Zhang, S.; Karthikeyan, R.; Fernando, S.D. Evaluating apoenzyme-coenzyme-substrate interactions of methane monooxygenase with an engineered active site for electron harvesting: a computational study. *Journal of Molecular Modeling* **2018**, *24*.
210. Ham, S.W.; Jeon, H.Y.; Kim, H. Verapamil augments carmustine- and irradiation-induced senescence in glioma cells by reducing intracellular reactive oxygen species and calcium ion levels. *Tumor Biology* **2017**, *39*, 101042831769224.
211. Rhee, S.G. Regulation of Phosphoinositide-Specific Phospholipase C. *Annual Review of Biochemistry* **2001**, *70*, 281–312.
212. Mroczko, B.; Groblewska, M.; Litman-Zawadzka, A.; Kornhuber, J.; Lewczuk, P. Amyloid β oligomers (AβOs) in Alzheimer's disease. *Journal of Neural Transmission* **2017**, *125*, 177–191.
213. Pinheiro, F.; Santos, J.; Ventura, S. AlphaFold and the amyloid landscape. *Journal of Molecular Biology* **2021**, p. 167059.
214. Hubbard, S.R. Insulin meets its receptor. *Nature* **2013**, *493*, 171–172.
215. Cao, J.; Yee, D. Disrupting Insulin and IGF Receptor Function in Cancer. *International Journal of Molecular Sciences* **2021**, *22*, 555.
216. Vigneri, R.; Squatrito, S.; Sciacca, L. Insulin and its analogs: actions via insulin and IGF receptors. *Acta Diabetologica* **2010**, *47*, 271–278.
217. 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.
218. Mervin, L.H.; Johansson, S.; Semenova, E.; Giblin, K.A.; Engkvist, O. Uncertainty quantification in drug design. *Drug Discovery Today* **2021**, *26*, 474–489.
219. Sam, E.; Athri, P. Web-based drug repurposing tools: a survey. *Briefings in Bioinformatics* **2017**, *20*, 299–316.
220. xin Hua, Q.; Nakagawa, S.H.; Jia, W.; Huang, K.; Phillips, N.B.; quan Hu, S.; Weiss, M.A. Design of an Active Ultrastable Single-chain Insulin Analog. *Journal of Biological Chemistry* **2008**, *283*, 14703–14716.
221. Weiss, M. Design of ultra-stable insulin analogues for the developing world. *Journal of Health Specialties* **2013**, *1*, 59.
222. Cheng, R.; Taleb, N.; Stainforth-Dubois, M.; Rabasa-Lhoret, R. The promising future of insulin therapy in diabetes mellitus. *American Journal of Physiology-Endocrinology and Metabolism* **2021**, *320*, E886–E890.
223. Varewijck, A.J.; Janssen, J.A.M.J.L. Insulin and its analogues and their affinities for the IGF1 receptor. *Endocrine-Related Cancer* **2012**, *19*, F63–F75.
224. Zhang, X.; Yu, D.; Sun, J.; Wu, Y.; Gong, J.; Li, X.; Liu, L.; Liu, S.; Liu, J.; Wu, Y.; Li, D.; Ma, Y.; Han, X.; Zhu, Y.; Wu, Z.; Wang, Y.; Ouyang, Q.; Wang, T. Visualization of Ligand-Bound Ectodomain Assembly in the Full-Length Human IGF-1 Receptor by Cryo-EM Single-Particle Analysis. *Structure* **2020**, *28*, 555–561.e4.
225. Zhang, X.; Yu, D.; Wang, T. Cryo-EM structure of the full-length human IGF-1R in complex with insulin, 2020.
226. Gallagher, E.J.; LeRoith, D. Minireview: IGF, Insulin, and Cancer. *Endocrinology* **2011**, *152*, 2546–2551.
227. Sandow, J. Growth effects of insulin and insulin analogues. *Archives of Physiology and Biochemistry* **2009**, *115*, 72–85.
228. Belfiore, A.; Malaguarnera, R.; Vella, V.; Lawrence, M.C.; Sciacca, L.; Frasca, F.; Morrione, A.; Vigneri, R. Insulin Receptor Isoforms in Physiology and Disease: An Updated View. *Endocrine Reviews* **2017**, *38*, 379–431.
229. Wan, L.Y.; Yuan, W.F.; Ai, W.B.; Ai, Y.W.; Wang, J.J.; Chu, L.Y.; Zhang, Y.Q.; Wu, J.F. An exploration of aptamer internalization mechanisms and their applications in drug delivery. *Expert Opinion on Drug Delivery* **2019**, *16*, 207–218.

 doi:10.20944/preprints202208.0213.v2




230. Jin, Y.; Schladetsch, M.A.; Huang, X.; Balunas, M.J.; Wiemer, A.J. Stepping forward in antibody-drug conjugate development. *Pharmacology & Therapeutics* **2022**, *229*, 107917.
231. Steiner, R. History and progress on accurate measurements of the Planck constant. *Reports on Progress in Physics* **2012**, *76*, 016101.
232. Pitre, L.; Plimmer, M.D.; Sparasci, F.; Himbert, M.E. Determinations of the Boltzmann constant. *Comptes Rendus Physique* **2019**, *20*, 129–139.
233. VIGOUREUX, P. Gyromagnetic Ratio of the Proton. *Nature* **1963**, *198*, 1188–1188.
234. Kibble, B.P.; Hunt, G.J. A Measurement of the Gyromagnetic Ratio of the Proton in a Strong Magnetic Field. *Metrologia* **1979**, *15*, 5–30.
235. Li, S.; Wang, Z.; Jia, X.; Niu, T.; Zhang, J.; Yin, G.; Zhang, X.; Zhu, Y.; Ji, G.; Sun, F. ELI trifocal microscope: a precise system to prepare target cryo-lamellae for in situ cryo-ET study. *Nature Methods* **2023**, *20*, 276–283.
236. Eisenstein, M. Catching proteins at play: the method revealing the cell's inner mysteries. *Nature* **2023**, *621*, 646–648.
237. Marsh, D. *Handbook of Lipid Bilayers*; CRC Press, 2013.
238. Kumar, S.; Nussinov, R. Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **2002**, *3*, 604–617.
239. Box, K.; Bevan, C.; Comer, J.; Hill, A.; Allen, R.; Reynolds, D. High-throughput measurement of pKa values in a mixed-buffer linear pH gradient system. *Analytical chemistry* **2003**, *75*, 883–892.
240. Avdeef, A. pH-metric log P. II: Refinement of partition coefficients and Ionization constants of multiprotic substances. *J. Pharm. Sci.* **1993**, *82*, 183–190.
241. Ishihama, Y.; Nakamura, M.; Miwa, T.; Kajima, T.; Asakawa, N. A rapid method for pKa determination of drugs using pressure-assisted capillary electrophoresis with photodiode array detection in drug discovery. *J. Pharm. Sci.* **2002**, *91*, 933–942.
242. 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.
243. Pelton, J.; Torchia, D.; Meadow, N.; Roseman, S. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Science* **2008**, *2*, 543–558.
244. Bezencon, J.; Wittwer, M.B.; Cutting, B.; Smiesko, M.; Wagner, B.; Kansy, M.; Ernst, B. pKa determination by (1)H NMR spectroscopy-an old methodology revisited. *J. Pharm. Biomed. Anal.* **2014**, *1*, 147–155.
245. Cai, Z.; Luo, F.; Wang, Y.; Li, E.; Huang, Y. Protein pKa Prediction with Machine Learning. *ACS Omega* **2021**, *6*, 34823–34831.
246. Xiong, J.; Li, Z.; Wang, G.; Fu, Z.; Zhong, F.; Xu, T.; Liu, X.; Huang, Z.; Liu, X.; Chen, K.; Jiang, H.; Zheng, M. Multi-instance learning of graph neural networks for aqueous pKa prediction. *Bioinformatics* **2021**, *38*, 792–798.
247. Costa, B.; Pozzo, E.D.; Giacomelli, C.; Barresi, E.; Taliani, S.; Settimo, F.D.; Martini, C. TSPO ligand residence time: a new parameter to predict compound neurosteroidogenic efficacy. *Scientific Reports* **2016**, *6*.
248. Copeland, R.A. The drug-target residence time model: a 10-year retrospective. *Nature Reviews Drug Discovery* **2015**, *15*, 87–95.
249. Zimmermann, A.; Sutter, A.; Shooter, E.M. Monoclonal antibodies against beta nerve growth factor and their effects on receptor binding and biological activity. *Proceedings of the National Academy of Sciences* **1981**, *78*, 4611–4615.
250. Enomoto, M.; Mantyh, P.W.; Murrell, J.; Innes, J.F.; Lascelles, B.D.X. Anti-nerve growth factor monoclonal antibodies for the control of pain in dogs and cats. *Veterinary Record* **2019**, *184*, 23–23.
251. Kalnins, G. mature recombinant horse NGF, 2020.
252. Sheng, X.; Ye, D.; Zhou, A.; Yao, X.; Luo, H.; He, Z.; Wang, Z.; Zhao, Y.; Ji, Z.; Zou, Q.; He, C.; Guo, J.; Tu, X.; Liu, Z.; Shi, B.; Liu, B.; Chen, P.; Wei, Q.; Hu, Z.; Zhang, Y.; Jiang, K.; Zhou, F.; Wu, D.; Fu, C.; Li, X.; Wu, B.; Wang, L.; Qin, S.; Li, G.; Liu, Y.; Guo, H.; Chen, K.; Zhang, D.; Wang, G.; Ding, L.; Wang, Y.; Yuan, X.; Guo, J. Efficacy and safety of vorolanib plus everolimus in metastatic renal cell carcinoma: A three-arm, randomised, double-blind, multicentre phase III study (CONCEPT). *European Journal of Cancer* **2023**, *178*, 205–215.

253. Urva, S.; O'Farrell, L.; Du, Y.; Loh, M.T.; Hemmingway, A.; Qu, H.; Alsina-Fernandez, J.; Haupt, A.; Milicevic, Z.; Coskun, T. The novel GIP, GLP-1 and glucagon receptor agonist retatrutide delays gastric emptying. *Diabetes, Obesity and Metabolism* **2023**, *25*, 2784–2788.
254. Arkin, M.R.; Tang, Y.; Wells, J.A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chemistry & Biology* **2014**, *21*, 1102–1114.
255. Erlanson, D.A.; Braisted, A.C.; Raphael, D.R.; Randal, M.; Stroud, R.M.; Gordon, E.M.; Wells, J.A. Site-directed ligand discovery. *Proceedings of the National Academy of Sciences* **2000**, *97*, 9367–9372.
256. Gonzalez, K.J.; Huang, J.; Criado, M.F.; Banerjee, A.; Tompkins, S.; Mousa, J.J.; Strauch, E.M. A general computational design strategy for stabilizing viral class I fusion proteins **2023**.
257. Shi, W.; Cai, Y.; Zhu, H.; Peng, H.; Voyer, J.; Rits-Volloch, S.; Cao, H.; Mayer, M.L.; Song, K.; Xu, C.; Lu, J.; Zhang, J.; Chen, B. Cryo-EM structure of SARS-CoV-2 postfusion spike in membrane. *Nature* **2023**, *619*, 403–409.
258. Wang, Y.; Xiong, J.; Xiao, F.; Zhang, W.; Cheng, K.; Rao, J.; Niu, B.; Tong, X.; Qu, N.; Zhang, R.; Wang, D.; Chen, K.; Li, X.; Zheng, M. LogD7.4 prediction enhanced by transferring knowledge from chromatographic retention time, microscopic pKa and logP. *Journal of Cheminformatics* **2023**, *15*.
259. Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics* **2009**, *1*.
260. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3-25. 1. *Advanced Drug Delivery Reviews* **2001**, *46*, 3–26.
261. Zheng, M.; Liu, X.; Xu, Y.; Li, H.; Luo, C.; Jiang, H. Computational methods for drug design and discovery: focus on China. *Trends in Pharmacological Sciences* **2013**, *34*, 549–559.
262. Zheng, M.; Zhao, J.; Cui, C.; Fu, Z.; Li, X.; Liu, X.; Ding, X.; Tan, X.; Li, F.; Luo, X.; Chen, K.; Jiang, H. Computational chemical biology and drug design: Facilitating protein structure, function, and modulation studies. *Medicinal Research Reviews* **2018**, *38*, 914–950.
263. Zhou, S.F.; Zhong, W.Z. Drug Design and Discovery: Principles and Applications. *Molecules* **2017**, *22*, 279.
264. Subramaniam, S.; Kleywegt, G.J. A paradigm shift in structural biology. *Nature Methods* **2022**, *19*, 20–23.
265. Bonvin, A.M.J.J. 50 years of PDB: a catalyst in structural biology. *Nature Methods* **2021**, *18*, 448–449.
266. Zhang, X.; Gao, H.; Wang, H.; Chen, Z.; Zhang, Z.; Chen, X.; Li, Y.; Qi, Y.; Wang, R. PLANET: A Multi-objective Graph Neural Network Model for Protein-Ligand Binding Affinity Prediction. *Journal of Chemical Information and Modeling* **2023**.
267. 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.
268. 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.
269. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S.A.A.; Ballard, A.J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A.W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
270. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.R.; Smith, J.C.; Kasson, P.M.; van der Spoel, D.; others. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 1–10.
271. Kendon, V. Quantum computing using continuous-time evolution. *Interface Focus* **2020**, *10*, 20190143.
272. Cookson, C. Taking **bold** bets: new UK agency prepares to fund breakthrough technologies, 2023. Accessed: (August 30, 2023).