Review

Application of Computational Biology and Artificial Intelligence in Drug Design

Yue Zhang [1,2,3,†], Mengqi Luo [1,4,†], Peng Wu [5], Song Wu [4], Tzong-Yi Lee [1,3,*] and Chen Bai [1,3,*]

1 School of Life and Health Sciences, School of Medicine, The Chinese University of Hong Kong, Shenzhen 518172, China
2 School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026, China
3 Warshel Institute for Computational Biology, Shenzhen 518172, China
4 South China Hospital, Health Science Center, Shenzhen University, Shenzhen 518116, China
5 School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518055, China
* Correspondence: leetzongyi@cuhk.edu.cn (T.-Y.L.); baichen@cuhk.edu.cn (C.B.)
† These authors contributed equally to this work.

Abstract: Traditional drug design requires a great amount of research time and developmental expense. Booming computational approaches, including computational biology, computer-aided drug design, and artificial intelligence, have the potential to expedite the efficiency of drug discovery by minimizing the time and financial cost. In recent years, computational approaches are being widely used to improve the efficacy and effectiveness of drug discovery and pipeline, leading to the approval of plenty of new drugs for marketing. The present review emphasizes on the applications of these indispensable computational approaches in aiding target identification, lead discovery, and lead optimization. Some challenges of using these approaches for drug design are also discussed. Moreover, we propose a methodology for integrating various computational techniques into new drug discovery and design.

Keywords: computational biology; computer-aided drug design (CADD); artificial intelligence-aided drug design (AIDD); deep learning

1. Introduction

Drug research and development is a multistep process that includes drug discovery, clinical testing, and approval for production (Figure 1). Drug discovery is a lengthy, expensive, and complicated process that spans years and costs millions of dollars [1,2]. This process consists of target identification, lead discovery, lead optimization, and preclinical testing (Figure 1) [3–5]. Traditional drug discovery begins with the identification of a specific disease, suitable target identification, effective molecule identification (including molecular synthesis and bioactivity testing), and preclinical testing. Despite investing large amounts of money and time, the success rate of clinical testing is below 15% [6]. The cause of failure in approximately 50% of drug discovery is poor pharmacokinetic properties (absorption, distribution, metabolism, excretion, and toxicity [ADMET]) [7]. The speed and success rate of drug discovery have tremendously increased with the development of computational approaches [8].

Nowadays, computational and deep learning approaches play an increasingly vital role in drug discovery. The fast evolvement of methods and algorithms has shortened the time and financial costs in finding the drug candidates.

In drug discovery, contributions of computational biology include the characterization of ligand-binding molecular mechanisms, the identification of binding/active sites and structure refinement of binding poses of the ligand-target. Most of these approaches indicate that binding/active sites on the target protein should be well determined. Specific

residues of these binding sites could be used to guide the modification and optimization of the initial lead compound and generate new ligand–target protein interactions. In some cases, engagement of the active site is inadequate for exploring the pathologic activity. Mutations away from the active site, conformational transitions, drug resistance, and expression levels are also known to induce pathosis. Computational biology, especially biomacromolecular simulation, is a powerful method for revealing the molecular mechanism of the target protein and providing new perspectives for drug design.
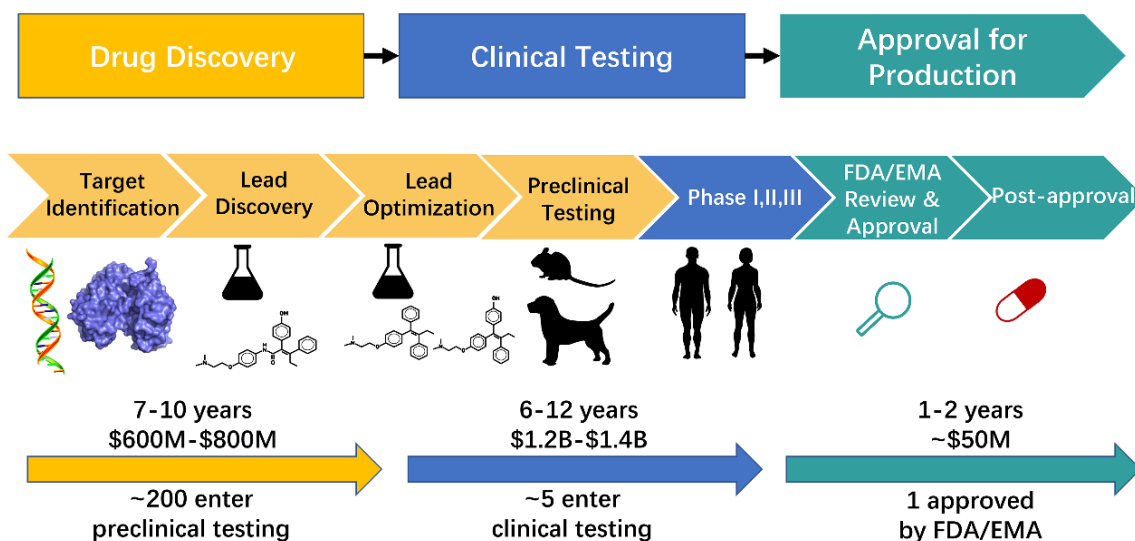

Figure 1. The process of drug research and development. The details in drug development have been improved in the past forty years. Nowadays, the complete process in drug research includes drug discovery, clinical testing, and approval for production. The process of drug discovery includes target identification, lead discovery, lead optimization, and preclinical testing. This process usually takes 7–10 years and $600 M–$800 M. Then approximately 200 compounds enter preclinical testing step while about 5 compounds enter clinical testing process. This process includes three steps, phase I, II and III clinical trials, respectively. It is a long and expensive process that costs 6–12 years and billions of dollars. The compounds that have passed clinical testing enter the process of approval for production. Approved compounds by FDA/EMA can commercialize on the market. This process takes 1–2 years and about $50 M.

According to Newtonian Mechanics, molecular dynamics (MD) simulations, which have been widely employed in drug discovery, can capture the position and motion of each atom in a system [9]. This approach can reveal the details of binding, unbinding, and conformational changes of the target protein, which provides complementary information to experiments [10]. Moreover, MD simulations can provide the thermodynamics, kinetics, and free energy profiles of target–ligand interactions [11]. This information can be useful in improving the binding affinity of the lead compound [9]. Due to the availability of more reliable binding affinity results, MD simulations are used to validate the accuracy of docking results [12]. Moreover, quantum mechanics (QM) approaches, such as density functional theory (DFT) [13,14] and ab initio calculation methods [14,15] can be applied to virtual screening (VS) by exploring atomic-electronic interactions between the ligand and target [16]. But these QM approaches are computationally extremely expensive and not always applied to VS in industry.

Computer-aided drug design (CADD) is typically used to discover, develop, and analyze drug candidates and active molecules having similar biochemical properties [4,17–20]; accelerate drug discovery, and reduce costs and failure rates [21]. CADD-discovered drug candidates are usually from small-molecule libraries. These discoveries are made using various methods, including molecular docking, pharmacophore modeling, VS, and quantitative structure–activity relationship (QSAR) [4,17–20]. Among these approaches, VS

is the major contributor applied to screen new hit compounds with required properties from large chemical databases. VS is classified as structure-based VS (SBVS) [22,23] and ligand-based VS (LBVS) [24,25]. VS is used to accelerate drug discovery and shorten the number of compounds to be tested in the wet lab. Additionally, VS also plays an important role in drug repurposing or repositioning, optimizing the drug candidates quickly, which accelerates the process of drug design and development [26,27].

Artificial intelligence (AI) has recently been proposed as a promising technique in learning and discovering pharmacological big data in drug discovery that has boosted the success rates of drug identification [28]. Using the extensive datasets from biomedical research, AI can learn and discover further rules for translating the data into accessible knowledge. Leading pharmaceutical companies have applied AI to enhance the efficacy of their drug candidates, thereby saving time and costs on unnecessary synthesis and tests. Machine learning [29–31], a subfield of AI, and its subfield, deep learning [32–34], have been combined with the VS process [35–37] to improve the efficiency of similarity searching and the reliability of mining screening data in the ligand-based VS process and enhance the accuracy of scoring functions in structure-based VS [36–38]. The approaches also contribute to the generation of novel compounds [32,34].

This review discusses the application of powerful computational approaches to drug discovery and overviews various computational techniques applied to drug discovery including MD simulations, developed Coarse-Grained models, QM, molecular docking, VS, pharmacophore modeling, QSAR, machine learning, and deep learning. Additionally, it discusses how to exploit various computational techniques along with VS to extend the chemical space of novel lead compounds and accelerate the process of drug discovery.

2. Computational Biology in Drug Design

Modern drug discovery begins with target identification. Various approaches at the crossroads such as structural biology, molecular biology, cell biology, genomics, proteomics, computational biology, and bioinformatics [39–45] are being explored to identify the target or related targets and investigate the pathogenesis. Understanding the pathogenesis is also vital for drug discovery and therapies. Computational chemistry techniques, such as MM, QM and MD simulations are widely used in computational biology and medicinal chemistry. MD simulations, DFT, and QM are efficient methods for exploring the pathogenic mechanism and drug resistance [39,46–57]. We herein summarize the applications of these methods in drug discovery, specifically in the study of pathogenic mechanism, molecular docking, and lead optimization.

2.1. Application of Molecular Mechanics in Drug Design

Molecular mechanics (MM) is an approach which approximately treats the molecules with the laws of classical mechanics and saves the computational resources required for quantum mechanical calculations [58]. Over the past decades, MM approach plays an important role in understanding the ligand-protein structures, interactions and optimizing leads. It is achieved by MM potential energy function, which represents the sum of different energy terms, referred as "force fields" [59]. MM potential energy functions are used in various sampling methods, such as MD and MC (Monte Carlo). MD simulations are often utilized in drug discovery [60,61]. MD is one of the most popular algorithms for sampling. It utilizes various integration algorithms, such as Verlet's Algorithm, Leap-frog Algorithm and Beeman's Algorithm, to interpret classical Newton's equation of motion to analyze the trajectories, movements and interactions in a given molecular system [61]. Time-dependent properties can be obtained from MD [62]. The system is generally a biomacromolecule, such as a protein for example an enzyme, with a solvent environment. For this protein or enzyme system, the initial protein structure is resolved by experiments [63]. Then, the structure could be modelled by different methods. After that, simulations start with the prepared model. X-ray crystallography is used as an experimental method for obtaining the three-dimensional protein structure [64]. However, X-ray requires the protein to form stable

crystals and the crystal quality determines the resolution of the structure, which limits the obtainment of high-quality protein structures, especially of membrane proteins [65]. Cryo-EM addresses the problem without the need to form crystals [66,67]. Cryo-EM can determine even quite unstable and intractable membrane protein structures [67]. However, Cryo-EM is not a panacea. In cryo-EM, sample quality is still the most critical factor for determining the high-resolution structure [54]. If no experimental structure is available, modeling or predicting the structure is necessary. Homology modeling [68,69] and AlphaFold developed by DeepMind [70–72] are preferred techniques for acquiring the initial protein structure. In the molecular dynamics simulation, atoms and molecules of the system interact during the fixed time, providing the dynamic features of the system. Atom trajectories are generally determined by Newton's laws of motion. Molecular mechanics methods with various force fields [73–78] are employed to calculate the energies of the system.

2.1.1. Application in Investigating the Mechanism of the Target Protein

Target protein can be regulated by drugs to cure the disease or relieve the symptoms. The overall process is dynamic and usually accompanied by the conformational changes of the target protein. Target protein conformation has an essential role in drug design. Even minor changes, as well as the motions of residues, may affect the target–ligand interactions. MD simulations can provide dynamic information about the target protein and the ligand in terms of drug design, which cannot be obtained through experimental methods. Compared with experiments, MD simulations can provide detailed information about the target protein folding process and describe the conformational changes of the protein with environmental changes such as temperature, pH, and residue mutations, with detailed energetic information. At present, MD simulations have been broadly applied to study the molecular mechanisms of the target protein to aid drug design.

For example, Horikoshi et al. revealed the molecular mechanisms underlying the loss of activity in the most severe glucose-6-phosphate dehydrogenase (G6PD) deficiency [79]. It is triggered by the long-distance propagation of structural defects at the dimer interface. The findings indicated that a promising drug can possibly be discovered and developed by correcting these structural defects. While studying pathogenic mutations in the kinesin-3 motor KIF1A by using MD simulations, Budaitis et al. found that these mutations were linked to neurodevelopmental and neurodegenerative disorders that impaired neck linker docking and dramatically reduced the KIF1A force generation [80]. Zanetti-Domingues's work revealed autoinhibition mechanisms in dimers and oligomers of the epidermal growth factor receptor (EGFR) and supported that dysregulated species bear populations of symmetric and asymmetric kinase dimers coexisting in an equilibrium [81]. The structural feature of the assembly inspires the related drug design. Based on MD simulations, Zhu's lab elaborated on the genotype-determined EGFR-RTK heterodimerization and its drug resistance mechanism in lung cancer caused by a tighter EGFR-RTK crosstalk [82]. The study promotes drug design against the tighter crosstalk of the genotype determined. Understanding the dynamic behaviors of sirtuins, which have several ligand-binding sites [83], may provide perspectives for the design of selective inhibitors or activators. Polymyxin resistance was found to be induced by lipopolysaccharides and outer membrane vesicles in the multidrug-resistant Pseudomonas aeruginosa [84]. Based on this mechanism, an intelligent strategy for designing lipopeptide antibiotics against polymyxin resistance was developed [84]. The strategy may be suitable for the discovery of other classes of bacterial pathogen-targeting antibiotics. In addition to regular MM approach, coarse-grained models can be used to investigate the molecular mechanism of the target. More details are shown in Section 2.1.4.

2.1.2. Application in Molecular Docking

In molecular docking, according to the space complementarity and energy match, compounds are docked in the specific site. Then, the docking poses are scored and ranked based on the scores [85]. On the basis of molecular docking, VS has been indispensable to

drug discovery [86,87]. VS saves time and costs for drug discovery and efficiently obtains various molecule scaffolds [88–92]. The complete VS process consists of pre-processing compound libraries, molecular docking, and the selection of pretest compounds [22,93,94]. In general, the enrichment factor greatly determines the success of VS. The enrichment factor is a validation tool that evaluates the effectiveness of VS by computing the ratio of active molecules among the tested molecules in the initial library. For each VS step, different strategies are used to enhance the enrichment factor [38,95]. The VS results depend on the rationality of the docking poses between the target and ligand, and the accuracy of binding affinity [39,96–103].

After VS step, filtering promising candidates may then be sampled with MD simulations. The use of MD simulations can improve the flexibility in conformational sampling, which increases the number of degrees of freedom of the system and consequently in the computational effort [104]. For MM MD simulations, one of the most time-consuming parts is the calculation of the interactions between each atom in the system, which cost more than 90% of the total simulation time. This is mainly related to the calculation giving rise to O(N^2) computational complexity (N represents the number of atoms in the system) [105–107]. The cutoff method applied to treat the interactions between atoms can reduce the computational complexity to O(N) [105,106]. Compared with the three-dimensional structure of the target protein obtained through X-ray or Cryo-EM, MD simulations take the flexibility of the target protein into account. The experimental structure is in the specific crystalized condition, which is possibly different from the real binding conformation with the ligand. A set of conformations can be obtained by modeling and simulations, especially the crucial intermediate or transition state that may contribute to the ligand–target protein binding process. MD simulations used to sample the specific conformation can provide more rational docking poses and achieve a higher enrichment factor. In addition to the conformation optimization of the target and ligand, MD simulations combined with binding free energy calculations are applied to assess the binding affinity of the ligand with the target. MM-PBSA and MM-GBSA are general approaches used to calculate binding free energy. Based on the trajectories from MD simulations, electrostatic energy, solvation energy, and van der Waals energy are calculated. Entropy change can be obtained through normal mode analysis. Then, the binding free energy can be obtained [108]. The binding free energy calculations are great and useful to augment the accuracy of the binding affinity of docking poses and improve the enrichment factor. But these high-cost sampling calculations are often used on an even smaller subset of potential hits.

2.1.3. Application in Lead Optimization

The optimal binding mode and the accurate binding affinity are vital for understanding the ligand–target interactions and guiding the modification of screened compounds. The ligand–target thermodynamical data, such as entropy change (ΔS) and free energy change (ΔG), can be determined through experiments and are used to distinguish between active and inactive compounds. However, the lack of details about target–ligand interactions limits further structural modifications of the compounds. MD simulation is a powerful approach for precisely evaluating the ligand–target binding modes. It can describe the detailed ligand–target interactions and determine the free energy contribution of each residue in the binding sites. The information can provide guidance for lead optimization.

Using the combination of MD simulations and VS, Patel's lab optimized bedaquiline to decrease the severity of its adverse side effects and discovered that the compound CID 15947587 with low cardiotoxicity has the highest binding free energy (−85.27 kcal/mol) and an optimal docking score (−5.64) with mycobacterial ATP synthase enzyme [109]. Castillo's group optimized AKT inhibitors by using MD simulations, thereby improving the binding affinity of the 2,4,6-trisubstituted pyridine scaffold in the ATP pocket of PKB/AKT and interacting well with glutamates/aspartates in ATP-binding sites [110]. Zhang et al. screened the new inhibitor against phosphodiesterase-2A (PDE2A). With the guidance of MD simulations, they obtained the optimized lead, LHB-8, forming an extra hydrogen

bond with D808 and a hydrophobic interaction with T768, in addition to the interactions with Q859 and F862 of the PDE2A target [111].

2.1.4. Application of Coarse-Grained Models in Drug Design

All-atom MD simulations present the limitation while exploring the dynamic process of the large-scale target protein or long-time scale. Coarse-grained (CG) models help overcome the limitation well. When using CG models, the main chain of residues is in the all-atom state, but the side chain is a simplified united atom. Compared with all-atom MD simulations, CG simulations decrease the number of particles and make the potential energy surface smoother. Thus, the longer time and larger scale are available using CG models. Martini is a classical force filed to employ CG simulations. Martini is currently applied to study the mechanism and oligomerization of membrane proteins and self-assembly of proteins, predict conformational changes, and study binding and pore formation in membranes [112–114].

The CG model consistently developed by Warshel et al. [77,78,115] is advantageous in investigating the molecular mechanism of different biophysical systems, such as SARS-CoV-2 [116], GPCR [117], ATPase [118,119], and kinase [120]. This model can accurately describe the electrostatic term [121], which usually is the major contributor compared with other types of interactions in proteins. The CG profile can determine the dynamic information of the reaction in proteins, including the reaction energy barrier, rate-determining step, and the transition state. These results offer energetic details for understanding the working mechanism of proteins and guide rational drug discovery and development.

We are currently attempting to apply the CG model to VS to obtain more effective compounds. CG simulations can provide vital dynamic information and details about the transition state and energy barrier. The transition state is induced in the molecular docking of VS, and the energy barrier is considered a rule in the score and rank. Then, the selected compounds are used to construct the training set of AI to generate new molecules.

2.2. Application of QM in Drug Design

Structural studies have shown that the details of the potential drug target are valuable not only for lead discovery and optimization but also for the later steps of drug design, such as toxicity determination and prediction of binding modes of the leads and drug targets. During drug discovery, the molecular docking or pharmacophore model is used for predicting the binding modes in a short time. MD simulations can be employed to obtain flexible and rational docking modes. They can also guide drug design by exploring ligand–target interactions, such as studying the active site for extra electrostatic, hydrophilic, or hydrophobic interactions that can increase binding affinity [39,122,123]. Although MD simulations improve the accuracy of scoring and docking [124–126], concerns still exist, especially in enzymes or metal-containing drug targets, in which valence electron transfer occurs [127,128].

QM is considered the potential solution for the aforementioned concerns, which can explore drug target details at the electronic level [52,123,128]. At present, QM is increasingly applied to enzymes or metal-containing proteins that are considered drug targets, such as HIV-1 protease [129], human DHFR [130], and GPCR [131], and clarify the molecular mechanism for drug design [132–135]. QM is also used for designing novel drugs, including the high-affinity ligands of FKBP12 [136] and novel inhibitors of human DHFR [137].

Additionally, researchers have attempted to improve scoring by inducing QM approaches, especially QM-polarized ligand docking [138], and QMScore, a semiempirical QM (SQM) scoring function [139]. QM in combination with molecular mechanics (MM) has been developed to enhance the accuracy of docking and VS [128,140–142]. Fong et al. applied a series of QM/MM scoring functions to six HIV-1 proteases and found that parts of QM/MM functions were superior to MM functions [143]. Kim et al. [144] proposed a new strategy of using QM/MM with the implicit solvation model to rescore docking poses and improve the docking performance on 40 GPCR–ligand complexes. A significant im-

provement was observed over the conventional docking method. Chaskar et al. developed a QM/MM on-the-fly docking method to deal with polarization and metal interactions in docking and observed a significant improvement in scoring [145]. Compared to MD simulations, QM calculations are even more expensive. For example, the Hartree-Fock recovers approximately 99% of the total electronic energy and requires diagonalizing the M by M Fock matrix at O(M^3) cost (M represents the number of basis functions) [146]. By Shor's factoring algorithm, the complexity of quantum calculations is O($(\log_2 N)^3$) (N represents the number of atoms in the system) [147]. Moreover, QM calculations are restricted to systems of up to a few hundred atoms in contrast to MD simulations, which has evolved from simulating tens of thousands of atoms to handling over 100 million atoms comprising an entire cell organelle [148,149].

3. Computer-Aided Drug Design

CADD has until now led to the discovery of more than 70 approved drugs [4], from Captopril in 1981 [150,151] to Remdesivir in 2021 [152]. Two important categories of CADD, structure-based drug design (SBDD) and ligand-based drug design (LBDD), are highlighted in this review. These two categories have been widely used in lead discovery during drug discovery (Figure 2). SBDD depends on the three-dimensional structure of the target and active sites to determine ligand–target interactions [153]. On the other hand, LBDD is used when the three-dimensional structure of the target is unknown. It begins with a single molecule or a set of molecules effective against the target and depends on the structure–activity relationship [153].
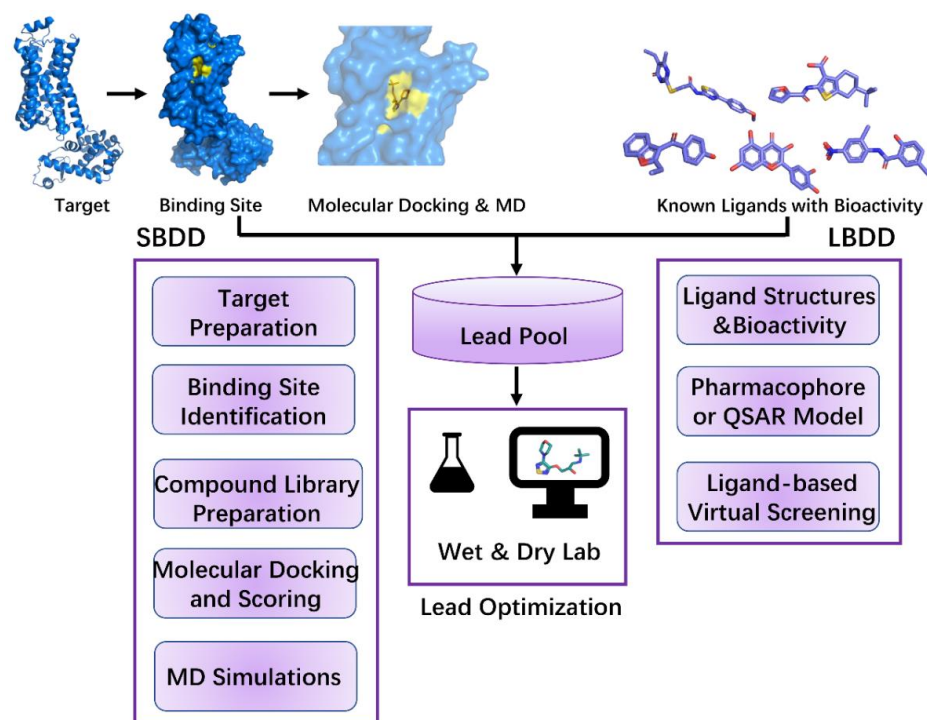

Figure 2. The workflow of structure-based drug design (SBDD) and ligand-based drug design (LBDD). For SBDD, it starts with target identification. Then, binding site of the target requires identifying and compound library needs to be prepared. Next, dock each compound from the library into the identified binding site evaluate the score. In molecular docking, MD simulations can be utilized to obtain more flexible target and rescore for the docking process. Additionally, MD simulations can be applied to lead optimization through ligand-target interactions. Through these steps, leads are obtained primarily. For LBDD, it starts with known ligands with bioactivity. Then, extracting the chemical features of these ligands and build pharmacophore or QSAR model. Next, according to the information of known ligands (e.g., ligand similarity), ligand-based virtual screening is performed in the compound library and leads are screened. These leads are further optimized in wet and dry lab.

3.1. Structure-Based Drug Design

SBDD is an efficient approach for lead discovery and optimization. The most frequently used methods in SBDD, that is, MD simulations, molecular docking, and structure-based VS, are applied to evaluate binding affinity and ligand–target interactions and explore conformational changes in the target. Using SBDD, some approved drugs, such as Imatinib (an abltyrosine kinase inhibitor) [154], Indinavir (Crixivan, the inhibitor of HIV-1 protease) [155], Nilotinib (Tasigna, a selective tyrosine kinase receptor inhibitor used in the treatment of chronic myelogenous leukemia) [156], and Lifitegrast (the LFA-1 antagonist that blocks binding of ICAM-1 to LFA-1) [157], were discovered. SBDD mainly includes target preparation, binding site identification, compound library preparation, molecular docking and scoring, and MD simulations (Figure 2).

3.1.1. Target Preparation

With advancements in structural biology, increasing structures of the target proteins are available and deposited in the PDB. Because of the limitations of experimental approaches, some target structures have not been obtained yet [5,158,159]. Computational approaches such as homology modeling, AlphaFold, and ab initio protein structure prediction can predict target structures according to their sequences [68,69,71,72,160]. Homology modeling selects an appropriate template structure to construct the target structure. AlphaFold is an AI technique developed by DeepMind that predicts three-dimensional protein structures according to their amino acid sequences, which can achieve approximate accuracy as experiments [72]. The ab initio protein structure prediction is considered suitable when the template structure is unavailable in the PDB. This technique considers global optimization, which can help find the tertiary structure with minimum energy based on the primary structure of the specific target [160].

3.1.2. Binding Site Identification

Binding site determination is an essential prerequisite for performing molecular docking. Information about the binding sites of target proteins can be obtained from site-directed mutation and the co-crystallized complex structures of proteins with ligands [161]. When any prior knowledge of the binding pocket is unknown, blind blocking is required to predict the binding sites [162–164]. For blind docking method, docking has to be performed on the entire protein surface to find the most probable binding mode. The whole process includes several trials (>100 times recommended by Hetenyi, and Van Der Spoel [165]) and energy evaluations (at least 10 million times per trial recommended by Hetenyi, and Van Der Spoel [165]) to obtain the favorable ligand-target complex pose [163–165]. Compared with regular docking, although blind docking is less reliable and limited due to inadequate sampling at the docking space, blind docking is meaningful to discover unexpected interactions that may exist in the unidentified binding modes [164,166,167]. Some tools are developed to predict the binding sites of target proteins by blind docking, including DeepSite [168], DoGSiteScorer [169], POCASA [170], Fpocket [171,172], RaptorX-Binding Site [173], COACH [174], and PocketDepth [175].

3.1.3. Compound Library Preparation

Compounds used for VS are selected from compound libraries such as the REAL library of Enamine (1.4 billion make-on-demand compounds) [176], ZINC (750 million purchasable compounds in ready-to-dock) [89,177,178], MCULE (122 million synthetic compounds) [179], PubChem (112 million bioactive compounds) [180,181], DrugBank (14,528 approved drug molecules or experimental drugs) [182,183], ChEMBL (approximately 2.2 million bioactive molecules with drug-like properties) [184,185], and ChemDB (approximately 5 million commercially available small molecules) [186]. The compounds were filtered on the basis of Lipinski's "Rule of Five" [187], Veber criteria [188], ADMET, and other specific properties (such as carcinogenicity and hepatotoxicity) [158]. Lipinski's "Rule of Five" and Veber criteria indicate that the compound can be recognized to be orally

bioactive if its molecular weight (MW) is <500 Da, hydrogen bond donors (HBD) $\leq$ 5, hydrogen bond acceptors (HBA) $\leq$ 10, octanol–water partition coefficient logP $\leq$ 5, rotatable bonds (RotB) $\leq$ 10, and topological polar surface area $\leq$ 140 [187,188]. Moreover, the synthetic accessibility of the compounds should be considered. After filtering ligands from libraries, the optimized 3D structure of the ligand should be modelled.

3.1.4. Molecular Docking and Scoring

Molecular docking is currently used in combination with VS to simplify the search process in the presence of a three-dimensional target structure [87]. It is used to assess ligand–target interactions at the molecular level and rank the ligands according to their binding affinity by using scoring functions [189]. The most frequently used molecular docking tools include Autodock [190], AutoDock Vina [191], CDOCKER [192], GLIDE [193], DOCK6 [194], GOLD [195], FLEXX [196], and SwissDock [197]. Regarding the flexibility of the ligand and target, molecular docking approaches include (1) rigid docking wherein the structures of the ligand and target are both rigid; (2) semi-flexible docking, which is the most commonly used approach, wherein the ligand structure is flexible and the target is rigid; and (3) flexible docking wherein the ligand and target structures are both flexible [198]. Different search algorithms are applied to deal with flexible ligands, such as systematic search algorithms, random or stochastic algorithms, and simulation algorithms [199]. To treat the flexible protein, molecular dynamic methods and Monte Carlo methods are usually applied [200,201]. The accuracy of molecular docking relies on scoring functions, which are applied to determine binding affinity and ligand–target binding modes and identify the potential drug candidates [202]. Physics-based, empirical, knowledge-based, and machine learning-based scoring functions are available [202]. Additionally, new deep learning methods such as EquiBind, GNINA, DiffDock are developed to predict the binding mode between the ligand and a specific protein target [203–205]. Especially, Equibind and Diffdock have the potential to significantly change the VS landscape. EquiBind, an SE(3)-equivariant geometric deep learning model, can perform direct-shot prediction of the receptor binding location (blind docking) and the ligand binding pose and orientation [203]. This method significantly speeds up with better quality compared to traditional docking methods [203]. DiffDock is a diffusion generative model over the non-Euclidean manifold of ligand poses, which has fast inference times and provides confidence estimates with high selective accuracy outperforming the previous traditional docking and deep learning methods [205]. GNINA utilizes an ensemble of convolutional neural networks (CNNs) as a scoring function and improves the quality of scoring and ranking binding poses for protein-ligand complexes [204]. This method significantly outperforms SMINA/Vina in all cases including redocking, cross-docking, flexible docking, and whole protein docking tasks [204].

3.1.5. MD Simulations

MD simulations have been extensively used in SBDD. In molecular docking, MD simulations can improve the flexibility of the target protein and obtain target conformations with well-defined binding cavities and flexibility for molecular docking [206,207]. Moreover, MD simulations can be applied for docking scoring and lead optimization. Combined with free energy calculations, MD simulations can accurately assess binding affinity and improve the accuracy of ranking the compounds. In lead optimization, MD simulations can be employed on the small sets of compounds (no more than a few hundred), and ligand–target interactions can be determined, which provide guidance for the further development of ligands [208]. More details about MD simulations are provided in Section 2.1 MD Simulations.

3.2. Ligand-Based Drug Design

In drug discovery, target structures may not be available, but some compounds against the specific target may be known. In this situation, LBDD is applied. LBDD begins with a

single compound or a set of active compounds against the specific target protein. Then, compounds with physicochemical and structural properties responsible for the given biological activity are identified, which is based on structural similarities related to similarities of biological activities [5,209]. According to structure–activity relationships (SARs), the properties of the compounds are improved by designing appropriate analogs [209]. Designing can be performed in terms of structural similarity or properties. Commonly used approaches in LBDD include pharmacophore modeling and QSAR.

3.2.1. Pharmacophore Modeling

Chemical features from a set of bioactive conformations of known ligands were extracted to employ the pharmacophore model. These conformations contain information about the vital interactions of ligands with the specific target [5,210]. The chemical features comprise hydrogen bond acceptors/donors, hydrophobic regions, positively/negatively charged groups, and aromatic ring regions [211]. Generation of the pharmacophore model generally includes the following steps: [212] (1) selecting a set of bioactive ligands against the specific target as the training set; (2) creating the conformation space for each ligand in the training set to characterize the conformation flexibility of the ligand; and (3) aligning the ligands in the training set and determining the chemical features to construct the pharmacophore model. Various pharmacophore model generators have been developed, such as Catalyst [213], LigandScout [214], MOE [215], PharmMapper [216], PharmaGist [217], Phase [218], Quasi [219], and UNITY [220]. Ligands with different scaffolds but similar interactions can be selected through pharmacophore-based VS. The pharmacophore model can also be combined with QSAR when aligning the ligands [221].

Classical pharmacophore models also have limitations: The models are static but are used to represent the dynamic systems. Interactions in the pharmacophore model are restricted to simple geometric features. The dynophore method [222], a combination of the pharmacophore model and MD simulations, can address these limitations. This method provides the details of ligand binding out of the traditional spherical geometry and offers statistics of different binding modes and feature frequencies during the trajectory [222].

3.2.2. Quantitative Structure–Activity Relationship

QSAR, a modeling approach, unravels the relationship between bioactivities and structural properties of ligands based on the principle that bioactivities are related to structural properties [223]. Bioactivities refer to pharmacokinetic properties, including ADMET and other properties. Structural properties refer to the physicochemical properties of ligands. QSAR can rank numerous compounds according to their bioactivities, and therefore, it is extensively used in lead discovery and optimization during drug discovery. The statistical model is used for predicting the bioactivity of new ligands [223]. A reliable QSAR should meet the following requirements: [5,224] (1) obtaining the dataset of sufficient ligands (≥20 compounds) with bioactivities from the conventional experimental protocol; (2) selecting appropriate compounds to construct the training set and testing set; (3) no autocorrelation among the descriptors of the ligands (describing the chemical features of the molecule in a numerical form) that induces overpredicting or overfitting; and (4) validating the final QSAR model through internal/external validation to assure model reliability.

Based on the method of deriving descriptors, dimension-based QSAR methods are classified as [188] (1) 1D-QSAR, relating bioactivity to global molecular physiochemical properties such as logP and pKa; (2) 2D-QSAR, relating bioactivity to structural features of the ligands, such as connectivity indices, without regard to three-dimensional representations of the features; (3) 3D-QSAR, relating bioactivity to noncovalent interactions around the ligand; (4) 4D-QSAR, additionally containing the ensemble of ligand conformations on the basis of 3D-QSAR; (5) 5D-QSAR, describing different induced-fit models of 4D-QSAR; and (6) 6D-QSAR, further combining different solvation models of 5D-QSAR. Moreover, based on the techniques of constructing the relationship between bioactivities and structural properties of ligands, QSAR methods are categorized as linear and nonlinear [225].

Linear methods consist of linear regression, multiple linear regression, partial least squares, and principal component analysis/regression [225]. Nonlinear methods include artificial neural networks, k-nearest neighbors, and bayesian neural nets [225]. To meet the QSAR requirements in drug design, various QSAR-related tools were developed, such as Cloud 3D-QSAR [226], Web-4D-QSAR [227], and DPubChem [228]. Although QSAR has its advantages, some challenges exist related to QSAR application to drug design. For example, the limitation of high-quality datasets makes it difficult to construct a reliable QSAR model. Another challenge is the limitation of descriptors used for constructing QSAR models. To address this problem, new descriptors are integrated to accurately extract structural characterizations.

4. De Novo Drug Design by Artificial Intelligence

The above-mentioned machine learning based target identification, binding site identification, docking prediction, develop ability predictions, affinity predictions, etc., are the whole or part of the work of drug discovery by screening the desired drug from existing compound data. EquiBind was developed to predict receptor binding location and ligand's bound pose and orientation by applying geometric deep learning [203] ; additionally, Diff-Dock completed the molecular docking task by a diffusion generative approach [205]. In the following sections, we focus on reviewing the machine learning based frameworks which are relevant to de novo drug design. In other words, the state-of-the-art AI-based approaches which are able to generate novel molecules with desired properties are depicted.

Drug discovery using AI is an innovative process in which candidate molecules with desired chemical properties are created [4,229–231]. The number of compounds that belong to the drug would be in the 10^{23}–10^{60} from the chemical space aspect, which makes computation for the mining of novel compounds a challenge task [229,232,233]. Meanwhile, since a molecule binds to a particular protein pocket so that it can inhibit or activate cellular biological functions, balancing multiple structural and physicochemical parameters is crucial in drug discovery [234]. Machine learning techniques have considerably accelerated the process of drug discovery, which can handle the complex relationship between input and output variables for high-dimensional data [235,236]. Advances in deep learning models have recently resulted in a significant progress in molecule generation [234]. While machine learning models have been used for molecular property prediction, they presented a big step forward in bridging the gap between chemical entities and drug-like properties [234,237]. In particular, combining generative techniques with various statistics and probabilistic methods is the state-of-the-art approach in this task. The goal of generative modeling is consistent with the aforementioned drug design: sample novel molecules with intersection of multiple property constraints [238].

4.1. Overview of the Machine Learning Based de Novo Drug Design

Machine learning-based drug design involves a sequence of processes from data selection and representation to generative model construction. Figure 3 presents an overview of the machine learning-based de novo drug design procedure. In the beginning, appropriate data are selected from publicly available data sources, and property-based filtering and classification are performed to obtain molecules having desired properties for subsequent model learning. Then, sophisticated feature representation methods, such as those based on simplified molecular input line entry system (SMILES) and graphs [34,239,240], are applied to learn and represent the structures and properties of molecules. Finally, the optimal generative model is selected for de novo molecule generation based on the learned representation [229,236,238,241]. Furthermore, at the appropriate time, the generative model is optimized by combining reinforcement learning strategy and property prediction models [233,242,243].
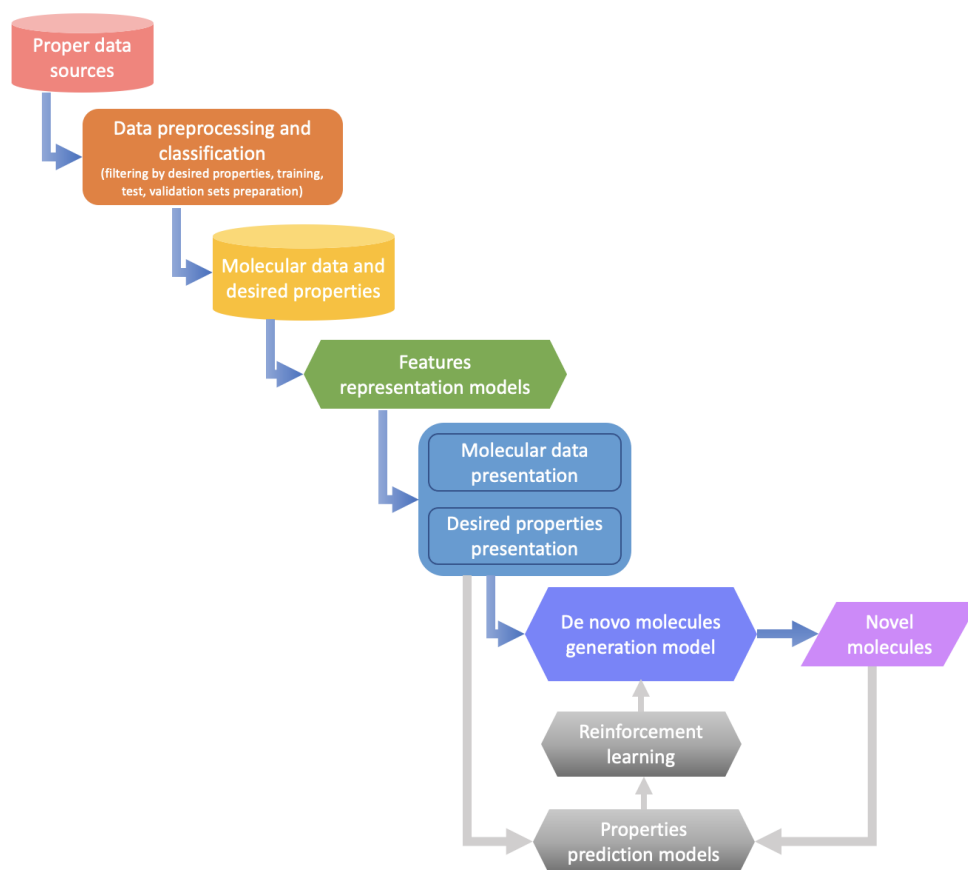
Figure 3. Overview of the machine learning-based de novo drug design procedure, from left-top to right-bottom: appropriate data selection; data filtering and classification; molecular data and de-sired properties storage; feature representation for molecules and properties; molecule generation by machine learning methods; generation model optimization by reinforcement learning strategy and property prediction models; de novo molecules generation.

4.2. Overview of de Novo Molecule Generation

The recent massive increase in AI-based de novo drug design research can be attributed primarily to the generative approach. This approach leverages deep learning strategy to learn the probability distribution of molecular data and produces continuous or discrete latent representation for molecules with property optimization. It finally maps learned probability distribution and molecule representation into novel molecules while optimizing molecular properties through the tuning of parameters of latent codes [230,231,244–246]. The generative approach is effective in property-based design, LBDD, and SBDD by generating both 2D and 3D molecules [239,247]. The generated novel drugs should potentially interact with therapeutic target proteins on the specific docking site, for which structural information about both proteins and molecules must be extracted during model training [248]. Hence, research efforts have been made on "structure-oriented generation" for drug design. Accordingly, "ligand-oriented generation"-based drug design focuses on structural and property information of molecules.

4.2.1. Structure-Oriented Generation

Structure-oriented generation creates novel molecules that bind to specific target proteins in drug design [240]. These target proteins could be receptors, enzymes, or other structural or functional proteins, and their structural information is required for the generation task [249]. In addition to the generation process, structural information of target proteins is crucial for exploring potential receptor–ligand interactions in drug design. Besides, the generation process depends on the structures of molecules, which are usually

represented by scaffolds of chemical compounds [239]. The structural information of the ligand and target protein can help explore the interaction between them [250].

In structure based de novo drug design, deep generative models are applied to learn and score docking of protein–ligand by exploring their structures and functions, and this information is used to generate de novo compound structures. The generation process applies a "fragment-based" strategy: given the initial chemical scaffold embedded in the binding site of the target protein, the pre-trained model generates molecules by sequentially adding, deleting, inserting, or replacing and linking fragments for it in an iterative manner [251,252]. In addition, with the availability of structural features of both target proteins and molecules, structure-oriented generation would allow better binding of designed drugs to target proteins [253].

In recent 5 years, the structure-based de novo drug design has applied generative machine learning models and made use of feature information such as ligand-protein complexes, protein binding sites, and bioactivity. Since the structure information of both protein and ligand are available, it is desirable to generate molecules in terms of 3D representations. A series of valuable tools for 3D molecule generation have emerged, including DeepLigBuilder, G-SchNet, RELATION, and Pocket2Mol, which employed multiple generative strategies. DeepLigBuilder applied graph generative model named ligand neural network (L-Net) to generate 3D molecules by iteratively refining existing structures, and it further combined a reinforcement learning method called Monte Carlo tree search (MCTS) to optimize the binding affinity [250]. The generative neural network (G-SchNet) learns the conditional distribution of 3D molecular structures and chemical properties [254]. RELATION also paid attention on conditional distribution by applying the variational autoencoder architecture, which consists of a 3D convolutional encoder, a LSTM-based captioning decoder and a bidirectional transfer learning module in order to transfer the features of protein-ligand complexes to latent space, for molecule generation [251]. It is not difficult to find out that the reinforcement learning and transfer learning are appropriate approaches for learning features information like binding sites, protein structure and docking scores, which were used to process molecule property optimization [234]. The Pocket2Mol has learned a probability distribution of atoms and bond types inside the pocked based on exiting atoms by adopting an auto-regression strategy, and used a graph neural network to capture features of atoms in binding pockets. For new drug sampling, this research considers the structures and geometrical constraints of protein pockets in drug design [255]. Another 3D generative model applied auto-regressive for novel molecule sampling can be found in study [240], similarly, it also used a neural network architecture to learn probability distribution of occurrences of atoms. Other studies also explored generating molecule from 1D and 2D aspects by exploiting SMILES representation of ligand and graph representation of protein binding sites, such as study [253], and they combined bioactivity affinity prediction model for generative model optimization.

4.2.2. Ligand-Oriented Generation

The designed novel molecules have high binding affinity to specific proteins but low binding affinity to other proteins [233]. When compounds in the employed data are already known to bind to the target proteins, the ligand-oriented approach is used for novel structure generation [252]. Ligand-oriented approaches focus on the molecules themselves, thereby generating compounds with new chemical hypotheses while optimizing the desired properties. For instance, some approaches use the known actives of a compound for a specific target receptor for latent chemical space retrieval [256].

The properties of chemical ligands are also optimized during generation, such as ADMET, binding affinity, logP, QED, solubility, easy to synthesize, and clearance. Properties can be optimized in two ways: one is property-based generation, wherein models would learn the chemical space of molecules with desirable properties, and then, the novel molecules are generated within a desired property space [245,257]. Autoencoder is a typical artificial neural network for property-based generation, which encodes molecular data

along with corresponding properties into latent space. Many de novo drug design models used the similar concept, such as CogMol applies the variational autoencoder [233], the junction tree variational autoencoder applies graph message passing network for molecular graph representation and chemical validity maintenance [230]; similarly, the graph generative model in study [238] also used message passing neural for graph generation and property control. The autoencoder in [231] then constructed a predictor to estimate chemical properties from the latent continuous representation when exploiting SMILES strings directly for encoding and decoding. A similar multilayer perceptron-based model for mapping between latent vectors and molecular properties was also trained in the popular molecule generation model named MolFlow [229]. In another method, a prediction function is applied for the desired properties of molecules and the generation model is fine-tuned in terms of a particular property by using the reinforcement learning strategy, such as constructing QSAR models, as reward functions [249,256]. GFlowNet and ReLeaSE are the well-known novel molecules generation models ap-plying reinforcement learning for property optimization [243,258]. Another common approach is to use transfer learning strategy to finetuned the model for properties such as activity optimization [259]. Whereas the influential research work, the generative model named GENTRL, not only learned a mapping from discrete molecular graphs with partially known properties to continuous latent space, but also applied reinforcement learning in the generating stage for property optimization. One of the major challenges of the ligand-based generation is the difficulty in actually synthesizing, that is, to ensure the high synthetic accessibility structures for the generated molecules [252,260]. Although some reactants and reaction rules can be used as the templates to guide the generation processes, such as RECAP [261] and SYNOPSIS [262], the structural novelty and diversity of new molecules have to be reduced.

5. Approaches and Techniques in Artificial Intelligence Based de Novo Drug Design

5.1. Datasets in AI-Based de Novo Drug Design

The first step in AI-based drug design is to learn the structures and properties of source chemical compounds to generate novel molecules that meet the requirements and expectations for desired properties such as proper binding energy, QED, and logP. Hence, appropriate molecular data are crucial for constructing the drug generative model [263]. High-throughput screening assays can act as a rich resource of bioactive molecular data, including molecular structures, chemical properties of molecular structures, molecular descriptors, side effects, clinical information, targets, and activity measurements [264]. Table 1 shows the widely used databases for AI-based drug design in recent years.

5.2. Descriptors/Feature Representation

Molecular descriptors/representations usually represent geometry, chemical structure, and physiochemical properties and biological activities of compounds. They are numerical vectors that are used as input for generative models [264]. Deep generative models produce molecule candidates by learning the underlying distribution of molecules on the basis of these descriptors/representation [240]. The way for representation plays a crucial role in the generative model, since efforts of deep neural networks (DNNs) have been focused on molecule representation learning. The strength of neural networks depends on their power and ability to make transformation among input, latent, and output representations [34,246].

Four categories of representation have been established: one-hot embedding-based, SMILES-based, graph-based, and 3D-based. Among these representations, the most commonly used feature representations for training DNN models are the SMILES representation and the molecular graph representation [234].

Table 1. Databases for AI based drug design.

Compound Database	Description	Number of Compounds
ChEMBL [264,265]	Drug discovery database provides bioactive molecules with drug-like properties knowledge.	2,157,379
ZINC [178,231–233,254]	Database enables access to compounds for drug discovery.	750 million + 230 million (3D)
PubChem [233,264,266]	Public chemical database at the National Library of Medicine (NLM) collects chemical information from different data sources.	112 million
DrugBank [183,231,232,254,264]	Web resource contains drug-related information.	14,528
STITCH [264,267]	Database contains interaction information between different chemicals.	0.5 million
BindingDB [264,268]	Database for molecular recognition, which supports drug discovery related work.	1.1 million
SIDER [264,269]	Resource contains drug reactions information.	1430 + 55,730
DCDB [264,270]	Drug Combination Database	1363
GDB-11 [231,232,254,271]	Database collects and generates molecules with up to 11 atoms of C, N, O, and F by considering simple valency, chemical stability, and synthetic feasibility rules.	26.4 million
GDB-13 [231,232,254,272]	Database upgrading from GDB-11, it enumerates in a similar manner small organic molecules containing up to 13 atoms of C, N, O, S, and Cl.	970 million
GDB-17 [231,232,254,273]	Chemical universe database covers drugs and typical for lead compounds for molecules with up to 17 atoms of C, N, O, S, and halogens.	166 billion

In one-hot embedding-based representation, binary vectors are used for atoms and bonds as molecular descriptors [229,249,264]. SMILES-based representation processes SMILES of a molecule directly by applying natural language processing techniques such as deep learning methods, including recurrent neural networks and long short-term memory (LSTM), which regard molecule generation as a Seq2Seq problem [34,234]. Seq2Seq is generally implemented on the basis of the encoder-decoder (encoder-decoder) framework, which is good at dealing with global information of long sequence and synthesizing contextual information of single token, to predict the corresponding alternative sequence [274,275]. The encoder transfers the SMILES strings representation into latent encodings, and the decoder transfers it into output SMILES strings by iteratively predicting probabilities of representation of particular tokens on the basis of previous information. In these methods, SMILES was firstly tokenized into a sequence of tokens, and each token was embedded into a space vector by different encoding methods such as one-hot encoding, pre-trained models [243,276]. Thus, grammar information of SMILES strings is learned for participating in the subsequent construction of generative models such as the transformer model, language model, variational autoencoder model, and generative adversarial networks [245,253]. There are recent studies used SMILES strings directly for molecule generation. The classical and representative ways are applying RNN on SMILES to generate scaffolds and corresponding attachment of molecules [276–278], applying self-attention mechanism model on SMILES for de novo drug design [279], and training molecule properties predictor by using SMILES [233]. Table 2 in next section shows the descriptions of the related deep learning methods.

Table 2. Deep learning techniques in molecules generation.

Deep Learning Techniques	Description	Applications
Recurrent neural networks (RNN)	Recurrent neural networks are similar to Markov chains with memory and feedback loops, each neuron in it would receive information from both actual time input and the previous neural [232,241,264]	SMILES strings representation [234,241,243]; generating novel and valid SMILES strings [34]; learn model autoregression [241]; construct encoder to convert discrete representations of molecules to multidimensional continuous representation [231]; estimate the probabilities of molecular data [259]
Long Short-Term Memory (LSTM)	LSTM is one kind of recurrent neural work with attention mechanism, which aims to solve the vanishing gradient problem for Recurrent Neural Networks (RNNs) [280,281]	encode SMILES strings [282]; builds sequence-to-sequence neural network for autoencoder [249]
Gated Recurrent Neural Network (Gated RNN)	One type of RNN with gated recurrent unit (GRU) containing forget gate and updating gate [283,284]	constructing encoder and decoder for SMILES sequence translation [253,285]; junction tree message passing [230]
Convolutional neural networks (CNN)	Convolutional neural networks contain sequential layers of convolution and pooling, and among them the convolution layers extract features by moving a window over the input tensors (arrays) and the pooling layers sub-sample the features [241,264,286]	construct graph convolutional neural networks [239]; grid-based 3D CNNs to predict protein–ligand binding affinity by constructing [287]
Multilayer perceptron networks (MPL)	Deep neural networks consist of multilayer perceptions, which are fully connected networks with activation functions [241]	chemical properties from latent codes [231]; mapping between latent vectors and molecular properties [229]
Multi-head attention networks	Contains encoder and decoder both with stacked self-attention and fully-connected layer inside, and the attention blocks in the network are all in the form of multi-head for receiving inputs of query, key, and value [288,289]	extract 3D conditional information of molecule [236] embed the active site graphs of target [253]
Message passing neural networks (MPNN)	A state-of-the-art and typical model for learning nodes and edges information in graph: a target node's representation come from its directly connected nodes through a multilayer neural network (or one layer), and the message passing between nodes in the graph is a circular iteration process [290].	parameterize atom graph encoding [243] encode connected motifs information of molecule [237] graph message passing network to represent the junction tree and molecular graph into latent codes [193] learn molecular graph and rationale distribution [238]
Graph neural network (GNN)	Regarding atoms as nodes and bonds as edges, this network applies convoluting operations for graphs encoding [40,264]	atoms and bonds information representation [240,245] parameterized the encoder and decoder for atoms and bonds types [239] spherical message passing graph neural networks to extract 3D conditional information of molecule [236]

A trend for adopting graph-based generation techniques in drug design has been increasing [233,237]. Graph-based models generate molecules regressively, sequentially predicting atoms or fragments for adding next time by learned probabilities [240]. Hence, information about atoms and bonds would be learned and represented, which correspond to nodes and edges in a graph. In graph-based representation, molecules are processed as graphs with nodes and edges corresponding to atoms and bonds by DNNs [246]. The three-dimensional information of a molecule is crucial for determining many molecular properties [236]. In the three-dimension-based representation, molecular graphs are translated into three-dimensional conformations including coordinates and distance of different

connected fragments [246,254]. Details of graph-based approaches and corresponding models are discussed in the next section since most of them applying deep learning methods.

5.3. Deep Learning Methods for Molecule Generation

DNNs are effective and efficient in drug discovery because they have a great learning capacity and a relatively less number of parameters [241]. Deep learning models learn the distribution of molecular structures, map the structures into continuous or discrete latent vectors, and finally generate novel molecules by picking up vectors in the latent space [236,238]. The complete generation procedure applies various deep learning techniques and generative frameworks. Table 2 summarizes popular deep learning techniques used in the molecule generation step of drug design.

Notably, graph neural networks (GNNs) can learn the structural information of atoms and their neighbors and perform well during both molecule generation and property prediction. These features are attributable to the message passing networks inside GNNs. Figure 4 shows a schema explaining how this model processes and learns graphical information.
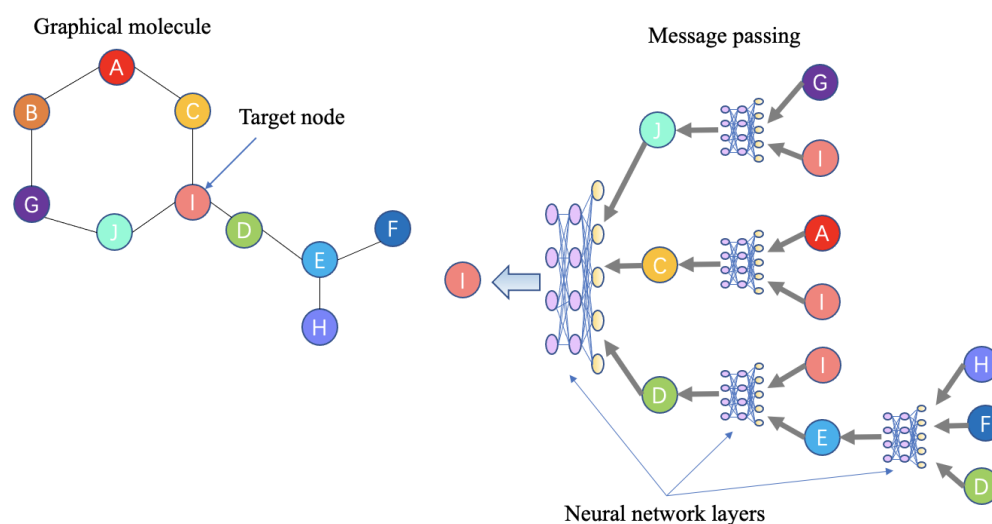

Figure 4. Example of a message passing neural network. The left side is an example of a graphical molecule. Ten atoms are regarded as nodes, with each node connecting with one or more nodes. The right side shows the message passing procedure for the target node I through a multilayer neural network (or one layer). The information from the directly connected nodes J, C, and D is passed to the target node I. As nodes J, C, and D have their own directly connected nodes (G, I, A, E), message is also passed by each of their neural networks. Message passing between nodes in a graph is a circular iteration process.

Other notable GNNs for de novel drug design include graph convolutional networks, graph attention networks, and graph generative networks [229,237,239,291].

As mentioned above, in current deep learning-based de novo drug design, models learn the probability distribution of molecular data and the mapping functions for encoding input molecular data into latent codes. They then generate new molecules by picking up vectors in the latent space based on probability distribution. This complete process can be accomplished by generative models. These models rely on machine learning-based approaches such as encoder–decoder (autoencoder), probability distribution learning, conditional distribution learning, transfer learning, and reinforcement learning [232,239,253]. The next paragraphs focus on sophisticated and classical molecule generation models.

Variational autoencoder (VAE) is a deep learning-based generative model that has been widely used in molecule design [249]. Being a probabilistic model, VAE can learn the distribution of given data and generate new meaningful data with more intra-class variations [249]. It consists of an encoder and a decoder. The encoder maps input molecular data x into latent codes z by parameterizing a posterior distribution $q_\varnothing(z|x)$, and the

decoder reconstructs molecular data from the learned distribution $p_\theta(x|z)$ [210,248,249]. Figure 5 illustrates the VAE architecture, which aims to maximize the likelihood of training data $p_\theta(x)$, which is expressed as Formula (1) [250]:

$$\log p_\theta(x^{(i)}) = E_z \left[\log p_\theta(x^{(i)}|z)\right] - D_{KL}(q_\varnothing(z|x^{(i)})||p_\theta(z)) + D_{KL}(q_\varnothing(z|x^{(i)})||p_\theta(z|x^{(i)})) \tag{1}$$
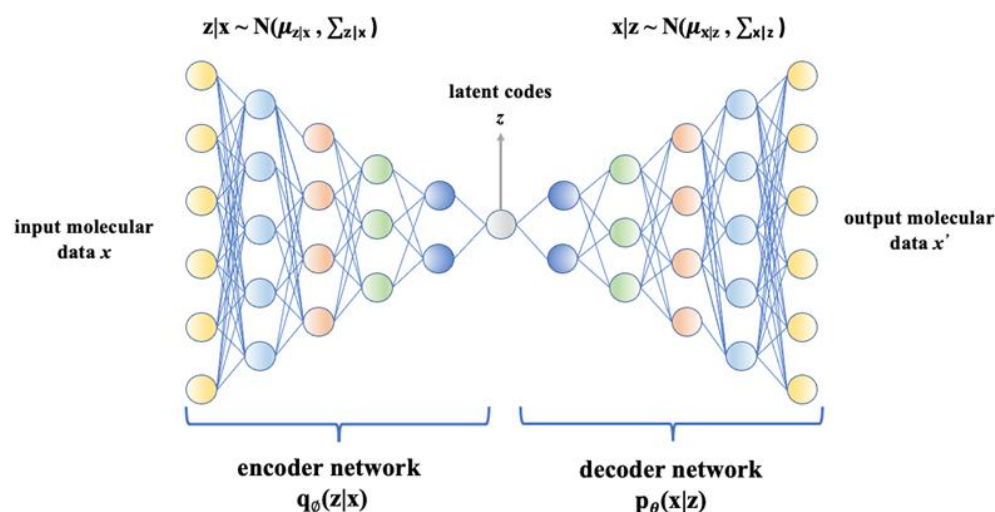

Figure 5. Variational autoencoder architecture. It consists of an encoder and a decoder, and they are deep neural networks. In general, the encoder maps input molecular data *x* into latent codes *z* by parameterizing a posterior distribution $q_\varnothing(z|x)$, and the decoder reconstructs molecular data from the learned distribution $p_\theta(x|z)$.

Kullback–Leibler (KL) divergence is used for measuring the difference between two probability distributions in the same space. The KL divergence $D_{KL}(q_\varnothing(z|x^{(i)})||p_\theta(z|x^{(i)}))$ measures difference between posterior distribution of latent variable and its prior distribution; however in VAE, it is impossible to process the posterior distribution $q_\varnothing(z|x^{(i)})$, $p_\theta(z|x^{(i)})$, which is introduced to approximate it [292,293]. Among Formula (1), as $p_\theta(z|x^{(i)})$ is intractable but Kullback–Leibler (KL) divergence $D_{KL}(q_\varnothing(z|x^{(i)})||p_\theta(z|x^{(i)}))$ is always larger than 0, the aim is to maximize the evidence lower bound (ELBO): $E_z \left[\log p_\theta(x^{(i)}|z)\right]$ $- D_{KL}(q_\varnothing(z|x^{(i)})||p_\theta(z))$. As $- D_{KL}(q_\varnothing(z|x^{(i)})||p_\theta(z))$ represents the negative KL divergence between the variational approximation distribution $q_\varnothing(z|x^{(i)})$ and the distribution of the latent variable z. In order to maximize the final result, this KL divergence has to be as small as possible. This KL divergence also prevents high consistency degree between the distribution of input x and distribution of output x'.

VAE-based drug design models can exploit both SMILES strings and graphs for molecular data representation and generation [253,292,294,295]. A popular VAE strategy-based drug design is the junction tree variational autoencoder (JT-VAE). JT-VAE applies the graph message passing the network to represent the junction tree and molecular graph as latent codes. Then, it generates valid chemical substructures with the learned maximal log-likelihood to form a tree-structured scaffold, and finally assembles these substructures into the molecule [230]. Jin et al. [237] proposed a VAE-based hierarchical graph encoder–decoder that applies the message passing the neural network for graphical representation, in which each layer performs graph convolutions iteratively conditioned on the results of the last layer. Another VAE-based molecule generation model employing graph information is GENTRL. GENTRL combines variational autoencoder, reinforcement learning, and tensor decompositions. It learns mapping from discrete molecular graphs with partially known properties to the continuous latent space parameterized by distribution. Moreover, the relationships between molecular structures and their properties are encoded by the tensor decomposition method, and reinforcement learning is applied in the generating stage [242]. DeepScaffold also applies VAE for constructing a scaffold-based molecular generative

model [239]. GraphVAE generates a probabilistic fully connected graph from continuous embedding and applies a graph matching algorithm to align the generated graph to ground truth [296]. For models utilizing SMILES strings, CogMol is a popular VAE-based molecule generation model that applies VAE to learn the latent space of SMILES representation along with properties such as QED. It generates novel molecules with desired properties by using a conditional latent (attribute) space sampling scheme [233]. Additionally, three-dimensional molecular structures were utilized in the VAE-based model G-SchNet that learns the conditional distribution of these structures, chemical properties, and sample molecules with target properties [254].

Generative adversarial networks (GANs) are another emerging simple but very efficient technology in drug design [297]. Compared to VEA's traditional training method of learning through loss functions, GAN uses a more realistic comparison method to implement adversarial training, which is more interpretable. The GAN model contains two major components: a generator G, which transforms latent vectors that are sampled from a prior distribution such as Gaussian into novel molecular data samples, similar to the training samples, and a discriminator D, which distinguishes fake molecular data points generated by G from the actual points sampled from the distribution of training data with the learning boundaries between them. Hence, the generator's task is to fool the well-trained discriminator by generating novel molecules, whereas the discriminator's task is to improve its ability to distinguish between real and fake molecules [246]. Figure 6 presents the flowchart of drug design using GANs.
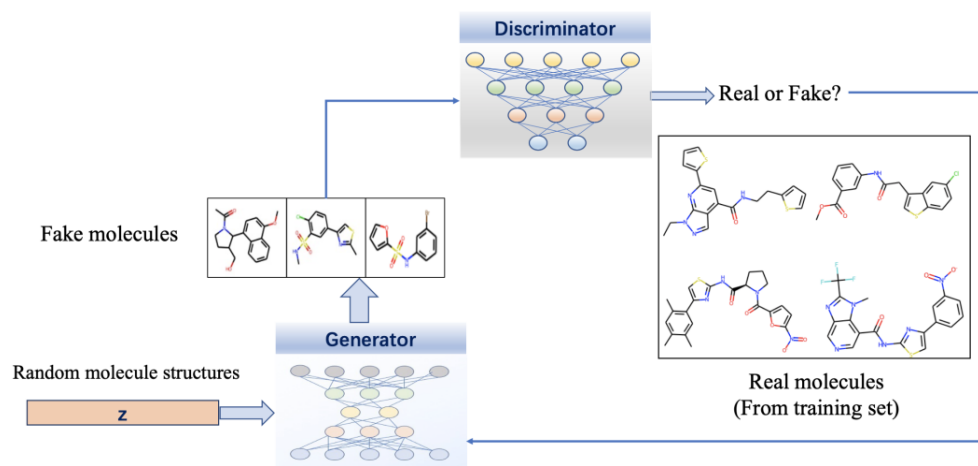

Figure 6. Flowchart of drug design using GANs. It contains a generator and a discriminator, and they are all deep neural networks. The generator transforms latent vectors that are sampled from a prior distribution such as Gaussian into novel molecular data samples, and the discriminator distinguishes fake molecular data generated by the generator from the actual points sampled from the distribution of training data and gives feed-back.

Combined with GNNs, the power of GANs can be further increased in molecule generation. One of the most widely used GAN-based drug design model is MolGAN. This model adopts GANs to directly process graphical molecular data and combines reinforcement learning to optimize specific desired chemical properties with generated molecules [298]. Since constructing sequences and graphs requires backpropagating the gradient, the training of GAN is more challenging than VEA. There are some drawbacks that are difficult to avoid for GAN: generators widely ignore random vectors leads the mappings of training data to output data are singularly deterministic; the two conflicting goals of generator and discriminator lead to a continuous drift of the learning parameters, resulting in varying degrees of distortion in the output.

Normalizing Flow is a method proposed to overcome the shortcomings of GAN and VAE through invertible functions. Normalizing flow is a probabilistic generative model

that uses simple probability distributions to simulate complex probability distributions. Composing invertible functions, the normalizing flow model learns a series of invertible transformations from the prior distribution of molecular data to simple distribution such as Gaussian. It finally converts the simple distribution into high-dimensional molecular data for de novo drug design [245,246]. Figure 7 illustrates the architecture of the normalizing flow model in molecule generation for drug design. Comparing to VAE and GAN, normalizing flow does not need any noisy data in the output, thus it allows for more robust local variance models; it is more stable during training process; and it is easier to converge. It also has its own disadvantages, namely poor interpretability, and hard to ensure the synthesizability of the generated molecules.
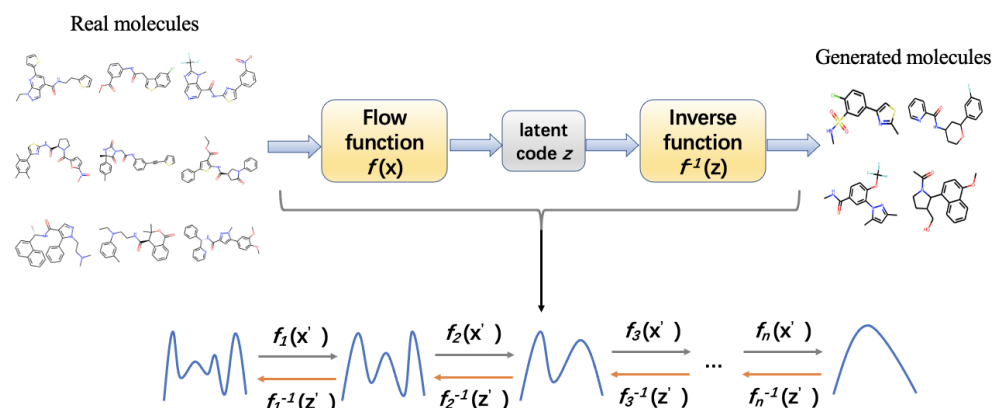
Figure 7. Architecture of normalizing flows. It contains a series of invertible functions for transforming molecular data into simple distribution and converting the distribution into high-dimensional molecular data for de novo drug design. Each of the functions optimizes the data distribution.

The normalizing flow architecture has been used in certain molecule generation models. MolFlow is a flow-based graph generative model that exploits normalizing flow to map molecular graphs and latent representations. It trains a model to generate bonds and a novel graph conditional flow to generate atoms on the basis of the bonds by leveraging graph convolution operations. Finally, the bonds and atoms are assembled in bond-valence constraints. They also train a multilayer perceptron model for mapping between latent vectors and molecular properties [229]. The hierarchical normalizing flow model, called MolGrow, generates a molecule from a single node and recursively splits every node into two; these operations are invertible and use graphical representation for node and edge attributes and feed them into an L invertible level architecture, wherein the generated latent codes fit into Gaussian distribution. Each level contains multiple blocks and linear transformations for noise separation and node merging; inside each block, three channel-wise transformations and two RealNVP layers are present [245]. GraphDF is a discrete latent variable model that applies normalizing flow for molecule generation. These discrete latent variables are sampled from multinomial distributions, and the model uses invertible modulo shift transform to sequentially map discrete latent variables to graph nodes and edges [236]. GraphNVP is a normalizing flow-based molecular graph generation model that represents the molecular graph by an adjacency tensor and a feature matrix of node attributes. It applies a continuous density model to learn probability distributions and two types of reversible affine coupling layers to transform the adjacency tensor and feature matrix into latent representations. This model first generates a graph structure and then generates node attributes [299].

In addition to normalizing flows, autoregressive model is another neural density estimator, in which a variable is predicted by the previous variable because the model decomposes the joint density as a product of conditionals [300]. The autoregressive model creates an explicit density model that can maximize the likelihood function of training

data. Several autoregressive generative models are used in drug design. They build the molecular graph by refining its intermediate structure in an iterative fashion [236,301].

In addition to the above architectures, a variety of pretrained neural network models were developed for handling drug discovery related problems. The pretrained graph neural networks model 3D Infomax is an example, it can predict latent 3D and quantum information by utilizing the 2D molecular graphical data, which is beneficial the down-stream tasks of molecule generation and molecular property prediction [302]. In recent years, some studies explored combining techniques of natural language processing for molecular data representation model training. The popular transformer architecture of language model BERT was applied to learn molecular representations, and the resulted model MolBert can be used for drug discovery related tasks [303]. Similar models that trained by transformer-based architecture on SMILES strings are SMILES-BERT, ChemBERTa, MegaMolBART [304–306], and so on. Another natural language processing related model, namely MolT5, was trained by a self- supervised learning framework on large amount of natural language text and molecule strings, which can be fine-tuned for molecule generation from natural language and molecule captioning [307]. Having such pre-trained molecular data representation models can largely improve the effectiveness and efficiency of de novo drug design.

5.4. Machine Learning Methods for Molecular Properties Optimization

In many drug design procedures, optimization of molecular properties (e.g., high drug-likeliness, synthetic accessibility, or solubility) is a critical step [308]. Various machine learning techniques can be applied to the input feature space, latent space, and output space. The probabilistic autoencoder can transform the features of molecular properties into latent variables [233]. Bayesian is the most popular method applied to the latent space to retrieve optimal latent solutions in the continuous latent space [34,233]. For property optimization on the output space, the most widely used strategy is to combine reinforcement learning with prediction machine learning models [233]. Statistical machine learning methods and deep learning methods help to build the classification or regression model that can predict molecular properties [263,309]. With the aid of prediction models, reinforcement learning maximizes the reward derived from the predicted scores of properties and biases the generative models, which allows the molecule generation model to achieve a high success rate in meeting the desired constraints [234,238,253]. Other optimization strategies have also been described in existing research. Modof, a generative model, applies message passing networks for encoding the difference between molecules before and after optimization at one disconnection site to connect changing fragments of a molecule and properties [308]. The Expectation–Maximization algorithm was employed in a hierarchical generative model to optimize molecular properties that mimic human experts [310].

For structure-oriented optimization, studies have been conducted to improve docking scores and activities of generated molecules for binding to specific targets. QSAR A is a classical model that is trained on docking scores from a chemical library [311]. EQUIBIND, a geometric and graph deep learning model, exploits graph matching networks, three-dimensional coordinates, and distance information-based graph neural networks (GNNs) for predicting the ligand–receptor complex structure [203]. Other studies have applied random forest, logistic regression, DNNs, and gradient boosting trees to predict the activity of generated molecules on the biological target by using molecular descriptors [259].

5.5. Evaluation

To understand the quality of generated molecules, evaluations are necessary. For different drug designs, different measurements from different aspects are used to evaluate the generated molecules. Table 3 displays crucial evaluation metrics in de novo drug design.

Table 3. Evaluation metrics in drug design.

Evaluation Metrics	Descriptions
LogP	The oil-water partition coefficient, also called the hydrophobic constant; the larger the LogP value, the more lipophilic the drug is; conversely, the smaller the LogP value, the more hydrophilic the drug is [233,239,243,245].
QED	Quantitative estimate of drug-likeness, and the value it is between 0 and 1 [239,240,250,309].
Synthesizability	The probability of the generated drug to be synthesized [233,240,277,309].
Binding affinity	The magnitude of the interaction force between receptor and ligand. It can be expressed by free binding energy [240,253].
Diversity	Generated molecules are similar in terms of the desired properties but with variety of forms [237–240,245].
Maximum Mean Discrepancy	Maximum Mean Discrepancy values between generated molecules and real molecules [236,239,245,250].
Docking score	To measure the probability of the mutual recognition between ligand and receptor through the matching principle [242,245,250].
Novelty	The quality for generated molecules to be different from existed molecules, new and unusual [238,312].
Validity	An inherent property of a drug, it represents the performance of drug in prevention, treatment, diagnosis of diseases and regulation of physiological functions [230,236,237,313].
Similarity	Similarity between generated molecules and real molecules, such as Tanimoto Similarity between molecular fingerprints [233,234].
Toxicity	The degree of poisonous or harmful that the drug would be [233,314,315].

6. Conclusions and Perspectives

Computational biology approaches have been extensively used to facilitate drug design in target discovery, mechanism study, VS, and lead optimization. These approaches have a solid theoretical foundation, and most training data required by deep learning methods are generated using the computational biology methods. MD simulations, including force field-based simulations and ab initio simulations, continue to play an indispensable role in molecular mechanism studies, as well as thermodynamic and kinetic property research. In drug design, accurate calculation of binding energy or free energy change of ligand–target, as well as capturing structural and dynamical features of targets continue to rely on MD simulations. Compared with the ab initio method, molecular force field-based simulations can be extended to a larger scale but lacks accuracy. The QM/MM method compensates for this defect and is gradually applied for drug exploration. However, a large number of computational tasks makes it difficult for MD simulation to expand to larger scales, which limits the wide application of ab initio MD simulation. To address the computational cost concern, the CG method is designed and applied in many cases. Overall, the emergence and current wide applications of CADD, molecular docking, VS, and QSAR have accelerated drug design. Researchers have recently used AI methods to accelerate the traditional drug design paradigm and have made considerable progress. In molecular generation, generative models based on molecular graphs or strings such as SMILES or SELFIES have become the mainstream as they exhibited excellent performance in various molecular optimization tasks [258,316–318].

Although computational modeling of complex protein machines and AI methods have demonstrated their superior capability in drug design, several challenges remain to be solved in the current AI-based designing framework. Similar to many other areas in machine learning, molecular generator evaluation is governed by certain compound datasets. Indicators such as novelty, validity, and uniqueness are commonly used to measure model performance. Numerous generators have achieved an excellent score among these datasets.

However, as some have suggested, it is difficult to confirm if the model actually learn the patterns in the training dataset. Moreover, most datasets are of low quality and cannot satisfy all demands in real drug discovery [319]. Ideal benchmark datasets should include diverse metrics for different tasks and consider practical applications; synthetic accessibility can be set as a general indicator. In addition to benchmarks and metrics in model evaluation, molecular representations play a vital role in molecular learning and generation. Two-dimensional graphs are the most conventional method to represent molecules, and such representations can be easily processed using GNNs. A typical drawback is that messages passing based on GNNs are unable to distinguish different configurations of molecules and some non-isomorphic graphs [320]. To capture spatial information, three-dimensional representations (point clouds, three-dimensional graphs, and three-dimensional grids) have recently gained considerable attention. Researchers must introduce additional units in generators to catch Euclidean symmetries in the three-dimensional space, including rotational, translational, and reflectional symmetries. Such generators are suitable for small molecular systems because of the increased complexity of macromolecules [321]. Besides, language models have received great attention in several challenging generative tasks as they learn complex molecular distributions. Language models can scale multi-modal distribution and generate larger molecules [322], while graph generative models are more interpretable. Integration of the interpretability of a graph model and the flexibility of a language model into a unified framework remains a challenge.

We suggest three potential future directions. (1) Similar to the representation learning models in natural language processing and computer vision, such as BERT, GPT3, and ViT, pre-trained molecular representation models have substantiated their potential in various downstream tasks [323,324]. (2) Domain knowledge contains a high degree of abstraction and summarization of natural phenomena and is essential for physics-informed neural networks. In computational biology, domain knowledge has been used to compute binding free energy and molecular potential and perform MD simulations [325–327]. To accurately estimate the binding affinity or binding free energy, rigorous but expensive methods, such as FEP or Linear Response Approximation, and their variants need to be used. Machine learning-based methods have made substantial progress in predicting binding affinity [325,328,329]. The combination of these differentiable modules with molecular generators is promising and may appreciably accelerate new drug development. (3) In addition to structural data of proteins and drugs, the availability of omics or clinical data can support drug discovery or repurposing. AI models can find hidden patterns and relations and offer more accurate prediction by using big data with different scales and types [330].

Author Contributions: Conceptualization, C.B., T.-Y.L. and S.W.; methodology, C.B. and T.-Y.L.; investigation and writing—original draft preparation, Y.Z., M.L. and P.W.; writing—review and editing, C.B., T.-Y.L. and S.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated and analyzed during this study are included in this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. DiMasi, J.A.; Grabowski, H.G.; Hansen, R.W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **2016**, *47*, 20–33. [CrossRef] [PubMed]

2. Song, C.M.; Lim, S.J.; Tong, J.C. Recent advances in computer-aided drug design. *Brief. Bioinform.* **2009**, *10*, 579–591. [CrossRef] [PubMed]

3. Kalyan, K. Pharmaceutical Medicine and Translational Clinical Research. *Curr. Sci.* **2018**, *115*, 1403.

4. Sabe, V.T.; Ntombela, T.; Jhamba, L.A.; Maguire, G.E.; Govender, T.; Naicker, T.; Kruger, H.G. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *Eur. J. Med. Chem.* **2021**, *224*, 113705. [CrossRef]

5. Gurung, A.B.; Ali, M.A.; Lee, J.; Farah, M.A.; Al-Anazi, K.M. An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19. *BioMed Res. Int.* **2021**, *2021*, 8853056. [CrossRef] [PubMed]

6. Zhong, F.; Xing, J.; Li, X.; Liu, X.; Fu, Z.; Xiong, Z.; Lu, D.; Wu, X.; Zhao, J.; Tan, X. Artificial intelligence in drug design. *Sci. China Life Sci.* **2018**, *61*, 1191–1204. [CrossRef]

7. Hou, T.; Xu, X. Recent development and application of virtual screening in drug discovery: An overview. *Curr. Pharm. Des.* **2004**, *10*, 1011–1033. [CrossRef]

8. Hill, R.G.; Richards, D. *Drug Discovery and Development E-Book: Technology in Transition*; Elsevier Health Sciences: Amsterdam, The Netherlands, 2021.

9. Durrant, J.D.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71. [CrossRef]

10. Zhao, H.; Caflisch, A. Molecular dynamics in drug design. *Eur. J. Med. Chem.* **2015**, *91*, 4–14. [CrossRef]

11. Huang, D.; Caflisch, A. The free energy landscape of small molecule unbinding. *PLoS Comput. Biol.* **2011**, *7*, e1002002. [CrossRef]

12. Honarparvar, B.; Govender, T.; Maguire, G.E.; Soliman, M.E.; Kruger, H.G. Integrated approach to structure-based enzymatic drug design: Molecular modeling, spectroscopy, and experimental bioactivity. *Chem. Rev.* **2014**, *114*, 493–537. [CrossRef] [PubMed]

13. Labanowski, J.K.; Andzelm, J.W. *Density Functional Methods in Chemistry*; Springer: Berlin/Heidelberg, Germany, 2012.

14. Hafner, J. *Ab-initio* simulations of materials using VASP: Density-functional theory and beyond. *J. Comput. Chem.* **2008**, *29*, 2044–2078. [CrossRef] [PubMed]

15. Chivian, D.; Robertson, T.; Bonneau, R.; Baker, D. Ab initio methods. *Meth. Biochem. Anal.* **2003**, *44*, 547–558. [CrossRef]

16. Sebastiani, D.; Rothlisberger, U. Advances in Density-Functional-Based Modeling Techniques-Recent Extensions of the Car-Parrinello Approach. *Meth. Princ. Med. Chem.* **2003**, *17*, 5–40. [CrossRef]

17. Veselovsky, A.; Ivanov, A. Strategy of computer-aided drug design. *Curr. Drug Targets-Infect. Disord.* **2003**, *3*, 33–40. [CrossRef]

18. Surabhi, S.; Singh, B. Computer aided drug design: An overview. *J. Drug Deliv. Ther.* **2018**, *8*, 504–509. [CrossRef]

19. Kore, P.P.; Mutha, M.M.; Antre, R.V.; Oswal, R.J.; Kshirsagar, S.S. Computer-aided drug design: An innovative tool for modeling. *Open J. Med. Chem.* **2012**, *2*, 139–148. [CrossRef]

20. Baig, M.H.; Ahmad, K.; Roy, S.; Ashraf, J.M.; Adil, M.; Siddiqui, M.H.; Khan, S.; Kamal, M.A.; Provazník, I.; Choi, I. Computer aided drug design: Success and limitations. *Curr. Pharm. Des.* **2016**, *22*, 572–581. [CrossRef]

21. Yu, W.; MacKerell, A.D. Computer-aided drug design methods. In *Antibiotics*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 85–106.

22. Lyne, P.D. Structure-based virtual screening: An overview. *Drug Discov. Today* **2002**, *7*, 1047–1055. [CrossRef]

23. Lionta, E.; Spyrou, G.; Vassilatis, D.K.; Cournia, Z. Structure-based virtual screening for drug discovery: Principles, applications and recent advances. *Curr. Top. Med. Chem.* **2014**, *14*, 1923–1938. [CrossRef]

24. Banegas-Luna, A.-J.; Cerón-Carrasco, J.P.; Pérez-Sánchez, H. A review of ligand-based virtual screening web tools and screening algorithms in large molecular databases in the age of big data. *Future Med. Chem.* **2018**, *10*, 2641–2658. [CrossRef] [PubMed]

25. Lavecchia, A.; Di Giovanni, C. Virtual screening strategies in drug discovery: A critical review. *Curr. Med. Chem.* **2013**, *20*, 2839–2860. [CrossRef] [PubMed]

26. Sohraby, F.; Bagheri, M.; Aryapour, H. Performing an in silico repurposing of existing drugs by combining virtual screening and molecular dynamics simulation. In *Computational Methods for Drug Repurposing*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 23–43.

27. Issa, N.T.; Kruger, J.; Byers, S.W.; Dakshanamurthy, S. Drug repurposing a reality: From computers to the clinic. *Expert Rev. Clin. Pharmacol.* **2013**, *6*, 95–97. [CrossRef] [PubMed]

28. Duch, W.; Swaminathan, K.; Meller, J. Artificial intelligence approaches for rational drug design and discovery. *Curr. Pharm. Des.* **2007**, *13*, 1497–1508. [CrossRef] [PubMed]

29. Gertrudes, J.C.; Maltarollo, V.G.; Silva, R.; Oliveira, P.R.; Honorio, K.M.; Da Silva, A. Machine learning techniques and drug design. *Curr. Med. Chem.* **2012**, *19*, 4289–4297. [CrossRef]

30. Lima, A.N.; Philot, E.A.; Trossini, G.H.G.; Scott, L.P.B.; Maltarollo, V.G.; Honorio, K.M. Use of machine learning approaches for novel drug discovery. *Expert Opin. Drug Discov.* **2016**, *11*, 225–239. [CrossRef]

31. Klambauer, G.; Hochreiter, S.; Rarey, M. Machine Learning in Drug Discovery. *J. Chem. Inf. Model.* **2019**, *59*, 945–946. [CrossRef]

32. Zhang, L.; Tan, J.; Han, D.; Zhu, H. From machine learning to deep learning: Progress in machine intelligence for rational drug discovery. *Drug Discov. Today* **2017**, *22*, 1680–1685. [CrossRef]

33. Gawehn, E.; Hiss, J.A.; Schneider, G. Deep learning in drug discovery. *Mol. Inform.* **2016**, *35*, 3–14. [CrossRef]

34. Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **2018**, *23*, 1241–1250. [CrossRef]

35. Carpenter, K.A.; Huang, X. Machine learning-based virtual screening and its applications to Alzheimer's drug discovery: A review. *Curr. Pharm. Des.* **2018**, *24*, 3347–3358. [CrossRef] [PubMed]

36. Mendolia, I.; Contino, S.; Perricone, U.; Ardizzone, E.; Pirrone, R. Convolutional architectures for virtual screening. *BMC Bioinform.* **2020**, *21*, 310. [CrossRef] [PubMed]

37. Bohr, H. Drug discovery and molecular modeling using artificial intelligence. In *Artificial Intelligence in Healthcare*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 61–83.

38. Gimeno, A.; Ojeda-Montes, M.J.; Tomás-Hernández, S.; Cereto-Massagué, A.; Beltrán-Debón, R.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. The light and dark sides of virtual screening: What is there to know? *Int. J. Mol. Sci.* **2019**, *20*, 1375. [CrossRef]

39. Liu, X.; Shi, D.; Zhou, S.; Liu, H.; Liu, H.; Yao, X. Molecular dynamics simulations and novel drug discovery. *Expert Opin. Drug Discov.* **2018**, *13*, 23–37. [CrossRef] [PubMed]

40. Yan, T.; Yu, L.; Zhang, N.; Peng, C.; Su, G.; Jing, Y.; Zhang, L.; Wu, T.; Cheng, J.; Guo, Q. The advanced development of molecular targeted therapy for hepatocellular carcinoma. *Cancer Biol. Med.* **2022**, *19*, 802–817. [CrossRef] [PubMed]

41. Boomsma, W.; Nielsen, S.V.; Lindorff-Larsen, K.; Hartmann-Petersen, R.; Ellgaard, L. Bioinformatics analysis identifies several intrinsically disordered human E3 ubiquitin-protein ligases. *PeerJ* **2016**, *4*, e1725. [CrossRef] [PubMed]

42. Xia, X. Bioinformatics and drug discovery. *Curr. Top. Med. Chem.* **2017**, *17*, 1709–1726. [CrossRef]

43. Huang, C.; Luo, H.; Huang, Y.; Fang, C.; Zhao, L.; Li, P.; Zhong, C.; Liu, F. AURKB, CHEK1 and NEK2 as the Potential Target Proteins of Scutellaria barbata on Hepatocellular Carcinoma: An Integrated Bioinformatics Analysis. *Int. J. Gen. Med.* **2021**, *14*, 3295. [CrossRef]

44. Gordon, D.E.; Hiatt, J.; Bouhaddou, M.; Rezelj, V.V.; Ulferts, S.; Braberg, H.; Jureka, A.S.; Obernier, K.; Guo, J.Z.; Batra, J. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **2020**, *370*, eabe9403. [CrossRef]

45. Chen, X.; Wang, Y.; Ma, N.; Tian, J.; Shao, Y.; Zhu, B.; Wong, Y.K.; Liang, Z.; Zou, C.; Wang, J. Target identification of natural medicine with chemical proteomics approach: Probe synthesis, target fishing and protein identification. *Signal Transduct. Target. Ther.* **2020**, *5*, 72. [CrossRef]

46. Guo, J.; Ning, L.; Ren, H.; Liu, H.; Yao, X. Influence of the pathogenic mutations T188K/R/A on the structural stability and misfolding of human prion protein: Insight from molecular dynamics simulations. *Biochim. Biophys. Acta Gen. Subj.* **2012**, *1820*, 116–123. [CrossRef] [PubMed]

47. Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190–196. [CrossRef]

48. Hasan, A.S.; Mohammed, F.Q.; Takz, M.M. Design and synthesis of graphene oxide-based glass substrate and its antimicrobial activity against MDR Bacterial Pathogens. *J. Mech. Eng. Res. Dev.* **2020**, *43*, 11–17.

49. Barone, V.; Improta, R.; Rega, N. Computation of protein pK's values by an integrated density functional theory/polarizable continuum model approach. *Theor. Chem. Acc.* **2004**, *111*, 237–245. [CrossRef]

50. Souza, P.C.; Thallmair, S.; Marrink, S.J.; Mera-Adasme, R. An allosteric pathway in copper, zinc superoxide dismutase unravels the molecular mechanism of the G93A amyotrophic lateral sclerosis-linked mutation. *J. Phys. Chem. Lett.* **2019**, *10*, 7740–7744. [CrossRef] [PubMed]

51. Moreira, C.; Ramos, M.J.; Fernandes, P.A. Reaction mechanism of *Mycobacterium Tuberculosis* glutamine synthetase using quantum mechanics/molecular mechanics calculations. *Chem. A Eur. J.* **2016**, *22*, 9218–9225. [CrossRef]

52. Peters, M.B.; Raha, K.; Merz, K. Quantum mechanics in structure-based drug design. *Curr. Opin. Drug Discov. Dev.* **2006**, *9*, 370.

53. Barman, T.K.; Hazarika, A.K.; Kalita, U.; Dhar, R.; Borah, L.; Chetri, S.; Ghosh, S.S. *Epidemiology of Anti-HIV Drug Resistance: Quantum Mechanics (Qm) and Molecular Mechanics (MM) Studies into the Binding of 6-Aminoquinoline Molecules within HIV Protein (1AJX) and Its Economic Implication*; Elsevier: Amsterdam, The Netherlands, 2022.

54. Xu, D.; Zhou, Y.; Xie, D.; Guo, H. Antibiotic Binding to Monozinc CphA β-Lactamase from *Aeromonas hydropila*: Quantum Mechanical/Molecular Mechanical and Density Functional Theory Studies. *J. Med. Chem.* **2005**, *48*, 6679–6689. [CrossRef]

55. Hassan, A.U.; Sumrra, S.H. Exploring the bioactive sites of new sulfonamide metal chelates for multi-drug resistance: An experimental versus theoretical design. *J. Inorg. Organomet. Polym. Mater.* **2022**, *32*, 513–535. [CrossRef]

56. Ode, H.; Matsuyama, S.; Hata, M.; Hoshino, T.; Kakizawa, J.; Sugiura, W. Mechanism of drug resistance due to N88S in CRF01_AE HIV-1 protease, analyzed by molecular dynamics simulations. *J. Med. Chem.* **2007**, *50*, 1768–1777. [CrossRef]

57. Amusengeri, A.; Tata, R.B.; Bishop, Ö.T. Understanding the pyrimethamine drug resistance mechanism via combined molecular dynamics and dynamic residue network analysis. *Molecules* **2020**, *25*, 904. [CrossRef]

58. Vanommeslaeghe, K.; Guvench, O. Molecular mechanics. *Curr. Pharm. Des.* **2014**, *20*, 3281–3292. [CrossRef] [PubMed]

59. Bekono, B.D.; Sona, A.N.; Eni, D.B.; Owono, L.C.; Megnassan, E.; Ntie-Kang, F. Molecular mechanics approaches for rational drug design: Forcefields and solvation models. *Phys. Sci. Rev.* **2021**, 20190128.

60. Williams-Noonan, B.J.; Yuriev, E.; Chalmers, D.K. Free Energy Methods in Drug Design: Prospects of "Alchemical Perturbation" in Medicinal Chemistry. *J. Med. Chem.* **2018**, *61*, 638–649. [CrossRef]

61. Allen, M.P. Introduction to molecular dynamics simulation. *Comput. Soft Matter: Synth. Polym. Proteins* **2004**, *23*, 1–28.

62. Moroy, G.; Sperandio, O.; Rielland, S.; Khemka, S.; Druart, K.; Goyal, D.; Perahia, D.; Miteva, M.A. Sampling of conformational ensemble for virtual screening using molecular dynamics simulations and normal mode analysis. *Future Med. Chem.* **2015**, *7*, 2317–2331. [CrossRef] [PubMed]

63. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]

64. Whittig, L.; Allardice, W. X-ray diffraction techniques. *Meth. Soil Anal. Part 1 Phys. Mineral. Meth.* **1986**, *5*, 331–362. [CrossRef]

65. Javier, G.N.; Christopher, G.T. Cryo-Electron Microscopy: Moving Beyond X-Ray Crystal Structures for Drug Receptors and Drug Development. *Ann. Rev. Pharmacol. Toxicol.* **2020**, *60*, 51–71. [CrossRef]

66. Bai, X.-C.; McMullan, G.; Scheres, S.H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **2015**, *40*, 49–57. [CrossRef]

67. Weissenberger, G.; Henderikx, R.J.M.; Peters, P.J. Understanding the invisible hands of sample preparation for cryo-EM. *Nat. Methods* **2021**, *18*, 463–471. [CrossRef]

68. Krieger, E.; Nabuurs, S.B.; Vriend, G. Homology modeling. *Meth. Biochem. Anal.* **2003**, *44*, 509–524. [CrossRef]

69. Cavasotto, C.N.; Phatak, S.S. Homology modeling in drug discovery: Current trends and applications. *Drug Discov. Today* **2009**, *14*, 676–683. [CrossRef] [PubMed]

70. AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **2019**, *35*, 4862–4865. [CrossRef] [PubMed]

71. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [CrossRef]

72. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]

73. MacKerell, A.D., Jr. Atomistic models and force fields. In *Computational Biochemistry and Biophysics*; CRC Press: Boca Raton, FL, USA, 2001; pp. 19–50.

74. Robertson, M.J.; Tirado-Rives, J.; Jorgensen, W.L. Improved peptide and protein torsional energetics with the OPLS-AA force field. *J. Chem. Theory Comput.* **2015**, *11*, 3499–3509. [CrossRef]

75. Senftle, T.P.; Hong, S.; Islam, M.; Kylasa, S.B.; Zheng, Y.; Shin, Y.K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M.J.; Aktulga, H.M. The ReaxFF reactive force-field: Development, applications and future directions. *NPJ Comput. Mater.* **2016**, *2*, 15011. [CrossRef]

76. Dauber-Osguthorpe, P.; Hagler, A.T. Biomolecular force fields: Where have we been, where are we now, where do we need to go and how do we get there? *J. Comput. Aided Mol. Des.* **2019**, *33*, 133–203. [CrossRef]

77. Lee, M.; Kolev, V.; Warshel, A. Validating a Coarse-Grained Voltage Activation Model by Comparing Its Performance to the Results of Monte Carlo Simulations. *J. Phys. Chem. B* **2017**, *121*, 11284–11291. [CrossRef]

78. Vorobyov, I.; Kim, I.; Chu, Z.T.; Warshel, A. Refining the treatment of membrane proteins by coarse-grained models. *Proteins: Struct. Funct. Bioinform.* **2016**, *84*, 92–117. [CrossRef] [PubMed]

79. Horikoshi, N.; Hwang, S.; Gati, C.; Matsui, T.; Castillo-Orellana, C.; Raub, A.G.; Garcia, A.A.; Jabbarpour, F.; Batyuk, A.; Broweleit, J. Long-range structural defects by pathogenic mutations in most severe glucose-6-phosphate dehydrogenase deficiency. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2022790118. [CrossRef] [PubMed]

80. Budaitis, B.G.; Jariwala, S.; Rao, L.; Yue, Y.; Sept, D.; Verhey, K.J.; Gennerich, A. Pathogenic mutations in the kinesin-3 motor KIF1A diminish force generation and movement through allosteric mechanisms. *J. Cell Biol.* **2021**, *220*, e202004227. [CrossRef]

81. Zanetti-Domingues, L.C.; Korovesis, D.; Needham, S.R.; Tynan, C.J.; Sagawa, S.; Roberts, S.K.; Kuzmanic, A.; Ortiz-Zapater, E.; Jain, P.; Roovers, R.C. The architecture of EGFR's basal complexes reveals autoinhibition mechanisms in dimers and oligomers. *Nat. Commun.* **2018**, *9*, 4325. [CrossRef] [PubMed]

82. Zhu, M.; Wang, D.D.; Yan, H. Genotype-determined EGFR-RTK heterodimerization and its effects on drug resistance in lung Cancer treatment revealed by molecular dynamics simulations. *BMC Mol. Cell Biol.* **2021**, *22*, 34. [CrossRef] [PubMed]

83. Rahnasto-Rilla, M.; Tyni, J.; Huovinen, M.; Jarho, E.; Kulikowicz, T.; Ravichandran, S.; Bohr, V.A.; Ferrucci, L.; Lahtela-Kakkonen, M.; Moaddel, R. Natural polyphenols as sirtuin 6 modulators. *Sci. Rep.* **2018**, *8*, 4163. [CrossRef] [PubMed]

84. Jiang, X.; Han, M.; Tran, K.; Patil, N.A.; Ma, W.; Roberts, K.D.; Xiao, M.; Sommer, B.; Schreiber, F.; Wang, L. An Intelligent Strategy with All-Atom Molecular Dynamics Simulations for the Design of Lipopeptides against Multidrug-Resistant *Pseudomonas aeruginosa*. *J. Med. Chem.* **2022**, *65*, 10001–10013. [CrossRef]

85. Cavasotto, C.; Orry, A.J.W. Ligand docking and structure-based virtual screening in drug discovery. *Curr. Top. Med. Chem.* **2007**, *7*, 1006–1014. [CrossRef]

86. Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949. [CrossRef]

87. Meng, X.Y.; Zhang, H.X.; Mezei, M.; Cui, M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* **2011**, *7*, 146–157. [CrossRef]

88. Shoichet, B.K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865. [CrossRef] [PubMed]

89. Irwin, J.J.; Shoichet, B.K. ZINC− a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182. [CrossRef] [PubMed]

90. Polgár, T.; Keseru, G.M. Integration of virtual and high throughput screening in lead discovery settings. *Comb. Chem. High Throughput Screen.* **2011**, *14*, 889–897. [CrossRef] [PubMed]

91. Hu, X.; Vujanac, M.; Southall, N.; Stebbins, C.E. Inhibitors of the Yersinia protein tyrosine phosphatase through high throughput and virtual screening approaches. *Bioorganic Med. Chem. Lett.* **2013**, *23*, 1056–1062. [CrossRef]

92. Lee, H.; Mittal, A.; Patel, K.; Gatuz, J.L.; Truong, L.; Torres, J.; Mulhearn, D.C.; Johnson, M.E. Identification of novel drug scaffolds for inhibition of SARS-CoV 3-Chymotrypsin-like protease using virtual and high-throughput screenings. *Bioorganic Med. Chem.* **2014**, *22*, 167–177. [CrossRef]

93. Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S.H. Structure-based virtual screening for drug discovery: A problem-centric review. *AAPS J.* **2012**, *14*, 133–141. [CrossRef]

94. Sastry, G.M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221–234. [CrossRef]

95. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W. Computational methods in drug discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395. [CrossRef]

96. Taylor, R.; Jewsbury, P.; Essex, J. A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151–166. [CrossRef]

97. Schulz-Gasch, T.; Stahl, M. Binding site characteristics in structure-based virtual screening: Evaluation of current docking tools. *J. Mol. Model.* **2003**, *9*, 47–57. [CrossRef]

98. Cross, J.B.; Thompson, D.C.; Rai, B.K.; Baber, J.C.; Fan, K.Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455–1474. [CrossRef] [PubMed]

99. Kumar, V.; Kancharla, S.; Jena, M.K. In silico virtual screening-based study of nutraceuticals predicts the therapeutic potentials of folic acid and its derivatives against COVID-19. *VirusDisease* **2021**, *32*, 29–37. [CrossRef] [PubMed]

100. Pantsar, T.; Poso, A. Binding affinity via docking: Fact and fiction. *Molecules* **2018**, *23*, 1899. [CrossRef] [PubMed]

101. Macip, G.; Garcia-Segura, P.; Mestres-Truyol, J.; Saldivar-Espinoza, B.; Ojeda-Montes, M.J.; Gimeno, A.; Cereto-Massagué, A.; Garcia-Vallvé, S.; Pujadas, G. Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition. *Med. Res. Rev.* **2022**, *42*, 744–769. [CrossRef] [PubMed]

102. Murugan, N.A.; Podobas, A.; Gadioli, D.; Vitali, E.; Palermo, G.; Markidis, S. A review on parallel virtual screening softwares for high-performance computers. *Pharmaceuticals* **2022**, *15*, 63. [CrossRef]

103. Rastelli, G.; Pinzi, L. Refinement and Rescoring of Virtual Screening Results. *Front. Chem.* **2019**, *7*, 498. [CrossRef]

104. Salmaso, V.; Moro, S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Front. Pharmacol.* **2018**, *9*, 923. [CrossRef]

105. Liu, W.; Schmidt, B.; Voss, G.; Müller-Wittig, W. Molecular dynamics simulations on commodity GPUs with CUDA. In Proceedings of the International Conference on High-Performance Computing, Goa, India, 18–21 December 2007; pp. 185–196.

106. Ramalingam, G.; Reps, T. On the computational complexity of dynamic graph problems. *Theor. Comput. Sci.* **1996**, *158*, 233–277. [CrossRef]

107. Salomon-Ferrer, R.; Gotz, A.W.; Poole, D.; Le Grand, S.; Walker, R.C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888. [CrossRef]

108. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461. [CrossRef]

109. Ahmad, I.; Jadhav, H.; Shinde, Y.; Jagtap, V.; Girase, R.; Patel, H. Optimizing Bedaquiline for cardiotoxicity by structure based virtual screening, DFT analysis and molecular dynamic simulation studies to identify selective MDR-TB inhibitors. *Silico Pharmacol.* **2021**, *9*, 23. [CrossRef] [PubMed]

110. Sanabria-Chanaga, E.E.; Betancourt-Conde, I.; Hernández-Campos, A.; Téllez-Valencia, A.; Castillo, R. In silico hit optimization toward AKT inhibition: Fragment-based approach, molecular docking and molecular dynamics study. *J. Biomol. Struct. Dyn.* **2019**, *37*, 4301–4311. [CrossRef] [PubMed]

111. Zhang, C.; Feng, L.-J.; Huang, Y.; Wu, D.; Li, Z.; Zhou, Q.; Wu, Y.; Luo, H.B. Discovery of Novel Phosphodiesterase-2A Inhibitors by Structure-Based Virtual Screening, Structural Optimization, and Bioassay. *J. Chem. Inf. Model.* **2017**, *57*, 355–364. [CrossRef] [PubMed]

112. Ruskamo, S.; Yadav, R.P.; Sharma, S.; Lehtimäki, M.; Laulumaa, S.; Aggarwal, S.; Simons, M.; Bürck, J.; Ulrich, A.S.; Juffer, A.H. Atomic resolution view into the structure–function relationships of the human myelin peripheral membrane protein P2. *Acta Crystallogr. Sect. D: Biol.Crystallogr.* **2014**, *70*, 165–176. [CrossRef]

113. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A.E.; Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016**, *116*, 7898–7936. [CrossRef]

114. Singh, N.; Li, W. Recent advances in coarse-grained models for biomolecules and their applications. *Int. J. Mol. Sci.* **2019**, *20*, 3774. [CrossRef]

115. Vicatos, S.; Rychkova, A.; Mukherjee, S.; Warshel, A. An effective Coarse-grained model for biological simulations: Recent refinements and validations. *Proteins: Struct. Funct. Bioinform.* **2014**, *82*, 1168–1185. [CrossRef]

116. Bai, C.; Wang, J.; Chen, G.; Zhang, H.; An, K.; Xu, P.; Du, Y.; Ye, R.D.; Saha, A.; Zhang, A.; et al. Predicting Mutational Effects on Receptor Binding of the Spike Protein of SARS-CoV-2 Variants. *J. Am. Chem. Soc.* **2021**, *143*, 17646–17654. [CrossRef]

117. Bai, C.; Wang, J.; Mondal, D.; Du, Y.; Ye, R.D.; Warshel, A. Exploring the Activation Process of the β2AR-Gs Complex. *J. Am. Chem. Soc.* **2021**, *143*, 11044–11051. [CrossRef]

118. Mukherjee, S.; Warshel, A. Electrostatic origin of the mechanochemical rotary mechanism and the catalytic dwell of F1-ATPase. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20550–20555. [CrossRef]

119. Bai, C.; Warshel, A. Revisiting the protomotive vectorial motion of F₀-ATPase. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19484–19489. [CrossRef] [PubMed]

120. Zhang, Z.; Thirumalai, D. Dissecting the kinematics of the kinesin step. *Structure* **2012**, *20*, 628–640. [CrossRef] [PubMed]

121. Warshel, A.; Sharma, P.K.; Kato, M.; Parson, W.W. Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta Proteins Proteom.* **2006**, *1764*, 1647–1676. [CrossRef] [PubMed]

122. Marrone, T.J.; Briggs, A.J.M.; McCammon, J.A. Structure-based drug design: Computational advances. *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71–90. [CrossRef] [PubMed]

123. Raha, K.; Peters, M.B.; Wang, B.; Yu, N.; Wollacott, A.M.; Westerhoff, L.M.; Merz, K.M. The role of quantum mechanics in structure-based drug design. *Drug Discov. Today* **2007**, *12*, 725–731. [CrossRef]

124. Enyedy, I.J.; Egan, W.J. Can we use docking and scoring for hit-to-lead optimization? *J. Comput. Aided Mol. Des.* **2008**, *22*, 161–168. [CrossRef]

125. Guterres, H.; Im, W. Improving protein-ligand docking results with high-throughput molecular dynamics simulations. *J. Chem. Inf. Model.* **2020**, *60*, 2189–2198. [CrossRef]

126. Berishvili, V.; Kuimov, A.; Voronkov, A.; Radchenko, E.; Kumar, P.; Choonara, Y.; Pillay, V.; Kamal, A.; Palyulin, V. Discovery of novel tankyrase inhibitors through molecular docking-based virtual screening and molecular dynamics simulation studies. *Molecules* **2020**, *25*, 3171. [CrossRef]

127. Halgren, T.A.; Damm, W. Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242. [CrossRef]

128. Arodola, O.A.; Soliman, M.E. Quantum mechanics implementation in drug-design workflows: Does it really help? *Drug Des. Dev. Ther.* **2017**, *11*, 2551. [CrossRef]

129. Ribeiro, A.J.; Santos-Martins, D.; Russo, N.; Ramos, M.J.; Fernandes, P.A. Enzymatic flexibility and reaction rate: A QM/MM study of HIV-1 protease. *ACS Catal.* **2015**, *5*, 5617–5626. [CrossRef]

130. Uddin, N.; Ahmed, S.; Khan, A.; Hoque, M.M.; Halim, M.A. Halogenated derivatives of methotrexate as human dihydrofolate reductase inhibitors in cancer chemotherapy. *J. Biomol. Struct. Dyn.* **2019**, *38*, 901–917. [CrossRef] [PubMed]

131. Nakliang, P.; Lazim, R.; Chang, H.; Choi, S. Multiscale molecular modeling in G protein-coupled receptor (GPCR)-ligand studies. *Biomolecules* **2020**, *10*, 631. [CrossRef]

132. Zhou, Y.; Wang, S.; Zhang, Y. Catalytic reaction mechanism of acetylcholinesterase determined by Born-Oppenheimer ab initio QM/MM molecular dynamics simulations. *J. Phys. Chem. B* **2010**, *114*, 8817–8825. [CrossRef] [PubMed]

133. Chen, X.; Fang, L.; Liu, J.; Zhan, C.-G. Reaction pathway and free energy profile for butyrylcholinesterase-catalyzed hydrolysis of acetylcholine. *J. Phys. Chem. B* **2011**, *115*, 1315–1322. [CrossRef]

134. Cheng, Y.; Cheng, X.; Radić, Z.; McCammon, J.A. Acetylcholinesterase: Mechanisms of covalent inhibition of wild-type and H447I mutant determined by computational analyses. *J. Am. Chem. Soc.* **2007**, *129*, 6562–6570. [CrossRef] [PubMed]

135. Liu, J.; Zhang, Y.; Zhan, C.-G. Reaction pathway and free-energy barrier for reactivation of dimethylphosphoryl-inhibited human acetylcholinesterase. *J. Phys. Chem. B* **2009**, *113*, 16226–16236. [CrossRef]

136. Olivieri, L.; Gardebien, F. Structure-affinity properties of a high-affinity ligand of FKBP12 studied by molecular simulations of a binding intermediate. *PLoS ONE* **2014**, *9*, e114610. [CrossRef]

137. Tosso, R.D.; Andujar, S.A.; Gutierrez, L.; Angelina, E.; Rodriguez, R.; Nogueras, M.; Baldoni, H.; Suvire, F.D.; Cobo, J.; Enriz, R.D. Molecular modeling study of dihydrofolate reductase inhibitors. Molecular dynamics simulations, quantum mechanical calculations, and experimental corroboration. *J. Chem. Inf. Model.* **2013**, *53*, 2018–2032. [CrossRef]

138. Cho, A.E.; Guallar, V.; Berne, B.J.; Friesner, R. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26*, 915–931. [CrossRef]

139. Raha, K.; Merz, K.M. Large-scale validation of a quantum mechanics based scoring function: Predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J. Med. Chem.* **2005**, *48*, 4558–4575. [CrossRef] [PubMed]

140. Gleeson, M.P.; Gleeson, D. QM/MM calculations in drug discovery: A useful method for studying binding phenomena? *J. Chem. Inf. Model.* **2009**, *49*, 670–677. [CrossRef] [PubMed]

141. Zhang, X.; Zhao, Y.; Lu, G. Recent development in quantum mechanics/molecular mechanics modeling for materials. *Int. J. Multiscale Comput. Eng.* **2012**, *10*, 65–82. [CrossRef]

142. Cavasotto, C.N.; Adler, N.S.; Aucar, M.G. Quantum chemical approaches in structure-based virtual screening and lead optimization. *Front. Chem.* **2018**, *6*, 188. [CrossRef] [PubMed]

143. Fong, P.; McNamara, J.P.; Hillier, I.H.; Bryce, R.A. Assessment of QM/MM scoring functions for molecular docking to HIV-1 protease. *J. Chem. Inf. Model.* **2009**, *49*, 913–924. [CrossRef]

144. Kim, M.; Cho, A.E. Incorporating QM and solvation into docking for applications to GPCR targets. *Phys. Chem. Chem. Phys.* **2016**, *18*, 28281–28289. [CrossRef] [PubMed]

145. Chaskar, P.; Zoete, V.; Röhrig, U.F. On-the-Fly QM/MM Docking with Attracting Cavities. *J. Chem. Inf. Model.* **2017**, *57*, 73–84. [CrossRef] [PubMed]

146. Whitfield, J.D.; Love, P.J.; Aspuru-Guzik, A. Computational complexity in electronic structure. *Phys. Chem. Chem. Phys.* **2013**, *15*, 397–411. [CrossRef]

147. Orús, R.; Latorre, J.I. Universality of entanglement and quantum-computation complexity. *Phys. Rev. A* **2004**, *69*, 052308. [CrossRef]

148. Senn, H.M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229. [CrossRef]

149. Wilson, E.; Vant, J.; Layton, J.; Boyd, R.; Lee, H.; Turilli, M.; Hernández, B.; Wilkinson, S.; Jha, S.; Gupta, C. Large-Scale Molecular Dynamics Simulations of Cellular Compartments. In *Structure and Function of Membrane Proteins*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 335–356.

150. Hoque, I.; Chatterjee, A.; Bhattacharya, S.; Biswas, R. An approach of computer-aided drug design (CADD) tools for in silico pharmaceutical drug design and development. *Int. J. Adv. Res. Biol. Sci.* **2017**, *4*, 60–71. [CrossRef]

151. McKenna, F.; Martin, M.; Bird, H.; Wright, V. Captopril. *Br. Med. J.* **1983**, *287*, 1299. [CrossRef] [PubMed]

152. Dos Santos Nascimento, I.J.; De Aquino, T.M.; Da Silva-Júnior, E.F. Drug repurposing: A strategy for discovering inhibitors against emerging viral infections. *Curr. Med. Chem.* **2021**, *28*, 2887–2942. [CrossRef]

153. Huang, H.-J.; Yu, H.W.; Chen, C.-Y.; Hsu, C.-H.; Chen, H.-Y.; Lee, K.-J.; Tsai, F.-J.; Chen, C.Y.-C. Current developments of computer-aided drug design. *J. Taiwan Instit. Chem. Eng.* **2010**, *41*, 623–635. [CrossRef]

154. Druker, B.J.; Lydon, N.B. Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J. Clin. Investig.* **2000**, *105*, 3–7. [CrossRef] [PubMed]

155. Van Drie, J.H. Computer-aided drug design: The next 20 years. *J. Comput. Aided Mol. Des.* **2007**, *21*, 591–601. [CrossRef]

156. Athanasiou, C.; Cournia, Z. From computers to bedside: Computational chemistry contributing to FDA approval. In *Biomolecular Simulations in Structure-Based Drug Discover*; Gervasio, F.L., Spiwok, V., Eds.; WILEY-VCH: Weinheim, Germany, 2018; Volume 75, pp. 163–203.

157. Macalino, S.J.Y.; Basith, S.; Clavio, N.A.B.; Chang, H.; Kang, S.; Choi, S. Evolution of in silico strategies for protein-protein interaction drug discovery. *Molecules* **2018**, *23*, 1963. [CrossRef]

158. Wang, X.; Song, K.; Li, L.; Chen, L. Structure-based drug design strategies and challenges. *Curr. Top. Med. Chem.* **2018**, *18*, 998–1006. [CrossRef]

159. Lerner, E.; Barth, A.; Hendrix, J.; Ambrose, B.; Birkedal, V.; Blanchard, S.C.; Börner, R.; Chung, H.S.; Cordes, T.; Craggs, T.D. FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *Elife* **2021**, *10*, e60416. [CrossRef]

160. Lee, J.; Freddolino, P.L.; Zhang, Y. Ab initio protein structure prediction. In *From Protein Structure to Function with Bioinformatics*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 3–35.

161. Pan, L.; Gardner, C.L.; Pagliai, F.A.; Gonzalez, C.F.; Lorca, G.L. Identification of the tolfenamic acid binding pocket in PrbP from Liberibacter asiaticus. *Front. Microbiol.* **2017**, *8*, 1591. [CrossRef]

162. Hetényi, C.; van der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.* **2006**, *580*, 1447–1450. [CrossRef] [PubMed]

163. Hassan, N.M.; Alhossary, A.A.; Mu, Y.; Kwoh, C.-K. Protein-ligand blind docking using QuickVina-W with inter-process spatio-temporal integration. *Sci. Rep.* **2017**, *7*, 15451. [CrossRef] [PubMed]

164. Liu, Y.; Grimm, M.; Dai, W.-t.; Hou, M.-c.; Xiao, Z.-X.; Cao, Y. CB-Dock: A web server for cavity detection-guided protein-ligand blind docking. *Acta Pharmacol. Sin.* **2020**, *41*, 138–144. [CrossRef] [PubMed]

165. Hetényi, C.; van der Spoel, D. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **2002**, *11*, 1729–1737. [CrossRef] [PubMed]

166. Iorga, B.; Herlem, D.; Barré, E.; Guillou, C. Acetylcholine nicotinic receptors: Finding the putative binding site of allosteric modulators using the "blind docking" approach. *J. Mol. Model.* **2006**, *12*, 366–372. [CrossRef]

167. Liang, J.; Karagiannis, C.; Pitsillou, E.; Darmawan, K.K.; Ng, K.; Hung, A.; Karagiannis, T.C. Site mapping and small molecule blind docking reveal a possible target site on the SARS-CoV-2 main protease dimer interface. *Comput. Biol. Chem.* **2020**, *89*, 107372. [CrossRef] [PubMed]

168. Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A.S.; De Fabritiis, G. DeepSite: Protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics* **2017**, *33*, 3036–3042. [CrossRef]

169. Volkamer, A.; Kuhn, D.; Grombacher, T.; Rippmann, F.; Rarey, M. Combining global and local measures for structure-based druggability predictions. *J. Chem. Inf. Model.* **2012**, *52*, 360–372. [CrossRef]

170. Yu, J.; Zhou, Y.; Tanaka, I.; Yao, M. Roll: A new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics* **2010**, *26*, 46–52. [CrossRef]

171. Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tufféry, P. Fpocket: Online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res.* **2010**, *38*, W582–W589. [CrossRef]

172. Bruno, C.; Cavalluzzi, M.M.; Rusciano, M.R.; Lovece, A.; Carrieri, A.; Pracella, R.; Giannuzzi, G.; Polimeno, L.; Viale, M.; Illario, M. The chemosensitizing agent lubeluzole binds calmodulin and inhibits Ca^{2+}/calmodulin-dependent kinase II. *Eur. J. Med. Chem.* **2016**, *116*, 36–45. [CrossRef] [PubMed]

173. Wang, S.; Li, W.; Liu, S.; Xu, J. RaptorX-Property: A web server for protein structure property prediction. *Nucleic Acids Res.* **2016**, *44*, W430–W435. [CrossRef] [PubMed]

174. Yang, J.; Roy, A.; Zhang, Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **2013**, *29*, 2588–2595. [CrossRef] [PubMed]

175. Kalidas, Y.; Chandra, N. PocketDepth: A new depth based algorithm for identification of ligand binding sites in proteins. *J. Struct. Biol.* **2008**, *161*, 31–42. [CrossRef] [PubMed]

176. Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.-F.; Fischer, P.D.; Coote, P.W.; Das, K.M.P.; Malets, Y.S.; Radchenko, D.S.; Moroz, Y.S.; Scott, D.A. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **2020**, *580*, 663–668. [CrossRef] [PubMed]

177. Irwin, J.J.; Sterling, T.; Mysinger, M.M.; Bolstad, E.S.; Coleman, R.G. ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model* **2012**, *52*, 1757–1768. [CrossRef]

178. Sterling, T.; Irwin, J.J. ZINC 15—Ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337. [CrossRef] [PubMed]

179. Kiss, R.; Sandor, M.; Szalai, F.A. http://Mcule.com: A public web service for drug discovery. *J. Cheminform.* **2012**, *4*, P17. [CrossRef]

180. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A. PubChem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [CrossRef]

181. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109. [CrossRef]

182. Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901–D906. [CrossRef] [PubMed]

183. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [CrossRef] [PubMed]

184. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [CrossRef] [PubMed]

185. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945. [CrossRef] [PubMed]

186. Chen, J.; Swamidass, S.J.; Dou, Y.; Bruand, J.; Baldi, P. ChemDB: A public database of small molecules and related chemoinformatics resources. *Bioinformatics* **2005**, *21*, 4133–4139. [CrossRef]

187. Lipinski, C.A. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today: Technol.* **2004**, *1*, 337–341. [CrossRef]

188. Whitty, A.; Zhong, M.; Viarengo, L.; Beglov, D.; Hall, D.R.; Vajda, S. Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs. *Drug Discov. Today* **2016**, *21*, 712–717. [CrossRef]

189. Huang, S.-Y.; Zou, X. Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* **2010**, *11*, 3016–3034. [CrossRef]

190. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [CrossRef]

191. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef]

192. Wu, G.; Robertson, D.H.; Brooks, C.L., III; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549–1562. [CrossRef] [PubMed]

193. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [CrossRef] [PubMed]

194. Allen, W.J.; Balius, T.E.; Mukherjee, S.; Brozell, S.R.; Moustakas, D.T.; Lang, P.T.; Case, D.A.; Kuntz, I.D.; Rizzo, R.C. DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015**, *36*, 1132–1156. [CrossRef]

195. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748. [CrossRef] [PubMed]

196. Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins: Struct. Funct. Bioinform.* **1999**, *37*, 228–241. [CrossRef]

197. Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **2011**, *39*, W270–W277. [CrossRef] [PubMed]

198. Berry, M.; Fielding, B.; Gamieldien, J. Practical considerations in virtual screening and molecular docking. *Emerg. Trends Comput. Biol. Bioinform. Syst. Biol.* **2015**, *1762*, 487–502. [CrossRef]

199. Yadava, U. Search algorithms and scoring methods in protein-ligand docking. *Endocrinol. Metab. Int. J.* **2018**, *6*, 359–367. [CrossRef]

200. Hart, T.N.; Read, R.J. A multiple-start Monte Carlo docking method. *Proteins: Struct. Funct. Bioinform.* **1992**, *13*, 206–222. [CrossRef]

201. Oshiro, C.M.; Kuntz, I.D.; Dixon, J.S. Flexible ligand docking using a genetic algorithm. *J. Comput. Aided Mol. Des.* **1995**, *9*, 113–130. [CrossRef]

202. Li, J.; Fu, A.; Zhang, L. An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdiscip. Sci.: Comput. Life Sci.* **2019**, *11*, 320–328. [CrossRef] [PubMed]

203. Stärk, H.; Ganea, O.; Pattanaik, L.; Barzilay, R.; Jaakkola, T. Equibind: Geometric deep learning for drug binding structure prediction. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 20503–20521.

204. McNutt, A.T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; Koes, D.R. GNINA 1.0: Molecular docking with deep learning. *J. Cheminformatics* **2021**, *13*, 43. [CrossRef] [PubMed]

205. Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. *arXiv* **2022**, arXiv:2210.01776.

206. Salsbury, F.R., Jr. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr. Opin. Pharmacol.* **2010**, *10*, 738–744. [CrossRef] [PubMed]

207. Ferreira, L.G.; Dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular docking and structure-based drug design strategies. *Molecules* **2015**, *20*, 13384–13421. [CrossRef] [PubMed]

208. Śledź, P.; Caflisch, A. Protein structure-based drug design: From docking to molecular dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93–102. [CrossRef]

209. Philip, P.; Anshuman, D.; Anil, K.S. Computer-aided drug design: Integration of structure-based and ligand-based approaches in drug design. *Curr. Comput. Aided Drug Des.* **2007**, *3*, 133–148. [CrossRef]

210. Schaller, D.; Šribar, D.; Noonan, T.; Deng, L.; Nguyen, T.N.; Pach, S.; Machalz, D.; Bermudez, M.; Wolber, G. Next generation 3D pharmacophore modeling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, e1468. [CrossRef]

211. Van Drie, J.H. Generation of three-dimensional pharmacophore models. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 449–464. [CrossRef]

212. Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: Challenges and recent advances. *Drug Discov. Today* **2010**, *15*, 444–450. [CrossRef]

213. Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563–571. [CrossRef] [PubMed]

214. Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 160–169. [CrossRef]

215. Chen, I.-J.; Foloppe, N. Conformational sampling of druglike molecules with MOE and catalyst: Implications for pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* **2008**, *48*, 1773–1791. [CrossRef] [PubMed]

216. Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper server: A web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* **2010**, *38*, W609–W614. [CrossRef] [PubMed]

217. Schneidman-Duhovny, D.; Dror, O.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PharmaGist: A webserver for ligand-based pharmacophore detection. *Nucleic Acids Res.* **2008**, *36*, W223–W228. [CrossRef] [PubMed]

218. Dixon, S.L.; Smondyrev, A.M.; Rao, S.N. PHASE: A novel approach to pharmacophore modeling and 3D database searching. *Chem. Biol. Drug Des.* **2006**, *67*, 370–372. [CrossRef]

219. Mallik, B.; Morikis, D. Development of a quasi-dynamic pharmacophore model for anti-complement peptide analogues. *J. Am. Chem. Soc.* **2005**, *127*, 10967–10976. [CrossRef]

220. Langer, T.; Wolber, G. Pharmacophore definition and 3D searches. *Drug Discov. Today: Technol.* **2004**, *1*, 203–207. [CrossRef]

221. Melo-Filho, C.C.; Braga, R.C.; Andrade, C.H. 3D-QSAR approaches in drug design: Perspectives to generate reliable CoMFA models. *Curr. Comput. Aided Drug Des.* **2014**, *10*, 148–159. [CrossRef]

222. Sydow, D. *Dynophores: Novel Dynamic Pharmacophores*; Humboldt-Universität zu Berlin, Lebenswissenschaftliche Fakultät: Berlin, Germany, 2015.

223. Verma, J.; Khedkar, V.M.; Coutinho, E.C. 3D-QSAR in drug design—A review. *Curr. Top. Med. Chem.* **2010**, *10*, 95–115. [CrossRef]

224. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R. QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977–5010. [CrossRef] [PubMed]

225. Patel, H.M.; Noolvi, M.N.; Sharma, P.; Jaiswal, V.; Bansal, S.; Lohan, S.; Kumar, S.S.; Abbot, V.; Dhiman, S.; Bhardwaj, V. Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Med. Chem. Res.* **2014**, *23*, 4991–5007. [CrossRef]

226. Wang, Y.-L.; Wang, F.; Shi, X.-X.; Jia, C.-Y.; Wu, F.-X.; Hao, G.-F.; Yang, G.-F. Cloud 3D-QSAR: A web tool for the development of quantitative structure–activity relationship models in drug discovery. *Brief. Bioinform.* **2020**, *22*, bbaa276. [CrossRef] [PubMed]

227. Martins, J.P.A.; de Oliveira, M.A.R.; de Queiroz, M.S.O. Web-4D-QSAR: A web-based application to generate 4D-QSAR descriptors. *J. Comput. Chem.* **2018**, *39*, 917–924. [CrossRef]

228. Soufan, O.; Ba-Alawi, W.; Magana-Mora, A.; Essack, M.; Bajic, V.B. DPubChem: A web tool for QSAR modeling and high-throughput virtual screening. *Sci. Rep.* **2018**, *8*, 9110. [CrossRef] [PubMed]

229. Zang, C.; Wang, F. MoFlow: An invertible flow model for generating molecular graphs. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, San Diego, CA, USA, 6–10 July 2020; pp. 617–626.

230. Jin, W.; Barzilay, R.; Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. In Proceedings of the 35th International Conference on MachineLearning, Stockholm, Sweden, 10–15 July 2018; pp. 2323–2332.

231. Gómez-Bombarelli, R.; Wei, J.N.; Duvenaud, D.; Hernández-Lobato, J.M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T.D.; Adams, R.P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276. [CrossRef]

232. Wang, M.; Wang, Z.; Sun, H.; Wang, J.; Shen, C.; Weng, G.; Chai, X.; Li, H.; Cao, D.; Hou, T. Deep learning approaches for de novo drug design: An overview. *Curr. Opin. Struct. Biol.* **2022**, *72*, 135–144. [CrossRef]

233. Chenthamarakshan, V.; Das, P.; Hoffman, S.; Strobelt, H.; Padhi, I.; Lim, K.W.; Hoover, B.; Manica, M.; Born, J.; Laino, T. CogMol: Target-specific and selective drug design for COVID-19 using deep generative models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 4320–4332.

234. Krishnan, S.R.; Bung, N.; Bulusu, G.; Roy, A. Accelerating de novo drug design against novel proteins using deep learning. *J. Chem. Inf. Model.* **2021**, *61*, 621–630. [CrossRef]

235. Coley, C.W.; Green, W.H.; Jensen, K.F. Machine learning in computer-aided synthesis planning. *Acc. Chem. Res.* **2018**, *51*, 1281–1289. [CrossRef]

236. Luo, Y.; Yan, K.; Ji, S. Graphdf: A discrete flow model for molecular graph generation. In Proceedings of the 38th International Conference on Machine Learning, Online, 18–24 July 2021; pp. 7192–7203.

237. Jin, W.; Barzilay, R.; Jaakkola, T. Hierarchical generation of molecular graphs using structural motifs. In Proceedings of the 37th International Conference on MachineLearning, Online, 13–18 July 2020; pp. 4839–4848.

238. Jin, W.; Barzilay, R.; Jaakkola, T. Multi-objective molecule generation using interpretable substructures. In Proceedings of the 37th International Conference on MachineLearning, Online, 13–18 July 2020; pp. 4849–4859.

239. Li, Y.; Hu, J.; Wang, Y.; Zhou, J.; Zhang, L.; Liu, Z. Deepscaffold: A comprehensive tool for scaffold-based de novo drug discovery using deep learning. *J. Chem. Inf. Model.* **2019**, *60*, 77–91. [CrossRef] [PubMed]

240. Luo, S.; Guan, J.; Ma, J.; Peng, J. A 3D generative model for structure-based drug design. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 6229–6239.

241. Pandey, M.; Fernandez, M.; Gentile, F.; Isayev, O.; Tropsha, A.; Stern, A.C.; Cherkasov, A. The transformational role of GPU computing and deep learning in drug discovery. *Nat. Mach. Intell.* **2022**, *4*, 211–221. [CrossRef]

242. Zhavoronkov, A.; Ivanenkov, Y.A.; Aliper, A.; Veselov, M.S.; Aladinskiy, V.A.; Aladinskaya, A.V.; Terentiev, V.A.; Polykovskiy, D.A.; Kuznetsov, M.D.; Asadulaev, A. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **2019**, *37*, 1038–1040. [CrossRef]

243. Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **2018**, *4*, eaap7885. [CrossRef]

244. Tong, X.; Liu, X.; Tan, X.; Li, X.; Jiang, J.; Xiong, Z.; Xu, T.; Jiang, H.; Qiao, N.; Zheng, M. Generative models for De Novo drug design. *J. Med. Chem.* **2021**, *64*, 14011–14027. [CrossRef]

245. Kuznetsov, M.; Polykovskiy, D. MolGrow: A graph normalizing flow for hierarchical molecular generation. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; pp. 8226–8234.

246. Bilodeau, C.; Jin, W.; Jaakkola, T.; Barzilay, R.; Jensen, K.F. Generative models for molecular discovery: Recent advances and challenges. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2022**, *12*, e1608. [CrossRef]

247. Adams, K.; Coley, C.W. Equivariant Shape-Conditioned Generation of 3D Molecules for Ligand-Based Drug Design. *arXiv* **2022**, arXiv:2210.04893.

248. Batool, M.; Ahmad, B.; Choi, S. A structure-based drug discovery paradigm. *Int. J. Mol. Sci.* **2019**, *20*, 2783. [CrossRef]

249. Sattarov, B.; Baskin, I.I.; Horvath, D.; Marcou, G.G.; Bjerrum, E.J.; Varnek, A. De novo molecular design by combining deep autoencoder recurrent neural networks with generative topographic mapping. *J. Chem. Inf. Model.* **2019**, *59*, 1182–1196. [CrossRef]

250. Li, Y.; Pei, J.; Lai, L. Structure-based de novo drug design using 3D deep generative models. *Chem. Sci.* **2021**, *12*, 13664–13675. [CrossRef] [PubMed]

251. Wang, M.; Hsieh, C.-Y.; Wang, J.; Wang, D.; Weng, G.; Shen, C.; Yao, X.; Bing, Z.; Li, H.; Cao, D. RELATION: A Deep Generative Model for Structure-Based De Novo Drug Design. *J. Med. Chem.* **2022**, *65*, 9478–9492. [CrossRef] [PubMed]

252. Wang, M.; Sun, H.; Wang, J.; Pang, J.; Chai, X.; Xu, L.; Li, H.; Cao, D.; Hou, T. Comprehensive assessment of deep generative architectures for de novo drug design. *Brief. Bioinform.* **2022**, *23*, bbab544. [CrossRef]

253. Krishnan, S.R.; Bung, N.; Vangala, S.R.; Srinivasan, R.; Bulusu, G.; Roy, A. De novo structure-based drug design using deep learning. *J. Chem. Inf. Model.* **2021**. [CrossRef] [PubMed]

254. Gebauer, N.W.A.; Gastegger, M.; Hessmann, S.S.P.; Müller, K.-R.; Schütt, K.T. Inverse design of 3d molecular structures with conditional generative neural networks. *Nat. Commun.* **2022**, *13*, 973. [CrossRef]

255. Peng, X.; Luo, S.; Guan, J.; Xie, Q.; Peng, J.; Ma, J. Pocket2Mol: Efficient Molecular Sampling Based on 3D Protein Pockets. *arXiv* **2022**, arXiv:2205.07249.

256. Xie, W.; Wang, F.; Li, Y.; Lai, L.; Pei, J. Advances and Challenges in De Novo Drug Design Using Three-Dimensional Deep Generative Models. *J. Chem. Inf. Model.* **2022**, *62*, 2269–2279. [CrossRef]

257. Thomas, M.; Smith, R.T.; O'Boyle, N.M.; De Graaf, C.; Bender, A. Comparison of structure-and ligand-based scoring functions for deep generative models: A GPCR case study. *J. Cheminformatics* **2021**, *13*, 39. [CrossRef]

258. Bengio, E.; Jain, M.; Korablyov, M.; Precup, D.; Bengio, Y. Flow network based generative models for non-iterative diverse candidate generation. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 27381–27394. [CrossRef]

259. Segler, M.H.; Kogej, T.; Tyrchan, C.; Waller, M.P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **2018**, *4*, 120–131. [CrossRef]

260. Chen, H. Can Generative-Model-Based Drug Design Become a New Normal in Drug Discovery? *J. Med. Chem.* **2021**, *65*, 100–102. [CrossRef] [PubMed]

261. Lewell, X.Q.; Judd, D.B.; Watson, S.P.; Hann, M.M. Recap retrosynthetic combinatorial analysis procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522. [CrossRef] [PubMed]

262. Vinkers, H.M.; de Jonge, M.; Daeyaert, F.F.; Heeres, J.; Koymans, L.M.; van Lenthe, J.H.; Lewi, P.J.; Timmerman, H.; Van Aken, K.; Janssen, P.A. SYNOPSIS: SYNthesize and OPtimize system in silico. *J. Med. Chem.* **2003**, *46*, 2765–2773. [CrossRef]

263. Schneider, P.; Walters, W.P.; Plowright, A.T.; Sieroka, N.; Listgarten, J.; Goodnow, R.A.; Fisher, J.; Jansen, J.M.; Duca, J.S.; Rush, T.S. Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **2020**, *19*, 353–364. [CrossRef] [PubMed]

264. Rifaioglu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Doğan, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases. *Brief. Bioinform.* **2019**, *20*, 1878–1912. [CrossRef]

265. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [CrossRef]

266. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [CrossRef]

267. Szklarczyk, D.; Santos, A.; Von Mering, C.; Jensen, L.J.; Bork, P.; Kuhn, M. STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **2016**, *44*, D380–D384. [CrossRef]

268. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [CrossRef]

269. Kuhn, M.; Letunic, I.; Jensen, L.J.; Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **2016**, *44*, D1075–D1079. [CrossRef]

270. Liu, Y.; Wei, Q.; Yu, G.; Gai, W.; Li, Y.; Chen, X. DCDB 2.0: A major update of the drug combination database. *Database* **2014**, *2014*, bau124. [CrossRef] [PubMed]

271. Fink, T.; Reymond, J.-L. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: Assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *J. Chem. Inf. Model.* **2007**, *47*, 342–353. [CrossRef]

272. Blum, L.C.; Reymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* **2009**, *131*, 8732–8733. [CrossRef] [PubMed]

273. Ruddigkeit, L.; Van Deursen, R.; Blum, L.C.; Reymond, J.L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875. [CrossRef] [PubMed]

274. Kim, M.; Park, K.; Kim, W.; Jung, S.; Cho, A.E. Target-specific drug design method combining deep learning and water pharmacophore. *J. Chem. Inf. Model.* **2020**, *61*, 36–45. [CrossRef] [PubMed]

275. Mouchlis, V.D.; Afantitis, A.; Serra, A.; Fratello, M.; Papadiamantis, A.G.; Aidinis, V.; Lynch, I.; Greco, D.; Melagraki, G. Advances in de novo drug design: From Conventional to Machine Learning Methods. *Int. J. Mol. Sci.* **2021**, *22*, 1676. [CrossRef]

276. Gupta, A.; Müller, A.T.; Huisman, B.J.; Fuchs, J.A.; Schneider, P.; Schneider, G. Generative recurrent networks for de novo drug design. *Mol. Inf.* **2018**, *37*, 1700111. [CrossRef]

277. Arús-Pous, J.; Patronov, A.; Bjerrum, E.J.; Tyrchan, C.; Reymond, J.-L.; Chen, H.; Engkvist, O. SMILES-based deep generative scaffold decorator for de-novo drug design. *J. Cheminform.* **2020**, *12*, 38. [CrossRef]

278. Liu, X.; IJzerman, A.P.; van Westen, G.J. Computational approaches for de novo drug design: Past, present, and future. *Artif. Neural Netw.* **2021**, 139–165.

279. Mandhana, V.; Taware, R. De novo drug design using self attention mechanism. In Proceedings of the 35th Annual ACM Symposium on Applied Computing, Online, 30 March–3 April 2020; pp. 8–12.

280. Graves, A. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45. [CrossRef]

281. Yasonik, J. Multiobjective de novo drug design with recurrent neural networks and nondominated sorting. *J. Cheminform.* **2020**, *12*, 14. [CrossRef]

282. Arús-Pous, J.; Johansson, S.V.; Prykhodko, O.; Bjerrum, E.J.; Tyrchan, C.; Reymond, J.-L.; Chen, H.; Engkvist, O. Randomized SMILES strings improve the quality of molecular generative models. *J. Cheminform.* **2019**, *11*, 71. [CrossRef] [PubMed]

283. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.

284. Zheng, S.; Yan, X.; Gu, Q.; Yang, Y.; Du, Y.; Lu, Y.; Xu, J. QBMG: Quasi-biogenic molecule generator with deep recurrent neural network. *J. Cheminform.* **2019**, *11*, 5. [CrossRef] [PubMed]

285. Barshatski, G.; Radinsky, K. Unpaired Generative Molecule-to-Molecule Translation for Lead Optimization. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Online, 14–18 August 2021; pp. 2554–2564.

286. Rezaei, M.A.; Li, Y.; Wu, D.; Li, X.; Li, C. Deep learning in drug design: Protein-ligand binding affinity prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *19*, 407–417. [CrossRef] [PubMed]

287. Francoeur, P.G.; Masuda, T.; Sunseri, J.; Jia, A.; Iovanisci, R.B.; Snyder, I.; Koes, D.R. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *J. Chem. Inf. Model.* **2020**, *60*, 4200–4215. [CrossRef]

288. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 1–11.

289. Cheng, Z.; Yan, C.; Wu, F.-X.; Wang, J. Drug-target interaction prediction using multi-head self-attention and graph attention network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *19*, 2208–2218. [CrossRef]

290. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural message passing for quantum chemistry. In Proceedings of Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1263–1272.

291. Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* **2019**, *63*, 8749–8760. [CrossRef]

292. Griffiths, R.-R.; Hernández-Lobato, J.M. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chem. Sci.* **2020**, *11*, 577–586. [CrossRef]

293. Asperti, A.; Trentin, M. Balancing reconstruction error and Kullback-Leibler divergence in Variational Autoencoders. *IEEE Access* **2020**, *8*, 199440–199448. [CrossRef]

294. Wei, R.; Mahmood, A. Recent advances in variational autoencoders with representation learning for biomedical informatics: A survey. *IEEE Access* **2020**, *9*, 4939–4956. [CrossRef]

295. Liu, Q.; Allamanis, M.; Brockschmidt, M.; Gaunt, A. Constrained graph variational autoencoders for molecule design. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, QC, Canada, 3–8 December 2018.

296. Simonovsky, M.; Komodakis, N. Graphvae: Towards generation of small graphs using variational autoencoders. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; pp. 412–422.

297. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]

298. De Cao, N.; Kipf, T. MolGAN: An implicit generative model for small molecular graphs. *arXiv* **2018**, arXiv:1805.11973.

299. Madhawa, K.; Ishiguro, K.; Nakago, K.; Abe, M. Graphnvp: An invertible flow model for generating molecular graphs. *arXiv* **2019**, arXiv:1905.11600.

300. Papamakarios, G.; Pavlakou, T.; Murray, I. Masked autoregressive flow for density estimation. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 1–10.

301. Xiong, J.; Xiong, Z.; Chen, K.; Jiang, H.; Zheng, M. Graph neural networks for automated de novo drug design. *Drug Discov. Today* **2021**, *26*, 1382–1393. [CrossRef] [PubMed]

302. Stärk, H.; Beaini, D.; Corso, G.; Tossou, P.; Dallago, C.; Günnemann, S.; Liò, P. 3D infomax improves gnns for molecular property prediction. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 20479–20502.

303. Fabian, B.; Edlich, T.; Gaspar, H.; Segler, M.; Meyers, J.; Fiscato, M.; Ahmed, M. Molecular representation learning with language models and domain-relevant auxiliary tasks. *arXiv* **2020**, arXiv:2011.13230.

304. Wang, S.; Guo, Y.; Wang, Y.; Sun, H.; Huang, J. SMILES-BERT: Large scale unsupervised pre-training for molecular property prediction. In Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Niagara Falls, NY, USA, 7–10 September 2019; pp. 429–436.

305. Reidenbach, D.; Livne, M.; Ilango, R.K.; Gill, M.; Israeli, J. Improving Small Molecule Generation using Mutual Information Machine. *arXiv* **2022**, arXiv:2208.09016.

306. Chithrananda, S.; Grand, G.; Ramsundar, B. ChemBERTa: Large-scale self-supervised pretraining for molecular property prediction. *arXiv* **2020**, arXiv:2010.09885.

307. Edwards, C.; Lai, T.; Ros, K.; Honke, G.; Ji, H. Translation between Molecules and Natural Language. *arXiv* **2022**, arXiv:2204.11817.

308. Chen, Z.; Min, M.R.; Parthasarathy, S.; Ning, X. A deep generative model for molecule optimization via one fragment modification. *Nat. Mach. Intell.* **2021**, *3*, 1040–1049. [CrossRef]

309. Button, A.; Merk, D.; Hiss, J.A.; Schneider, G. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. *Nat. Mach. Intell.* **2019**, *1*, 307–315. [CrossRef]

310. Chen, B.; Wang, T.; Li, C.; Dai, H.; Song, L. Molecule optimization by explainable evolution. In Proceedings of the International Conference on Learning Representation (ICLR 2021), Online, 3–7 May 2021; pp. 1–15.

311. Gentile, F.; Agrawal, V.; Hsing, M.; Ton, A.-T.; Ban, F.; Norinder, U.; Gleave, M.E.; Cherkasov, A. Deep docking: A deep learning platform for augmentation of structure based drug discovery. *ACS Cent. Sci.* **2020**, *6*, 939–949. [CrossRef] [PubMed]

312. Chen, H.; Engkvist, O. Has drug design augmented by artificial intelligence become a reality? *Trends Pharmacol. Sci.* **2019**, *40*, 806–809. [CrossRef]

313. Klebe, G. On the validity of popular assumptions in computational drug design. *J. Cheminform.* **2011**, *3*, O18. [CrossRef]

314. Garrido, A.; Lepailleur, A.; Mignani, S.M.; Dallemagne, P.; Rochais, C. hERG toxicity assessment: Useful guidelines for drug design. *Eur. J. Med. Chem.* **2020**, *195*, 112290. [CrossRef] [PubMed]

315. Hessler, G.; Baringhaus, K.-H. Artificial intelligence in drug design. *Molecules* **2018**, *23*, 2520. [CrossRef] [PubMed]

316. Liu, M.; Luo, Y.; Uchino, K.; Maruhashi, K.; Ji, S. Generating 3D Molecules for Target Protein Binding. *arXiv* **2022**, arXiv:2204.09410.

317. Xie, Y.; Shi, C.; Zhou, H.; Yang, Y.; Zhang, W.; Yu, Y.; Li, L. Mars: Markov molecular sampling for multi-objective drug discovery. *arXiv* **2021**, arXiv:2103.10432.

318. Eckmann, P.; Sun, K.; Zhao, B.; Feng, M.; Gilson, M.K.; Yu, R. LIMO: Latent Inceptionism for Targeted Molecule Generation. *arXiv* **2022**, arXiv:2206.09010.

319. Renz, P.; Van Rompaey, D.; Wegner, J.K.; Hochreiter, S.; Klambauer, G. On failure modes in molecule generation and optimization. *Drug Discov. Today: Technol.* **2019**, *32–33*, 55–63. [CrossRef]

320. Battaglia, P.W.; Hamrick, J.B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R. Relational inductive biases, deep learning, and graph networks. *arXiv* **2018**, arXiv:1806.01261.

321. Atz, K.; Grisoni, F.; Schneider, G. Geometric deep learning on molecular representations. *Nat. Mach. Intell.* **2021**, *3*, 1023–1032. [CrossRef]

322. Flam-Shepherd, D.; Zhu, K.; Aspuru-Guzik, A. Language models can learn complex molecular distributions. *Nat. Commun.* **2022**, *13*, 3293. [CrossRef] [PubMed]

323. Lu, W.; Wu, Q.; Zhang, J.; Rao, J.; Li, C.; Zheng, S. TANKBind: Trigonometry-Aware Neural NetworKs for Drug-Protein Binding Structure Prediction. *bioRxiv* **2022**. [CrossRef]

324. Zhou, G.; Gao, Z.; Ding, Q.; Zheng, H.; Xu, H.; Wei, Z.; Zhang, L.; Ke, G. Uni-Mol: A Universal 3D Molecular Representation Learning Framework. *ChemRxiv* **2022**. [CrossRef]

325. Ding, X.; Zhang, B. Computing absolute free energy with deep generative models. *J. Phys. Chem. B* **2020**, *124*, 10166–10172. [CrossRef]

326. Lahey, S.L.J.; Rowley, C.N. Simulating protein–ligand binding with neural network potentials. *Chem. Sci.* **2020**, *11*, 2362–2368. [CrossRef]

327. Zhang, L.; Han, J.; Wang, H.; Car, R.; Weinan, E. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001. [CrossRef]

328. Liu, X.; Feng, H.; Wu, J.; Xia, K. Persistent spectral hypergraph based machine learning (PSH-ML) for protein-ligand binding affinity prediction. *Brief. Bioinform.* **2021**, *22*, bbab127. [CrossRef] [PubMed]

329. Steinbrecher, T.B.; Dahlgren, M.K.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.A.; Friesner, R.; Sherman, W. Accurate binding free energy predictions in fragment optimization. *J. Chem. Inf. Model.* **2015**, *55*, 2411–2420. [CrossRef]

330. Tayara, H.; Abdelbaky, I.; Chong, K.T. Recent omics-based computational methods for COVID-19 drug discovery and repurposing. *Brief. Bioinform.* **2021**, *22*, bbab339. [CrossRef]