Review

Past, Present, and Future Perspectives on Computer-Aided Drug Design Methodologies

Davide Bassani [1,2] and Stefano Moro [2,*]

1   Pharmaceutical Research & Early Development, Roche Innovation Center Basel, F. Hoffmann—La Roche Ltd.,
    4070 Basel, Switzerland; davide.bassani@roche.com
2   Molecular Modeling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences,
    University of Padova, Via Marzolo 5, 35131 Padova, Italy
*   Correspondence: stefano.moro@unipd.it

Abstract: The application of computational approaches in drug discovery has been consolidated in the last decades. These families of techniques are usually grouped under the common name of "computer-aided drug design" (CADD), and they now constitute one of the pillars in the pharmaceutical discovery pipelines in many academic and industrial environments. Their implementation has been demonstrated to tremendously improve the speed of the early discovery steps, allowing for the proficient and rational choice of proper compounds for a desired therapeutic need among the extreme vastness of the drug-like chemical space. Moreover, the application of CADD approaches allows the rationalization of biochemical and interactive processes of pharmaceutical interest at the molecular level. Because of this, computational tools are now extensively used also in the field of rational 3D design and optimization of chemical entities starting from the structural information of the targets, which can be experimentally resolved or can also be obtained with other computer-based techniques. In this work, we revised the state-of-the-art computer-aided drug design methods, focusing on their application in different scenarios of pharmaceutical and biological interest, not only highlighting their great potential and their benefits, but also discussing their actual limitations and eventual weaknesses. This work can be considered a brief overview of computational methods for drug discovery.

Keywords: CADD; computational; chemistry; drug; design; AI; molecular docking; molecular dynamics; learning

1. Introduction: The Benefits of Computational Methods for Drug Discovery

1.1. The Drug Discovery Pipeline and the Problem of Candidate Selection

The drug discovery process is an extremely money- and time-consuming procedure, which is necessary to guarantee the safety and the quality of novel therapeutical entities entering the market. It has been reported that a single novel small molecule can require up to 14 years and more than one billion dollars in the several steps from target assessment to regulatory approval [1,2]. Moreover, the failure risk in the pharmaceutical scenario is known to be one of the highest in the industry. Indeed, it has been estimated that just one or two out of 10,000 screened molecules can effectively become drugs [3].

Another major challenge in the medicinal field is the tremendously extended chemical space forming the "drug-like" environment. It has been calculated that the number of small molecules included in such a concept would be roughly 10^{60} [4], which is a number higher than the seconds of life in the entire Universe. Such a vast chemical space is unfeasible to explore, and this is even more true from an experimental perspective. Indeed, even if the most technological high-throughput screening (HTS) methods today can evaluate the on-target activity of hundreds of thousands of compounds/week [5], their capacity would never reach the order of magnitude of the potential candidates for that specific biological entity. This limitation can be overcome by medicinal chemists by bringing the candidate

selection problem from a laboratory setup to a "virtual environment". Specifically, one of the first ideas that came up was to exploit computers to perform molecular "virtual" screenings before the experimental ones. This approach, which was called "high-throughput virtual screening" (HTVS), still represents one of the main applications of computational methodologies in drug discovery [6]. Indeed, the capacity of the virtual screening depends essentially on the computer power of the infrastructure exploited for the purpose, and it is much faster and cheaper than the preparation and execution of experimental assays. As a demonstration of this, with the actual combinations of software and hardware, the evaluation of several millions of compounds/day is achievable [7–9] (even billions, as stated by the researchers at the Oak Ridge National Laboratory working on the SUMMIT supercomputer [10], which was recently exploited for an ultralarge GPU-accelerated virtual screening against SARS-CoV-2 main protease [11]). In recent decades, many academic and industrial groups have extensively put efforts into the improvement of these methods, making them among of the pillars in the current drug discovery pipeline, especially in the early discovery phases.

1.2. The Application of Computational Methods in Drug Discovery

To go a little bit more into detail, the drug discovery process can be divided into five main steps [12], according to the United States Food and Drug Administration (US FDA). The first is the "discovery and development" phase, which includes hit identification, hit-to-lead (H2L), and lead optimization. The first of these consists of not only highlighting some molecular candidates with a good activity profile against the desired target, but also presenting pharmacokinetic (PK) or pharmacodynamics (PD) limitations. Talking in terms of on-target potency, the hit compounds are usually in the micromolar (μM) range of activity, and they are tendentially not so selective. Even with all these pharmaceutical problems, the hit compounds are very important to give some hints to the drug design teams, being very useful starting points for further modification [13,14].

The second passage consists of the hit-to-lead optimization phase. In this step, the hit compounds are modified with different methodologies to improve their on-target activity and selectivity, always keeping their PK/PD profile under strict evaluation [15]. After this process, the optimized molecular candidates take the name of "lead" compounds and are usually very active (in the nanomolar range for what concerns potency) and reasonably selective. These compounds then enter the second main step, which is the preclinical experimental phase, where they are tested in animal and organoid models to assess their safety and efficacy. The third phase, which is also the longest, consists of human clinical trials. These are divided into three main sections (I, II, and III), each with a different endpoint and with increasing patients enrolled in the tests. Just after a positive outcome of the therapy with the new candidate in the clinical phase III, the commercialization of the drug can be asked the regulatory agencies (e.g., EMEA for Europe and FDA for the USA), determining the opening of the fourth phase of drug development. After this, the fifth and last step consists of post-market drug safety monitoring.

Even if the preclinical and clinical trials constitute the longest and most expensive part of the drug discovery pipeline, not so much can be implemented to reduce them, mainly because of the extremely delicate outcomes in terms of safety and efficacy that they provide. On the other hand, the steps that lead from the hits up to the lead compoundscan be extensively optimized, and that is the space in which computational design approaches have found their main application. Indeed, other than exponentially improving the number of virtual compounds that can be evaluated daily, these methodologies offer also the possibility to deeply analyze the patterns in the chemical data under evaluation; furthermore, they make the rational design of such entities much more than achievable [16]. Indeed, the advances in spectroscopic techniques, together with the tremendous improvement in computer graphics, allowed the visual inspection of proteins, ligands, and biologically relevant complexes in the routines of drug design groups [17]. With such computational approaches, it is possible to effectively design new molecular candidates; for this reason, the

techniques of this family have been gathered under the common name of "computer-aided drug design" (CADD).

1.3. The Main Methodology Branches in CADD

How a computational chemist approaches a pharmacological problem can be multifaceted, but the main underlying factor discriminating the procedure is the quantity of data available on the topic examined (Figure 1). Indeed, one of the most important determinants is the presence of experimental structural information on the target of interest [16], which can currently be obtained using various techniques, among which the most relevant are certainly nuclear magnetic resonance (NMR), X-ray crystallography (XR), and cryogenic electron microscopy (cryo-EM) [18]. If such data are available, then the approach chosen by the scientist usually addresses the application of an ensemble of computational methods, which take advantage of this, such as molecular docking and molecular dynamics. Because of this, those techniques fall under the family of the "structure-based drug design" (SBDD) approaches. On the other hand, when no experimental information about the target three-dimensional structure is available, then two main possibilities are open to CADD scientists. The first consists of searching for close homologs of the target of interest to create a computational model of it (also known as the *homology model*), which would then be tested for structural reliability and used with SBDD techniques [19]. In the latest years, a huge revolution in the field of protein structure prediction was represented by the creation of AlphaFold [20], which is now also available in version 2.0. This algorithm, developed by the company DeepMind, exploits artificial intelligence (AI) approaches to predict the three-dimensional structure of a biological entity on the basis of its sequence, which is also associated with a confidence score in its different functional regions. Another limitation of this method, which still necessitates the homology models created ad hoc by the scientists, is represented by the fact that it predicts only one conformational state (usually the inactive one) of the targets of interest (this was partially resolved with the recent implementation of AphaFill [21]), and that not all the proteins are still included in the AlphaFold database (e.g., many viral proteins are still lacking) [22].

Another main option for the CADD scientist lacking the structural information of the biological entity of interest is to exploit just the information coming from the ligands which have been tested on it, extrapolating from them enough information to build reliable quantitative structure-activity relationship (QSAR) models [23]. These approaches were among the first used in rational drug design and now fall into the category of the "ligand-based drug design" (LBDD) techniques. This family comprises methods such as pharmacophore search [24] and matched molecular pair analysis [25]; even if their application over the years has lowered, giving more and more space to the SBDD techniques, they are still widely used in several scenarios. Lastly, the latest advances in computer science, together with the exponential increase in the application of machine learning (ML) and artificial intelligence approaches, provided another powerful instrument to CADD scientists [26]. Indeed, when huge amounts of data are available for a defined context (regarding both the target and the ligands), these approaches can be proficiently used for the proper prediction of molecular properties of pharmaceutical relevance. Moreover, the creation and maintenance of an "intelligent" algorithm in the most recent years have allowed such "computational brains" to create novel chemical structures, through an approach known as de novo drug design [27].
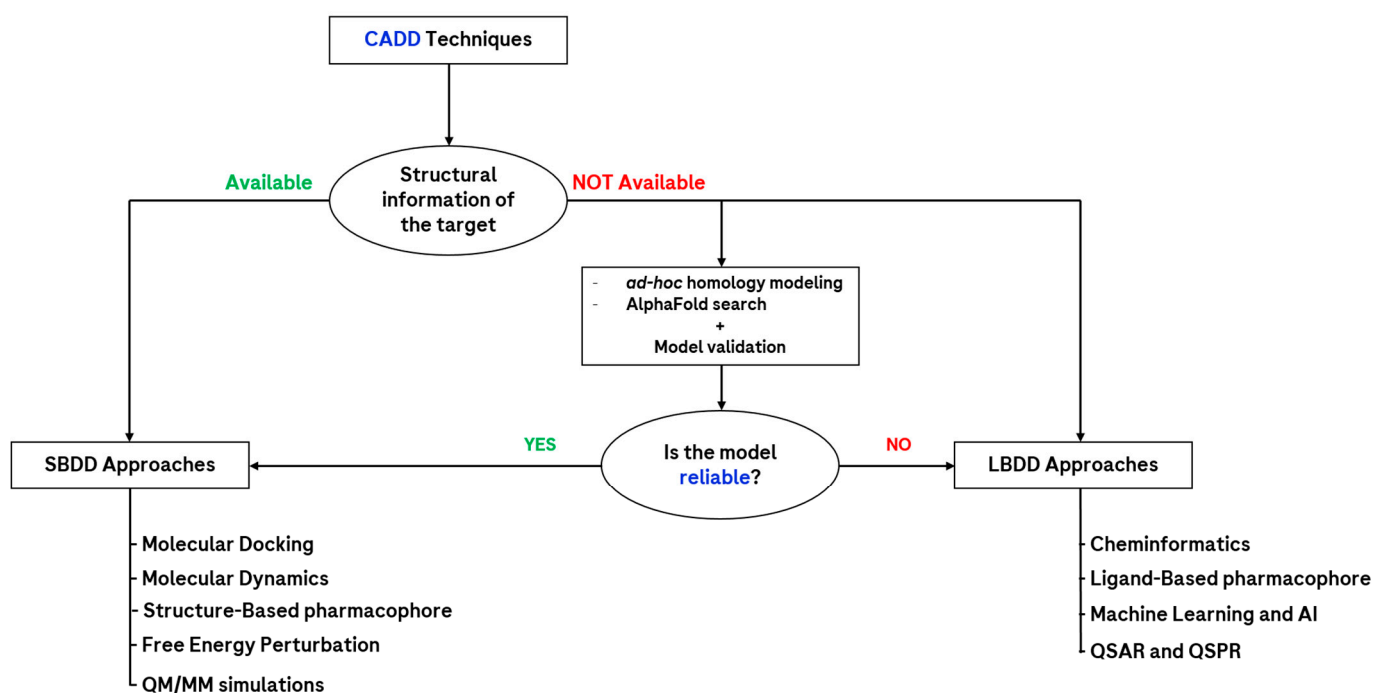
Figure 1. Scheme representing the main computational approaches available to the CADD scientist in drug discovery. As can be seen, a key factor is represented by the availability of structural information about the target. Abbreviations: QM/MM = quantum mechanics/molecular mechanics; AI = artificial intelligence; QSAR = quantitative structure–activity relationship; QSPR = quantitative structure–property relationship.

2. Discussion

2.1. Ligand-Based Drug Design (LBDD)

By far the most used approaches in the dawn of rational drug design, these techniques rely only on the structural information of the molecular structures tested on the desired target. The main goal of such methods is to identify patterns in the data that can be extrapolated to guide the further steps to take in terms of drug discovery. Those patterns are usually identified as "quantitative structure–activity relationship" (QSAR) models and should allow scientists to obtain a discrete and quantitative correlation between chemical moieties and pharmacological outcomes. Some of the techniques that are mostly used for ligand-based determination of these interconnections are cheminformatics [28] (e.g., matched molecular pair analysis), ligand-based pharmacophore search, and Free–Wilson analysis [29]. Some very famous equations, which are fundamental for the birth and the rise of QSAR modeling, were advanced by Hansch, Hammet, and Taft [30]. Even if some of the cited methods are still widely used, their main disadvantage remains related to the low level of generalization that they can provide. Indeed, they tend to work only on highly congeneric series of ligands; in other cases, they require great amounts of experimental data to provide reliable results. Moreover, these approaches do not take into consideration the conformational freedom of the ligands, focusing only on the 2D representation of the molecules considered. Together with the rising importance of structure-based approaches, and of "three-dimensionality" in general, the conformational properties of the ligands have also been taken into strong consideration by LBDD methods. One main example is the generation of "3D pharmacophores", which take into account both the atomic and the conformational features of the molecules to build proper "3D-QSAR" models [31].

Quantitative Structure–Activity Relationship (QSAR) Modeling and Cheminformatics

As already mentioned, one of the earliest historical needs for medicinal chemists was to be able to correlate molecular modification with biological activity data. In this perspective,

the first design efforts were focused only on ligand small molecules, trying to properly tune their properties simply by following the results of experiments in a target-agnostic way. This specific field of research was the so-called "quantitative structure–activity relationship" modeling, much more commonly called by its acronym "QSAR" [32]. In the second half of the last century, this strategy was known for great methodological advances, which led to its increasing application in the drug design scenario [33–35]. When the exploitation of computational approaches and the tools related to them started to spread, these methods were coupled with a series of other techniques in the already known and greater concept of "cheminformatics". Indeed, this term was defined by Gasteiger and Engel as "the application of informatics methods to solve chemical problems" [36]. This field has greatly expanded over the years [37], passing from the earliest simple data analysis methodologies applicable to chemical data to the most actual implementation of cheminformatics suites and packages in widely used programming languages. In this respect, very relevant is the creation of RDKit [38], a famous and versatile cheminformatics package for Python, which is in continuous development and whose application is well documented in the scientific literature [39–41]. Some of the many tasks that are executable with RDKit are molecular clustering, substructure search, compound fragmentation, chemical reaction handling, shape and structural similarity analysis [42], etc.

With the development of molecular modeling, more and more relevance has been given to the three-dimensional representation of chemical entities, which is now routinely analyzed together with the more classical two-dimensional depiction [43,44]. Indeed, the modern cheminformatics tools (RDKit included) have implemented different approaches for conformer generation and prioritization, given the great importance this aspect has both in chemical research and, even more so, in drug design [45].

SBDD methods have spread exponentially in the pharmaceutical scenario; nevertheless, cheminformatics and 3D-QSAR are still widely used, as many recent papers have assessed [46,47]. Moreover, these tools and methods have been coupled with actual machine learning approaches, resulting in algorithms able to autonomously detect structural patterns in chemical data, as well as automatically create novel QSAR models [48].

2.2. Structure-Based Drug Design (SBDD)

With the exponential increase in the availability of three-dimensional structures of proteins and nucleic acids, which started roughly in the 2000s, the trend in the methodologies in computational drug design moved toward other techniques, which could also take into account the three-dimensional interactive features of the molecules with respect to the target. Indeed, the prior knowledge of the biological entity of interest conferred a huge advantage to the scientists, which could develop novel chemical species on the basis of its binding site characteristics. All methods based on this kind of data fall into the family of "structure-based drug design" (SBDD) [49], which are by far the most used approaches in computational drug discovery. Moreover, while some complex membrane protein structures were not considered feasible to determine experimentally 20 years ago, the modern technology of cryo-EM allowed the reliable resolution of some of those complex systems [18,50], further extending the applicability domain of SBDD. An overview of the main SBDD methods available today is reported in Figure 2.

2.2.1. Molecular Docking

Maybe the most exploited technique in modern computer-aided drug design, molecular docking, consists of the determination of the best conformation with which a molecule binds to another to form a stable complex. The name of the technique comes from the very first program which exploited it, which was called "DOCK", proposed by Kuntz et al. in 1982 [51]. In a pharmaceutical context, these methods are extensively used to screen millions or billions of small molecules against a biological target of interest (e.g., a protein or a nucleic acid). It is important to mention that molecular docking requires prior knowledge of the binding site location on the target. A molecular docking algorithm consists of two

main parts: the conformational search algorithm and the scoring function [52]. The first has the purpose to search the conformational space of the ligand considered to find a state that fits within the binding site, while the second ranks the different conformations to prioritize the most reliable [53]. Scoring functions operate on the basis of equations that take into account the conformational strain, the electrostatics, and the steric hindrance of the ligand in its bound state. Three main types of scoring functions are available today, force field-based scoring functions, empirical scoring functions, and knowledge-based scoring functions. In the first, the energy of the system is evaluated using a force field [54]. On the other hand, empirical scoring functions consist of different terms representing different intermolecular interactions, where each term is modeled using experimental values for the interaction related to it. Specifically, empirical scoring functions are based on three main pillars: descriptors for the binding event, a database of ligand–target complexes with associated experimental activity data, and an algorithm establishing a relationship between the binding descriptors and the experimental affinity [55]. The top-ranked poses by these functions are those closest to the experimental values of reference. Lastly, knowledge-based scoring functions rely on statistical analyses of the most observed interactions between a specific ligand's atom type and a particular protein's atom type. These functions are developed by extracting structural information from high-quality X-Ray databases (usually the Protein Data Bank and Cambridge Structural Database), and then transforming atom pair preferences into distance-dependent pairwise potentials using the Boltzmann law [56]. The top-ranked poses are those more similar to what is statistically retrievable in the experimental databases [57].
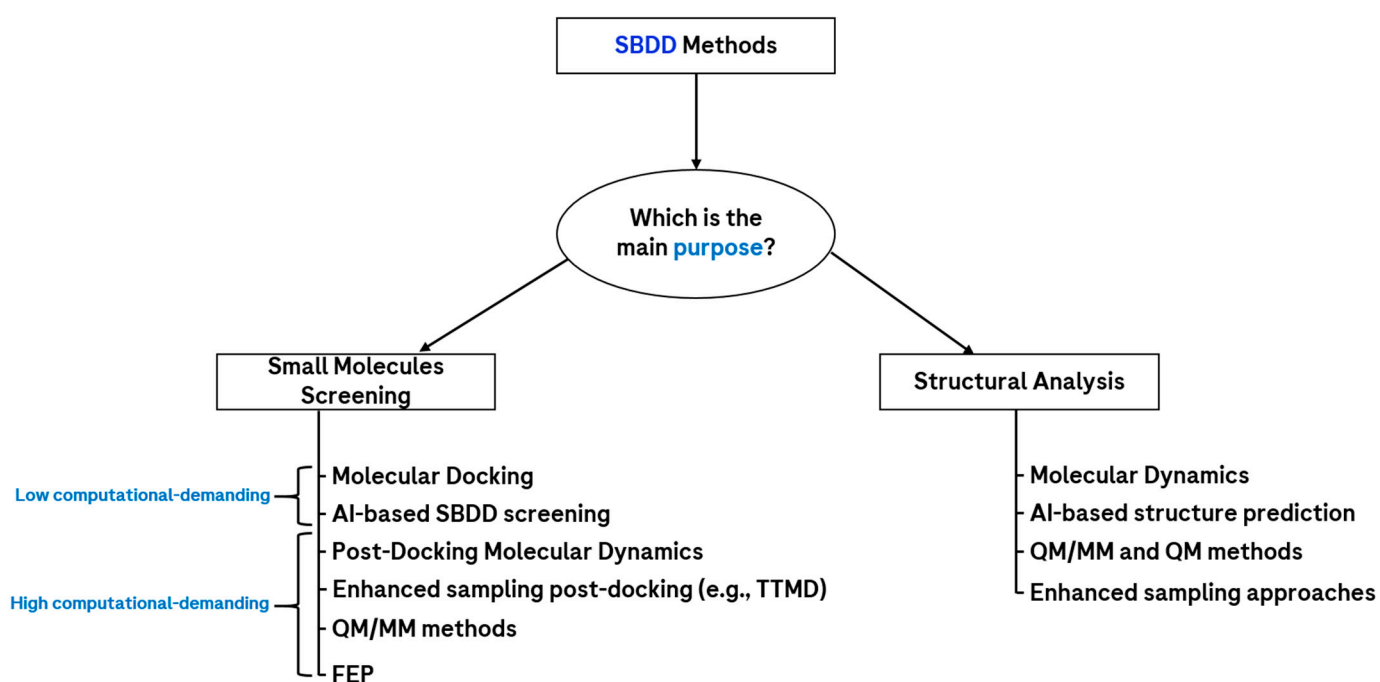

Figure 2. Scheme showing the main SBDD approaches, classified on the basis of their main purpose. If the main goal is to screen small molecules against a desired biological target, another important factor to consider is the computational power available, which allows roughly discriminating the techniques on the basis of the number of molecules/day screened with the same computational infrastructure, which of course plays a huge role in this perspective. Abbreviations: TTMD = thermal titration molecular dynamics; AI = artificial intelligence; QM/MM = quantum mechanics/molecular mechanics; FEP = free-energy perturbation.

Hundreds of different molecular docking algorithms exist today, each coupling different search algorithms and scoring functions. Even if they are different, they can be grouped into different families on the basis of how they operate to find the "bound" state of the

ligand. Some famous families are represented by genetic algorithms (among which the program GOLD is one of the most known and used [58]), systematic algorithms (such as the program Glide, developed and distributed by the company Schrödinger [59]), and ant colony optimization algorithms (such as the program PLANTS [60–62]. Some docking programs have also gained popularity due to both the robustness of their algorithms and the choice of their creators to freely distribute the software without restrictions. This is the case of AutoDock [63] (latest version 4.2.6, updated 14 April 2023) and AutoDock VINA [64] (latest version 1.2.0, updated 14 April 2023), which were both developed by the Scripps Research Institute. Both these programs have been successfully applied in small molecule research, as assessed in the literature [65].

Another classification for molecular docking is related to the degrees of freedom considered in the calculation. Indeed, in "rigid docking", both the ligand and the protein are kept rigid. The "flexible ligand docking" approach, on the other hand, allows the ligand to explore different conformational states, keeping the target rigid [66]. Then, the "semi-flexible" or "induced fit" approach consists of taking into account the conformational spaces of both the ligand and the binding site residues [67], avoiding the scenario that small clashes with a rigid side chain could impair the selection of reasonable docking poses. In the last method, which is "ensemble" docking, molecular docking is executed against an ensemble of protein conformations, usually coming from molecular dynamics simulations. In this way, the full flexibility of the protein can be indirectly taken into account [68]. A less computationally demanding strategy to obtain conformational ensembles relies on the exploitation of multiple experimentally resolved structures of the target of interest, captured in different conformational states.

As already mentioned, molecular docking is extensively used in the early phases of drug discovery, from hit identification up to lead optimization [69]. It is applied both to identify novel chemical candidates for pharmacological testing and to help rationalize experimental data to a molecular level. Nevertheless, molecular docking is essentially a "static" approach, which considers only the final state of the ligand–target system and is performed mainly in a vacuum [70]. Indeed, even if water molecules are included in the docking calculations, this information has to be given explicitly [71], and this requires structural information coupled with molecular dynamics data. As a result of all of this, the main problem of molecular docking is the high false-positive ratio of the compounds selected by the scoring functions [72], which has induced CADD scientists to investigate ways to filter the poses produced by molecular docking algorithms using other approaches. Today, many of these methods are available. One of the simplest is called "consensus docking", which relies on the principle that, by exploiting different docking programs based on different algorithms, the molecules prioritized by them will have a lower probability of being false positives [73,74]. The success of such an approach has been extensively demonstrated in the literature [75]. Other methods, called "post-docking" techniques, further filter the poses produced by molecular docking on the basis of certain molecular features [76]. One example of the first case is the implementation of a three-dimensional pharmacophore, in which the most relevant 3D features for the interaction with the target are embedded. In this case, only the docking poses which can respect these boundaries are kept, discarding the others [77]. Regarding the second kind of technique, one example is the implementation of molecular dynamics as a "post-docking" approach [78]. The poses in which the molecules can keep the interaction pattern with the protein for a longer simulation time are referred to as more "kinetically stable", while the others are deprioritized. A novel, simple, and effective technique exploiting this approach, as well as implementing a temperature increase with simulation time, was recently developed, known as "thermal titration molecular dynamics" (TTMD) [79,80], which is further discussed in the next section.

2.2.2. Molecular Dynamics

Molecular dynamics (MD) is a computational technique used to investigate the dynamic behavior of a chemical and/or biological entity over time [81]. The method consists of the iterative resolution of Newton's equations of motion to continuously predict the atomic positions of the molecules considered with respect to each other during the simulation time [82]. Molecular dynamics is used for various purposes in drug discovery, from the simple dynamic evaluation of a system to the mechanistic understanding of a molecular process [83], or as a well-known post-docking filtration system [84] (Figure 3). The main advantage of this method is that, in contrast to molecular docking, the system considered is free to move and, thus, more "realistic" if we consider the environment in which the real biochemical processes will happen. Moreover, molecular dynamics can be executed using explicit solvent models (e.g., TIP3P [85]), in which the contribution of each single water molecule is taken into account. This allows a better understanding of the role of water molecules in target stabilization, as well as in ligand–target recognition. The first and main drawback of molecular dynamics is certainly related to the computational power required for its implementation in the pipeline. Indeed, depending on the simulation time and on the dimensions of the system to evaluate, MD can require tens to thousands of time/molecule more than molecular docking; for this reason, its application is usually limited to a lower number of compounds (e.g., in the post-docking approaches). Furthermore, MD simulations rely on molecular mechanics and force fields in order to extrapolate atomic velocities with time; if this exponentially increases the speed of the simulations with respect to the quantum-based methods, it also carries several approximations that have to be known and taken into account by CADD scientists [86]. First, polarizability is not conceived in force-field-based MD simulations, whereby every molecule of the system has fixed bond lengths and partial charges, and no bond can be created or broken (except for QM/MM methods, which includes a focused region calculated at the QM level [87]). In recent years, the continuous increase in computational power of modern hardware architecture is allowing quantum-mechanical calculations to be more and more affordable, possibly leading to a new era in computational drug discovery [88].
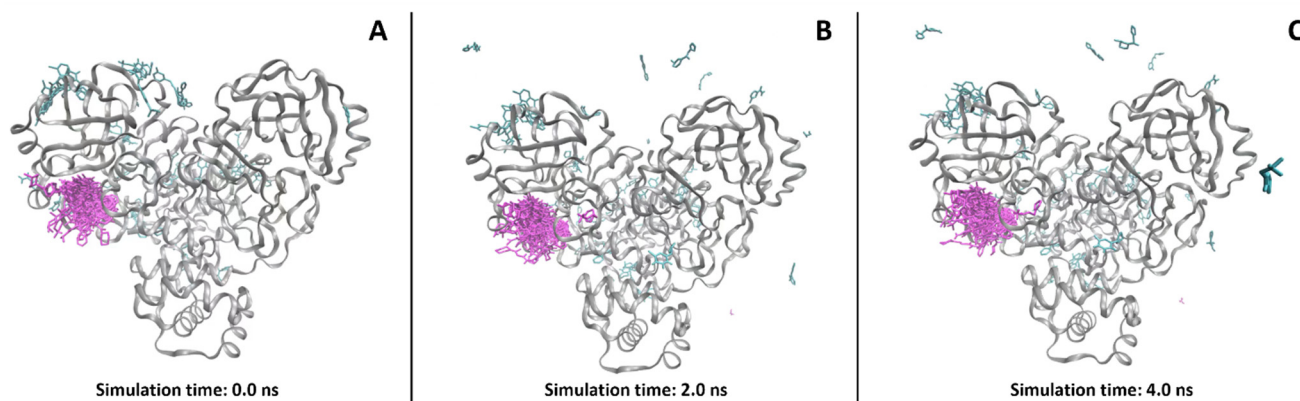

Figure 3. Example of the evolution of a system (from **A**–**C**) using molecular dynamics simulations, taken from a recent study published by our lab [89]. As can be seen, each of the SARS-CoV-2 M^{pro} crystallographic ligands starts from a defined position (the crystallographic one) at the beginning of the simulation. Then, after the MD is started, the molecules outside the catalytic pocket (colored in cyan), which are tendentially more exposed to the solvent, are more prone to lose the initial conformation and, eventually, the binding site itself. In contrast, the compounds which are crystallized in the catalytic pocket (depicted in magenta) are more prone to keep their initial position during the simulation, being more strongly bound to the protein.

Enhanced Sampling Methods in Molecular Dynamics

One of the major issues in simulating a molecular event is represented by the differences in timescales between the biological and the virtual environments. Indeed, episodes

such as ligand unbinding may require hundreds or even thousands of nanoseconds of simulation. Ligand binding is usually even more demanding if unbiased, and events such as major conformational changes of biological entities would rarely take less than several microseconds of simulations to be sampled [90,91]. These ranges in timescales come mainly from the fact that the event to be captured is considered to be "rare" and, by definition, harder to sample [92]. In the years, many different methods have been introduced and developed to increase this sampling, favored by the introduction of some kinds of biases in the simulated system [93]. Some remarkable examples are steered MD [94], scaled MD [95], replica exchange MD [96], metadynamics approaches [97], and Gaussian accelerated molecular dynamics [98]. The first of these techniques, which is mainly used for ligand unbinding sampling, relies on the introduction of a coordinate-defined force that "guides" the ligand away from its initial placement in the binding site. The reconstruction of the free-energy profile is then possible from the Jarzynski equality, but tendentially only if the forces introduced are limited in magnitude. Scaled MD, on the other hand, involves the introduction of a scaling factor for the regulation of the potential energy of the solutes' degrees of freedom in the simulation. Replica exchange MD is based on the swap of atomic positions between parallel simulations carried out at different temperatures, by employing independent Monte Carlo random walks. This allows a fairly augmented sampling of the system's events. Lastly, metadynamics relies on the iterative "filling" of the potential energy in the simulation with a series of Gaussian curves, to force the system to explore different minima and, hence, improve the sampling of "rare" events. Lastly, the Gaussian accelerated molecular dynamics (GaMD) approach consists of enhancing the conformational sampling of molecules by smoothening the potential energy surface through the addition of a harmonic boost potential that follows Gaussian distribution. Recently, Yu et al. applied multiple-replica Gaussian accelerated molecular dynamics (MR-GaMD) for the analysis of the mutation-induced conformational changes in the GTPase NRAS [99]. A very similar approach was implemented in a very recent study by Chen et al., in which GaMD was exploited for the investigation of S-adenosyl-l-methionine (SAM)-responsive riboswitches [100]. Together with these methods, some other enhanced sampling techniques have been developed in the years, eventually allowing a proficient sampling of the desired biological event with lower to no bias introduction in the system. This is the case of supervised molecular dynamics (SuMD) [101] and thermal titration molecular dynamics (TTMD) [79], which are discussed in the next section.

Molecular Dynamics as a Post-Docking Approach

As already mentioned, one of the main applications of molecular dynamics in the drug discovery pipeline consists of its exploitation in the discrimination of molecules after a molecular docking run. Indeed, while molecular docking gives only a static representation of the binding event, focusing on the final state only, MD is able to evaluate the dynamic stability of such conformation in the binding site. In post-docking MD, the poses resulting from the docking calculations are then used as the starting point for molecular dynamics, which samples diverse short simulations (usually very few nanoseconds long) for the complex considered [102]. The parameters usually tackled from this perspective are of a geometric kind, such as the root-mean-square deviation (RMSD) and the root-mean-square fluctuation (RMSF) of atomic positions. While the first describes how much a molecular entity (e.g., the ligand) displaces from its initial position during the simulation, the second quantifies the magnitude of the displacement from the most frequent position, indicating the "fluctuation" of the entity itself. Even if such parameters are easy to calculate and compare, they often offer a poor description of the binding quality. Indeed, small changes in RMSD can bring the ligand to completely lose its interaction pattern during the simulation, while higher deviations in RMSD could be due to some flexible moieties that are exposed to the solvent and, thus, able to freely fluctuate in it. To overcome these limitations, other metrics should be considered to evaluate MD-based post-docking replicas. One of the examples of this is the tackling of protein–ligand interaction fingerprints (PLIFs), which

can be compared for all the MD frames, allowing for evaluation of the effective quality of the interactions described by the molecular docking poses [53]. In this perspective, the molecule with the most proficient interaction patterns keeps the PLIFs during the MD simulations, while more weakly interacting compounds tendentially lose the contacts which stabilize their bound conformation.

Free-Energy Perturbation (FEP) and Thermodynamic Integration (TI)

In recent years, the increase in hardware performance (especially looking at the power of graphics processing units—GPUs), together with the advances in parallel computing, has allowed the spreading of the application of free-energy perturbation (FEP) methodology [103]. Specifically, this technique allows, though the exploitation of the thermodynamic free energy cycle, to extrapolate of the relative binding free energy of a series of co-generic ligands (the changes among the tested small molecules should be restricted to <10 atoms) [104]. This approach is gaining exponentially increasing interest in the world of drug design, with applications for targets ranging from protein kinases [105] to viral proteases [106] and GPCRs [107]. Specifically, FEP can be discussed in two main terms, which are absolute binding free energy (ABFE) and relative binding free energy (RBFE). The first refers to the calculation of the binding energy of a solvated ligand to a target, while the second regards the relative free energy of binding between two ligands and a target [108,109].

The advances in the predictive accuracy of these methods are mainly attributable to the improved force fields (e.g., the latest releases of the OPLS force field [110], implemented in the FEP+ application from Schrödinger [111]) and novel advances in sampling algorithms [112]. Nevertheless, it is crucial to remember that much importance has to be given to the system setup before performing an FEP calculation, to allow it to return reliable $\Delta\Delta G$ values. Indeed, the positions of the binding site waters should be accurately defined, the co-generic ligands should be docked in a way that their bound conformation is almost superimposable (at least for the common scaffold), and some MD simulations should be executed on the starting ligand, to ensure that its binding mode is stable [113].

With FEP, it is also feasible to understand the importance that each molecular portion has for binding, weighting it in quantitative terms according to the binding free energy. In its latest implementation, uncertainties <1 kcal/mol were reached, comparable to the experimental error associated with the measurement of the actual values [110]. Free-energy perturbation methods have already been applied successfully to different scenarios in drug discovery [114,115].

Thermodynamic integration (TI), on the other hand, can be referred to as a method to compare the difference in free energy between two given states. These states are characterized by different potential energies, and these have different dependences on the spatial coordinates of the entities involved in the simulated system [116,117]. Unlike FEP, which relies strongly on MD or Monte Carlo simulations, TI calculates the difference in free energy between these states by integrating over ensemble-averaged enthalpy along an alchemical path.

As in the case of FEP, the integration step is dictated by the coupling parameter λ; in TI, the potential energy of the first state is associated with $\lambda = 0$, while the final state has $\lambda = 1$ [118]. TI has successfully been applied to molecular biology [119] and drug discovery [120–122] for the prediction of the binding free energy between chemical and biological entities, and its implementations come often together with other techniques, resulting in approaches such as independent trajectories thermodynamic integration (IT-TI, in which the "independent trajectories" term is similar to a replica-exchange method) [123], density field thermodynamic integration (DFTI) [124], and umbrella integration (UI) [125].

With both FEP and TI, the main limitations are related to the fact that the high reliability of alchemical methods, other than requiring a deep prior structural and functional knowledge of the simulated system, is unequivocally bound to the molecular similarities of the compounds considered, which all have to belong to the same chemical series.

Thermal Titration Molecular Dynamics (TTMD)

One of the limitations of MD-based post-docking methods is related to the fact that, in many situations, some nanoseconds of simulation are not enough to discriminate potentially good from potentially weak binders [126]. Indeed, in many scenarios, both kinds of ligands keep low RMSD and RMSF values through the MD replicas, and their PLIFs are also equally maintained. To overcome this limitation, our group recently proposed a new method, named "thermal titration molecular dynamics" (TTMD), which proficiently takes advantage of the "concept of temperature" in molecular mechanics and molecular dynamics to classify ligands on the basis of their on-target affinity [79]. Indeed, in classic MD, the temperature is simply a value used to scale the atomic velocities with time and is not related to the "real-life" concept of temperature, which heavily influences all biochemical processes. This allows MD simulations to be performed at temperatures above 350 K or even 400 K without seeing any unfolding event taking place; in the experimental setup, these values would totally compromise the experiment. Accordingly, the TTMD method consists, starting from a protein–ligand complex, of executing MD replicas in which, for every TTMD step (T_i), the temperature of the system is increased to a certain defined number of degrees. While this process takes place, the PLIFs are monitored, along with the RMSD of the protein backbone (which can be used as a metric to effectively assess the protein integrity during simulations). The TTMD experiment may end in two different ways: when the PLIFs are lost, or after a user-defined simulation time. Examples of the results of two TTMD experiments are given in Figure 4.

In its first implementation, this method was applied to four different case studies (CK1δ, CK2, PDK2, and SARS-CoV-2 M^pro), taking five different crystal structures each. The potencies of the ligands in such complexes were known experimentally. The TTMD experiment was set up with a temperature ramp from 300 K to 450 K, with an increase of 10 K every 10 ns of simulation. In all cases considered, TTMD was effectively able to clearly distinguish the potent nanomolar ligands from weaker micro- to millimolar ones (Figure 5). The potential of this simple approach was also more recently confirmed with its application in MD-based post-docking approaches, given its intrinsic advantage to overcome the eventual inability of classical MD to discriminate ligands on a reduced timescale [80].

2.2.3. Supervised Molecular Dynamics

Supervised molecular dynamics (SuMD) consists of an MD approach that is designed to describe a "rare" molecular binding event on a reduced timescale [101]. In the specific case of SuMD, the event considered is the target–ligand recognition process, which would require simulations on the timescales of microseconds to be described by a classical MD simulation. This is due to the huge amount of time that the ligand would spend simply fluctuating in the free solvent, without any contact with the target and, even less, with the binding site. Indeed, classical MD is referred to as a "low-sampling" approach, because the force fields can only very partially sample the potential energy surface of the system. Usually, to overcome these limitations, computational approaches include Markov state models and the enhanced sampling techniques, as discussed previously. In the first case, the MD simulation is considered an ensemble of microstates, which are independent of one another. The algorithm then calculates the transition probability matrix, which allows deriving the probability of the system passing from one microstate to another [129]. On the other hand, enhanced sampling methods consist of the perturbation of the potential energy surface of the system, allowing an escape from local minima [130].
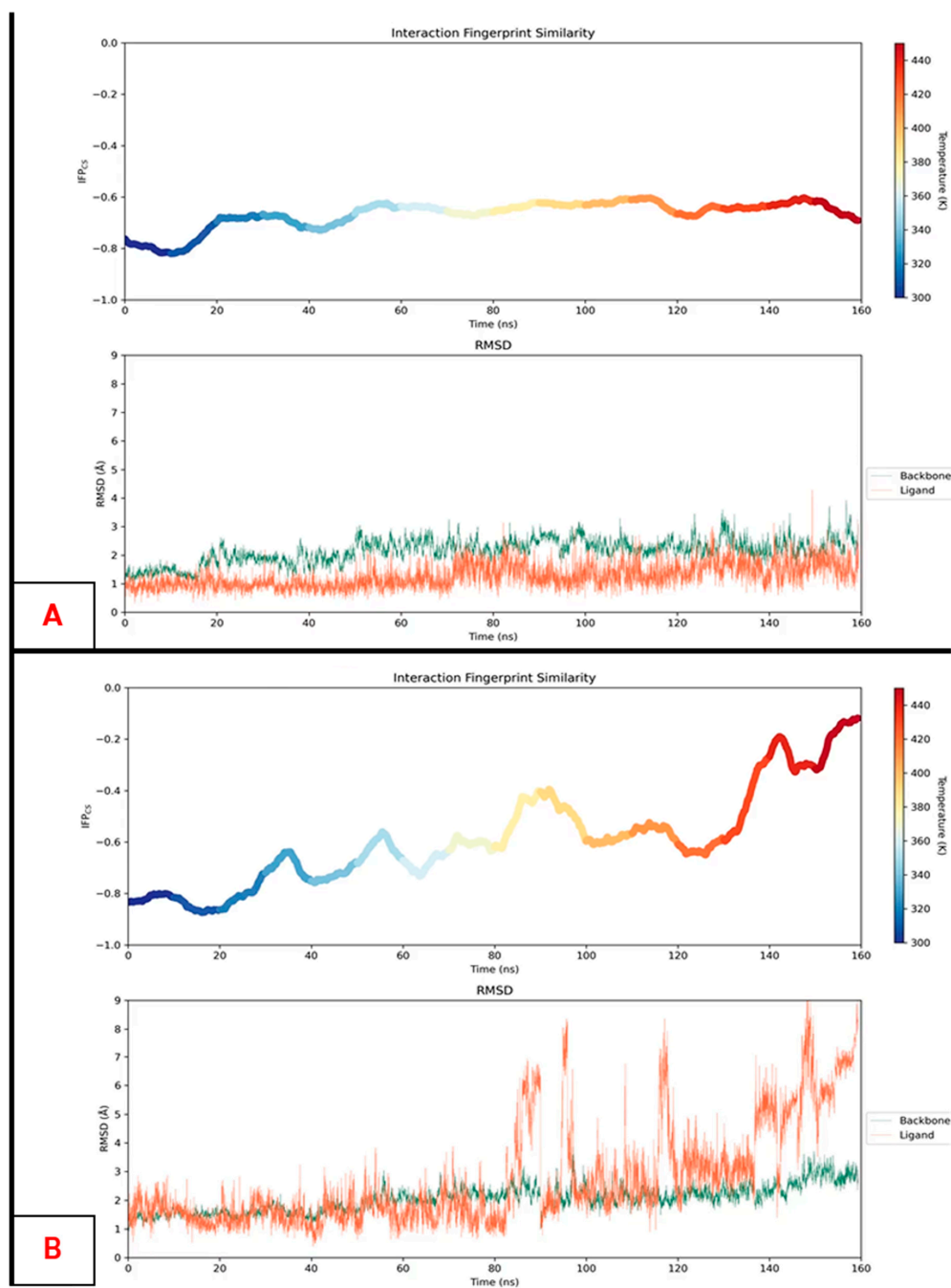
Figure 4. Examples of two TTMD profiles. On the upper side of each panel, is it possible to depict the IFP$_{CS}$ score change during the simulation, where higher scores represent a loss in the initial protein–ligand interaction fingerprints and, consequently, a loss of the binding mode. The lower part of each panel shows the root-mean-square deviations (RMSDs) of the protein backbones (in green) and the ligand (in orange) against the simulation time. These last plots allow assessing that the increase in temperature does not affect the protein folding in a relevant fashion. **Panel A** represents the results of the TTMD simulation for the ligand PF670462 in the pocket of casein kinase 1δ (CK1δ), starting from the bound conformation of the crystal 3UZP [127], while **panel B** depicts the results of the TTMD experiment for an epiblastin A brominated derivative bound to CK1δ, coming from the crystal with PDB code 5IH6 [128]. As can be observed, the nanomolar ligand PF670462 keeps its protein–ligand interaction fingerprint during the simulation, while the micromolar epiblastin A derivative progressively loses the initial binding mode, assessing the instability of its contact with the protein.
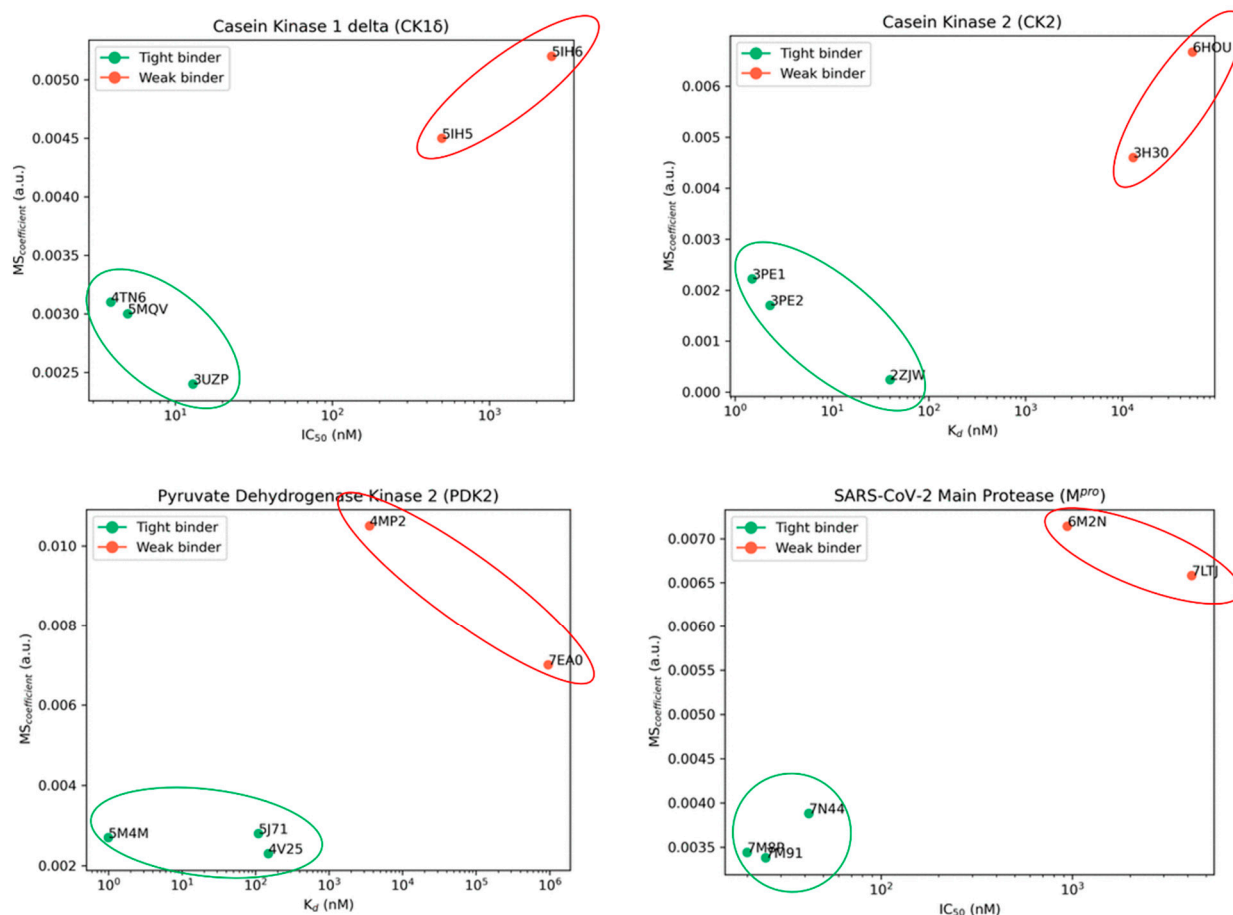
Figure 5. The results of the first published application of thermal titration molecular dynamics (TTMD). As can be seen, in all case studies considered, the method was able to efficiently discriminate nanomolar ligands (indicated with the green dots and highlighted with the green circles) from micro- and millimolar ones (depicted with red dots and circled in red). The MS coefficient, which was used for the classification, depends on the ability of each molecule to preserve its protein–ligand interaction fingerprints (PLIFs) during the TTMD experiment (a more detailed and mathematical explanation of its derivation was reported in [79]).

SuMD, on the other hand, is a technique in which no perturbation of the potential energy surface takes place. It is based on a supervision algorithm that evaluates the distance between the ligand and its binding site on the target in an iterative fashion. Specifically, the ligand is placed at a distance from the binding site in which there is no possibility of interaction at the beginning of the simulation (at least 40 Å away), and then a series of small MD simulations are sampled. At the end of each one of these, the distance between the ligand and the binding site is calculated; if the value is lower than the distance at the beginning of the small MD simulation, only the final coordinates are kept, and another MD is started from those. Otherwise, the initial coordinates are restored, and the MD simulation is repeated. When the ligand reaches a defined distance threshold (5 Å, by default), the supervision is shut down, and the simulation continues as a classical MD, allowing the ligand to relax into the binding site. In the end, the final "SuMD trajectory" is obtained by merging all small MD simulations [131]. The great advantage of this method is that it can describe an event such as the target–ligand recognition process on the nanosecond timescale rather than microseconds (typical of classical MD), accomplished without the introduction of energetic biases. This technique has already been extensively used on different targets, such as G-protein-coupled receptors (GPCRs) [132,133] (see Figure 6), proteases [134], and kinases [135], considered as ligands, small molecules, peptides [136],

and aptamers [83]. The SuMD analysis allows getting a visual representation of the most relevant residues for the interaction with the target at each step of the simulation, providing very relevant information to how the approaching process influences the binding, as well as regarding the eventual "meta-binding sites" present on the target.
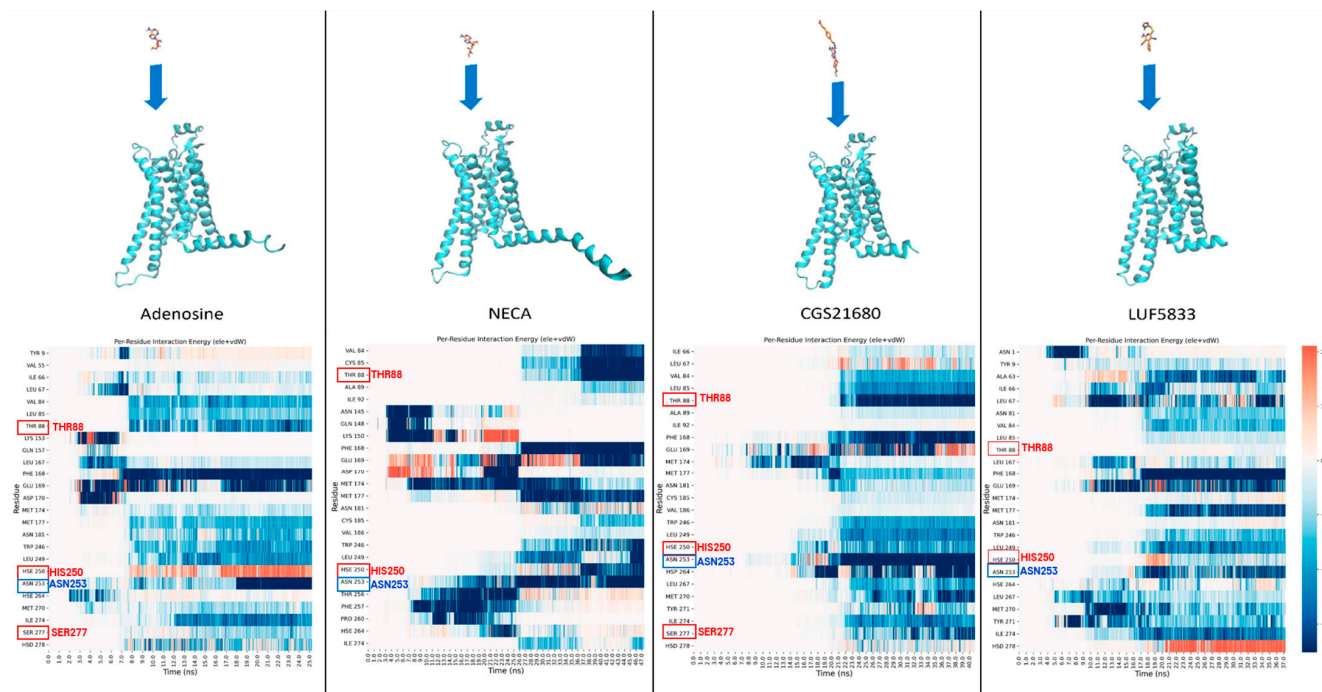

Figure 6. Results of the application of supervised molecular dynamics (SuMD) to the elucidation of the differences in the ligand binding paths between full agonists (adenosine, NECA, and CGS21680) and a non-ribosidic partial agonist (LUF5833) of the adenosine $A_{2A}R$ receptor. The panels in the upper part of the figure represent the initial configuration of each SuMD simulation, with the ligand placed away from the orthosteric binding site. The plots in the lower part depict the outcomes of the time-based per-residue interaction analysis, in which the summation of electrostatic and van der Waals contributions for each of the 25 most contacted protein residues is reported for each frame of the trajectories produced. In this study, SuMD highlighted the main differences between the ligand–protein recognition paths of full agonists and LUF5833, which can exert a partial agonism to $A_{2A}R$, even if important residues such as Thr88 and Ser277 (which are labeled in the plots in red) are not recruited directly by this specific compound. An explanation of the molecular reasons behind this behavior was reported in the original publication by Bolcato et al. [133].

3. Conclusions and Future Perspectives

In this review, we summarized the advantages and disadvantages of some of the main computational methods applied for computer-aided drug design in the past and today, looking at their origin, rationale, application, and future perspectives. The continuous advances in both methodological development and informatics infrastructure are now motivating another push through the boundaries of drug discovery. Indeed, AI-related approaches are attracting great importance in the fields of molecular property and behavior prediction [137] wherever enabled by the amount of data. Several different fields of computational chemistry have already experienced the benefits given by artificial intelligence, with a special focus on the early discovery environment [138,139]. One very relevant example is represented by on-target and off-target effect predictions in computational toxicology [140–142], which is configured in the family of AI-based methods for target prediction based only on ligand chemical data. On the structure-based side, delta-learning [143], deep learning-based 3D pocket mapping [144], and AI rescoring techniques

have been developed and documented in the recent years. These latest methods have gone beyond the prioritization performance offered by the classical scoring approaches [145,146].

On the other hand, quantum mechanics methods are becoming more and more feasible [147]. The application of AI for QM property prediction is already established in the CADD field [148,149], but it is possible to foresee that, with the advent of quantum computing, the massive calculation of such attributes will become routine.

This review can be helpful to scientists approaching the world of CADD and computational chemistry applied to pharmaceutical development, serving as a helpful tool for gaining an understanding of the possibilities that these strategies have in improving the success rate of drug discovery pipelines.

References

1. Martin, L.; Hutchens, M.; Hawkins, C. Clinical trial cycle times continue to increase despite industry efforts. *Nat. Rev. Drug Discov.* **2017**, *16*, 157. [CrossRef] [PubMed]
2. Simoens, S.; Huys, I. R&D Costs of New Medicines: A Landscape Analysis. *Front. Med.* **2021**, *8*, 760762. [CrossRef]
3. The Pharmaceutical Industry in Figures. Available online: https://www.efpia.eu/media/602709/the-pharmaceutical-industry-in-figures-2021.pdf (accessed on 2 March 2023).
4. Bohacek, R.S.; McMartin, C.; Guida, W.C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **1996**, *16*, 3–50. [CrossRef]
5. Szymański, P.; Markowicz, M.; Mikiciuk-Olasik, E. Adaptation of High-Throughput Screening in Drug Discovery—Toxicological Screening Tests. *Int. J. Mol. Sci.* **2011**, *13*, 427–452. [CrossRef]
6. Moro, S.; Bacilieri, M.; Deflorian, F. Combining ligand-based and structure-based drug design in the virtual screening arena. *Expert Opin. Drug Discov.* **2007**, *2*, 37–49. [CrossRef]
7. Petrović, D.; Scott, J.S.; Bodnarchuk, M.S.; Lorthioir, O.; Boyd, S.; Hughes, G.M.; Lane, J.; Wu, A.; Hargreaves, D.; Robinson, J.; et al. Virtual Screening in the Cloud Identifies Potent and Selective ROS1 Kinase Inhibitors. *J. Chem. Inf. Model.* **2022**, *62*, 3832–3843. [CrossRef]
8. Luttens, A.; Gullberg, H.; Abdurakhmanov, E.; Vo, D.D.; Akaberi, D.; Talibov, V.O.; Nekhotiaeva, N.; Vangeel, L.; De Jonghe, S.; Jochmans, D.; et al. Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J. Am. Chem. Soc.* **2022**, *144*, 2905–2920. [CrossRef]
9. 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.; et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **2020**, *580*, 663–668. [CrossRef]
10. Gupta, G. Racing the Clock, COVID Killer Sought among a Billion Molecules. Available online: https://blogs.nvidia.com/blog/2020/05/26/covid-autodock-summit-ornl/ (accessed on 2 March 2023).
11. LeGrand, S.; Scheinberg, A.; Tillack, A.F.; Thavappiragasam, M.; Vermaas, J.V.; Agarwal, R.; Larkin, J.; Poole, D.; Santos-Martins, D.; Solis-Vasquez, L.; et al. GPU-Accelerated Drug Discovery with Docking on the Summit Supercomputer. In Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Virtual Event, 21–24 September 2020; pp. 1–10. [CrossRef]
12. FDA. The Drug Development Process. Available online: https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process (accessed on 2 March 2023).
13. Zhu, T.; Cao, S.; Su, P.-C.; Patel, R.; Shah, D.; Chokshi, H.B.; Szukala, R.; Johnson, M.E.; Hevener, K.E. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. *J. Med. Chem.* **2013**, *56*, 6560–6572. [CrossRef]

14. Hughes, J.P.; Rees, S.; Kalindjian, S.B.; Philpott, K.L. Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249. [CrossRef]

15. Keseru, G.M.; Makara, G.M. Hit discovery and hit-to-lead approaches. *Drug Discov. Today* **2006**, *11*, 741–748. [CrossRef]

16. Leelananda, S.P.; Lindert, S. Computational methods in drug discovery. *Beilstein J. Org. Chem.* **2016**, *12*, 2694–2718. [CrossRef]

17. Dauter, Z.; Wlodawer, A. Progress in protein crystallography. *Protein Pept. Lett.* **2016**, *23*, 201–210. [CrossRef]

18. Benjin, X.; Ling, L. Developments, applications, and prospects of cryo-electron microscopy. *Protein Sci.* **2020**, *29*, 872–882. [CrossRef]

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

20. 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] [PubMed]

21. Hekkelman, M.L.; de Vries, I.; Joosten, R.P.; Perrakis, A. AlphaFill: Enriching AlphaFold models with ligands and cofactors. *Nat. Methods* **2023**, *20*, 205–213. [CrossRef]

22. David, A.; Islam, S.; Tankhilevich, E.; Sternberg, M.J. The AlphaFold Database of Protein Structures: A Biologist's Guide. *J. Mol. Biol.* **2021**, *434*, 167336. [CrossRef]

23. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R.; et al. QSAR Modeling: Where Have You Been? Where Are You Going to? *J. Med. Chem.* **2014**, *57*, 4977–5010. [CrossRef]

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

25. Kramer, C.; Fuchs, J.E.; Whitebread, S.; Gedeck, P.; Liedl, K.R. Matched Molecular Pair Analysis: Significance and the Impact of Experimental Uncertainty. *J. Med. Chem.* **2014**, *57*, 3786–3802. [CrossRef] [PubMed]

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

27. 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]

28. Begam, B.; Kumar, J.S. A Study on Cheminformatics and its Applications on Modern Drug Discovery. *Procedia Eng.* **2012**, *38*, 1264–1275. [CrossRef]

29. Kubinyi, H. Free Wilson Analysis. Theory, Applications and its Relationship to Hansch Analysis. *Quant. Struct. Relatsh.* **1988**, *7*, 121–133. [CrossRef]

30. Silakari, O.; Singh, P.K. QSAR: Descriptor calculations, model generation, validation and their application. In *Concepts and Experimental Protocols of Modelling and Informatics in Drug Design*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 29–63. [CrossRef]

31. Ragno, R.; Esposito, V.; Di Mario, M.; Masiello, S.; Viscovo, M.; Cramer, R.D. Teaching and Learning Computational Drug Design: Student Investigations of 3D Quantitative Structure–Activity Relationships through Web Applications. *J. Chem. Educ.* **2020**, *97*, 1922–1930. [CrossRef]

32. Dearden, J.C. The History and Development of Quantitative Structure-Activity Relationships (QSARs). *Int. J. Quant. Struct. Relatsh.* **2016**, *1*, 1–44. [CrossRef]

33. TSoares, T.A.; Nunes-Alves, A.; Mazzolari, A.; Ruggiu, F.; Wei, G.-W.; Merz, K. The (Re)-Evolution of Quantitative Structure–Activity Relationship (QSAR) Studies Propelled by the Surge of Machine Learning Methods. *J. Chem. Inf. Model.* **2022**, *62*, 5317–5320. [CrossRef]

34. Golbraikh, A.; Wang, X.S.; Zhu, H.; Tropsha, A. Predictive QSAR Modeling: Methods and Applications in Drug Discovery and Chemical Risk Assessment. In *Handbook of Computational Chemistry*; Springer: Dordrecht, The Netherlands, 2012; pp. 1309–1342. [CrossRef]

35. Verma, J.; Khedkar, V.M.; Coutinho, E.C. 3D-QSAR in Drug Design-A Review. *Curr. Top. Med. Chem.* **2010**, *10*, 95–115. [CrossRef]

36. Gasteiger, J.; Engel, T. *Chemoinformatics: A Textbook*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2003.

37. Muchmore, S.W.; Edmunds, J.J.; Stewart, K.D.; Hajduk, P.J. Cheminformatic Tools for Medicinal Chemists. *J. Med. Chem.* **2010**, *53*, 4830–4841. [CrossRef]

38. Landrum, G. RDKit: Open-Source Cheminformatics. 2010. Available online: https://www.rdkit.org/ (accessed on 2 March 2023).

39. Dalke, A.; Hert, J.; Kramer, C. mmpdb: An Open-Source Matched Molecular Pair Platform for Large Multiproperty Data Sets. *J. Chem. Inf. Model.* **2018**, *58*, 902–910. [CrossRef] [PubMed]

40. Bolcato, G.; Heid, E.; Boström, J. On the Value of Using 3D Shape and Electrostatic Similarities in Deep Generative Methods. *J. Chem. Inf. Model.* **2022**, *62*, 1388–1398. [CrossRef] [PubMed]

41. Spiegel, J.O.; Durrant, J.D. AutoGrow4: An open-source genetic algorithm for de novo drug design and lead optimization. *J. Cheminform.* **2020**, *12*, 25. [CrossRef]

42. Riniker, S.; Landrum, G.A. Similarity maps-a visualization strategy for molecular fingerprints and machine-learning methods. *J. Cheminform.* **2013**, *5*, 43. [CrossRef]

43. Daré, J.K.; Freitas, M.P. Is conformation relevant for QSAR purposes? 2D Chemical representation in a 3D-QSAR perspective. *J. Comput. Chem.* **2022**, *43*, 917–922. [CrossRef]

44. Nikonenko, A.; Zankov, D.; Baskin, I.; Madzhidov, T.; Polishchuk, P. Multiple Conformer Descriptors for QSAR Modeling. *Mol. Inform.* **2021**, *40*, 2060030. [CrossRef]

45. Günther, S.; Senger, C.; Michalsky, E.; Goede, A.; Preissner, R. Representation of target-bound drugs by computed conformers: Implications for conformational libraries. *BMC Bioinform.* **2006**, *7*, 293. [CrossRef]

46. Bosc, N.; Atkinson, F.; Felix, E.; Gaulton, A.; Hersey, A.; Leach, A.R. Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *J. Cheminform.* **2019**, *11*, 4. [CrossRef]

47. Neves, B.J.; Braga, R.C.; Melo-Filho, C.C.; Moreira-Filho, J.T.; Muratov, E.N.; Andrade, C.H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front. Pharmacol.* **2018**, *9*, 1275. [CrossRef]

48. Kwon, S.; Bae, H.; Jo, J.; Yoon, S. Comprehensive ensemble in QSAR prediction for drug discovery. *BMC Bioinform.* **2019**, *20*, 521. [CrossRef]

49. Anderson, A.C. The Process of Structure-Based Drug Design. *Chem. Biol.* **2003**, *10*, 787–797. [CrossRef] [PubMed]

50. Pavan, M.; Bassani, D.; Sturlese, M.; Moro, S. From the Wuhan-Hu-1 strain to the XD and XE variants: Is targeting the SARS-CoV-2 spike protein still a pharmaceutically relevant option against COVID-19? *J. Enzyme Inhib. Med. Chem.* **2022**, *37*, 1704–1714. [CrossRef]

51. Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288. [CrossRef]

52. 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]

53. Pavan, M.; Menin, S.; Bassani, D.; Sturlese, M.; Moro, S. Implementing a Scoring Function Based on Interaction Fingerprint for Autogrow4: Protein Kinase CK1δ as a Case Study. *Front. Mol. Biosci.* **2022**, *9*, 909499. [CrossRef]

54. Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* **2015**, *55*, 475–482. [CrossRef]

55. Guedes, I.A.; Pereira, F.S.S.; Dardenne, L.E. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges. *Front. Pharmacol.* **2018**, *9*, 1089. [CrossRef]

56. Shen, Q.; Xiong, B.; Zheng, M.; Luo, X.; Luo, C.; Liu, X.; Du, Y.; Li, J.; Zhu, W.; Shen, J.; et al. Knowledge-Based Scoring Functions in Drug Design: 2. Can the Knowledge Base Be Enriched? *J. Chem. Inf. Model.* **2010**, *51*, 386–397. [CrossRef]

57. 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]

58. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. *J. Mol. Biol.* **1997**, *267*, 727–748. [CrossRef]

59. 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.; et al. 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]

60. Korb, O.; Stützle, T.; Exner, T.E. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In *Ant Colony Optimization and Swarm Intelligence*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4150, pp. 247–258. [CrossRef]

61. Pecoraro, C.; De Franco, M.; Carbone, D.; Bassani, D.; Pavan, M.; Cascioferro, S.; Parrino, B.; Cirrincione, G.; Dall'acqua, S.; Moro, S.; et al. 1,2,4-Amino-triazine derivatives as pyruvate dehydrogenase kinase inhibitors: Synthesis and pharmacological evaluation. *Eur. J. Med. Chem.* **2023**, *249*, 115134. [CrossRef] [PubMed]

62. Torres, P.H.M.; Sodero, A.C.R.; Jofily, P.; Silva, F.P., Jr. Key Topics in Molecular Docking for Drug Design. *Int. J. Mol. Sci.* **2019**, *20*, 4574. [CrossRef] [PubMed]

63. 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] [PubMed]

64. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. [CrossRef]

65. Cosconati, S.; Forli, S.; Perryman, A.L.; Harris, R.; Goodsell, D.S.; Olson, A.J. Virtual screening with AutoDock: Theory and practice. *Expert Opin. Drug Discov.* **2010**, *5*, 597–607. [CrossRef]

66. Huang, S.-Y. Comprehensive assessment of flexible-ligand docking algorithms: Current effectiveness and challenges. *Brief. Bioinform.* **2018**, *19*, 982–994. [CrossRef]

67. Sotriffer, C.A. Accounting for Induced-Fit Effects in Docking: What is Possible and What is Not? *Curr. Top. Med. Chem.* **2011**, *11*, 179–191. [CrossRef]

68. Amaro, R.E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J.A.; Miao, Y.; Smith, J.C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114*, 2271–2278. [CrossRef]

69. Spinaci, A.; Buccioni, M.; Catarzi, D.; Cui, C.; Colotta, V.; Ben, D.D.; Cescon, E.; Francucci, B.; Grieco, I.; Lambertucci, C.; et al. Dual Anta-Inhibitors' of the A2A Adenosine Receptor and Casein Kinase CK1delta: Synthesis, Biological Evaluation, and Molecular Modeling Studies. *Pharmaceuticals* **2023**, *16*, 167. [CrossRef]

70. Sartore, G.; Bassani, D.; Ragazzi, E.; Traldi, P.; Lapolla, A.; Moro, S. In silico evaluation of the interaction between ACE2 and SARS-CoV-2 Spike protein in a hyperglycemic environment. *Sci. Rep.* **2021**, *11*, 22860. [CrossRef]

71. Roberts, B.C.; Mancera, R.L. Ligand−Protein Docking with Water Molecules. *J. Chem. Inf. Model.* **2008**, *48*, 397–408. [CrossRef]

72. Deng, N.; Forli, S.; He, P.; Perryman, A.; Wickstrom, L.; Vijayan, R.S.K.; Tiefenbrunn, T.; Stout, D.; Gallicchio, E.; Olson, A.J.; et al. Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease. *J. Phys. Chem. B* **2015**, *119*, 976–988. [CrossRef]
73. Poli, G.; Tuccinardi, T. Consensus Docking in Drug Discovery. *Curr. Bioact. Compd.* **2020**, *16*, 182–190. [CrossRef]
74. Houston, D.R.; Walkinshaw, M.D. Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. *J. Chem. Inf. Model.* **2013**, *53*, 384–390. [CrossRef]
75. Bolcato, G.; Cescon, E.; Pavan, M.; Bissaro, M.; Bassani, D.; Federico, S.; Spalluto, G.; Sturlese, M.; Moro, S. A Computational Workflow for the Identification of Novel Fragments Acting as Inhibitors of the Activity of Protein Kinase CK1δ. *Int. J. Mol. Sci.* **2021**, *22*, 9741. [CrossRef]
76. Rastelli, G.; Pinzi, L. Refinement and Rescoring of Virtual Screening Results. *Front. Chem.* **2019**, *7*, 498. [CrossRef]
77. Peach, M.L.; Nicklaus, M.C. Combining docking with pharmacophore filtering for improved virtual screening. *J. Cheminform.* **2009**, *1*, 6. [CrossRef]
78. Shivanika, C.; Kumar, D.; Ragunathan, V.; Tiwari, P.; Sumitha, A. Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of natural compounds against the SARS-CoV-2 main-protease. *J. Biomol. Struct. Dyn.* **2022**, *40*, 585–611. [CrossRef]
79. Pavan, M.; Menin, S.; Bassani, D.; Sturlese, M.; Moro, S. Qualitative Estimation of Protein–Ligand Complex Stability through Thermal Titration Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2022**, *62*, 5715–5728. [CrossRef]
80. Menin, S.; Pavan, M.; Salmaso, V.; Sturlese, M.; Moro, S. Thermal Titration Molecular Dynamics (TTMD): Not Your Usual Post-Docking Refinement. *Int. J. Mol. Sci.* **2023**, *24*, 3596. [CrossRef] [PubMed]
81. Pavan, M.; Bassani, D.; Bolcato, G.; Bissaro, M.; Sturlese, M.; Moro, S. Computational Strategies to Identify New Drug Candidates against Neuroinflammation. *Curr. Med. Chem.* **2022**, *29*, 4756–4775. [CrossRef] [PubMed]
82. Hollingsworth, S.A.; Dror, R.O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143. [CrossRef] [PubMed]
83. Pavan, M.; Bassani, D.; Sturlese, M.; Moro, S. Investigating RNA–protein recognition mechanisms through supervised molecular dynamics (SuMD) simulations. *NAR Genom. Bioinform.* **2022**, *4*, lqac088. [CrossRef] [PubMed]
84. De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061. [CrossRef]
85. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
86. Durrant, J.D.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71. [CrossRef]
87. Tzeliou, C.E.; Mermigki, M.A.; Tzeli, D. Review on the QM/MM Methodologies and Their Application to Metalloproteins. *Molecules* **2022**, *27*, 2660. [CrossRef]
88. Gorgulla, C.; Jayaraj, A.; Fackeldey, K.; Arthanari, H. Emerging frontiers in virtual drug discovery: From quantum mechanical methods to deep learning approaches. *Curr. Opin. Chem. Biol.* **2022**, *69*, 102156. [CrossRef]
89. Bassani, D.; Pavan, M.; Bolcato, G.; Sturlese, M.; Moro, S. Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro. *Pharmaceuticals* **2022**, *15*, 180. [CrossRef]
90. Lotz, S.D.; Dickson, A. Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. *J. Am. Chem. Soc.* **2018**, *140*, 618–628. [CrossRef]
91. Shaw, D.E.; Adams, P.J.; Azaria, A.; Bank, J.A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J.A.; Correia, T.; et al. Anton 3. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, St. Louis, MO, USA, 14–19 November 2021; pp. 1–11. [CrossRef]
92. Hartmann, C.; Banisch, R.; Sarich, M.; Badowski, T.; Schütte, C. Characterization of Rare Events in Molecular Dynamics. *Entropy* **2013**, *16*, 350–376. [CrossRef]
93. Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int. J. Mol. Sci.* **2020**, *21*, 6339. [CrossRef]
94. Patel, J.S.; Berteotti, A.; Ronsisvalle, S.; Rocchia, W.; Cavalli, A. Steered Molecular Dynamics Simulations for Studying Protein–Ligand Interaction in Cyclin-Dependent Kinase 5. *J. Chem. Inf. Model.* **2014**, *54*, 470–480. [CrossRef]
95. Sinko, W.; Miao, Y.; de Oliveira, C.A.F.; McCammon, J.A. Population Based Reweighting of Scaled Molecular Dynamics. *J. Phys. Chem. B* **2013**, *117*, 12759–12768. [CrossRef]
96. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151. [CrossRef]
97. Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2020**, *2*, 200–212. [CrossRef]
98. Miao, Y.; Feher, V.A.; McCammon, J.A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584–3595. [CrossRef]
99. Yu, Z.; Su, H.; Chen, J.; Hu, G. Deciphering Conformational Changes of the GDP-Bound NRAS Induced by Mutations G13D, Q61R, and C118S through Gaussian Accelerated Molecular Dynamic Simulations. *Molecules* **2022**, *27*, 5596. [CrossRef]
100. Chen, J.; Zeng, Q.; Wang, W.; Sun, H.; Hu, G. Decoding the Identification Mechanism of an SAM-III Riboswitch on Ligands through Multiple Independent Gaussian-Accelerated Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2022**, *62*, 6118–6132. [CrossRef]

101. Sabbadin, D.; Moro, S. Supervised Molecular Dynamics (SuMD) as a Helpful Tool To Depict GPCR–Ligand Recognition Pathway in a Nanosecond Time Scale. *J. Chem. Inf. Model.* **2014**, *54*, 372–376. [CrossRef] [PubMed]

102. Alonso, H.; Bliznyuk, A.A.; Gready, J.E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531–568. [CrossRef] [PubMed]

103. Fratev, F.; Sirimulla, S. An Improved Free Energy Perturbation FEP+ Sampling Protocol for Flexible Ligand-Binding Domains. *Sci. Rep.* **2019**, *9*, 16829. [CrossRef] [PubMed]

104. Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937. [CrossRef] [PubMed]

105. Lovering, F.; Aevazelis, C.; Chang, J.; Dehnhardt, C.; Fitz, L.; Han, S.; Janz, K.; Lee, J.; Kaila, N.; McDonald, J.; et al. Imidazotriazines: Spleen Tyrosine Kinase (Syk) Inhibitors Identified by Free-Energy Perturbation (FEP). *ChemMedChem* **2016**, *11*, 217–233. [CrossRef]

106. Ngo, S.T.; Nguyen, H.M.; Huong, L.T.T.; Quan, P.M.; Truong, V.K.; Tung, N.T.; Vu, V.V. Assessing potential inhibitors of SARS-CoV-2 main protease from available drugs using free energy perturbation simulations. *RSC Adv.* **2020**, *10*, 40284–40290. [CrossRef]

107. Deflorian, F.; Perez-Benito, L.; Lenselink, E.B.; Congreve, M.; Van Vlijmen, H.W.T.; Mason, J.S.; De Graaf, C.; Tresadern, G. Accurate Prediction of GPCR Ligand Binding Affinity with Free Energy Perturbation. *J. Chem. Inf. Model.* **2020**, *60*, 5563–5579. [CrossRef]

108. Gapsys, V.; Yildirim, A.; Aldeghi, M.; Khalak, Y.; van der Spoel, D.; de Groot, B.L. Accurate absolute free energies for ligand–protein binding based on non-equilibrium approaches. *Commun. Chem.* **2021**, *4*, 61. [CrossRef]

109. Azimi, S.; Khuttan, S.; Wu, J.Z.; Pal, R.K.; Gallicchio, E. Relative Binding Free Energy Calculations for Ligands with Diverse Scaffolds with the Alchemical Transfer Method. *J. Chem. Inf. Model.* **2022**, *62*, 309–323. [CrossRef]

110. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M.K.; Greenwood, J.; et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703. [CrossRef]

111. Abel, R.; Wang, L.; Harder, E.D.; Berne, B.J.; Friesner, R.A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50*, 1625–1632. [CrossRef]

112. LMartins, L.C.; Cino, E.A.; Ferreira, R.S. PyAutoFEP: An Automated Free Energy Perturbation Workflow for GROMACS Integrating Enhanced Sampling Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4262–4273. [CrossRef]

113. Mey, A.S.J.S.; Allen, B.K.; Macdonald, H.E.B.; Chodera, J.D.; Hahn, D.F.; Kuhn, M.; Michel, J.; Mobley, D.L.; Naden, L.N.; Prasad, S.; et al. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2020**, *2*, 18378. [CrossRef]

114. Wu, D.; Zheng, X.; Liu, R.; Li, Z.; Jiang, Z.; Zhou, Q.; Huang, Y.; Wu, X.-N.; Zhang, C.; Huang, Y.-Y.; et al. Free energy perturbation (FEP)-guided scaffold hopping. *Acta Pharm. Sin. B* **2021**, *12*, 1351–1362. [CrossRef]

115. Steinbrecher, T.B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W. Accurate Binding Free Energy Predictions in Fragment Optimization. *J. Chem. Inf. Model.* **2015**, *55*, 2411–2420. [CrossRef]

116. Resat, H.; Mezei, M. Studies on free energy calculations. I. Thermodynamic integration using a polynomial path. *J. Chem. Phys.* **1993**, *99*, 6052–6061. [CrossRef]

117. Bruckner, S.; Boresch, S. Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2010**, *32*, 1320–1333. [CrossRef]

118. Zhang, Q.; Yang, Y.; Gong, X.; Zhao, N.; Zhang, Y.; Liu, H. Thermodynamic integration combined with molecular dynamic simulations to explore the cross-resistance mechanism of isoniazid and ethionamide. *Proteins Struct. Funct. Bioinform.* **2022**, *90*, 1142–1151. [CrossRef]

119. Mishra, S.K.; Calabró, G.; Loeffler, H.H.; Michel, J.; Koča, J. Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3333–3345. [CrossRef]

120. Huai, Z.; Shen, Z.; Sun, Z. Binding Thermodynamics and Interaction Patterns of Inhibitor-Major Urinary Protein-I Binding from Extensive Free-Energy Calculations: Benchmarking AMBER Force Fields. *J. Chem. Inf. Model.* **2020**, *61*, 284–297. [CrossRef]

121. Christ, C.D.; Fox, T. Accuracy Assessment and Automation of Free Energy Calculations for Drug Design. *J. Chem. Inf. Model.* **2014**, *54*, 108–120. [CrossRef] [PubMed]

122. Garbett, N.C.; Chaires, J.B. Thermodynamic studies for drug design and screening. *Expert Opin. Drug Discov.* **2012**, *7*, 299–314. [CrossRef] [PubMed]

123. Decherchi, S.; Cavalli, A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem. Rev.* **2020**, *120*, 12788–12833. [CrossRef] [PubMed]

124. Endter, L.J.; Smirnova, Y.; Risselada, H.J. Density Field Thermodynamic Integration (DFTI): A 'Soft' Approach to Calculate the Free Energy of Surfactant Self-Assemblies. *J. Phys. Chem. B* **2020**, *124*, 6775–6785. [CrossRef]

125. Kästner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *J. Chem. Phys.* **2005**, *123*, 144104. [CrossRef]

126. Pan, A.C.; Xu, H.; Palpant, T.; Shaw, D.E. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 3372–3377. [CrossRef]

127. ALong, A.; Zhao, H.; Huang, X. Structural Basis for the Interaction between Casein Kinase 1 Delta and a Potent and Selective Inhibitor. *J. Med. Chem.* **2012**, *55*, 956–960. [CrossRef]

128. Ursu, A.; Illich, D.J.; Takemoto, Y.; Porfetye, A.T.; Zhang, M.; Brockmeyer, A.; Janning, P.; Watanabe, N.; Osada, H.; Vetter, I.R.; et al. Epiblastin A Induces Reprogramming of Epiblast Stem Cells Into Embryonic Stem Cells by Inhibition of Casein Kinase 1. *Cell Chem. Biol.* **2016**, *23*, 494–507. [CrossRef]

129. Husic, B.E.; Pande, V.S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396. [CrossRef]

130. Bernardi, R.C.; Melo, M.C.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta (BBA)—Gen. Subj.* **2015**, *1850*, 872–877. [CrossRef]

131. Cuzzolin, A.; Sturlese, M.; Deganutti, G.; Salmaso, V.; Sabbadin, D.; Ciancetta, A.; Moro, S. Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *J. Chem. Inf. Model.* **2016**, *56*, 687–705. [CrossRef]

132. Sabbadin, D.; Ciancetta, A.; Deganutti, G.; Cuzzolin, A.; Moro, S. Exploring the recognition pathway at the human A 2A adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations. *Medchemcomm* **2015**, *6*, 1081–1085. [CrossRef]

133. Bolcato, G.; Pavan, M.; Bassani, D.; Sturlese, M.; Moro, S. Ribose and Non-Ribose A2A Adenosine Receptor Agonists: Do They Share the Same Receptor Recognition Mechanism? *Biomedicines* **2022**, *10*, 515. [CrossRef]

134. Pavan, M.; Bolcato, G.; Bassani, D.; Sturlese, M.; Moro, S. Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332. *J. Enzyme Inhib. Med. Chem.* **2021**, *36*, 1645–1649. [CrossRef]

135. Panday, S.K.; Sturlese, M.; Salmaso, V.; Ghosh, I.; Moro, S. Coupling Supervised Molecular Dynamics (SuMD) with Entropy Estimations To Shine Light on the Stability of Multiple Binding Sites. *ACS Med. Chem. Lett.* **2019**, *10*, 444–449. [CrossRef]

136. Salmaso, V.; Sturlese, M.; Cuzzolin, A.; Moro, S. Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach. *Structure* **2017**, *25*, 655–662.e2. [CrossRef]

137. Jayatunga, M.K.; Xie, W.; Ruder, L.; Schulze, U.; Meier, C. AI in small-molecule drug discovery: A coming wave? *Nat. Rev. Drug Discov.* **2022**, *21*, 175–176. [CrossRef]

138. Boniolo, F.; Dorigatti, E.; Ohnmacht, A.J.; Saur, D.; Schubert, B.; Menden, M.P. Artificial intelligence in early drug discovery enabling precision medicine. *Expert Opin. Drug Discov.* **2021**, *16*, 991–1007. [CrossRef]

139. Cavasotto, C.N.; Di Filippo, J.I. Artificial intelligence in the early stages of drug discovery. *Arch. Biochem. Biophys.* **2020**, *698*, 108730. [CrossRef]

140. Füzi, B.; Mathai, N.; Kirchmair, J.; Ecker, G.F. Toxicity prediction using target, interactome, and pathway profiles as descriptors. *Toxicol. Lett.* **2023**, *381*, 20–26. [CrossRef]

141. Lysenko, A.; Sharma, A.; Boroevich, K.A.; Tsunoda, T. An integrative machine learning approach for prediction of toxicity-related drug safety. *Life Sci. Alliance* **2018**, *1*, e201800098. [CrossRef] [PubMed]

142. Hao, Y.; Moore, J.H. TargetTox: A Feature Selection Pipeline for Identifying Predictive Targets Associated with Drug Toxicity. *J. Chem. Inf. Model.* **2021**, *61*, 5386–5394. [CrossRef] [PubMed]

143. Atz, K.; Isert, C.; Böcker, M.N.A.; Jiménez-Luna, J.; Schneider, G. Δ-Quantum machine-learning for medicinal chemistry. *Phys. Chem. Chem. Phys.* **2022**, *24*, 10775–10783. [CrossRef] [PubMed]

144. Pu, L.; Govindaraj, R.G.; Lemoine, J.M.; Wu, H.-C.; Brylinski, M. DeepDrug3D: Classification of ligand-binding pockets in proteins with a convolutional neural network. *PLoS Comput. Biol.* **2019**, *15*, e1006718. [CrossRef]

145. Ain, Q.U.; Aleksandrova, A.; Roessler, F.D.; Ballester, P.J. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5*, 405–424. [CrossRef]

146. Graff, D.E.; Shakhnovich, E.I.; Coley, C.W. Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem. Sci.* **2021**, *12*, 7866–7881. [CrossRef]

147. Pozzan, A. QM Calculations in ADMET Prediction. *Quantum Mech. Drug Discov.* **2020**, *2114*, 285–305. [CrossRef]

148. Isert, C.; Atz, K.; Jiménez-Luna, J.; Schneider, G. QMugs, quantum mechanical properties of drug-like molecules. *Sci. Data* **2022**, *9*, 273. [CrossRef]

149. Böselt, L.; Thürlemann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641–2658. [CrossRef]