

Review

An Overview of Molecular Modeling for Drug Discovery with Specific Illustrative Examples of Applications

Maral Aminpour ^{1,2,3,*}, Carlo Montemagno ^{1,2,4,†} and Jack A. Tuszynski ^{3,5,6} 

¹ Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada; chancellor@siu.edu

² Ingenuity Lab, Edmonton, AB T6G 2R3, Canada

³ Department of Oncology, University of Alberta, Edmonton, AB T6G 1Z2, Canada; jackt@ualberta.ca

⁴ Southern Illinois University, Carbondale, IL 62901, USA

⁵ Department of Physics, University of Alberta, Edmonton, AB T6G 2E1, Canada

⁶ Department of Mechanical Engineering and Aerospace Engineering (DIMEAS), Politecnico di Torino, 10129 Turin, Italy

* Correspondence: aminpour@ualberta.ca

† Deceased.

Received: 21 March 2019; Accepted: 23 April 2019; Published: 30 April 2019



Abstract: In this paper we review the current status of high-performance computing applications in the general area of drug discovery. We provide an introduction to the methodologies applied at atomic and molecular scales, followed by three specific examples of implementation of these tools. The first example describes in silico modeling of the adsorption of small molecules to organic and inorganic surfaces, which may be applied to drug delivery issues. The second example involves DNA translocation through nanopores with major significance to DNA sequencing efforts. The final example offers an overview of computer-aided drug design, with some illustrative examples of its usefulness.

Keywords: molecular modeling; biomaterial; DNA sequencing; drug discovery

1. Introduction to Molecular Modeling Methods

Computer experiments play an increasingly significant role in science today. The advent of high-performance computing has enabled virtual experimentation in silico as a tool which allows for interpolation between laboratory experiments and theory. Schulten introduced the term “computational microscope” to describe the role of computational simulations in augmenting experimental research when direct measurements are not possible. He believed that computational biophysics has progressed to the point where it presents a realistic view of intra-cellular components, often at a resolution not attainable through laboratory instruments, reaching atomic or even electronic dimensions [1]. Feynman presciently stated in 1964: “Certainly no subject or field is making more progress on so many fronts at the present moment than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jiggings and wiggings of atoms” [2]. Molecular dynamics (MD) is an important computational tool for understanding the physical basis of the structure, the dynamic evolution of the system, and the function of biological macromolecules. Fourteen years later, the first MD simulation of a biological macromolecule, namely, Bovine pancreatic trypsin inhibitor (BPTI), was published [3]. Although the relatively accurate X-ray structure of BPTI was available at the time, its physiological function was unknown.

One of the most significant resources for MD simulations is the Research Collaboratory for Structural Bioinformatics (RCSB, www.rcsb.org), which makes available 3D experimentally-determined biological macromolecular structural data. The RCSB Protein Data Bank (PDB) is a worldwide repository for processing and distribution of 3D structure data of macromolecules, such as proteins and nucleic acids [4], and is an essential resource for biomolecular modeling. X-ray crystallography has made the largest contribution to our understanding of protein structure. Crystallization for many proteins (e.g., membrane proteins) is a difficult task that may require considerable effort. An important measure of the accuracy of a crystallographic structure, meanwhile, is its resolution. Relaxing a protein structure using molecular mechanics (MM) energy minimization methods reduces the energy of the protein structure and leads to a subtly different structure in important ways. In high-resolution structures, the only dynamic property that can be obtained is the isotropic temperature factor (B factor or Debye factor). Nuclear magnetic resonance (NMR) is another method by which to determine the protein structure. In NMR, the magnetic spin properties of atomic nuclei are used to build up a list of distance constraints between atoms in an enzyme. This list is then used to build a model of the protein that shows the location of each atom. A major advantage of NMR spectroscopy is that it provides information on proteins in solution, as opposed to those forming a crystal. Direct determination of structure by NMR is generally restricted to smaller proteins, typically under 20 kDa in molecular weight. High-resolution X-ray powder diffraction has also been used to solve and refine protein structures [5]. This method shares the advantage of not requiring a protein crystal. It should be noted that the structures obtained from experiments require some post-processing to prepare them for simulations. Issues could be encountered due to missing residues, atom clashes, crystallographic waters, and alternate locations that should be resolved before MD simulations are performed. In structure preprocessing, missing residues and missing atoms such as hydrogen are added, and atomic clashes are eliminated. Ionization, protonation states, and tautomer states are assigned as well to ensure the best quality structure to start.

Quantum mechanics (QM) methods (e.g., ab initio molecular orbital or density functional theory calculations) can be used to study reactions for molecular systems consisting of hundreds of atoms. Biological systems are particularly challenging for ab initio quantum mechanics methods because of their large sizes, yet first principles calculations can now tackle problems of great biological interest that cannot be solved by other means. In this regard, a detailed atomistic investigation of a biological system requires knowledge of its electronic structure. For instance, enzymatic reactions involve bond-forming and bond-breaking, effects that need to be treated quantum-mechanically [1]. Quantum calculations can provide useful models of transition states and reaction intermediates. Another example relates to photoreceptors (such as rhodopsin [6–8]) and involves excited states and the interaction between the biomolecules and the electromagnetic field. QM-based approaches may also be instrumental in correctly describing polarization effects (e.g., in ion channels [9]) as well as metallo-proteins, where very subtle chemical phenomena (such as the fact that the metal ion ligand bond has a partially covalent nature) play a vital role. Finally, QM calculations are useful in comparing different variety of spectroscopic data, such as infrared spectroscopy (IR) [10,11], Raman [12,13] and NMR [14], which can be obtained from the electronic structure calculations without additional empirical assumptions [15]. Direct applications of first principle approaches to the study of biomolecules are still restricted to systems of up to a few hundred atoms, while the size and conformational complexity of biological systems calls for methods capable of treating up to several 100,000 atoms over time scales of tens of nanoseconds.

Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations significantly expand the scope of quantum mechanical calculations to much larger systems by partitioning the problem into two parts, each of which is treated with different computational methods. The part of the system directly participating in a given chemical reaction, such as catalysis, involves the active site, substrates, and directly participating amino acid residues and is treated with a QM level simulation. The remaining part of the enzyme, which does not participate directly in the reaction and typically encompasses a much larger number of atoms, is simulated using molecular mechanics with a biomolecular force

field (FF). QM/MM methodologies can differ notably in terms of features, such as: (1) the type of scheme used to calculate the QM/MM energy; (2) the different boundary regions chosen; (3) how the interaction between the QM and MM region is investigated; (4) how an appropriate computational method is selected; and (5) how the enzymatic reaction and the associated conformational flexibility are tackled. There are advantages and disadvantages to these different QM/MM methods, largely depending on the type of enzyme and reaction under study. The most common QM/MM methods are Car–Parrinello/Molecular Mechanics MD [16], empirical valence bond (EVB) Method [17], the cluster model [18], and QM/MM MD Methods.

MD uses Newton’s equations of motion to determine the net force and acceleration experienced by each atom from which it simulates the time evolution of a set of interacting atoms. Each atom i at position r_i is treated as a point with a mass m_i and a fixed charge, q_i . The atomic coordinates evolve according to the laws of Newtonian physics where F_i is the force exerted on the mass m_i and a_i is the acceleration of it, i.e.:

$$F_i = m_i a_i \quad (1)$$

↓

$$F_i(t) = m_i \frac{d^2 r_i(t)}{dt^2} \quad (2)$$

$F_i(t)$ can also be expressed as the gradient of the potential energy, where V is potential energy (i.e., FF).

$$F_i = -\nabla_i V \quad (3)$$

↓

$$-\nabla_i V = m_i \frac{d^2 r_i(t)}{dt^2} \quad (4)$$

The engine of an MD program, it should be noted, is its time integration algorithm. The most popular integration methods for MD calculations are Verlet, velocity Verlet, and Leap-Frog Algorithms [19]. In the Verlet algorithm, two third-order Taylor expansions are used for the positions $r(t)$, one forward and one backward in time. Denoting velocity as v , acceleration as a , and the third derivative of $r(t)$ with respect to t as b , we obtain the following expression:

$$r(t + \Delta t) = r(t) + v(t) \Delta t + (1/2)a(t)\Delta t^2 + (1/6)b(t)\Delta t^3 + O(\Delta t^4) \quad (5)$$

$$r(t - \Delta t) = r(t) - v(t) \Delta t + (1/2)a(t)\Delta t^2 - (1/6)b(t)\Delta t^3 + O(\Delta t^4) \quad (6)$$

Adding the two expressions gives:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t) \Delta t^2 + O(\Delta t^4) \quad (7)$$

Accordingly, $a(t)$ is the force divided by mass, such that:

$$a(t) = -(1/m) \nabla V (r(t)) \quad (8)$$

The time step used in MD calculations is approximately one order of magnitude smaller than the fastest motion (hydrogen molecule’s bond vibration), which is about 10 femtoseconds (fs).

The result of an MD simulation is a trajectory in a $6N$ -dimensional phase space ($3N$ positions and $3N$ momenta). However, such a trajectory is usually not particularly relevant in and of itself. MD is a statistical mechanics method and that generates a set of configurations distributed according to some statistical distribution function, or also known as a statistical ensemble. Three different ensembles are commonly used in MD simulations: the Microcanonical Ensemble (NVE), the Canonical ensemble (NVT), and the Isotherma-isobaric ensemble (NPT). These ensembles are used during

equilibration to achieve the desired temperature and pressure before changing to the constant-volume or constant-energy ensemble when data collection starts. Here, N stands for the number of particles, E for energy, V for volume, and P for pressure. Each of these denotes a value to be kept constant during simulation. Consequently, measuring quantities in MD usually entails performing time averages of physical properties over the system trajectory (averages over configurations). For instance, one can define the instantaneous value of a generic physical property A at time t as:

$$A(t) = f(r_1(t), \dots, r_N(t), v_1(t), \dots, v_N(t)) \quad (9)$$

We then obtain its average as:

$$\langle A \rangle = \frac{1}{N_T} \sum_{t=1}^{N_T} A(t) \quad (10)$$

where t is an index which runs over the time steps from 1 to the total number of steps, N_T .

Sufficiently long simulation times ensure that the phase space will be well sampled, such that the averaging process approximates the corresponding thermodynamic properties. The most commonly measured properties are potential energy, kinetic energy, total energy, temperature, pressure, and root-mean-square deviation (RMSD). RMSD is the measure of similarity between structures and is found as the average distance between the atoms (usually the backbone atoms) of superimposed proteins. Therefore, the lower RMSD is, the better the model will be in comparison to the target structure. RMSD is found as:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad (11)$$

where δ_i is the distance between atom, i , and a reference coordinate. This quantity is often calculated for the backbone heavy atoms or in some cases just the C_α atoms.

The heart of any MD scheme is the FF used to analytically describe the atomistic interactions. The atomic forces that govern molecular movement can be divided into those caused by interactions between atoms that are chemically bonded and those caused by interactions between atoms that are not bonded.

$$E_{total} = E_{bonded} + E_{nonbonded} \quad (12)$$

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} \quad (13)$$

$$E_{nonbonded} = E_{electrostatic} + E_{van\ der\ waals} \quad (14)$$

In other words, Force fields present the potential energy surface of the system represented by a closed set of analytical potential energy functions. Since the kinetic energy in force fields is also taken into account, the system is able to move across the energy barriers on the potential energy surface, which implies substantial changes (e.g., conformational) during the simulation. The results of simulations will be realistic only if the potential energy function mimics the forces experienced by the 'real' atoms.

An example of an equation used to approximate the atomic forces that govern molecular movement is depicted in Figure 1.

In a typical FF, chemical bonds and atomic angles are modeled using simple springs (quadratic energy functions) that do not allow bond breaking. The functional form for dihedral energy is highly variable, so dihedral angles (that is, rotations about a bond) are modeled using a sinusoidal function that approximates the energy differences between eclipsed and staggered conformations. Additional improper torsional terms may be added to impose the planarity of aromatic rings, and "cross-terms" may be used to describe coupling of different internal variables, such as angles and bond lengths. Some FFs also include explicit terms for hydrogen bonds. Non-bonded forces arise due to van der Waals interactions and are modeled using the Lennard-Jones potential, and charged (electrostatic) interactions are modeled using Coulomb's law.

$$E_{tot} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - r\theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

Figure 1. A typical force field (FF) model. Energy dependencies are related to (a) stretching or compressing a bonded pair of atoms (modeled by a simple spring) (b) Increasing or decreasing the bond angle (modeled by a simple spring) (c) dihedral angle rotations (modeled by a sinusoidal function) (d) Van der Waals interactions (modeled by Lennard–Jones potential) and (e) Electrostatic interactions (modeled by Coulomb’s law). (a–c) are caused by interactions between atoms that are chemically bonded to one another while (c–e) are caused by interactions between atoms that are not bonded.

There are major limitations present in all FFs, one being the lack of polarizability, since each atom is fixed, forbidding any change in polarizability over time. However, both non-bonded terms can be scaled by a constant factor to account for electronic polarizability and produce better agreement with experimental observations. Although polarizable FFs are very promising, they remain computationally very expensive, with their use and parameterization being less user-friendly than that of their fixed charge counterparts [20]. Another major limitation of FFs is that they cannot be used to study reactivity, since bonds cannot be broken or formed during the simulation. To overcome this, the aforementioned QM/MM methods can be employed. ReaxFF [21] is one of the emerging methods in the study chemical reactivity using classical MD. ReaxFF enables chemical reactions to be studied through a geometry-dependent parameterization of reactants and products. Since chemical reactions are not permitted in simulations using conventional FFs, certain protonation and tautomeric states must be assigned in advance to all system residues and maintained during the simulation.

FF parameters are determined from experiments in physics, chemistry, and/or electronic structure calculations (using QM). A set of parameters is defined for different types of atoms, chemical bonds, dihedral angles, and so on. For example, an FF would include distinct parameters for an oxygen atom in a carbonyl functional group and in a hydroxyl group. A typical parameter set includes values for atomic masses, van der Waals radii, and partial charges for individual atoms; equilibrium values of bond lengths, bond angles, and dihedral angles for pairs, triplets, and quadruplets of bonded atoms; and values corresponding to the effective spring constants for each potential. For example, a typical FF for propane contains 10 bond-stretching terms, 18 angle-bending terms, eighteen torsional terms, and 27 non-bonded interactions [22]. Some popular FFs used in classical MD simulations are AMBER, CHARMM, and GROMACS.

Biologically relevant macromolecules, such as proteins, do not operate as static, isolated entities. Contrarily, they are involved in numerous interactions with other species in such a highly specific manner and recognition. Binding between two interacting systems has both enthalpic (ΔH) and entropic ($-T\Delta S$) components and is associated with a negative Gibbs free energy of binding ($\Delta G = \Delta H - T\Delta S$), which may have differing thermodynamic signatures, varying from enthalpy- to entropy-driven. Thus, the understanding of the forces driving the recognition and interaction require a detailed description of the binding thermodynamics, and a correlation of the thermodynamic parameters with the structures of interacting partners [23]. More on free energy methods will be discussed in Section 4.

Length and Time Scale Limitations in Molecular Dynamics Simulations

Large system sizes and long timescales are two challenges frequently encountered in MD simulation. Moreover, MD simulations are computationally demanding for two reasons. First, the force calculation at each time step requires considerable computation. Second, the force calculation must be

repeated many times. For example, to produce sufficient sampling to study folding, 1012 timesteps are required to reach millisecond timescales. Individual steps are limited to a few femtoseconds (fs $\sim 10^{-15}$ s), so simulating a millisecond of physical time requires nearly one trillion timesteps. Most biomolecular events of interest, such as protein folding, protein-drug binding, and major conformational changes essential to protein function, typically take longer (microseconds to milliseconds) than the MD timescales, thus limiting the applicability of these simulations. Spatiotemporal resolution of molecular modeling techniques is illustrated in Figure 2.

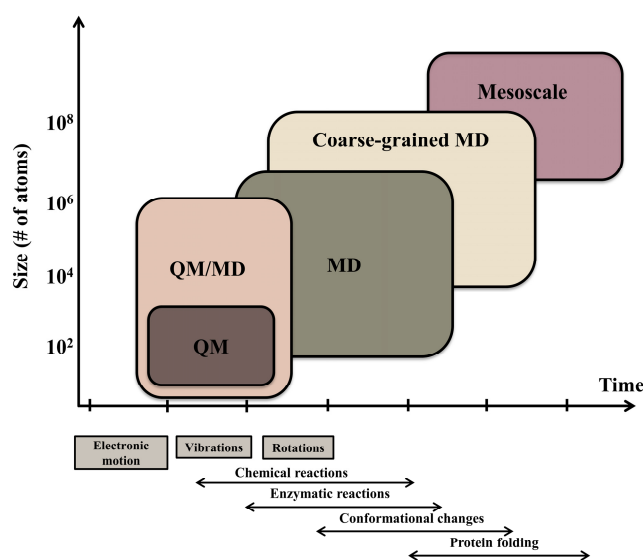


Figure 2. Spatiotemporal resolution of molecular modeling techniques: The temporal (abscissa) and spatial (ordinate) resolutions of each technique are indicated by colored boxes. The timescales of some fundamental molecular processes, as well as composite physiological processes, are indicated below the abscissa.

MD simulations have been performed on large macromolecular systems such as the ribosome [24,25] or entire viral capsids of viruses such as satellite tobacco mosaic [26,27] and HIV1 [28]. Moreover, protein folding simulations [29,30] and protein dynamics and functions [31] for smaller systems have been simulated for 10–100 μ s. The largest photosynthetic membrane simulation published thus far is that of a 100-million-atom MD simulation of a chromatophore from *Rb. sphaeroides* [32]. Various methods have been used to overcome the size and timescale limitations in MD. The coarse graining (CG) method simplifies and accelerates MD simulations [33–36]. CG employs mesoscale models, in which a group of atoms is treated as a single interaction site or a bead, this idea having been introduced by Levitt and Warshel in the 1970s [37,38]. Enhanced sampling methods also address the timescale issue, and these include Steered molecular dynamics (SMD), Umbrella sampling (US) [39], and Metadynamics [40]. US [39] is one notable equilibrium-collective variable-based enhanced sampling method, while SMD [41,42] and metadynamics [41] are the most popular nonequilibrium ones [43]. SMD has been used to accelerate the biomolecular simulations by applying external forces. It has been extensively used to calculate the potential of mean force along aquaporin channels. SMD has also been used to mimic forces that naturally arise in the context of atomic force microscopy (AFM) and optical tweezer experiments [44–49], and can be used to drag the ligand along the possible pathways predicted from electrostatic surface potential in drug design simulations [43].

The US pioneered the use of enhanced sampling methods. An energy term or a bias potential, mostly harmonic potential, is applied to the system along a reaction coordinate, and moves it from its initial state to its final state by varying, for example, the forces, distances, and angles manipulated in the simulation. MD, meanwhile, can be used to simulate the intermediate states. The weighted histogram analysis method (WHAM) is the most popular postprocessing method, and it analyzes a

series of umbrella sampling simulations [50]. WHAM is performed by unweighting and stitching together the underlying free energy function, leading to a potential of mean force (PMF) reconstruction. This methodology has been successfully applied to numerous drug discovery-relevant problems [51].

Metadynamics is a relatively new MD-enhanced sampling technique to efficiently sample the phase space and map out the underlying free energy landscape as a function of collective variables. Here, a history-dependent repulsive bias potential as a function of a set of collective variables is added to the Hamiltonian of the system in order to push the system away from its local energy minima. This can be achieved through the addition of a small Gaussian-shaped potential to the current bias to encourage the system to explore the regions of the phase space that are otherwise not sampled by conventional MD by escaping from a saddle point to a nearby local minimum, where the procedure is repeated. When all minima are occupied with Gaussians, the system diffuses to a barrier-free state along the collective variable, and the simulation can be stopped.

Finding a set of collective variables (coordination numbers, the number of hydrogen bonds, relative molecule orientation/rotation and bond lengths, angles or torsions) is a challenging task in Metadynamics simulations. Recently, metadynamics has been combined with docking calculations to study a number of ligand–target complexes, demonstrating the power of this method to characterize binding and unbinding paths, to treat conformation flexibility, and to compute free-energy profiles without requiring an excessive computational cost [52].

Among other challenges in MD simulations, like other simulation techniques, is the heavy computational time and production of the large data. The amount of data necessary to represent a rigid biomolecule (~100Kb) can increase 5 orders of magnitude (~10Gb) when simulated. As simulations have greatly increased in scale reaching cellular levels, the large amount of data generated by computer simulations intrinsically presents big data challenges. System sizes between 50,000 and 1 M atoms are common nowadays for simulations of single macromolecules or macromolecular complexes [53]. Recently, Molecular dynamics simulations have been extended to highly crowded heterogeneous cellular systems (100 M atoms for the cytoplasm of a bacterial cell) and simulations of entire cells in molecular detail will soon become reality [54]. The time scales covered by such simulations are now routinely reaching 1 μ s and in exceptional cases as much as 1 ms [55]. Depending on how often coordinates are saved, this means that a single simulation may generate data on terabyte to petabyte scales. The large amounts of data coupled with the high degree of complexity in many systems presents formidable challenges in managing, analyzing, and interpreting such big data in comparison with experiments that are being discussed. As traditional approaches to the analysis of simulations do not scale well to highly complex systems of macromolecules, a greater emphasis on automated machine learning and artificial intelligence will be required in the future [56,57]. On the other hand, the increased capabilities and flexibility of recent modern graphics processing units (GPUs) hardware combined with high level GPU programming languages such as CUDA and OpenCL has made computational power accessible to computational community. Many molecular modeling applications are well suited to GPUs since they adopt themselves to the design and implementation of data-parallel algorithms that scale to hundreds of tightly coupled processing units. One of the most time consuming calculations in a typical molecular dynamics simulation is the evaluation of forces between atoms that do not share bonds. The high degree of parallelism capability of GPUs can attain performance levels twenty times that of a single CPU core. The twenty-fold acceleration accessible by the GPU decreases the runtime for the non-bonded force evaluations such that it can be overlapped with bonded forces and PME long-range force calculations on the CPU. According to the latest Molecular dynamics GPU benchmarking report provided by NVIDIA-Tesla (<https://nvidianews.nvidia.com>), compared to CPUs, GPUs run common molecular dynamics, quantum chemistry, visualization, and docking applications more than 5 times faster.

In the era of petascale computing, large-scale MD simulations are having a profound impact in numerous and diverse scientific endeavors [58]. These have ranged from the treatment of disease and development of drugs [59,60] to biotechnological applications such as the fabrication

of novel biomaterials [61], DNA sequencing [62], and the creation of bio-based renewable energy sources [63]. Two notable applications of biomolecular modeling have been in bionanotechnology and structure-aided drug design fields. For bionanotechnology applications we discuss below two rapidly developing computational areas: organic–inorganic interface simulations to design smart novel materials, and modeling nanopores for DNA sequencing purposes.

2. Organic–Inorganic Interface Simulations for Smart Novel Material Discoveries

Hybrid organic-(bio)-inorganic materials play a major role in the development of advanced functional nanomaterials with nanobiotechnological applications. The development of these materials represents an emerging interdisciplinary topic at the interface of biology, material science, and nanotechnology. It was Thompson, in his book, *On Growth and Form* [64], who first described in detail the complex nature of inorganic crystal morphologies formed in association with living organisms and the presence of a pre-existing template controlling the growth of inorganic material. The mineralization process is usually controlled by biomacromolecules such as proteins and peptides, and the resulting hybrid structures have a physiological function. Examples of naturally occurring hybrid materials include seashells, bone, teeth, and cartilage. Adsorption and assembly of biomolecules on material surfaces through interactions on a nanometric scale form the basis for the preparation of bio-nanohybrid materials. The effects of selective adsorption of biomolecules in an active inorganic material is of critical importance in many applications, motivating the development and optimization of active surfaces for biosensors [65], bioactive nanoparticles [66], biocatalysis [67], bioanalytical systems for diagnostics and detection [68], solar cells [69], and bioseparations. Biomaterials have been designed for use in advanced drug delivery systems for over 60 years. Biomaterials for drug delivery are engineered to protect and then release molecules at desired rates [70]. Drug delivery systems has its own chemical, physical and morphological characteristics, and may have affinity for different drugs polarities through chemical interactions (e.g., covalent bonds and hydrogen bonds) or physical interactions (e.g., electrostatic and van der Waals interactions) [71]. All these factors influence the interaction of nanocarriers with biological systems [72]. To achieve optimal outcomes in drug delivery systems different parameters, such as the composition of the nanocarriers (e.g., organic, inorganic, and hybrid materials) and the form in which drugs are associated with them (such as core–shell system or matrix system) are also fundamental for understanding their drug delivery profile.

Current experimental methods are unable either to track the trajectory and follow the dynamics of bio-material interactions at the picosecond scale or to observe the surface morphology and growth at the nanoscale level. Computational modeling, however, has accelerated the discovery of hybrid materials by providing in-depth understanding at the interface between organics and inorganics and by reducing the experimental trials.

The rapidly growing interest in bio-material interfaces, particularly for bionanotechnology applications, necessitates proteins or peptides designed to recognize the inorganic surface with high specificity. Molecular modeling and simulation methods ranging from quantum mechanics, to atomistic, and coarse-grained simulations have been employed to investigate protein–surface interactions at different levels of time and length scales.

Penna and Biggs [73] offer up significant new molecular-level insight into the adsorption mechanism by considering results from over 240 MD simulations of up to 100 ns length in which one of two different peptides adsorb after starting from a distance beyond the range of the peptide–solid surface interaction. They asserted that their proposed peptide adsorption mechanism at the molecular level was generalizable for the case where the interaction between the surface and the solution above it is strong, such as would occur for metal surfaces. They proposed a putative three-phase adsorption mechanism as illustrated in Figure 3: (1) biased diffusion of protein or peptide from bulk solution towards the interface; (2) ‘anchoring’ of the peptide or protein via a hydrophilic group of the peptide to the second water layer (WL), which is adjacent to the solid surface; and finally (3) formation of the fully adsorbed peptide, which occurs through a (‘lockdown’) process of stepwise rearrangement of

the peptide that is initiated by the anchor group popping into the WL immediately adjacent to the surface [74].

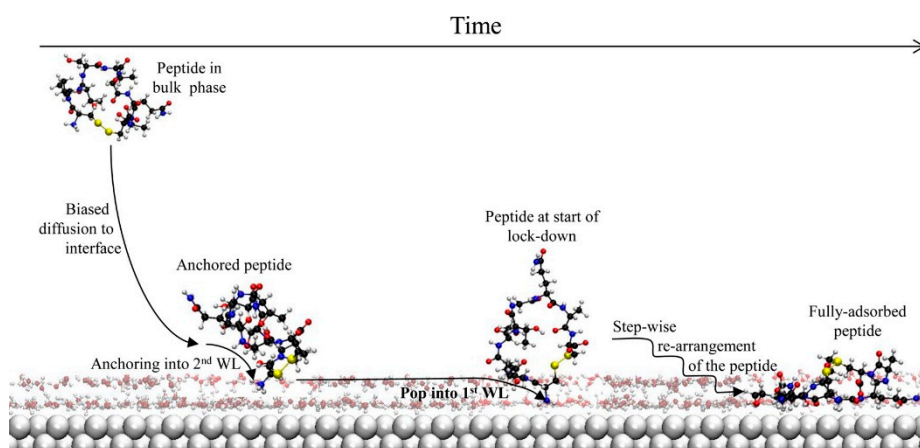


Figure 3. This illustration shows the simulations of proteins adsorbed on different surfaces. [74]. The first step is the biased diffusion towards the interface. The second step is the ‘anchoring’ of the peptide via a hydrophilic group of the peptide to the second water layer (WL) that occurs adjacent to the solid surface and finally the third step is the formation of the fully adsorbed peptide through a (‘lockdown’) process of stepwise rearrangement of the peptide that is initiated by the anchor group popping into the WL immediately adjacent to the surface.

Simulating the interactions of peptide–inorganics requires a comprehensive understanding of protein and the surface properties, such as surface morphology, size, shape, chemical composition, and adsorption characteristics. Most commonly studied surfaces interacting with proteins and peptides are metals and metal oxides due to their applications in nanobiotechnology. Peptide interactions with metals have been extensively studied due to their inertness (e.g., noble metals) and various applications in sensors, bioelectronic devices, medical implants, and catalysts. Exploiting the driving forces for molecular recognition, crystal growth, and shape development promises to yield a useful strategy for material design and performance predictions. Two examples of inorganic–organic models are shown in Figure 4.

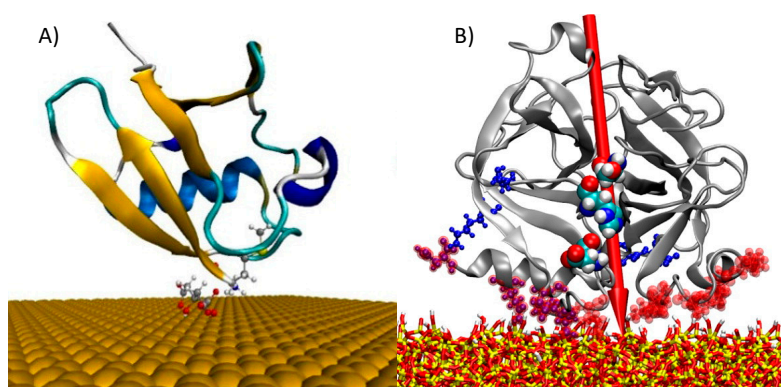


Figure 4. Modeling of proteins on different surfaces (a) ab initio Quantum Mechanics (QM), Molecular Dynamics (MD) and Brownian dynamics is used to study Ubiquitin–Au complex [75]. (b) The adsorption of α -chymotrypsin and hen egg white lysozyme on amorphous silica is studied by means of MD simulations in comparison with adsorption experiments. [76].

While atomistic simulations and FF parameters are well established for describing biomolecules and inorganic materials, FF parameters of these simulations related to one side may not be applicable for the other one, and they are likely to fail for the interfaces. Finding the right dataset to parameterize

an FF for organic–inorganic hybrid systems is a challenging task, but it is necessary for geometry optimization and quantitative calculation of system properties. Another significant challenge in modeling of organic–inorganic interactions is the unavailability of FFs and software that adequately describe the inorganic surface, the biomolecule, and the cross-terms. The INTERFACE FF [77] currently includes the PCFF-INTERFACE, CHARMM-INTERFACE, and CVFF-INTERFACE distributions that can be used with Discover, Forcite (Materials Studio), LAMMPS [78], and NAMD [79]. Alternative FFs also include GOLP [80] and GOLP-CHARMM [81] FFs to account for polarizability. These models fail when metal atoms are not fixed and thus they are mostly applicable to idealized surfaces.

A large number of computational studies have been performed on the interface of proteins/peptides and metals such as Au(111), Au(100), Pd, Pt and Ni and Pd-Au bimetal [82–84]. Heinz et al. optimized Lennard Jones (12-6 and 9-6) parameters for several face-centered cubic (fcc) metals such as gold, palladium, and platinum and their interfaces with organic and inorganic molecules, water, and biopolymers, and thereby achieved the ability to accurately reproduce densities, surface energies, and interface energies. Furthermore, based on MD simulations and experimental results, it has been concluded that the mechanism of adsorption conforms to soft epitaxy (a crystalline overlayer on a crystalline substrate) for peptides on metal surfaces [85]; this mechanism is characterized by the coordination of polarizable atoms (C, N, O) in peptides with epitaxial fcc, and epitaxial hexagonal cubic packing (hcp) sites on the metal surface. Therefore, the molecular size and geometry, rather than the specific chemistry, determine the adsorption energy. Independent of the type of metal, simulations have shown a preference in adsorption of peptides on fcc(111) facets over (100) facets. This preference can be explained in terms of the available fcc lattice sites above the metal surface. A hexagonal spacing of ~ 1.6 Å between available lattice spacing favors aromatic rings, while a quadratic spacing of ~ 2.8 Å prefers only small molecules such as water (see Figure 5a.). Large molecules, such as common sp² (Arg, Trp, Gln, Tyr, Asn, and PPh₃) and hybridized sp³ in peptides adsorb most strongly since they are a very good fit to (111) metal surfaces, while short molecules with sp³ hybridized alkyl groups exhibit the least attraction. The phenyl ring of hexagonal symmetry, for example, can best coordinate on (111) surface (flat-on parallel conformation) rather than on (100) and (110) surfaces (tilted conformation). The strength of adsorption of peptides on metal surfaces is a result of competition between small water molecules with the peptides (see Figure 5b). Residues such as Phe, Arg, Tyr, Trp, His, and Asp (F, R, Y, W, H, and D) that contribute to binding are in direct contact with the metal surfaces, and, in contrast, less-binding residues are separated from the surface by one or two water layers, thus resulting in lower adsorption energy.

According to simulations, the binding differential of the peptides to (111) and (100) facets in the soft epitaxial adsorption can lead to the selective stabilization of crystal facets during nanoparticle growth from seed crystals, while various phenylalanine containing peptides are employed as shape directing templates for nanoparticles. It has been shown that the presence of the phenyl ring in peptide sequences with and without F is sufficient as a molecular switch to convert cuboctahedral or cubic nanocrystals into tetrahedral during the growth process [86]. In particular, decorated metal nanoparticles have applications in biomineralization [87], cell targeting, imaging, and therapeutic purposes [88,89]. Specific binding of peptides on graphene [90–94] has also been investigated both computationally and experimentally. Graphene and Graphite surfaces are hydrophobic in nature, which results in negative binding energy of peptides in aqueous solutions. Common residues with higher affinity for graphene and graphite involve H, Y, W, and F, as well as amide groups in Q and N [90–95]. The epitaxial interaction between graphite layers is expected to be weak (the cleavage energy is only 190 mJ m⁻²). Therefore, the possibility of pi-stacking interactions contributes to moderate adsorption of aromatic molecules.

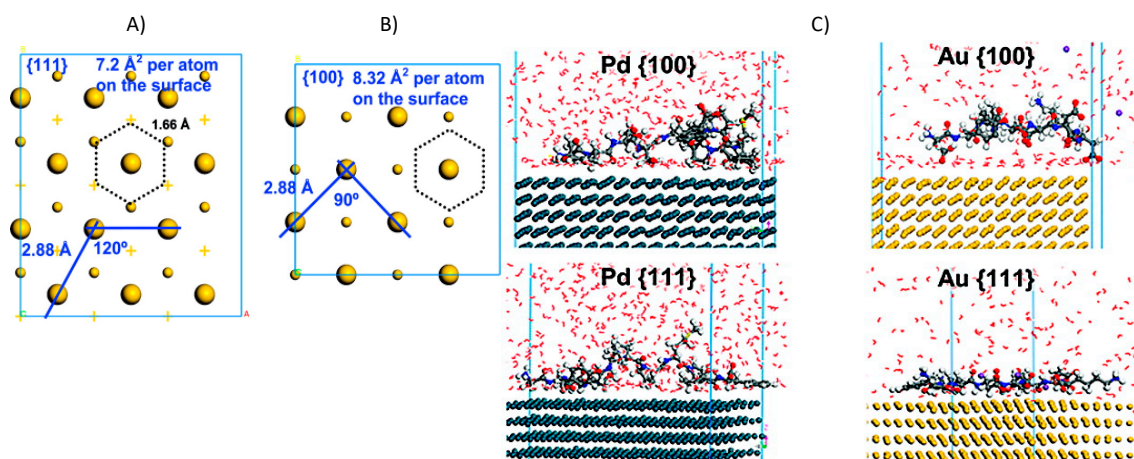


Figure 5. Comparison of the geometry (a) of the (111) surface and (b) of the (100) surface with numerical values for Au. Larger spheres represent the top atomic surface layer, smaller spheres represent the hcp sites and the crosses represent the fcc sites. Dotted black lines indicate energetically favorable orientations of aromatic rings such as in Tyr, Phe, and Trp. In (a), the ring or a sp²-hybridized group can occupy free fcc (111) “lattice” sites which contributes to high binding strength, while this is difficult to achieve in (100) lattice. Therefore, the small, flexible solvent molecules and possibly designed molecules with a rectangular repeat are of higher binding strength. (c) Representative snapshots of the Flg–Na₃ peptide on the pd and Au metal surfaces. Water interlayer is maintained between the peptide and the (100) surface, resulting in lower adsorption energy [85].

SiO₂ [96–99] and TiO₂ [100–102] have been widely studied due to their numerous potential applications. Silica is a crucial component in drug carriers [103,104], catalyst support [105] and filler in deformed modifier of polymer composites and hydrogels [106]], and is one of the most abundant oxides on Earth. Developing FFs for the interface between silica and water is very challenging due to the variety of the surface chemistry and the complex nature of the interface. Several FFs have been developed for bulk silica [107,108], silicon oxides [109], and alpha-quartz [110]. Rimola et al. developed FFs for both bulk and surface silica [111]. Ramakrishnan et al. parametrized the existing CVFF FF with silicon parameters for n+silicon (1 0 0) surface [93,112], and studied the specific binding of peptides selected by phage display on it. Strongly bonded peptides on n+ Si(100) surface, meanwhile, have been identified as M, W, D, T, H, S, and R [93]. The surface chemistry of the amorphous silica particles is, therefore, dependent on the surface topography; the distribution of Q₄, Q₃, and Q₂ environments; and the distribution of siloxane (Si–O–Si), silanol (Si–OH), and ionic siloxide (Si–O[−]Na⁺) groups.

Adsorption on oxide surfaces is governed by a different mechanism than are metals such as ion pairing, hydrogen bonds, hydrophobic interactions, and conformational changes of peptides. It has been shown that the ion pairing is the dominant mechanism of peptide binding when the surface charge of silica is significant. A higher pH value in the solution results in a higher negative charge density on the silica surface, and favors the adsorption toward positively-charged peptides, while it weakens the attraction of negatively charged peptides and reduces the influence of hydrogen bonds and hydrophobic interactions. pH has little effect on neutral peptides, which are not as strongly bound by hydrogen bonds and hydrophobic interactions. Interfacial hydrogen bonds involve oxygen and hydrogen atoms in silanol groups, siloxide ions, and lattice oxygen atoms on the silica surface in contact with alcohol groups, backbone amide groups, and aromatic heterocycles in peptides. Hydrogen bonds play a primary role when the charge of the silica is zero for all peptides, and peptides with cationic groups can predominantly adsorb to silica surface at any pH. In addition, Peptides containing hydrophobic residues increasingly interact with the silica surface. Residues such as F, W, L, I, and V can be effectively drawn to silica surfaces at lower pH, as well as to silica surfaces of lower area density of silanol groups. The interaction of hydrophobic residues originates from depletion forces to

avoid disruptions of the network of hydrogen bonds in the aqueous phase that would occur when the hydrophobic residues remain immersed. As a result, there is no intrinsic attraction of these groups to silica, and the driving force is rather the exclusion from water on less ionized silica surfaces. On increasingly ionized silica substrates, hydrophobic groups do not approach the surface because they would disrupt the hydration shells of siloxide ions and of cations in proximity of the surface. Another contribution to the adsorption mechanism arises from the conformation preference of the peptides on silica surfaces, especially for longer sequences of peptides. The effect of conformational changes on the adsorption mechanism is more dominant in metal oxides compared to the metals as the reduction in surface energy in metal oxides [113,114].

Titanium oxide (Titania) surfaces exhibit ionization of superficial terminal hydroxyl groups (Ti-oh) similar to silica, and have exhibited similar mechanisms of molecular recognition and binding. Steered MD simulations show that the ion pairing is the dominant binding mechanism between TiO- groups and positively charged R and K residues in the sample peptide (RKLPDA). The free energy profile of RKLPDA on an oxidized titanium surface has been calculated using metadynamics and replica exchange with solution tempering (REST). The typical conformations for the peptide on the surface involves flat as well as upright conformation with R and K residues bounded to the surface [61]. Furthermore, Monti et al. and Carravetta have parametrized the FF for the rutile TiO₂(110) [101,115].

Computer simulation of material surfaces and interfaces is still in the early stages of development, and most of the studies related to the interactions of surface and interface have been limited to electrostatic and Van der Waals interactions. However, the interactions occurring at surfaces and interfaces are not restricted to non-bonding interactions. These interactions constitute a complex process including adsorption, diffusion, and even breaking or forming covalent bonds. Some specific methods have been developed to meet such demands. ReaxFF, for instance, was developed to allow molecular bonds to be broken and re-formed and covalent chemical reactions to take place in a simulation without using quantum mechanics. [21,116], while the QM/MM method was developed to meet the demands both of serving large simulation systems and ensuring high accuracy [117,118].

3. Modeling of Nanopores for DNA Sequencing Applications

DNA sequencing technology provides an opportunity to identify the genetic risk factors associated with complex human diseases, and has the potential to establish a new era of personalized medicine in medical research and health care. The chain-termination method, proposed by Sanger and Gilbert in 1977 [119], was the first feasible method by which to detect nucleic acid sequences. However, according to the National Human Genome Research Institute, it costs at least \$10 million to sequence the enormous amount of 3 billion base pairs of the DNA found in the genomes of humans using Sanger-based sequencing. The growing need for faster and more cost-effective sequencing of the human genome has created enthusiasm for the development of new technologies that surpass the conventional Sanger chain-termination methods in terms of speed and economy. Nanopore sensors are promising DNA sequencing technologies poised to meet this demand. The basic idea of using nanopores for DNA sequencing was proposed by Church et al. in 1995 [120]. The basic idea is to disperse either double-strand (dsDNA) or single-strand (ssDNA) in a salt solution and apply an electric field in order to pass the DNA molecule through the pore. As the DNA translocates through the pore, the ionic flow will be blocked. Therefore, by monitoring the changes in ionic current as a function of time, it should be possible to determine which nucleotide or base is in the nanopore at the time. Despite multiple efforts and some promising developments, though, several crucial issues in DNA nanopore sequencing, such as the high speed of the translocation (temporal resolution) and the low single-nucleotide sensitivity of the ionic current (spatial resolution), have not yet been fully resolved. Experimental challenges encountered in DNA nanopore sequencing have stimulated significant computational studies to resolve these experimental issues. The computational investigations can directly relate the microscopic state of the system to the measured current, providing an important molecular picture of the mechanisms of DNA sequence detection in a manner not attainable through experimentation. MD simulations also

provide a means to study the translocation of biological molecules through nanoscale pores at levels of detail that are difficult to achieve with experimental methods alone [121–125]. Biological, solid-state, and the combination of both (hybrid) nanopores can be used for DNA sequencing. α -haemolysin and Mycobacterium smegmatis porin A (MspA) membrane pores, it should be noted, were the first biological pores to demonstrate the feasibility of DNA sequencing experimentally.

α -Hemolysin (α -HL) was the first and is most commonly used biological nanopore, representing a remarkable value in the field of DNA sequencing. Several properties of this mushroom-shaped heptamer make this membrane channel suitable for various biotechnological applications. First, the structure of α -HL remains functionally stable at temperatures close to 100 °C [126] within a wide pH range (pH 2–12) [126]. Furthermore, the inner diameter of the α -HL channel and the size of a single-stranded DNA (ssDNA) molecule are very close in size (diameter \sim 1.3 nm), such that the α -HL nanopore is able to discriminate single nucleotides using the ionic current inside the nanopore [127]. Since its transmembrane pore is open at normal conditions [128], α -hemolysin has the ability to spontaneously bind to various biological or synthetic lipid bilayers [129], and does not require specific ionic conditions. It has been experimentally observed that the interaction between ssDNA and the α -HL pore channel depends strongly on the orientation of the ssDNA molecules with respect to the pore. Remarkably, the voltage-free diffusion of the 3'-threaded DNA (in the trans-to-cis direction) is two times slower than the corresponding 5'-threaded DNA having the same poly(dA) sequence. All-atom molecular dynamics simulations of this system delineate a microscopic mechanism for the asymmetric behavior. In a confining pore of biological protein, the ssDNA straightens and its bases tilt toward the 5' end, assuming an asymmetric conformation. Consequently, the bases of a 5'-threaded DNA experience larger effective friction and forced reorientation that favors co-passing of ions. MD simulation results denote that the translocation process through a narrow pore is more complicated than previously believed and involves base tilting and stretching of ssDNA molecules inside the confining pore. It is worth noting that MD simulations have been performed without any a priori knowledge of the experimental data, (i.e., which orientation is faster or causes larger ion current blockade) [130].

Another commonly used biological nanopore is MspA. It has been experimentally shown that DNA strands immobilized inside the MspA pore produce ionic current blockades that permit identification of a single nucleotide substitution in the DNA sequence [131](see Figure 6). However, with a high transport velocity of the DNA within the nanopore, the ionic current cannot be used to distinguish signals within noise. The Aksimentiev group investigated, through all atom molecular dynamics (MD) simulations (\sim 100 μ s in total), the molecular origin of such extreme sensitivity of the ionic current to the sequence and orientation of DNA strands. The constriction of this nanopore, they found, contains positively asparagine substitutions. They performed several arginine mutations to examine the effect of the positively-charged residues on the DNA transportation speed, discovering that additional positively charged arginine substitutions near the constriction of MspA can cause a \sim 10- to 30-fold reduction in DNA translocation speed due to continuous contact between the arginine substitution and the DNA backbone, or due to base-stacking interactions while preserving the nucleotide-type specificity of the ionic current blockades. These results reveal the importance of nanopore geometry and charge location and may help to direct the modification of nanopores such as MspA to improve their utility for nanopore sequencing and other nanopore technologies. The same group coupled a MspA biological nanopore with a DNA processing enzyme, which opens the double-stranded DNA helix and slowly ratchets one of its strands through the pore channel in order to reduce the speed of DNA transportation through the MspA protein. Employing molecular dynamics simulations, they found that the quantity of the displaced water from the nanopore by the DNA strand regulates the nanopore ionic current, whereas the steric and base-stacking properties of the DNA nucleotides control the amount of water displaced. Unexpectedly, they found the effective force on DNA in MspA to undergo large fluctuations, which may produce insertion errors in the DNA sequence readout [132].

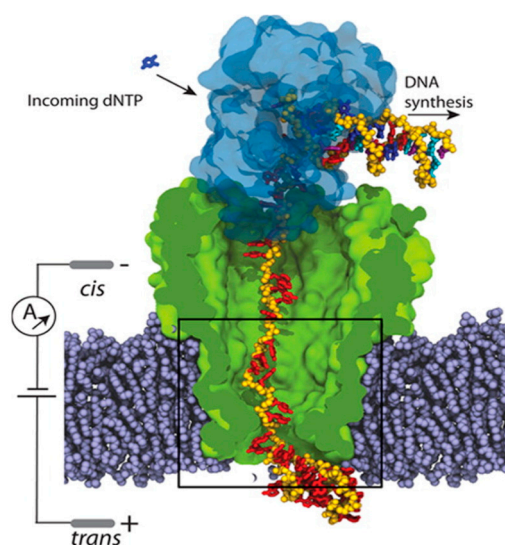


Figure 6. Molecular dynamics simulations of blockade currents in biological nanopores: All-atom model of MspA (green) suspended in a lipid bilayer membrane (purple spheres). A phi29 polymerase (blue semitransparent surface) is bound to the junction of single- and double-stranded DNA (the DNA backbone is represented by the yellow spheres [132]). The water molecules and ions are not shown in either figure.

Despite having shown much promise, the sensitivity of the lipid membrane used to fix biological nanopores to the temperature, pH value, and salt concentration, and the applied bias have been identified as major drawbacks with respect to the use of biological pores in practical applications. For this reason, solid-state nanopores, fabricated in membrane materials such as SiO₂, Si₃N₄, Al₂O₃, and plastic have emerged as a promising alternative to biological nanopores [133–135], as they not only are robust and durable within the given environment but also permit convenient, controllable, and reproducible manipulation of physical and chemical properties of nanopores, in addition to delivering the advantage of being readily integrated into semiconductor devices and chips.

There have been extensive studies conducted on double-stranded DNA (dsDNA) translocation [136–148], single-stranded DNA (ssDNA) translocation [138,149], and protein translocation [150,151] through solid-state pores. A wealth of interesting results have been obtained with respect to solid-state nanopores, such as translocation time as a function of DNA length [137], salt dependence on ion transport during DNA translocation [146,147,152], unzipping of DNA during translocation [141,153], and discrimination of ssDNA and dsDNA based on pore diameter [138,154]. However, the thickness of solid-state nanopores makes it difficult to detect individual base-specific modulation in ion currents as multiple base pairs interact with the nanopore channel simultaneously [155]. Single-layer materials such as graphene [133–135,156], MoS₂ [157], and BN [138,158,159] nanopores have been used as an alternative to conventional solid-state pores. A potential advantage of these ultrathin membranes is their thickness (~0.3 nm, which is comparable with the height of the nucleotide), which improves spatial resolution in sequencing measurements. Instead of actually building and testing the device experimentally, molecular dynamics simulations can assist and enable a bottom-up design of two-dimensional material nanopore devices by unveiling the atomic-level processes occurring during nanopore sensing.

Graphene is one of the most fascinating single-layer materials for DNA sequencing. It is impermeable to ions, and, due to its strength, it can form a freestanding membrane, enabling the ideal atomically thin membrane for nanopore measurements. Another important and advantageous property of graphene is that it is electrically conductive, which opens up the possibility to monitor an in-plane current through the membrane when the DNA molecule translocates. However, despite intensive research efforts in this area, the identification of individual bases has not yet been achieved by graphene nanopores, mainly because the speed of DNA translocation through the pore is too high to permit

individual bases to be distinguished. Furthermore, the conformational fluctuations of DNA inside the pore add significant noise to the measured signal. Computational simulations have been conducted to investigate whether indeed DNA sequencing is possible with ionic current detection through graphene nanopores. Sathe et al. studied the interaction of DNA with the pore while DNA translocates through a graphene pore in order to evaluate the manner in which this affects the ionic current [160–162]. Initially, they found that poly(AT) and poly(GC) can be distinguished at a bias voltage of 1 V [160]. However, the simulations also exposed some complications with the approach, as they showed that the bases move stochastically through the pore, leading to sequencing errors. Furthermore, the blockage in the current was predicted to be highly dependent on the local conformation of the DNA bases inside the pore, resulting in a strong overlap of the current blockades for the different bases [160,161]. Interestingly, hydrophobic adhesion of bases to the graphene surface right next to the pore was found to considerably reduce the possible single-stranded DNA conformations, leading to a reduction in the translocation speed of the DNA through the nanopore [161]. These simulations suggested that a three-layer graphene sheet produces the best ‘stepwise’ translocation pattern, such that collective binding and unbinding of the bases on both sides of the membranes is possible, while fluctuations in the DNA base conformations inside the pore are minimized [161].

Recently, quantum mechanics Green’s function-based transport calculations were used to calculate the transverse electronic conductance of the graphene sheet while the stretched ssDNA are translocating through the pore to record the intrinsic stepwise DNA motion (see Figure 7). The measurement scheme described paves the way to enhance the signal-to-noise ratio, not only by slowing the DNA translocation to provide sufficient time for base recognition, but also by stabilizing single DNA bases and thereby reducing thermal noise [163].

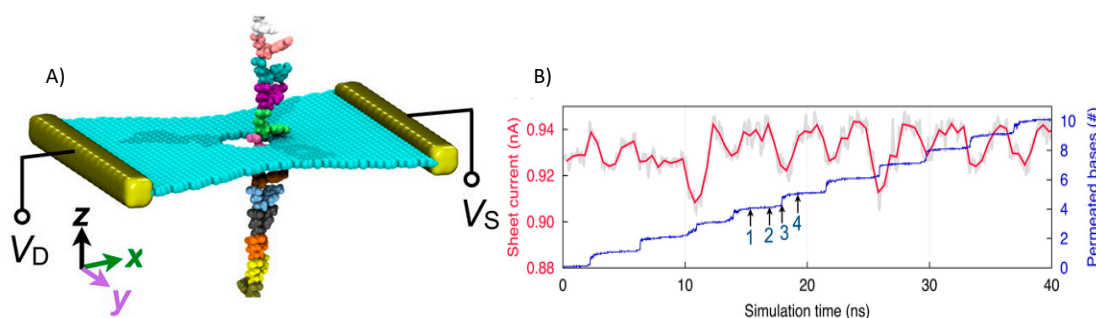


Figure 7. Electronic detection of stepwise motion of stretched ssDNA through graphene nanopore: (a) Schematic of ssDNA and a graphene nanopore, and (b) Transverse calculated sheet current through graphene shown together with the number of permeated nucleotides during ssDNA translocation [163].

To overcome the limitations of biological and solid state nanopores, it has been proposed to combine the best properties from both types of nanopores [164,165], leading to next-generation hybrid nanopores. The idea is to attach specific biological recognition groups inside the solid-state nanopores. This should substantially improve the chemical specificity, while still maintaining mechanical and chemical stability. Several hybrid nanopores have been created [164–167]; however, there are no reports in the literature of their having been used in DNA sequencing.

In summary, nanopore-based DNA sequencing offers an inexpensive, reliable technology which could thrust genomics into personal medicine and open a new frontier in gene detection. Instead of physically building and testing the DNA sequencing device experimentally, with the associated cost, molecular dynamics simulations can be employed to build a bottom-up design of biological or material nanopore devices. Simulation unveils the atomic-level processes occurring during nanopore sensing, guiding experiments to overcome significant challenges in the effort to reach single-base resolution, given the fast translocation times, the conformational fluctuations, the stochastic translocation of the bases, and the high noise levels. To address the abovementioned challenges, alternative materials can be used for nanopore structures because we can evidently not change the properties of the DNA.

Furthermore, the nanopore geometry (pore diameter and charge on the pore surface) can greatly modulate the drag forces, something that is not considered in most simulations. In addition, the error rate of DNA sequencing has not been assessed in most computational studies since the read lengths of the DNA molecules used in these studies have been very short. All these issues need to be attended in the future.

4. Computer-Aided Drug Design

One notable application of biomolecular modeling is Computer-Aided Drug Design (CADD). Drug discovery begins with target and lead discovery, followed by lead optimization and pre-clinical in vitro and in vivo studies to identify the best candidate compounds that satisfy the main criteria for drug development. [8] Drug design using in silico methods is cost-effective compared to in vivo and in vitro methods and it may also reduce the time it takes for a drug to reach the market. In this regard, an article in the October 5, 1981, issue of *Fortune*, entitled the “Next Industrial Revolution: Designing Drugs by Computer at Merck” (Van Drie, 2007 [168]), is sometimes referred to as heralding the inception of the field of computer-aided drug design (CADD).

CADD methods can be broadly classified into two groups, namely, ligand-based (LB) and structure-based (SB) drug discovery (see Figure 8). When the target structure is not experimentally determined (via X-ray Crystallography or NMR) or when it is found to be challenging to predict a structure using homology modeling or ab initio methods, ligand-based approaches are often used as an alternative. Also, LBDD methods can be performed even if the target structure is available. These methods, however, rely on information about known actives. Molecular similarity approaches, QSAR (quantitative structure–activity relationship) modeling, and pharmacophore modeling are some popular LBDD approaches. In LBDD molecular similarity approaches, common structural features of ligands (fingerprints) that bind to a target are used to carry out the screening [169]. The similarity search approach allows for the representation of a molecule in such a way that it can be effectively compared against other molecules [170]. Another notable approach is the quantitative structure–activity relationship (QSAR), which models the relationship between the structural features of the ligands that bind to a target and the corresponding biological activity effect [171]. Another field gaining ground in the area of computational ligand-based drug discovery is pharmacophore modeling, where common structural features of ligands that bind to a target are used to carry out the screening [172]. The 3D structure of various drug molecules, it should be noted, is now available in several large databases such as NIH [173], ZINC [174,175], and DrugBank [176].

In SBDD methods, the 3D structure of the target (protein receptors, enzymes) is known and can usually be obtained by X-ray crystallography or NMR experiments as mentioned above. If the crystal structure of the target is not known, then homology modeling can be used to build the 3D structure of the target. Knowing the target structure makes it possible to exploit effective tools such as structure-based virtual screening and direct docking methods on targets and possible drug molecules. In structure-based virtual screening, large databases of chemical structures are searched in order to identify the potential drug candidates that are most likely to bind to a drug target. In this process, the drug candidate will be docked into a protein target, and a scoring function will be applied to estimate the probability that the drug candidate will bind to the protein target with high affinity [177]. Geometry-based binding site identification algorithms can be used to predict the binding pocket sites of the protein. Binding pocket sites indicate the location where small molecules can bind to target structures, which are associated with diseases. In this regard the high affinity drug candidates are called “hits”. Molecular docking is a method that can predict the preferred orientation of one molecule (ligand) to a second (receptor, protein) when they bind and form a stable complex structure [178]. *De novo* ligand design is another method that makes it possible to build up ligand molecules that are drug-like within the constraints of the binding pocket by assembling small pieces in a stepwise manner with much less search space having to be explored. The biochemical and organic model builder (BOMB) program, as a *de novo* ligand design method, can be used to design ligands that bind to the target

without using ligand databases by adding substituents into a core structure of ligand molecules [179]. Inhibitors for *E. coli* RNS polymerase, for instance, have been designed using this method [180].

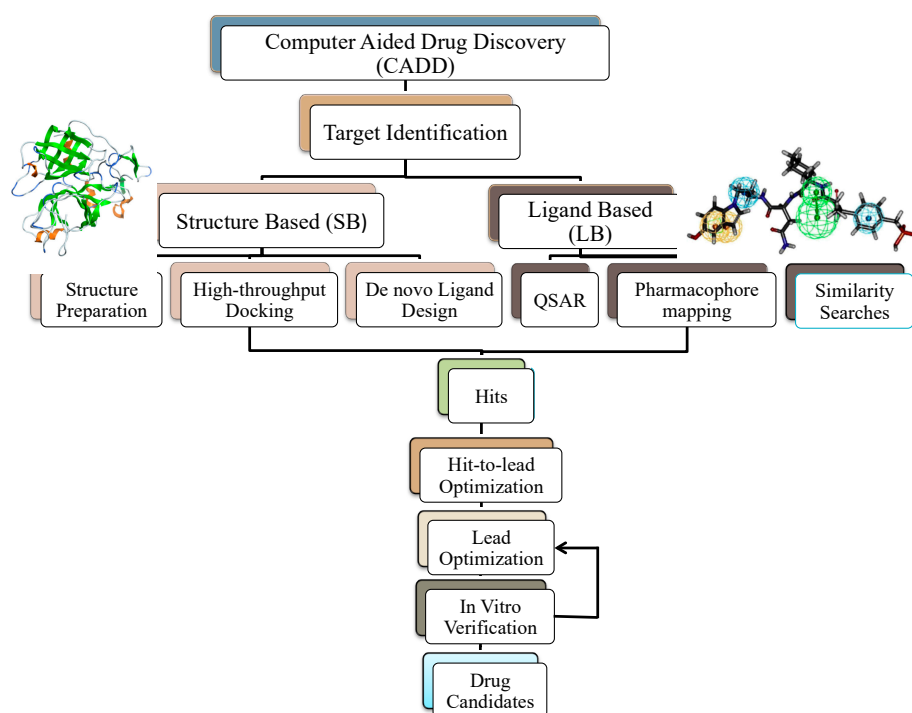


Figure 8. Schematic representation of a Computer-Aided Drug Design (CADD) pipeline whose methods are mainly classified into structure-based and ligand-based methods. Hits are identified, filtered, and optimized to obtain potential drug candidates to be experimentally tested [52].

High affinity hit molecules that have passed through the hit-to-lead optimization process are called leads. During lead optimization, the effectiveness of hits obtained is augmented to reach the desired affinity, drug safety, and absorption, distribution, metabolism, and excretion/elimination (ADME) properties. For more precise results, the affinity of hit molecules to targets can be evaluated by computing various estimates of binding free energies. The outcome selected lead molecules are tested in vitro for their activity. Lead optimisation is, by its very nature, an iterative process whereby the information between In vitro verification and lead optimization is interchanged, leading to candidate drug. This narrows down the drug candidates and allows experimentalists to focus their resources on examining compounds likely to have a given activity of interest.

Structure-based virtual screening processes are fast but are associated with an underlying problem, which is the plasticity of the target. In most cases the flexibility of the target molecules is limited or ignored. MD simulations are frequently used in drug design processes, and can be used to generate multiple receptor conformations for virtual screening purposes. Multiple snapshots can be extracted from MD trajectory as a different conformation variant of the binding pocket representative. MD can also be used as a post-docking tool to validate and/or refine docking solutions. In this case, MD is able to distinguish the bad docking poses from meaningful ones. First, the complex ligand-receptor structure is not stable and the ligand may even leave the binding site. In this regard, RMSD analysis can be used to check the stability of the ligand-receptor structure. MD in explicit solvent, meanwhile, can take into account the presence of structural water molecules within the binding site, which is important for correctly predicting ligand binding [181]. MD simulation is also an important tool in identifying the drug binding pathway for a number of reasons; first, insight into the binding pathway is crucial in gaining a thorough understanding of ligand binding kinetics [182]. Second, it is important for investigation of non-active site mutations which could possibly prevent drugs from entering binding sites by rupturing the structure of the binding pathway. Finally, the intermediate

states in the pathway of a ligand may themselves be additional binding sites to be considered in drug development. [183]. Enhanced sampling methods are often coupled to MD simulations to expedite the drug-binding processes and to extract useful thermodynamic and kinetic data. A number of techniques have been developed in this regard, among them free energy perturbation (FEP), umbrella sampling, SMD, and metadynamics, all of which have been briefly described above.

One of the thermodynamic observables that can be computed from MD trajectory is free energy. Given that the free energy of a system is a function of its state, the free energy difference found between the initial and final (binding/unbinding) states is independent of the path taken between them. Differences in binding energies can be evaluated by different methods. Free energy perturbation and thermodynamic integration are among the most popular free energy methods [184–186]. Relative binding affinity of a drug-target system can be calculated using a technique called “alchemical transformation”, although in drug design they are known as “free energy perturbation” (FEP) methods [187].

Alchemical methods calculate the work needed to move a system from one state to another through unphysical pathways. Alchemical methods are based on a non-physical thermodynamic cycle, where the binding free energy is computed as the sum of multiple steps during which the ligand is “inserted” or “removed” from different environments, such as bound and unbound states. FEP methods can be used to investigate the effect of a change of the system (e.g., an amino acid mutation or a ligand modification) on binding free energy by applying perturbations to see if the binding affinity is improved or diminished. The relative binding free energy, it should be noted, is an indirect measure of drug potency upon the chemical change. A typical thermodynamic cycle used to calculate relative binding free energy employing FEP method is shown in Figure 9.

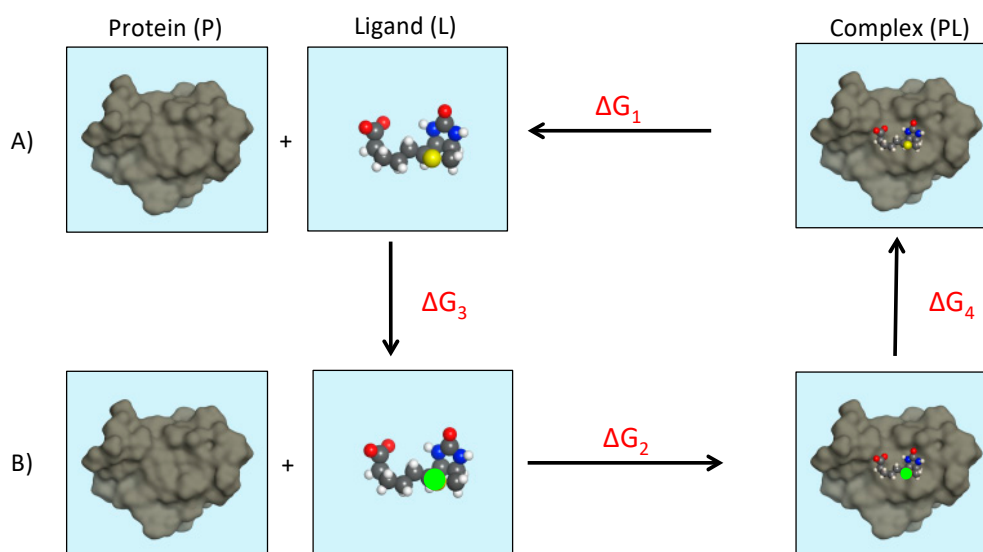


Figure 9. Schematic representation of a thermodynamic cycle to calculate the relative binding free energy upon a ligand modification. The binding free energy of the ligand or protein in state **A** (ΔG_1) and that in state **B** (ΔG_2) are the physical binding processes that need to be determined, whereas (ΔG_3) and (ΔG_4) (the vertical legs) indicate the unphysical transformation of ligand **A** (yellow) to ligand **B** (green). However, (ΔG_3) and (ΔG_4) values are more accessible to the free energy perturbation methods. According to the conservation of energy, we can derive ($\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$).

FEP/MD has been successfully employed to predict the stereoselective binding of a potent modified peptide inhibitor to the HIV-1 protease [188], with the results demonstrating that the theoretical model developed can be used reliably for the prediction of relative binding affinities and should be a useful tool for the design of feasible anti-AIDS therapeutics. Interestingly, the design of HIV-1 inhibitors has emerged as an excellent testing ground for the FEP methodology in general.

The thermodynamic integration (TI) method is an alternative to the FEP method for calculating the difference in free energy between two thermodynamic states which differ from one another according to their intermolecular or intramolecular interaction potentials. In this case, the interaction potential can be expressed as a function of a coupling parameter, λ , that determines the state of the system. The free energy difference of two states is then calculated by integrating the derivative of potential energy over all coupling parameters related to a series of unphysical intermediate states [189].

Absolute binding free energies, it should be noted, have been calculated with alchemical methods for a few protein–ligand systems (see Figure 10 for details). One of the most studied macromolecular systems has been the engineered binding pocket of T4 lysozyme. Mobley et al. studied the binding of thirteen single-ring fragment-like ligands to a L99A hydrophobic T4 lysozyme cavity mutant. The computed absolute binding free energies were found to have an RMS error of 1.9 kcal/mol relative to previously determined experimental values.

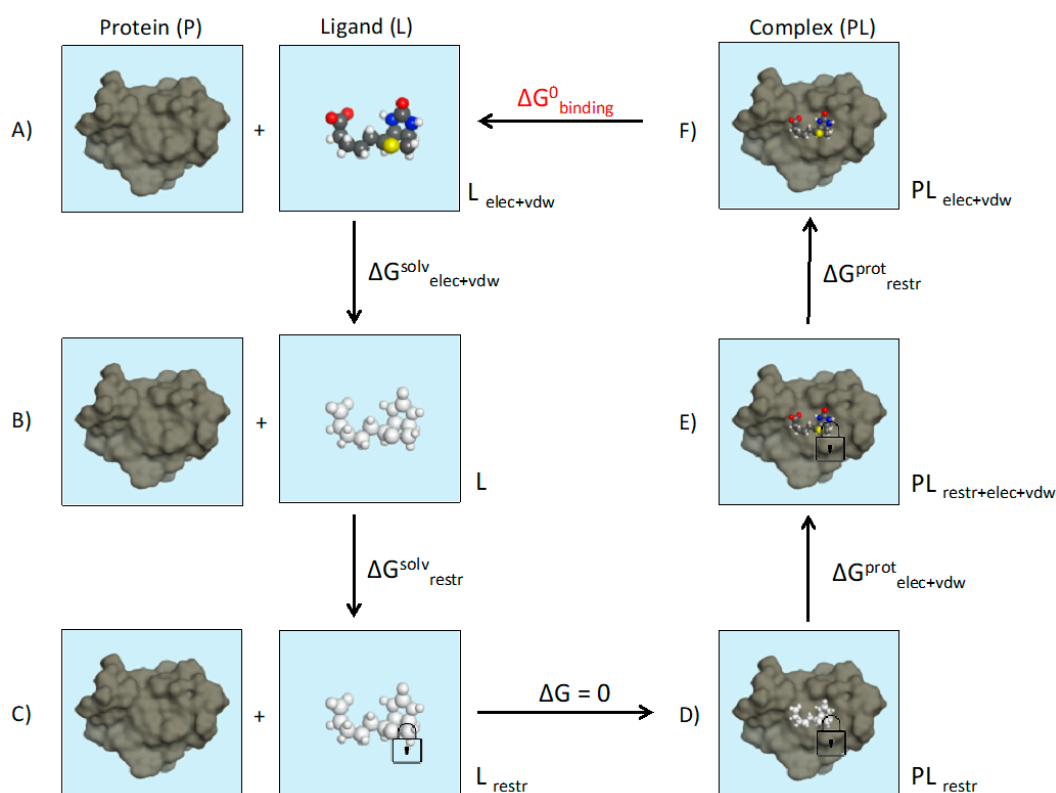


Figure 10. Scheme of the alchemical thermodynamic cycle used to obtain the absolute binding free energies. The restraints are represented as a lock fastening the ligand to the protein. The interacting ligand is depicted in color while the non-interacting one is transparent. The light-blue background represents the water environment. (A) Starting from the top-left corner, the interacting ligand (in color) is transformed to a non-interacting ligand (transparent) where its electrostatic and vdW interactions are switched off, releasing the term $\Delta G^{\text{solv}}_{\text{elec+vdw}}$. (B) Then a set of restraints is added to the non-interacting ligand, delivering the term, $\Delta G^{\text{solv}}_{\text{restr}}$, which is computed analytically using the protocol proposed by Boresch [190]. (C) This state is equivalent to having the non-interacting ligand restrained within the protein cavity. Here the restraint is still on in order to prevent the ligand leaving the binding pocket when the interactions are scaled to zero. (D) To bring back the restrained non-interacting ligand in complex with protein to the interaction mode, the charges are turned back on by switching the vdW and electrostatic interactions on ($\Delta G^{\text{prot}}_{\text{elec+vdw}}$). Finally, the restraints between the fully interacting ligand in complex with protein are eliminated. To calculate the absolute binding free energies (ΔG_{bind}) we sum up the energies of a series of intermediate simulations in a closed cycle. The standard state correction ($\Delta G = 0$) is usually added to this contribution.

In the past, applications of free energy simulations were typically limited to a very handful number of ligands that were analyzed after the experimental work. In this respect, all reports describe work that has been carried out in a retrospective manner, where ligand affinities are known and the binding conformations have usually been determined by X-ray crystallography. While retrospective studies often report good results (successful examples), the applicability of free-energy simulations to prospective work has been limited due to the lack of large-scale validation coupled with the technical challenges traditionally associated. Fortunately, in recent years more prospective free energy simulations (in some cases, combined with retrospective studies) have been published by different groups such as Janssen Pharmaceuticals [191–193], Jorgensen lab [179,194–199], Lovering et al. at Pfizer [200] and researches at Schrodinger and their academic collaborators [201,202] demonstrating the robustness and broad range of applicability of this approach, which can be used to drive decisions in virtual screening and lead optimization stages of drug discovery. For further reading, we refer the interested readers to other excellent recent review papers [203,204] on free energy methods.

Although recent results from MD-based drug discovery studies are very encouraging, it ought to be borne in mind that there are still major challenges to be overcome in order to strengthen the impact of MD-based methods on drug design. Further improvement in the current FFs is required to order to achieve more accurate simulations. To date *in silico* drug design methods have been of vast importance in target identification and in prediction of novel drugs. In conclusion, the fusion of genomics, bioinformatics, and computational power could do wonders in improving the success of computer-aided drug design strategies, assisting in identification of new targets, determination of their structures, and the forming of a quantitative picture of the interactions between macromolecule and the ligands. Computational methods have provided a powerful toolbox for target identification, discovery, and optimization of drug candidate molecules. Information technologies coupled with statistics and chemoinformatic tools, in turn, shed light on disease mechanisms and phenotypes, revealing potential drug targets to be further validated by high throughput screening technologies. Consecutively, multiple methods allow for the prediction and characterization of binding sites through study of the dynamic nature of drug targets, identifying and optimizing new active molecular entities. Today, large databases of commercially available compounds together with ligand chemical space exploration offer drug discovery scientists an enormous volume of data with which to work. Different methods based on readily available information on the biological system under study are evolving to assist with the manipulation and processing of this data. Moreover, integration of ‘-omics’ technologies and databases may facilitate the identification of novel drug targets or the design of network-based multi-target drugs. Structure- and ligand-based methods are the most commonly used methods within the drug discovery field; however, combinatorial techniques such as proteochemometrics are gaining prominence.

In future, after the completion of human genome project, pharmacogenomics, which evaluates the effect of genes and their polymorphism on drug response will revolutionize the drug discovery and development process based on the evaluation of the different genetic markers. Furthermore, medicine will be smarter, safer, more personalized and more efficacious based on pharmacogenomics approaches [205]. Another promising example in drug discovery is that deep learning methods will become a major computer-aided drug design (CADD) approach in the near future. Although machine-learning approaches (e.g., QSAR) in modeling studies are more popular, machine intelligence approaches has been replaced with the deep learning in recent years since it can deal with complex tasks based on large, heterogeneous, and high-dimensional data sets without the need for human input [206]. Also, an emerging paradigm in modern drug discovery is of the Poly pharmacology, which is the design or use of pharmaceutical agents that act on multiple targets or disease pathways. It is generally thought that complex diseases (ie. Cancer) may require complex therapeutic approaches. Therefore, polypharmacology suggests that more effective drugs can be developed by specifically modulating multiple targets. The computational strategies plays an important role in its progress [207].

5. Conclusions

Large-scale MD simulation of biomolecules and biomacromolecules is an interesting and rapidly developing area that is contributing increasingly to fundamental understanding of living organisms. In the era of petascale computing today, large-scale MD simulations are having a profound impact on numerous diverse scientific endeavors, from biotechnological applications such as the fabrication of novel smart biomaterials, to DNA sequencing and the treatment of disease and development of drugs. Although several challenges lie ahead with regard to the improvement of the molecular FF and sampling of the conformational space, considering the success of these applications thus far, there is little doubt that large-scale molecular dynamics simulations will play an even more crucial and expanding role in future work in this area.

The use of computational modeling to complement experiments is helping to bridge the gap between atomic-level properties with whole-organism function, an endeavor which cannot be accomplished by either approach alone. A combination of multiple computational techniques, covering a vast range of time and size scale, is optimal for efficiently capturing information across biological scales.

Funding: This research was funded by Province of Alberta, Alberta Innovates—Technology and the Natural Sciences and Engineering Research Council of Canada (NSERC).

Acknowledgments: To the memory of Carlo Montemagno, my advisor, my mentor and a great human being. This work was funded by Province of Alberta, Alberta Innovates—Technology and the Natural Sciences and Engineering Research Council of Canada (NSERC).

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

Abbreviations

MD, Molecular dynamics; BPTI, Bovine pancreatic trypsin inhibitor; RCSB, Research Collaboratory for Structural Bioinformatics; MM, molecular mechanics; NMR, Nuclear magnetic resonance; QM, Quantum mechanics; IR, infrared spectroscopy; EVB, empirical valence bond; FF, force field; NVE, Microcanonical Ensemble; NVT, Canonical ensemble; NPT, Isotherma-isobaric ensemble; RMSD, root-mean-square deviation; CG, coarse graining, SMD, Steered molecular dynamics; US, Umbrella sampling; AFM, atomic force microscopy; WHAM, weighted histogram analysis method; PMF, potential of mean force; GPU, graphics processing units; Titania, Titanium oxide; REST, replica exchange with solution tempering; dsDNA, double-strand DNA; ssDNA, single-stranded DNA; MspA, Mycobacterium smegmatis porin A; α -HL, α -Hemolysin; CADD, Computer-aided Drug Design; LB, ligand-based; SB, structure-based; QSAR, quantitative structure–activity relationship; BOMB, biochemical and organic model builder; ADME, absorption, distribution, metabolism, and excretion/elimination; FEP, free energy perturbation; TI, thermodynamic integration.

References

1. Pollack, L. Klaus Schulten Reflects on his 2015 National Lecture. Available online: <http://www.ks.uiuc.edu/events/NationalLecture2015/reflection/> (accessed on 21 March 2019).
2. Feynman, R.P.; Leighton, R.B.; Sands, M. *The Feynman Lectures in Physics*; Addison-Wesley: Reading, MA, USA, 1963; Volume I.
3. McCammon, J.A.; Gelin, B.R.; Karplus, M. Perspective on “Dynamics of folded proteins”. *Nature* **1977**, *267*, 585–590. [[CrossRef](#)]
4. 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](#)] [[PubMed](#)]
5. Von Dreele, R.B.; Stephens, P.W.; Smith, G.D.; Blessing, R.H. The first protein crystal structure determined from high-resolution X-ray powder diffraction data: A variant of T3R3 human insulin-zinc complex produced by grinding. *Acta Cryst. D* **2000**, *56*, 1549–1553. [[CrossRef](#)] [[PubMed](#)]
6. Röhrig, U.F.; Guidoni, L.; Laio, A.; Frank, I.; Rothlisberger, U. A Molecular Spring for Vision. *J. Am. Chem. Soc.* **2004**, *126*, 15328–15329. [[CrossRef](#)]
7. Rousseau, R.; Kleinschmidt, V.; Schmitt, U.W.; Marx, D. Modeling protonated water networks in bacteriorhodopsin. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1848–1859. [[CrossRef](#)]

8. Molteni, C.; Frank, I.; Parrinello, M. An Excited State Density Functional Theory Study of the Rhodopsin Chromophore. *J. Am. Chem. Soc.* **1999**, *121*, 12177–12183. [[CrossRef](#)]
9. Guidoni, L.; Carloni, P. Potassium permeation through the KcsA channel: A density functional study. *Biochim. Biophys. Acta* **2002**, *1563*, 1–6. [[CrossRef](#)]
10. Silvestrelli, P.L.; Bernasconi, M.; Parrinello, M. Ab initio infrared spectrum of liquid water. *Chem. Phys. Lett.* **1997**, *277*, 478–482. [[CrossRef](#)]
11. Zhu, Z.; Tuckerman, M.E. Ab Initio Molecular Dynamics Investigation of the Concentration Dependence of Charged Defect Transport in Basic Solutions via Calculation of the Infrared Spectrum. *J. Phys. Chem. B* **2002**, *106*, 8009–8018. [[CrossRef](#)]
12. Putrino, A.; Sebastiani, D.; Parrinello, M. Generalized variational density functional perturbation theory. *J. Chem. Phys.* **2000**, *113*, 7102–7109. [[CrossRef](#)]
13. Putrino, A.; Parrinello, M. Anharmonic Raman Spectra in High-Pressure Ice from Ab Initio Simulations. *Phys. Rev. Lett.* **2002**, *88*, 176401. [[CrossRef](#)]
14. Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. Ab Initio Molecular Dynamics-Based Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site. *J. Am. Chem. Soc.* **2001**, *123*, 8730–8737. [[CrossRef](#)]
15. Raugei, S.a.C. Scientific Highlight of the Month: “ab Initio Modeling of Biological Systems” Ab Initio Modeling of Biological Systems. p. 10. Available online: https://psi-k.net/download/highlights/Highlight_71.pdf (accessed on 24 April 2019).
16. Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474. [[CrossRef](#)]
17. Warshel, A. Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 5250–5254. [[CrossRef](#)]
18. Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Computational enzymatic catalysis clarifying enzymatic mechanisms with the help of computers. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12431–12441. [[CrossRef](#)]
19. Verlet, L. Computer Experiments on Classical Fluids.I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98. [[CrossRef](#)]
20. Huang, J.; Lopes, P.E.M.; Roux, B.; MacKerell, A.D. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *J. Phys. Chem. Lett.* **2014**, *5*, 3144–3150. [[CrossRef](#)]
21. Senftle, T.P.; Hong, S.; Islam, M.M.; Kylasa, S.B.; Zheng, Y.; Shin, Y.K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M.J.; Aktulga, H.M.; et al. The ReaxFF reactive force-field: Development, applications and future directions. *Npj Comput. Mat.* **2016**, *2*, 15011. [[CrossRef](#)]
22. Leach, A.R. *Molecular Modeling: Principles and Applications*; Addison Wesley Longman: Essex, UK, 1996.
23. Bronowska, A.K. *Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design, Thermodynamics—Interaction Studies—Solids, Liquids and Gases*; Moreno-Pirajan, J.C., Ed.; IntechOpen: London, UK, 2011. [[CrossRef](#)]
24. Villa, E.; Sengupta, J.; Trabuco, L.G.; LeBarron, J.; Baxter, W.T.; Shaikh, T.R.; Grassucci, R.A.; Nissen, P.; Ehrenberg, M.; Schulten, K.; et al. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1063–1068. [[CrossRef](#)]
25. Sanbonmatsu, K.Y.; Tung, C.S. High performance computing in biology: Multimillion atom simulations of nanoscale systems. *J. Struct. Biol.* **2007**, *157*, 470–480. [[CrossRef](#)]
26. Freddolino, P.L.; Arkhipov, A.S.; Larson, S.B.; McPherson, A.; Schulten, K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **2006**, *14*, 437–449. [[CrossRef](#)]
27. Zink, M.; Grubmuller, H. Mechanical Properties of the Icosahedral Shell of Southern Bean Mosaic Virus: A Molecular Dynamics Study. *Biophys. J.* **2009**, *96*, 1350–1363. [[CrossRef](#)]
28. Zhao, G.P.; Perala, J.R.; Yufenyuy, E.L.; Meng, X.; Chen, B.; Ning, J.Y.; Ahn, J.; Gronenborn, A.M.; Schulten, K.; Aiken, C.; et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013**, *497*, 643–646. [[CrossRef](#)]
29. Freddolino, P.L.; Liu, F.; Gruebele, M.; Schulten, K. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophys. J.* **2008**, *94*, L75–L77. [[CrossRef](#)]
30. Freddolino, P.L.; Schulten, K. Common Structural Transitions in Explicit-Solvent Simulations of Villin Headpiece Folding. *Biophys. J.* **2009**, *97*, 2338–2347. [[CrossRef](#)]

31. Klepeis, J.L.; Lindorff-Larsen, K.; Dror, R.O.; Shaw, D.E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struc. Biol.* **2009**, *19*, 120–127. [[CrossRef](#)]
32. Cartron, M.L.; Olsen, J.D.; Sener, M.; Jackson, P.J.; Brindley, A.A.; Qian, P.; Dickman, M.J.; Leggett, G.J.; Schulten, K.; Neil Hunter, C. Integration of energy and electron transfer processes in the photosynthetic membrane of *Rhodobacter sphaeroides*. *Biochim. Biophys. Acta (BBA) Bioenerg.* **2014**, *1837*, 1769–1780. [[CrossRef](#)]
33. Shih, A.Y.; Arkhipov, A.; Freddolino, P.L.; Schulten, K. Coarse Grained Protein–Lipid Model with Application to Lipoprotein Particles. *J. Phys. Chem. B* **2006**, *110*, 3674–3684. [[CrossRef](#)]
34. Arkhipov, A.; Freddolino, P.L.; Schulten, K. Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling. *Structure* **2006**, *14*, 1767–1777. [[CrossRef](#)]
35. Klein, M.L.; Shinoda, W. Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science* **2008**, *321*, 798. [[CrossRef](#)]
36. Yin, Y.; Arkhipov, A.; Schulten, K. Simulations of Membrane Tubulation by Lattices of Amphiphysin N-BAR Domains. *Structure* **2009**, *17*, 882–892. [[CrossRef](#)]
37. Levitt, M.; Warshel, A. Computer simulation of protein folding. *Nature* **1975**, *253*, 694–698. [[CrossRef](#)]
38. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **1976**, *104*, 59–107. [[CrossRef](#)]
39. Torrie, G.M.; Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199. [[CrossRef](#)]
40. Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566. [[CrossRef](#)]
41. Isralewitz, B.; Gao, M.; Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struc. Biol.* **2001**, *11*, 224–230. [[CrossRef](#)]
42. Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **1996**, *271*, 997. [[CrossRef](#)]
43. Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Mod.* **2001**, *19*, 13–25. [[CrossRef](#)]
44. Florin, E.L.; Moy, V.T.; Gaub, H.E. Adhesion forces between individual ligand-receptor pairs. *Science* **1994**, *264*, 415. [[CrossRef](#)]
45. Svoboda, K.; Schmidt, C.F.; Schnapp, B.J.; Block, S.M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **1993**, *365*, 721–727. [[CrossRef](#)]
46. Smith, S.B.; Cui, Y.; Bustamante, C. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* **1996**, *271*, 795. [[CrossRef](#)] [[PubMed](#)]
47. Cluzel, P.; Lebrun, A.; Heller, C.; Lavery, R.; Viovy, J.-L.; Chatenay, D.; Caron, F. DNA: An Extensible Molecule. *Science* **1996**, *271*, 792. [[CrossRef](#)]
48. Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J.M.; Gaub, H.E. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science* **1997**, *276*, 1109. [[CrossRef](#)] [[PubMed](#)]
49. Kellermayer, M.S.Z.; Smith, S.B.; Granzier, H.L.; Bustamante, C. Folding-Unfolding Transitions in Single Titin Molecules Characterized with Laser Tweezers. *Science* **1997**, *276*, 1112. [[CrossRef](#)]
50. Kumar, S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021. [[CrossRef](#)]
51. Ge, X.; Roux, B. Absolute Binding Free Energy Calculations of Sparsomycin Analogs to the Bacterial Ribosome. *J. Phys. Chem. B* **2010**, *114*, 9525–9539. [[CrossRef](#)] [[PubMed](#)]
52. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2013**, *66*, 334. [[CrossRef](#)] [[PubMed](#)]
53. Perilla, J.R.; Goh, B.C.; Cassidy, C.K.; Liu, B.; Bernardi, R.C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struc. Biol.* **2015**, *31*, 64–74. [[CrossRef](#)]
54. Yu, I.; Mori, T.; Ando, T.; Harada, R.; Jung, J.; Sugita, Y.; Feig, M. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife* **2016**, *5*, e19274. [[CrossRef](#)]

55. Shaw, D.E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Eastwood, M.P.; Bank, J.A.; Jumper, J.M.; Salmon, J.K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346. [[CrossRef](#)]
56. Feig, M.; Nawrocki, G.; Yu, I.; Wang, P.-H.; Sugita, Y. Challenges and opportunities in connecting simulations with experiments via molecular dynamics of cellular environments. *J. Phys. Conf. Series* **2018**, *1036*, 012010. [[CrossRef](#)] [[PubMed](#)]
57. Miwa, K.; Asahi, R. Molecular dynamics simulations with machine learning potential for Nb-doped lithium garnet-type oxide $\text{Li}_{7-x}\text{La}_3(\text{Zr}_{2-x}\text{Nb}_x)\text{O}_{12}$. *Phys. Rev. Mat.* **2018**, *2*, 105404.
58. Hollingsworth, S.A.; Dror, R.O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143. [[CrossRef](#)] [[PubMed](#)]
59. Zhao, H.; Caflisch, A. Molecular dynamics in drug design. *Eur. J. Med. Chem.* **2015**, *91*, 4–14. [[CrossRef](#)] [[PubMed](#)]
60. 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](#)] [[PubMed](#)]
61. Heinz, H.; Ramezani-Dakhel, H. Simulations of inorganic–bioorganic interfaces to discover new materials: insights, comparisons to experiment, challenges, and opportunities. *Chem. Soc. Rev.* **2016**, *45*, 412–448. [[CrossRef](#)] [[PubMed](#)]
62. Maffeo, C.; Yoo, J.; Comer, J.; Wells, D.B.; Luan, B.; Aksimentiev, A. Close encounters with DNA. *J. Phys.* **2014**, *26*, 413101. [[CrossRef](#)]
63. Adhvaryu, A.; Sharma, B.K.; Hwang, H.S.; Erhan, S.Z.; Perez, J.M. Development of Biobased Synthetic Fluids: Application of Molecular Modeling to Structure–Physical Property Relationship. *Ind. Eng. Chem. Res.* **2006**, *45*, 928–933. [[CrossRef](#)]
64. Thompson, D.W. *On Growth and Form*; Cambridge University Press: Cambridge, UK, 1942.
65. Maier, I.; Morgan, M.R.A.; Lindner, W.; Pittner, F. Optical Resonance-Enhanced Absorption-Based Near-Field ImmunoChip Biosensor for Allergen Detection. *Anal. Chem.* **2008**, *80*, 2694–2703. [[CrossRef](#)] [[PubMed](#)]
66. Xu, Y.; Cao, Q.; Svec, F.; Fréchet, J.M.J. Porous Polymer Monolithic Column with Surface-Bound Gold Nanoparticles for the Capture and Separation of Cysteine-Containing Peptides. *Anal. Chem.* **2010**, *82*, 3352–3358. [[CrossRef](#)]
67. Hartmann, M. Ordered Mesoporous Materials for Bioadsorption and Biocatalysis. *Chem. Mat.* **2005**, *17*, 4577–4593. [[CrossRef](#)]
68. Yguerabide, J.; Yguerabide, E.E. Resonance light scattering particles as ultrasensitive labels for detection of analytes in a wide range of applications. *J. Cellular Biochem.* **2001**, *84*, 71–81. [[CrossRef](#)]
69. Wright, M.; Uddin, A. Organic–inorganic hybrid solar cells: A comparative review. *Solar Energy Mat. Solar Cells* **2012**, *107*, 87–111. [[CrossRef](#)]
70. Fenton, O.S.; Olafson, K.N.; Pillai, P.S.; Mitchell, M.J.; Langer, R. Advances in Biomaterials for Drug Delivery. *Adv. Mater.* **2018**, *30*, 1705328. [[CrossRef](#)]
71. Patra, J.K.; Das, G.; Fraceto, L.F.; Campos, E.V.R.; Rodriguez-Torres, M.D.P.; Acosta-Torres, L.S.; Diaz-Torres, L.A.; Grillo, R.; Swamy, M.K.; Sharma, S.; et al. Nano based drug delivery systems: Recent developments and future prospects. *J. Nanobiotechnol.* **2018**, *16*, 71. [[CrossRef](#)]
72. Kinnear, C.; Moore, T.L.; Rodriguez-Lorenzo, L.; Rothen-Rutishauser, B.; Petri-Fink, A. Form Follows Function: Nanoparticle Shape and Its Implications for Nanomedicine. *Chem. Rev.* **2017**, *117*, 11476–11521. [[CrossRef](#)]
73. Penna, M.J.a.B.; Mark, J. The Binding Mechanism of an Experimentally Identified Platinum-binding Peptide by Molecular Dynamics Simulation. In Proceedings of the Chemeca 2010: Engineering at the Edge, Hilton Adelaide, Australia, 26–29 September 2010; Engineers Australia: Barton, Australia; pp. 289–298.
74. Penna, M.J.; Mijajlovic, M.; Biggs, M.J. Molecular-Level Understanding of Protein Adsorption at the Interface between Water and a Strongly Interacting Uncharged Solid Surface. *J. Am. Chem. Soc.* **2014**, *136*, 5323–5331. [[CrossRef](#)]
75. Brancolini, G.; Kokh, D.B.; Calzolari, L.; Wade, R.C.; Corni, S. Docking of Ubiquitin to Gold Nanoparticles. *ACS Nano* **2012**, *6*, 9863–9878. [[CrossRef](#)]
76. Hildebrand, N.; Köppen, S.; Derr, L.; Li, K.; Koleini, M.; Rezwan, K.; Colombi Ciacchi, L. Adsorption Orientation and Binding Motifs of Lysozyme and Chymotrypsin on Amorphous Silica. *J. Phys. Chem. C* **2015**, *119*, 7295–7307. [[CrossRef](#)]

77. Heinz, H.; Lin, T.-J.; Kishore Mishra, R.; Emami, F.S. Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field. *Langmuir* **2013**, *29*, 1754–1765. [[CrossRef](#)]
78. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19. [[CrossRef](#)]
79. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [[CrossRef](#)]
80. Iori, F.; Corni, S. Including image charge effects in the molecular dynamics simulations of molecules on metal surfaces. *J. Comput. Chem.* **2008**, *29*, 1656–1666. [[CrossRef](#)]
81. Wright, L.B.; Rodger, P.M.; Corni, S.; Walsh, T.R. GoIP-CHARMM: First-Principles Based Force Fields for the Interaction of Proteins with Au(111) and Au(100). *J. Chem. Theory Comp.* **2013**, *9*, 1616–1630. [[CrossRef](#)]
82. Monti, S.; Barcaro, G.; Sementa, L.; Carravetta, V.; Ågren, H. Dynamics and self-assembly of bio-functionalized gold nanoparticles in solution: Reactive molecular dynamics simulations. *Nano Res.* **2018**, *11*, 1757–1767. [[CrossRef](#)]
83. Watabe, M.; Yamada, H.; Miyakawa, T.; Morikawa, R.; Takasu, M.; Uchida, T.; Yamagishi, A. Structural Analysis of Metal-Binding Peptides Using Molecular Dynamics. *Conf. Proc. ICBBB* **2018**.
84. Ma, W.; Saccardo, A.; Roccatano, D.; Aboagye-Mensah, D.; Alkaseem, M.; Jewkes, M.; Di Nezza, F.; Baron, M.; Soloviev, M.; Ferrari, E. Modular assembly of proteins on nanoparticles. *Nature Commu.* **2018**, *9*, 1489. [[CrossRef](#)]
85. Heinz, H.; Farmer, B.L.; Pandey, R.B.; Slocik, J.M.; Patnaik, S.S.; Pachter, R.; Naik, R.R. Nature of Molecular Interactions of Peptides with Gold, Palladium, and Pd–Au Bimetal Surfaces in Aqueous Solution. *J. Am. Chem. Soc.* **2009**, *131*, 9704–9714. [[CrossRef](#)]
86. Ruan, L.; Ramezani-Dakhel, H.; Chiu, C.-Y.; Zhu, E.; Li, Y.; Heinz, H.; Huang, Y. Tailoring Molecular Specificity Toward a Crystal Facet: a Lesson From Biorecognition Toward Pt{111}. *Nano Lett.* **2013**, *13*, 840–846. [[CrossRef](#)]
87. Cetinel, S.; Shen, W.-Z.; Aminpour, M.; Bhomkar, P.; Wang, F.; Borujeny, E.R.; Sharma, K.; Nayeibi, N.; Montemagno, C. Biomining of MoS(2) with Peptide-based Smart Biomaterials. *Sci. Rep.* **2018**, *8*, 3374. [[CrossRef](#)]
88. Dreaden, E.C.; Alkilany, A.M.; Huang, X.; Murphy, C.J.; El-Sayed, M.A. The golden age: gold nanoparticles for biomedicine. *Chem. Soc. Rev.* **2012**, *41*, 2740–2779. [[CrossRef](#)]
89. Jain, P.K.; Huang, X.; El-Sayed, I.H.; El-Sayed, M.A. Noble Metals on the Nanoscale: Optical and Photothermal Properties and Some Applications in Imaging, Sensing, Biology, and Medicine. *Accounts Chem. Res.* **2008**, *41*, 1578–1586. [[CrossRef](#)]
90. Salzmann, C.G.; Ward, M.A.H.; Jacobs, R.M.J.; Tobias, G.; Green, M.L.H. Interaction of Tyrosine-, Tryptophan-, and Lysine-Containing Polypeptides with Single-Wall Carbon Nanotubes and Its Relevance for the Rational Design of Dispersing Agents. *J. Phys. Chem. C* **2007**, *111*, 18520–18524. [[CrossRef](#)]
91. Kim, S.N.; Kuang, Z.; Slocik, J.M.; Jones, S.E.; Cui, Y.; Farmer, B.L.; McAlpine, M.C.; Naik, R.R. Preferential Binding of Peptides to Graphene Edges and Planes. *J. Am. Chem. Soc.* **2011**, *133*, 14480–14483. [[CrossRef](#)]
92. Yang, Z.; Wang, Z.; Tian, X.; Xiu, P.; Zhou, R. Amino acid analogues bind to carbon nanotube via π - π interactions: Comparison of molecular mechanical and quantum mechanical calculations. *J. Chem. Phys.* **2012**, *136*, 025103. [[CrossRef](#)]
93. Ramakrishnan, S.K.; Martin, M.; Cloitre, T.; Firlej, L.; Gergely, C. Molecular Mechanism of Selective Binding of Peptides to Silicon Surface. *J. Chem. Informat. Modeling* **2014**, *54*, 2117–2126. [[CrossRef](#)]
94. Welch, C.M.; Camden, A.N.; Barr, S.A.; Leuty, G.M.; Kedziora, G.S.; Berry, R.J. Computation of the binding free energy of peptides to graphene in explicit water. *J. Chem. Phys.* **2015**, *143*, 045104. [[CrossRef](#)]
95. Kim, S.S.; Kuang, Z.; Ngo, Y.H.; Farmer, B.L.; Naik, R.R. Biotic–Abiotic Interactions: Factors that Influence Peptide–Graphene Interactions. *ACS Appl. Mat. Interfaces* **2015**, *7*, 20447–20453. [[CrossRef](#)]
96. Patwardhan, S.V.; Emami, F.S.; Berry, R.J.; Jones, S.E.; Naik, R.R.; Deschaume, O.; Heinz, H.; Perry, C.C. Chemistry of Aqueous Silica Nanoparticle Surfaces and the Mechanism of Selective Peptide Adsorption. *J. Am. Chem. Soc.* **2012**, *134*, 6244–6256. [[CrossRef](#)]
97. Tosaka, R.; Yamamoto, H.; Ohdomari, I.; Watanabe, T. Adsorption Mechanism of Ribosomal Protein L2 onto a Silica Surface: A Molecular Dynamics Simulation Study. *Langmuir* **2010**, *26*, 9950–9955. [[CrossRef](#)]

98. Rimola, A.; Sodupe, M.; Ugliengo, P. Affinity Scale for the Interaction of Amino Acids with Silica Surfaces. *J. Phys. Chem. C* **2009**, *113*, 5741–5750. [[CrossRef](#)]
99. Nonella, M.; Seeger, S. Investigating Alanine–Silica Interaction by Means of First-Principles Molecular-Dynamics Simulations. *ChemPhysChem* **2008**, *9*, 414–421. [[CrossRef](#)] [[PubMed](#)]
100. Sun, T.; Han, G.; Lindgren, M.; Shen, Z.; Laaksonen, A. Adhesion of lactoferrin and bone morphogenetic protein-2 to a rutile surface: dependence on the surface hydrophobicity. *Biomater. Sci.* **2014**, *2*, 1090–1099. [[CrossRef](#)]
101. Monti, S.; Carravetta, V.; Battocchio, C.; Iucci, G.; Polzonetti, G. Peptide/TiO₂ Surface Interaction: A Theoretical and Experimental Study on the Structure of Adsorbed ALA-GLU and ALA-LYS. *Langmuir* **2008**, *24*, 3205–3214. [[CrossRef](#)] [[PubMed](#)]
102. Carravetta, V.; Monti, S.; Zhang, W. Interaction of biomolecular systems with titanium-based materials: Computational investigations. *Theor. Chem. Acc.* **2009**, *123*, 299–309. [[CrossRef](#)]
103. Sanchez, C.; Julián, B.; Belleville, P.; Popall, M. Applications of hybrid organic–inorganic nanocomposites. *J. Mat. Chem.* **2005**, *15*, 3559–3592. [[CrossRef](#)]
104. Ambrogio, M.W.; Thomas, C.R.; Zhao, Y.-L.; Zink, J.I.; Stoddart, J.F. Mechanized Silica Nanoparticles: A New Frontier in Theranostic Nanomedicine. *Acc. Chem. Res.* **2011**, *44*, 903–913. [[CrossRef](#)] [[PubMed](#)]
105. Deng, Y.; Cai, Y.; Sun, Z.; Liu, J.; Liu, C.; Wei, J.; Li, W.; Liu, C.; Wang, Y.; Zhao, D. Multifunctional Mesoporous Composite Microspheres with Well-Designed Nanostructure: A Highly Integrated Catalyst System. *J. Am. Chem. Soc.* **2010**, *132*, 8466–8473. [[CrossRef](#)] [[PubMed](#)]
106. Jouault, N.; Vallat, P.; Dalmas, F.; Said, S.; Jestin, J.; Boué, F. Well-Dispersed Fractal Aggregates as Filler in Polymer–Silica Nanocomposites: Long-Range Effects in Rheology. *Macromolecules* **2009**, *42*, 2031–2040. [[CrossRef](#)]
107. Tangney, P.; Scandolo, S. An ab initio parametrized interatomic force field for silica. *J. Chem. Phys.* **2002**, *117*, 8898–8904. [[CrossRef](#)]
108. Kermodé, J.R.; Cereda, S.; Tangney, P.; De Vita, A. A first principles based polarizable O(N) interatomic force field for bulk silica. *J. Chem. Phys.* **2010**, *133*, 094102. [[CrossRef](#)]
109. Butenuth, A.; Moras, G.; Schneider, J.; Koleini, M.; Köppen, S.; Meißner, R.; Wright, L.B.; Walsh, T.R.; Ciacchi, L.C. Ab initio derived force-field parameters for molecular dynamics simulations of deprotonated amorphous-SiO₂/water interfaces. *Phys. Status Solidi B* **2012**, *249*, 292–305. [[CrossRef](#)]
110. Lopes, P.E.M.; Murashov, V.; Tazi, M.; Demchuk, E.; MacKerell, A.D. Development of an Empirical Force Field for Silica. Application to the Quartz–Water Interface. *J. Phys. Chem. B* **2006**, *110*, 2782–2792. [[CrossRef](#)]
111. Pedone, A.; Malavasi, G.; Menziani, M.C.; Segre, U.; Musso, F.; Corno, M.; Civalleri, B.; Ugliengo, P. FFSiOH: a New Force Field for Silica Polymorphs and Their Hydroxylated Surfaces Based on Periodic B3LYP Calculations. *Chem. Mat.* **2008**, *20*, 2522–2531. [[CrossRef](#)]
112. Ramakrishnan, S.K.; Estephan, E.; Martin, M.; Cloitre, T.; Gergely, C. Probing the mechanism of material specific peptides for optical biosensors. *Int. Soc. Opt. Photonics* **2013**. [[CrossRef](#)]
113. Heinz, H. Adsorption of biomolecules and polymers on silicates, glasses, and oxides: mechanisms, predictions, and opportunities by molecular simulation. *Curr. Opin. Chem. Eng.* **2016**, *11*, 34–41. [[CrossRef](#)]
114. Emami, F.S.; Puddu, V.; Berry, R.J.; Varshney, V.; Patwardhan, S.V.; Perry, C.C.; Heinz, H. Prediction of Specific Biomolecule Adsorption on Silica Surfaces as a Function of pH and Particle Size. *Chem. Mat.* **2014**, *26*, 5725–5734. [[CrossRef](#)]
115. Carravetta, V.; Monti, S. Peptide–TiO₂ Surface Interaction in Solution by Ab Initio and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006**, *110*, 6160–6169. [[CrossRef](#)]
116. Chenoweth, K.; van Duin, A.C.T.; Goddard, W.A. ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation. *J. Phys. Chem. A* **2008**, *112*, 1040–1053. [[CrossRef](#)]
117. van der Kamp, M.W.; Mulholland, A.J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728. [[CrossRef](#)]
118. Ghadari, R. A study on the interactions of amino acids with nitrogen doped graphene; docking, MD simulation, and QM/MM studies. *Phys. Chem. Chem. Phys.* **2016**, *18*, 4352–4361. [[CrossRef](#)] [[PubMed](#)]
119. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [[CrossRef](#)] [[PubMed](#)]
120. Akesson, M.; Branton, D.; Church, G.; Deamer, D.W. Characterization of individual polymer molecules based on monomer-interface interactions. U.S. Patent US20120160687A1, 28 June 2018.

121. Bond, P.J.; Guy, A.T.; Heron, A.J.; Bayley, H.; Khalid, S. Molecular Dynamics Simulations of DNA within a Nanopore: Arginine–Phosphate Tethering and a Binding/Sliding Mechanism for Translocation. *Biochemistry* **2011**, *50*, 3777–3783. [[CrossRef](#)] [[PubMed](#)]
122. Guy, A.T.; Piggot, T.J.; Khalid, S. Single-Stranded DNA within Nanopores: Conformational Dynamics and Implications for Sequencing; a Molecular Dynamics Simulation Study. *Biophys. J.* **2012**, *103*, 1028–1036. [[CrossRef](#)] [[PubMed](#)]
123. Parkin, J.; Khalid, S. Atomistic Molecular-Dynamics Simulations Enable Prediction of the Arginine Permeation Pathway through OccD1/OprD from *Pseudomonas aeruginosa*. *Biophys. J.* **2014**, *107*, 1853–1861. [[CrossRef](#)]
124. Bhattacharya, S.; Derrington, I.M.; Pavlenok, M.; Niederweis, M.; Gundlach, J.H.; Aksimentiev, A. Molecular Dynamics Study of MspA Arginine Mutants Predicts Slow DNA Translocations and Ion Current Blockades Indicative of DNA Sequence. *ACS Nano* **2012**, *6*, 6960–6968. [[CrossRef](#)]
125. Haider, S.; Neidle, S. Molecular Modeling and Simulation of G-Quadruplexes and Quadruplex-Ligand Complexes. In *G-Quadruplex DNA: Methods and Protocols*; Baumann, P., Ed.; Humana Press: Totowa, NJ, USA, 2010; pp. 17–37. [[CrossRef](#)]
126. Kang, X.-F.; Gu, L.-Q.; Cheley, S.; Bayley, H. Single Protein Pores Containing Molecular Adapters at High Temperatures. *Angew. Chem. Int. Ed.* **2005**, *44*, 1495–1499. [[CrossRef](#)]
127. Cherf, G.M.; Lieberman, K.R.; Rashid, H.; Lam, C.E.; Karplus, K.; Akeson, M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. *Nat. Biotech.* **2012**, *30*, 344–348. [[CrossRef](#)]
128. Menestrina, G. Ionic channels formed by *Staphylococcus aureus* alpha-toxin: Voltage-dependent inhibition by divalent and trivalent cations. *J. Mem. Biol.* **1986**, *90*, 177–190. [[CrossRef](#)]
129. Korchev, Y.E.; Alder, G.M.; Bakhrarov, A.; Bashford, C.L.; Joomun, B.S.; Sviderskaya, E.V.; Usherwood, P.N.R.; Pasternak, C.A. *Staphylococcus aureus* alpha-toxin-induced pores: Channel-like behavior in lipid bilayers and patch clamped cells. *J. Membrane Biol.* **1995**, *143*, 143–151. [[CrossRef](#)]
130. Mathé, J.; Aksimentiev, A.; Nelson, D.R.; Schulten, K.; Meller, A. Orientation discrimination of single-stranded DNA inside the α -hemolysin membrane channel. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12377–12382. [[CrossRef](#)]
131. Derrington, I.M.; Butler, T.Z.; Collins, M.D.; Manrao, E.; Pavlenok, M.; Niederweis, M.; Gundlach, J.H. Nanopore DNA sequencing with MspA. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16060. [[CrossRef](#)]
132. Bhattacharya, S.; Yoo, J.; Aksimentiev, A. Water Mediates Recognition of DNA Sequence via Ionic Current Blockade in a Biological Nanopore. *ACS Nano* **2016**, *10*, 4644–4651. [[CrossRef](#)]
133. Merchant, C.A.; Healy, K.; Wanunu, M.; Ray, V.; Peterman, N.; Bartel, J.; Fischbein, M.D.; Venta, K.; Luo, Z.; Johnson, A.T.C.; et al. DNA Translocation through Graphene Nanopores. *Nano Lett.* **2010**, *10*, 2915–2921. [[CrossRef](#)] [[PubMed](#)]
134. Schneider, G.F.; Kowalczyk, S.W.; Calado, V.E.; Pandraud, G.; Zandbergen, H.W.; Vandersypen, L.M.K.; Dekker, C. DNA Translocation through Graphene Nanopores. *Nano Lett.* **2010**, *10*, 3163–3167. [[CrossRef](#)]
135. Garaj, S.; Hubbard, W.; Reina, A.; Kong, J.; Branton, D.; Golovchenko, J.A. Graphene as a subnanometre trans-electrode membrane. *Nature* **2010**, *467*, 190–193. [[CrossRef](#)] [[PubMed](#)]
136. Li, J.; Stein, D.; McMullan, C.; Branton, D.; Aziz, M.J.; Golovchenko, J.A. Ion-beam sculpting at nanometre length scales. *Nature* **2001**, *412*, 166–169. [[CrossRef](#)] [[PubMed](#)]
137. Storm, A.J.; Chen, J.H.; Zandbergen, H.W.; Dekker, C. Translocation of double-strand DNA through a silicon oxide nanopore. *Phys. Rev. E* **2005**, *71*, 051903. [[CrossRef](#)]
138. Heng, J.B.; Ho, C.; Kim, T.; Timp, R.; Aksimentiev, A.; Grinkova, Y.V.; Sligar, S.; Schulten, K.; Timp, G. Sizing DNA Using a Nanometer-Diameter Pore. *Biophys. J.* **2004**, *87*, 2905–2911. [[CrossRef](#)] [[PubMed](#)]
139. Heng, J.B.; Aksimentiev, A.; Ho, C.; Marks, P.; Grinkova, Y.V.; Sligar, S.; Schulten, K.; Timp, G. Stretching DNA Using the Electric Field in a Synthetic Nanopore. *Nano Lett.* **2005**, *5*, 1883–1888. [[CrossRef](#)] [[PubMed](#)]
140. Heng, J.B.; Aksimentiev, A.; Ho, C.; Marks, P.; Grinkova, Y.V.; Sligar, S.; Schulten, K.; Timp, G. The Electromechanics of DNA in a Synthetic Nanopore. *Biophys. J.* **2006**, *90*, 1098–1106. [[CrossRef](#)] [[PubMed](#)]
141. Zhao, Q.; Comer, J.; Dimitrov, V.; Yemencioğlu, S.; Aksimentiev, A.; Timp, G. Stretching and unzipping nucleic acid hairpins using a synthetic nanopore. *Nucleic Acids Res.* **2008**, *36*, 1532–1541. [[CrossRef](#)]
142. Fologea, D.; Uplinger, J.; Thomas, B.; McNabb, D.S.; Li, J. Slowing DNA Translocation in a Solid-State Nanopore. *Nano Lett.* **2005**, *5*, 1734–1737. [[CrossRef](#)] [[PubMed](#)]
143. Chen, P.; Gu, J.; Brandin, E.; Kim, Y.-R.; Wang, Q.; Branton, D. Probing Single DNA Molecule Transport Using Fabricated Nanopores. *Nano Lett.* **2004**, *4*, 2293–2298. [[CrossRef](#)]

144. Gershow, M.; Golovchenko, J.A. Recapturing and trapping single molecules with a solid-state nanopore. *Nat. Nano* **2007**, *2*, 775–779. [[CrossRef](#)] [[PubMed](#)]
145. Wanunu, M.; Sutin, J.; McNally, B.; Chow, A.; Meller, A. DNA Translocation Governed by Interactions with Solid-State Nanopores. *Biophys. J.* **2008**, *95*, 4716–4725. [[CrossRef](#)] [[PubMed](#)]
146. Chang, H.; Kosari, F.; Andreadakis, G.; Alam, M.A.; Vasmatazis, G.; Bashir, R. DNA-Mediated Fluctuations in Ionic Current through Silicon Oxide Nanopore Channels. *Nano Lett.* **2004**, *4*, 1551–1556. [[CrossRef](#)]
147. Smeets, R.M.M.; Keyser, U.F.; Krapf, D.; Wu, M.-Y.; Dekker, N.H.; Dekker, C. Salt Dependence of Ion Transport and DNA Translocation through Solid-State Nanopores. *Nano Lett.* **2006**, *6*, 89–95. [[CrossRef](#)] [[PubMed](#)]
148. Trepagnier, E.H.; Radenovic, A.; Sivak, D.; Geissler, P.; Liphardt, J. Controlling DNA Capture and Propagation through Artificial Nanopores. *Nano Lett.* **2007**, *7*, 2824–2830. [[CrossRef](#)]
149. Fologea, D.; Gershow, M.; Ledden, B.; McNabb, D.S.; Golovchenko, J.A.; Li, J. Detecting Single Stranded DNA with a Solid State Nanopore. *Nano Lett.* **2005**, *5*, 1905–1909. [[CrossRef](#)]
150. Han, A.; Schürmann, G.; Mondin, G.; Bitterli, R.A.; Hegelbach, N.G.; de Rooij, N.F.; Stauffer, U. Sensing protein molecules using nanofabricated pores. *Appl. Phys. Lett.* **2006**, *88*, 093901. [[CrossRef](#)]
151. Fologea, D.; Ledden, B.; McNabb, D.S.; Li, J. Electrical characterization of protein molecules by a solid-state nanopore. *Appl. Phys. Lett.* **2007**, *91*, 053901. [[CrossRef](#)]
152. Fan, R.; Karnik, R.; Yue, M.; Li, D.; Majumdar, A.; Yang, P. DNA Translocation in Inorganic Nanotubes. *Nano Lett.* **2005**, *5*, 1633–1637. [[CrossRef](#)] [[PubMed](#)]
153. McNally, B.; Wanunu, M.; Meller, A. Electromechanical Unzipping of Individual DNA Molecules Using Synthetic Sub-2 nm Pores. *Nano Lett.* **2008**, *8*, 3418–3422. [[CrossRef](#)]
154. Harrell, C.C.; Choi, Y.; Horne, L.P.; Baker, L.A.; Siwy, Z.S.; Martin, C.R. Resistive-Pulse DNA Detection with a Conical Nanopore Sensor. *Langmuir* **2006**, *22*, 10837–10843. [[CrossRef](#)]
155. Bayley, H. Nanotechnology: Holes with an edge. *Nature* **2010**, *467*, 164–165. [[CrossRef](#)] [[PubMed](#)]
156. Yu, Y.-s.; Lu, X.; Ding, H.-m.; Ma, Y.-q. Computational investigation on DNA sequencing using functionalized graphene nanopores. *Phy. Chem. Chem. Phys.* **2018**, *20*, 9063–9069. [[CrossRef](#)] [[PubMed](#)]
157. Graf, M.; Liu, K.; Sarathy, A.; Leburton, J.-P.; Radenovic, A. Transverse Detection of DNA in a MoS₂ Nanopore. *Biophys. J.* **2018**, *114*, 180a. [[CrossRef](#)]
158. Li, H.; Carrion-Vazquez, M.; Oberhauser, A.F.; Marszalek, P.E.; Fernandez, J.M. Point mutations alter the mechanical stability of immunoglobulin modules. *Nat. Struct. Mol. Biol.* **2000**, *7*, 1117–1120.
159. Storm, A.J.; Chen, J.H.; Ling, X.S.; Zandbergen, H.W.; Dekker, C. Fabrication of solid-state nanopores with single-nanometre precision. *Nat. Mater.* **2003**, *2*, 537–540. [[CrossRef](#)]
160. Sathe, C.; Zou, X.; Leburton, J.-P.; Schulten, K. Computational Investigation of DNA Detection Using Graphene Nanopores. *ACS Nano* **2011**, *5*, 8842–8851. [[CrossRef](#)]
161. Wells, D.B.; Belkin, M.; Comer, J.; Aksimentiev, A. Assessing Graphene Nanopores for Sequencing DNA. *Nano Lett.* **2012**, *12*, 4117–4123. [[CrossRef](#)]
162. Liang, L.; Cui, P.; Wang, Q.; Wu, T.; Agren, H.; Tu, Y. Theoretical study on key factors in DNA sequencing with graphene nanopores. *RSC Adv.* **2013**, *3*, 2445–2453. [[CrossRef](#)]
163. Qiu, H.; Sarathy, A.; Leburton, J.-P.; Schulten, K. Intrinsic Stepwise Translocation of Stretched ssDNA in Graphene Nanopores. *Nano Lett.* **2015**, *15*, 8322–8330. [[CrossRef](#)]
164. Wanunu, M.; Meller, A. Chemically Modified Solid-State Nanopores. *Nano Lett.* **2007**, *7*, 1580–1585. [[CrossRef](#)] [[PubMed](#)]
165. Iqbal, S.M.; Akin, D.; Bashir, R. Solid-state nanopore channels with DNA selectivity. *Nat. Nano* **2007**, *2*, 243–248. [[CrossRef](#)] [[PubMed](#)]
166. Keyser, U.F. Controlling molecular transport through nanopores. *J. Royal Soc. Interface* **2011**, *8*, 1369. [[CrossRef](#)] [[PubMed](#)]
167. Hall, A.R.; Scott, A.; Rotem, D.; Mehta, K.K.; Bayley, H.; Dekker, C. Hybrid pore formation by directed insertion of [alpha]-haemolysin into solid-state nanopores. *Nat. Nano* **2010**, *5*, 874–877. [[CrossRef](#)]
168. Van Drie, J.H. Computer-aided drug design: the next 20 years. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 591–601. [[CrossRef](#)]
169. Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Disc. Today* **2010**, *15*, 444–450. [[CrossRef](#)]
170. Bajorath, J. (Ed.) *Chemoinformatics and Computational Chemical Biology*; Humana Press: Totowa, NJ, USA, 2011.

171. Jitender, V.; Vijay, M.K.; Evans, C.C. 3D-QSAR in Drug Design—A Review. *Curr. Topics Med. Chem.* **2010**, *10*, 95–115.
172. Lin, S.-K. Pharmacophore Perception, Development and Use in Drug Design. Edited by Osman F. Güner. *Molecules* **2000**, *5*, 987. [[CrossRef](#)]
173. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [[CrossRef](#)] [[PubMed](#)]
174. 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. Inform. Modeling* **2012**, *52*, 1757–1768. [[CrossRef](#)] [[PubMed](#)]
175. Irwin, J.J.; Shoichet, B.K. ZINC—A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inform. Modeling* **2005**, *45*, 177–182. [[CrossRef](#)] [[PubMed](#)]
176. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672. [[CrossRef](#)] [[PubMed](#)]
177. 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](#)]
178. Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **1996**, *6*, 402–406. [[CrossRef](#)]
179. Barreiro, G.; Kim, J.T.; Guimarães, C.R.W.; Bailey, C.M.; Damaoal, R.A.; Wang, L.; Anderson, K.S.; Jorgensen, W.L. From Docking False-Positive to Active Anti-HIV Agent. *J. Med. Chem.* **2007**, *50*, 5324–5329. [[CrossRef](#)]
180. Agarwal, A.K.; Johnson, A.P.; Fishwick, C.W.G. Synthesis of de novo designed small-molecule inhibitors of bacterial RNA polymerase. *Tetrahedron* **2008**, *64*, 10049–10054. [[CrossRef](#)]
181. Stephanie, B.A.d.B.; Nico, P.E.V.; Chris, O. The Role of Water Molecules in Computational Drug Design. *Curr. Topics Med. Chem.* **2010**, *10*, 55–66.
182. Pan, A.C.; Borhani, D.W.; Dror, R.O.; Shaw, D.E. Molecular determinants of drug–receptor binding kinetics. *Drug Dis. Today* **2013**, *18*, 667–673. [[CrossRef](#)]
183. Dror, R.O.; Pan, A.C.; Arlow, D.H.; Borhani, D.W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D.E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13118–13123. [[CrossRef](#)]
184. Zwanzig, R.W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *Chem. Phys.* **1954**, *22*, 1420–1426. [[CrossRef](#)]
185. Straatsma, T.P.; Berendsen, H.J.C.; Postma, J.P.M. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. *Chem. Phys.* **1986**, *85*, 6720–6727. [[CrossRef](#)]
186. Straatsma, T.P.; Berendsen, H.J.C. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *Chem. Phys.* **1988**, *89*, 5876–5886. [[CrossRef](#)]
187. Durrant, J.D.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71. [[CrossRef](#)]
188. Ferguson, D.M.; Radmer, R.J.; Kollman, P.A. Determination of the relative binding free energies of peptide inhibitors to the HIV-1 protease. *J. Med. Chem.* **1991**, *34*, 2654–2659. [[CrossRef](#)]
189. Merz, K.M.R.D.; Reynolds, C.H. *Drug Design: Structure-and Ligand-Based Approaches*; Cambridge University Press: New York, NY, USA, 2010.
190. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phy. Chem. B* **2003**, *107*, 9535–9551. [[CrossRef](#)]
191. Rombouts, F.J.R.; Tresadern, G.; Buijnsters, P.; Langlois, X.; Tovar, F.; Steinbrecher, T.B.; Vanhoof, G.; Somers, M.; Andrés, J.-I.; Trabanco, A.A. Pyrido[4,3-e][1,2,4]triazolo[4,3-a]pyrazines as Selective, Brain Penetrant Phosphodiesterase 2 (PDE2) Inhibitors. *ACS Med. Chem. Lett.* **2015**, *6*, 282–286. [[CrossRef](#)]
192. Ciordia, M.; Pérez-Benito, L.; Delgado, F.; Trabanco, A.A.; Tresadern, G. Application of Free Energy Perturbation for the Design of BACE1 Inhibitors. *J. Chem. Inform. Modeling* **2016**, *56*, 1856–1871. [[CrossRef](#)] [[PubMed](#)]

193. Keränen, H.; Pérez-Benito, L.; Ciordia, M.; Delgado, F.; Steinbrecher, T.B.; Oehlrich, D.; van Vlijmen, H.W.T.; Trabanco, A.A.; Tresadern, G. Acylguanidine Beta Secretase 1 Inhibitors: A Combined Experimental and Free Energy Perturbation Study. *J. Chem. Theory Comput.* **2017**, *13*, 1439–1453. [[CrossRef](#)] [[PubMed](#)]
194. Jorgensen, W.L. Computer-aided discovery of anti-HIV agents. *Bioorg. Med. Chem.* **2016**, *24*, 4768–4778. [[CrossRef](#)] [[PubMed](#)]
195. Jorgensen, W.L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K.S.; Hamilton, A.D. Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 663–667. [[CrossRef](#)] [[PubMed](#)]
196. Ruiz-Caro, J.; Basavapathruni, A.; Kim, J.T.; Bailey, C.M.; Wang, L.; Anderson, K.S.; Hamilton, A.D.; Jorgensen, W.L. Optimization of diarylamines as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 668–671. [[CrossRef](#)] [[PubMed](#)]
197. Kim, J.T.; Hamilton, A.D.; Bailey, C.M.; Domoal, R.A.; Wang, L.; Anderson, K.S.; Jorgensen, W.L. FEP-Guided Selection of Bicyclic Heterocycles in Lead Optimization for Non-Nucleoside Inhibitors of HIV-1 Reverse Transcriptase. *J. Am. Chem. Soc.* **2006**, *128*, 15372–15373. [[CrossRef](#)] [[PubMed](#)]
198. Jorgensen, W.L.; Bollini, M.; Thakur, V.V.; Domoal, R.A.; Spasov, K.A.; Anderson, K.S. Efficient Discovery of Potent Anti-HIV Agents Targeting the Tyr181Cys Variant of HIV Reverse Transcriptase. *J. Am. Chem. Soc.* **2011**, *133*, 15686–15696. [[CrossRef](#)] [[PubMed](#)]
199. Bollini, M.; Domoal, R.A.; Thakur, V.V.; Gallardo-Macias, R.; Spasov, K.A.; Anderson, K.S.; Jorgensen, W.L. Computationally-Guided Optimization of a Docking Hit to Yield Catechol Diethers as Potent Anti-HIV Agents. *J. Med. Chem.* **2011**, *54*, 8582–8591. [[CrossRef](#)]
200. 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](#)]
201. 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](#)]
202. Abel, R.; Mondal, S.; Masse, C.; Greenwood, J.; Harriman, G.; Ashwell, M.A.; Bhat, S.; Wester, R.; Frye, L.; Kapeller, R.; et al. Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. *Curr. Opin. Struct. Biol.* **2017**, *43*, 38–44. [[CrossRef](#)]
203. 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](#)]
204. Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inform. Modeling* **2017**, *57*, 2911–2937. [[CrossRef](#)]
205. Gupta, S.; Jhawar, V. Quality by design (QbD) approach of pharmacogenomics in drug designing and formulation development for optimization of drug delivery systems. *J. Control. Release* **2017**, *245*, 15–26. [[CrossRef](#)]
206. Zhang, L.; Tan, J.; Han, D.; Zhu, H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Disco. Today* **2017**, *22*, 1680–1685. [[CrossRef](#)]
207. Lavecchia, A.; Cerchia, C. In silico methods to address polypharmacology: current status, applications and future perspectives. *Drug Disco. Today* **2016**, *21*, 288–298. [[CrossRef](#)]

