Received: 21 December 2020 | Revised: 15 March 2021 | Accepted: 17 March 2021

DOI: 10.1002/wcms.1535

ADVANCED REVIEW

WIREs COMPUTATIONAL MOLECULAR SCIENCE WILEY

Expanding the boundaries of ligand–target modeling by exascale calculations

Viacheslav Bolnykh[1,2] | **Giulia Rossetti**[2,3,4] | **Ursula Rothlisberger**[1] | **Paolo Carloni**[5,6]

[1]Laboratory of Computational Chemistry and Biochemistry, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[2]Computational Biomedicine, Institute of Neuroscience and Medicine (INM-9)/Institute for Advanced Simulations (IAS-5), Forschungszentrum Jülich, Jülich, Germany

[3]Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, Jülich, Germany

[4]Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation University Hospital Aachen, RWTH Aachen University, Aachen, Germany

[5]Institute for Neuroscience and Medicine and Institute for Advanced Simulations (IAS-5/INM-9) "Computational Biomedicine", Forschungszentrum Jülich, Jülich, Germany

[6]JARA-Institute INM-11 "Molecular Neuroscience and Neuroimaging", Forschungszentrum Jülich, Jülich, Germany

Correspondence
Paolo Carloni, Institute for Neuroscience and Medicine and Institute for Advanced Simulations (IAS-5/INM-9) "Computational Biomedicine", Forschungszentrum Jülich, Jülich, Germany.
Email: p.carloni@fz-juelich.de

Funding information
Bioexcel-2; Helmholtz European Partnering; Human Brain Project SGA3, Grant/Award Number: 945539; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung; Horizon 2020 Framework Programme; European Union; Swiss National Science Foundation; Deutsche Forschungsgemeinschaft

Abstract

Molecular simulations and molecular docking are widely used tools to investigate ligand/target interactions and in drug design. High-performance computing (HPC) is boosting both the accuracy and predictive power of these approaches. With the advent of exascale computing, HPC may become standardly applied in many drug design campaigns and pharmacological applications. This review discusses how innovative HPC algorithms and hardware are being exploited in current simulations and docking codes, pointing also at some of the limitations of these approaches. The focus is on technical aspects which might not be all that familiar to the computational pharmacologist.

This article is categorized under:

Software > Molecular Modeling

Software > Simulation Methods

Structure and Mechanism > Computational Biochemistry and Biophysics

KEYWORDS

drug design, exascale computing, molecular dynamics, ligand-target modeling

1 | INTRODUCTION

Modeling plays a key role in drug discovery and design.[1,2] Developing a new drug molecule can cost up to $2.6 billion: the use of computational approaches may decrease such costs by 30%, as well as the research time by several months.[3]

The computational investigations of target/ligand complexes may help discover drug leads, compounds with at least micromolar affinity for their targets.[4–7] In particular, computational structure-based drug design (CSBDD) has played a significant role in discovering many FDA-approved drugs that reached the consumer market.[8–11]

Computational methods of ligand/target interactions include structure-based virtual screening (SBVS) and molecular dynamics (MD) simulations.[12–14]

SBVS docks large libraries of small molecules into the binding site of a selected therapeutic target.[15] Using ad hoc scoring functions, it predicts ligand binding poses, ranking their potency according to their relative binding affinities. This allows to identify quickly potential drug lead candidates,[16] lowering greatly the number of molecules to be tested experimentally.[17–22] For instance, the SBVS campaign against the M1 Acetylcholine Receptor, a known target against dementia, yielded approximately 1000 putative hits, including 1-(N-substituted piperidin-4-yl) benzimidazolone derivatives. These turned out to be orally administered, central nervous system penetrant, potent agonists of the receptor.[17] Another example is Biogen Idec's SBVS on the transforming growth factor-β1 receptor kinase.[23] The company identified 87 drug candidates, the best hit being identical in structure to the lead compound discovered through traditional high throughput approaches at Eli Lilly.[24] Thus, SBVS, involving diminished costs and workload, turned out to identify the same lead as a full-scale experimental High Throughput Screening (HTS).[25]

The second approach predicts more accurately, but also at a much larger computational cost, structural properties. In addition, and most importantly, it may provide the energetics of target proteins in complex with ligands.[26] Thus, it works very well as a refinement of SBVS. In most cases, it is based on atomistic parametrized empirical potential energy functions or force fields.[27] It can routinely run systems of 10^5 atoms for μs timescales. MD-based ligand binding free energies[13,28–32] allow to predict potency and residence times of drugs.[30,33–35] Examples of the impact of MD in pharmacology include (but are by no means limited to) the development of the FDA-approved HIV-1 lifesaving drugs nelfinavir and raltegravir,[36] the ongoing SARS-CoV-2 research[37–40] and the development of a variety of anti-cancer drugs.[41,42]

The development of supercomputers, combined with the power of parallel algorithms, can greatly boost the investigations of ligand/target complexes and CSBDD.[41] On one hand, this improves the global optimization procedures required in docking algorithms and allows to consider large conformational ensembles as well as the use of more accurate scoring functions.[43] On the other hand, in MD simulations, HPC not only allows to extend the system sizes, time scales, and accuracy[44–46] but it also fosters the use of hybrid approaches such as quantum mechanics/molecular mechanics (QM/MM) simulations.[47] These are particularly useful to investigate bond forming/breaking processes such as in the case of covalent inhibitor binding transition-metal based drugs[48,49] and enzymatic reactions.[49] The latter can be used as a basis to design transition state-analog inhibitors.[50] Furthermore, although generally of much higher computational cost, QM/MM codes may scale better than force field-based MD with the number of processors.[51]

Prompted by the importance of HPC for modern computational biochemistry and pharmacology, we have here compiled a review focusing on technical aspects of high-performance computing (HPC) aspects particularly relevant for MD and molecular docking. Obviously, because we deal mostly with methods rather than applications, the material can be useful also for readers interested in the use of these techniques for applications other than those dealt with here.

The review is organized as follows. Section 2 provides some basic concepts of HPC. Their implementation in MD and docking, along with specific examples of force-field-based and QM/MM codes, is offered in Section 3 together with a description of their actual implementation in specific HPC codes. Some of the major limitations of docking and molecular simulation approaches are summarized. Section 4 draws some conclusions and provides an outlook.

2 | HPC: THE BASICS

We survey general HPC techniques that may be applied to maximize the performance of a variety of applications, including (but not limited to) MD and molecular docking. Let us start with some key definitions.

The computational complexity of an algorithm or a problem is measured in FLOPs, which stands for Floating-Point Operations. The performance of a hardware or an algorithm is measured by the number of FLOPs executed per second. This can be obtained by dividing the total amount of FLOPs in the task by the time needed to perform the calculation:

$$\mathrm{FLOPS} = \frac{\mathrm{FLOPs}}{t(sec)}$$

Many biomolecular simulations may routinely reach the order of zetta (10^{21}) FLOPs.[52] If run on a single PC (whose computational power is few tera [10^{12}] FLOPS), these would take decades to solve (Serial Performance, Section 2.1). Therefore, optimizing and distributing the computational effort so that it can be performed simultaneously is crucial. This also applies to molecular docking because of the immensely large amount of library compounds and possible docking poses that need to be scored. HPC (Section 2.2) is a key technique in this respect.

2.1 | Single-node computations

2.1.1 | Serial performance

Modern Central Processing Units (CPUs) provide a wide range of tools designed to maximize the performance of an algorithm, executed by a single CPU even without any parallelization (serial performance). Particularly important here is the "vectorization": vector units (SIMD, Single Instruction, Multiple Data).[53] These allow to perform the same operation on multiple data elements instead of just one. To illustrate this, let us consider an example of the addition of two vectors each containing N elements (Figure 1(a)). Completing such an addition requires N operations. These take a fraction of ns. With $N = 10{,}000{,}000$, it may take 2 s to perform such an addition—an unfeasible time as this operation should be performed thousands or millions of times. In contrast, vectorization allows us to perform up to 32 additions (depending on the CPU and the precision) with a single instruction (Figure 1(b)): rendering the calculation much more feasible.

2.1.2 | Multi-threading

Modern CPUs contain several (from 2 to few dozen) independent processing units (or cores). The latter may execute simultaneously 2–4 threads (smallest independent sets of program instructions). The parallel execution of the code is then enabled by multi-threading across all the cores. In multi-CPU systems, many threads are further distributed between all the CPUs available in the computer.

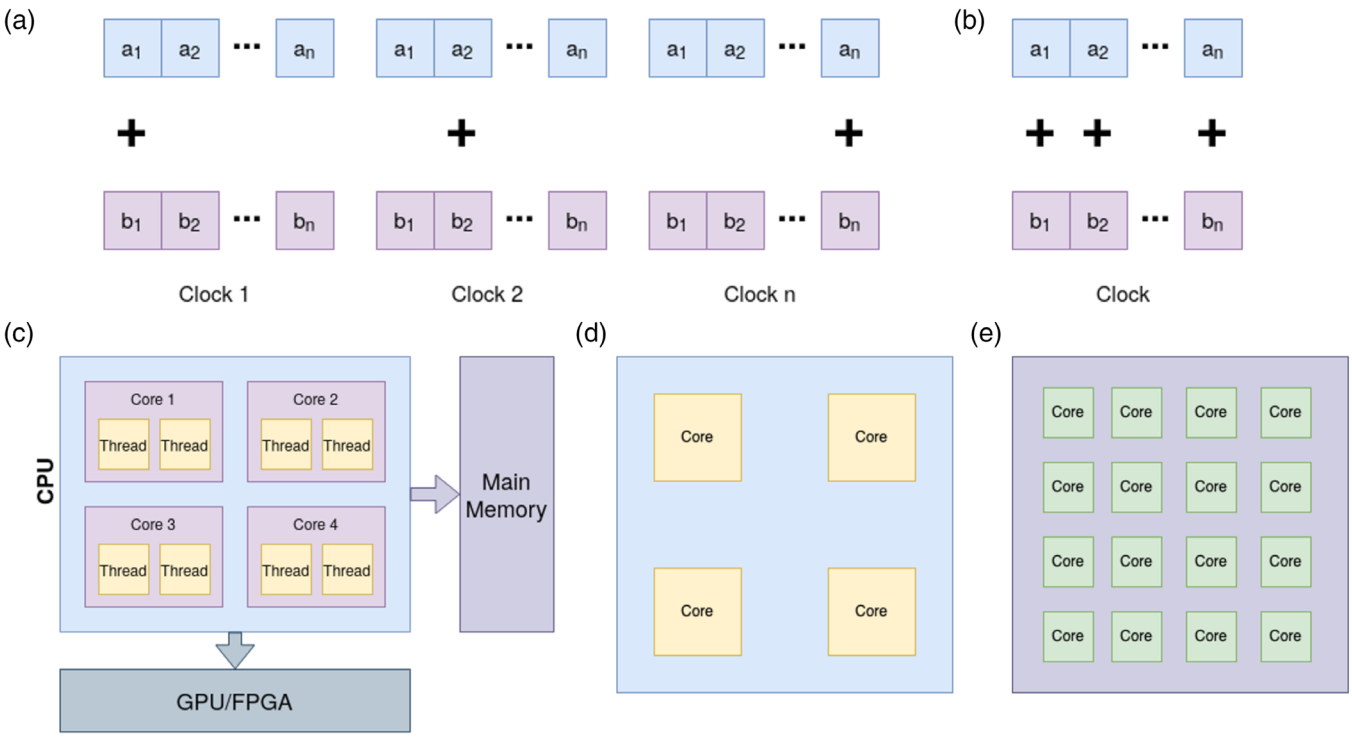

FIGURE 1 Addition with scalar operations (a) and with SIMD (b). Schematic of a computer system with multi-core CPU and with an accelerator (c). Schematic of a CPU (d) and a GPU (e) architecture

A few hundreds threads operating on a single CPU and a full utilization of vector units may speed-up calculations by up to three orders of magnitude.

2.1.3 | Accelerated computing

Graphical Processing Units or GPUs feature many very weak cores (Figure 1(c–e)), unlike CPUs, characterized by a small number of very powerful cores. This makes GPUs very useful tools for highly parallelizable applications, such as molecular simulations and docking. An alternative approach to accelerate calculations is through the use of Field Programmable Gate Arrays (*FPGAs*) that implement a specific algorithm on a hardware level. This avoids loading the program from the memory and it allows for optimal hardware implementation, greatly speeding up calculations relative to CPUs. Modern FPGAs can be reprogrammed on-the-fly, allowing them to be used as co-processors.[54]

2.2 | Parallel computing

The main idea here is to split the workload between multiple computational units in the form of independent computers connected via the Internet (Figure 2(a)) or parts of a large installation (compute cluster) joined by a local high-speed network (Figure 2(b)). The first approach is the foundation of a so-called distributed computing network, whereas the second one results in cluster or supercomputing.

Distributed systems represent a range of possible options: grid, cloud, and volunteer computing. They typically consist of a large number of (possibly) physically separated client machines and a coordinator node. The latter is responsible for distributing the computational workload between the clients. In a computer grid, a set of possibly heterogeneous clients may be designed to perform different types of tasks (depending on the way the coordinator is distributing the workload). In the case of a computational cloud, several clients (situated at the same or different locations) are connected to form a virtual "computer" with significantly more resources. Unlike the grid computers, clouds are typically more homogeneous in terms of hardware and tend to be more tightly coupled. Finally, volunteer computing projects like *folding@home* (https://foldingathome.org/) unite a large number of commodity PCs voluntarily connected by their users to a computational network. These represent a highly heterogeneous system with extreme possible hardware variability but they offer the benefit of low cost combined with high throughput and, in some cases, may exhibit collective performance higher than the most powerful supercomputers.

Clusters instead, connect a large number of relatively powerful computers (called nodes) with a high-speed and low-latency network fabric to create a localized integrated system. Such systems offer a huge amount of computational resources with very fast communication between the nodes. As such, they lie at the foundation of HPC.

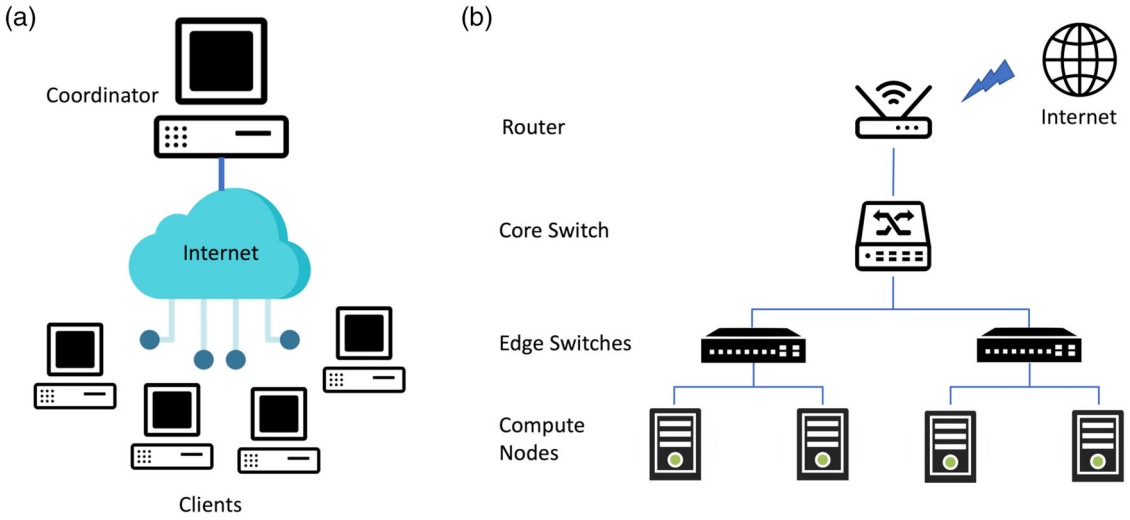

FIGURE 2 Schematic representation of a distributed network system (a) and a compute cluster (b)

Problems with little to no dependency between computational tasks (embarrassingly parallel problems) can be trivially distributed across the nodes. They exhibit a linearly scaling behavior. In most cases, however, the nodes need to exchange data. They do so indirectly, through a message passing mechanism, primarily the Message Passing Interface (MPI[55]). The process takes a few μs or more,[56] three or more orders of magnitude longer than the CPU clock cycle. Good scaling with the number of nodes is then achieved by hiding data exchange behind the computation ("overlapping" of computation with communication) or by minimizing the frequency of data exchange.

3 | HPC: APPLICATIONS

This section introduces very succinctly MD and SBVS, concentrating on their HPC aspects. For in-depth discussion of MD and docking methods, the reader is referred to a variety of excellent textbooks and reviews.[27,28,57–60]

3.1 | Molecular dynamics

3.1.1 | Introduction: Predictive power of MD and its limitations

All atom MD has revealed itself as a very valuable tool to study structure, dynamics, spectroscopic properties, and energetics of biological systems, including those of pharmaceutical relevance. The biomolecules are investigated here in explicit solvent. MD solves numerically Newton's second law, assuming that one has a suitable interatomic potential. In most applications, an effective potential (or *force field*) is used, leading to great computational efficiency. Nowadays, force field-based MD can routinely cover microsecond time scales for systems containing a few dozens or even hundreds of thousands of atoms. However, a variety of pharmacologically relevant processes, from enzymatic reactions,[61–66] to photochemistry-based processes[67–69] and metal-based drug/target interactions[70–73] can often not be described with standard potentials. For some of the aforementioned cases, one can conveniently use modified force fields.[74–76] These however might not be transferable to other systems. A more general description can be provided by an explicit quantum mechanical (QM) treatment of the electronic degrees of freedom. The way one can simulate biological systems at the QM level may vary widely. To treat entire biological systems (often in the range of several hundreds of thousands of atoms), one can resort to (semi-)empirical methods (e.g., INDO[77] or EVB[78]). In this way, one can reach relatively long-time scales, with remarkable accuracy. These methods require system-dependent parametrization and they may not be applied in all cases (for instance, most semiempirical methods encounter difficulties in describing transition metal ions[79,80]). Approaches based on first-principles QM have a broader scope and can be applied to virtually any system, although they come at a much higher computational cost. For this reason, they are rarely able to describe the whole biological system (e.g., a protein in water). The method of choice is then Quantum Mechanical/Molecular Mechanical (QM/MM)-based MD. Its relevance for complex systems has been recognized by the Nobel prize in Chemistry in 2013.[47] Here, one embeds the region treated at the QM level (e.g., an enzymatic active site) in the biomolecular frame and the solvent, treated with effective potentials. Most often, density functional theory (DFT), a method roughly as expensive as Hartree–Fock but including electron correlation,[27] is used to treat the QM part.

The predictive power of both force field- and QM/MM-based MD simulations is affected by two major issues. First, the configuration space sampling should be sufficient enough so that the results approximate an equilibrium distribution. Only then, the machinery of statistical mechanics can be applied, properties such as binding affinities can be calculated and proper comparison with in vitro experiments, which measure equilibrium properties, can be made. This issue is even more stringent in the case of QM/MM-based MD. Specialized hardware[81] or advanced simulations methods (such as enhanced sampling[60]) can alleviate the sampling problem. Although such advanced methods constitute excellent tools to study target/protein interactions, one may require some training before successfully applying them in the drug design field.

The second key issue is the quality of the interatomic potential. Standard biomolecular force-fields, widely used in CSBDD, are parametrized to reproduce a subset of experimental data, necessarily limiting their domain of applicability.[34] They neglect polarization effects and charge transfer, which may play a role in ligand/target interactions.[34] In QM/MM simulations, DFT is a computationally efficient method to treat fairly large systems with a good accuracy but that it can be difficult to attain chemical accuracy especially for transition states and systems with pronounced multireference character. One should keep these limitations in mind when using MD to predict ligand poses and affinities.

3.1.2 | HPC approaches

HPC parallelizes the computation of forces—the most expensive calculation in force-field, and, even more, QM/MM simulations. The numerical integration of the equations of motion can be done in a distributed way, too. The most common way of distributing computations in MD is the so-called *Domain Decomposition* (DD). It splits the simulation box into a set of smaller domains. Each of these domains is then assigned to a separate computational unit (node/task/thread as described in the previous section). In this way, the computational effort needed to perform a single time step is divided among all of the units involved in a simulation. The benefit of DD is the reduced memory consumption per computational unit (as they have to store only part of the whole system) and relatively low communication overhead. The DD type is defined by the shape of the subdomain. Typically, in MD simulations three types of DD can be found: (a) "Slab" DD—where the system is split along one dimension (Figure 3(a)); (b) "Pencil" DD—the system is split along two dimensions (Figure 3(b)); or (c) "Volumetric" DD—the system is split into a set of 3D domains along all the dimensions (Figure 3(c)). The optimal type of DD is dependent on the type of problem one is trying to solve. Typically, DDs of higher dimensions result in a better scaling performance.

3.1.3 | Examples

We describe here two codes, which differ in the interatomic potential used, from force fields (parametrized empirical potential energy functions)[27] to QM/MM.[47] Some examples of HPC-based MD codes may be found in Table 1. Here we focus on the widely used GROMACS code[82] as a representative case. Similar techniques may be found in other MD software packages.

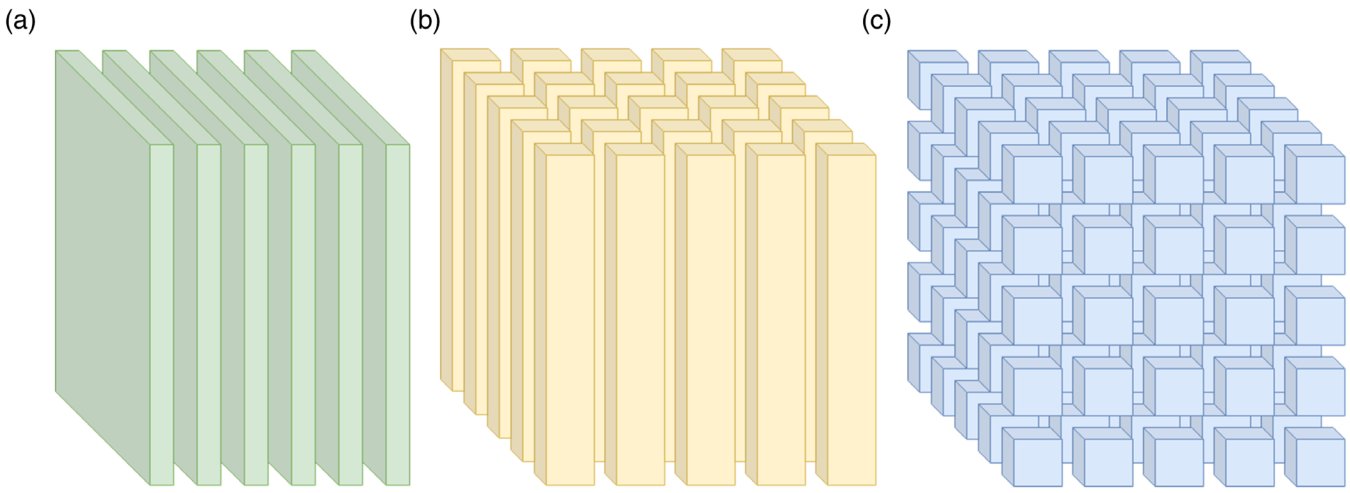

FIGURE 3 Representation of 1D (a), 2D (b), and 3D (c) DD approaches

TABLE 1 Major force-field-based and DFT-based MD programs. The name, the level of theory, the type of parallelization, and a link to the code are provided

Name	Level of theory	Parallelization	Link
GROMACS	MM	OpenMP/MPI/GPU	https://www.gromacs.org/
AMBER	MM	OpenMP/MPI/GPU	https://ambermd.org/
LAMMPS	MM	OpenMP/MPI/GPU	https://lammps.sandia.gov/
Orca	QM	MPI	https://orcaforum.kofo.mpg.de/app.php/portal
CPMD	QM	OpenMP/MPI/GPU(experimental)	https://www.cpmd.org/wordpress/
CP2K	QM	OpenMP/MPI/GPU	https://www.cp2k.org/
TeraChem	QM	GPU	http://www.petachem.com/products.html

GROMACS: *A force-field-based MD code*. This is a force-field-based MD code with high level of performance, free license, and an active developer community. The heterogeneous (combining, threads, MPI, and GPUs, see Sections 2.1.2., 2.1.3 and 2.2) parallelization strategy of GROMACS enables it to reach the limit of approximately 1 ms per time step for systems containing 100,000–1,000,000 atoms.[83] This strategy involves 3D DD of the simulation box with clusterization of atoms to improve the vectorization support. Clusterization is done in a way that it ensures the vectorization of computational kernels. Vectorization (Section 2.1.1) is done explicitly using architecture-specific intrinsic functions and currently, 14 SIMD architectures are supported. In addition, the advanced GPU offloading is used to boost the performance. Most of the force terms are computed on the GPU, while the CPU is busy building the pair lists.[83]

One of the most important applications of MD for Computer-Aided Drug Design (CADD) are free energy calculations. Many of them exploit HPC resources. These include temperature-induced methods such as simulated[84] and parallel tempering[85] to methods biasing potentials like metadynamics,[86] umbrella sampling,[87] or adaptively biased MD.[88] These methods often may benefit from the use of multiple replicas of the system to enhance sampling. This represents an additional level of parallelism (Section 2.2). Machine learning (ML) can further boost HPC-based free-energy calculations either by refining existing approaches[89] or by introducing new ones.[90–93]

A hybrid QM/MM interface. The Multiscale Modeling in Chemistry (or MiMiC)[51,94] interface joins GROMACS with the CPMD code.[95] The latter implements DFT-based MD with high levels of parallel performance using plane-wave (PW) basis sets. Several DFT terms are calculated in reciprocal space, where the algebraic expressions of some of the functionals are relatively simple and allow one to achieve high scalability of computations for HPC applications. The transition between real and reciprocal space is done via Fast Fourier Transforms (FFTs).[96] The electronic density is mapped onto a regular grid both in real and reciprocal space and this grid is distributed using slab DD. Although, the latter introduces quite a tight scaling limit (equal to the number of grid planes along the DD dimension, further scaling is achieved by distributing Kohn–Sham orbitals across task groups. Each group computes partial density (and potential) on the same grid and then it is summed across all the groups. Computations are distributed using a hybrid scheme combining the MPI parallelization (Section 2.2) with the threading approach (Section 2.1.2). The scaling limit is highly dependent on the system size as well as the type of exchange-correlation functional used. Typically, the main bottleneck preventing further scaling is the global communication in parallel 3D FFT. For small system sizes (10–100 QM atoms) the scaling limit is around 1000–10,000 cores, whereas, for larger systems, it can be in the area of millions.[97] Recently, CPMD was also coupled with ML in order to produce QM-quality MD trajectories at a significantly lower computational cost.[98] The use of Kernel Ridge Regression (KRR) allows one to achieve an extremely high sampling rate in comparison to pure DFT.

The parallelization of computations in MiMiC is inspired by the distributed computation scheme implemented in CPMD. It uses slab DD of the grid across MPI tasks (Section 2.2) with a subsequent splitting on slabs across threads (Section 2.1.2) assigned to each task. On top of that, task grouping is used (as in CPMD, see above) to divide MM atoms into subsets to further increase the scaling. With the use of this technique, we were able to use over 10,000 cores in a single simulation.[51] As a result, MiMiC may reach sub-ns time scale with current HPC architectures.[99] This may pave the way to highly efficient HPC-based hybrid simulations of pharmaceutically relevant enzymatic reactions, such as those performed on SARS-COV-2 proteins.[37]

3.2 | Structure-based virtual screening

3.2.1 | Introduction: Docking in SBVS

Molecular docking algorithms[100] predict binding-conformations of two interacting molecules. They use a searching algorithm that explores possible positions, orientations, and conformations of the potential drugs and target proteins. Then, they rank the ligand poses using either a physics-based, or an empirical, or a knowledge-based or a machine learning-based scoring function.[101,102] The latter is related to ligand/target binding strength.[103] Several dozens of different docking tools and programs for both academic and commercial use are currently available[104,105] (Table 2).

SBVS is based on state-of-the-art, flexible molecular docking algorithms.[100] Flexible docking algorithms consider both ligand and protein as flexible counterparts. They can predict more reliably binding-conformations of two interacting molecules than standard rigid docking approaches, at a higher computational cost.[104,117] In these approaches, the conformational degrees of freedom of the ligand are always included. To include receptor flexibility, which plays an

TABLE 2 Some of the many successful applications of docking codes

Code	l applications	Platform
FlexX	• Plasmepsin II and IV inhibitors[106] • Malaria Anthrax edema factor[107] • Pneumococcal peptidoglycan deacetylase inhibitors[108]	Linux Unit SunSolaris HP-UX Windows
Glide	• Aurora kinases inhibitors[109] • Falcipain inhibitors[110] • Cytochrome P450 inhibitors[111]	Linux Unit IBM AIX
Dock	• FK506 immunophilin[112] • BCL6, oncogene in B-cell lymphomas[113]	Linux Unit Windows Max OS X
Gold	• Peroxisome Proliferator-Activated Receptor γ Agonists.[114]	Linux Unit Windows SunSolaris IBM AIX
Fred	• inhibitors of C2 domain of factor V[115]	Linux Unit OS X Windows Solaris AIX HP-UX Tru64UNIX
AutoDock	• inhibitors of tyrosine phosphatase[116]	Linux Unit Windows Max OS X

important role in drug discovery,[118] extensive sampling of the protein degrees of freedom would be required. In principle, predicting binding poses accurately would require extensive sampling of the protein degrees of freedom. However, this is not yet feasible for SBVS campaigns[119] and one has to resort to approximations. The first ones were the "Soft Docking"[120] and the "SideChain Flexibility",[121] which used a single protein conformation as an input. In the first, the Van der Waals repulsion term (described by a Lennard–Jones [LJ] potential energy function in the force-field based scoring functions) is reduced allowing for closer ligand–protein interactions. Following this idea, for instance, the scaling factor of the LJ potential is treated as a variable input parameter in codes such as Dock,[122] Glide,[123] and AutoDock.[124] Soft Docking can simulate induced-fit, albeit, in a very approximate manner[119,125] and its main drawback is that it could implicate unreal poses.[125] The SideChain Flexibility approach instead, introduces alternative conformations for protein side chains in the binding site,[121] under the assumption that ligand binding induces only side-chain motions,[126] in most of the cases. This is achieved usually by discrete sampling, that is, by exploiting databases of rotamer libraries (i.e., Autodock Vina[127,128]), or by sampling some degrees of freedom within the search engine of the software (like in GOLD[129]), or by a posteriori optimizing the side chains in the presence of the rigid ligand (i.e., Glide[123]). Discrete sampling limits side-chain motion to a small set of energetically accessible conformations reducing computational time. Obviously, backbone flexibility and huge conformational variations of the protein are neglected by these approaches.

Multiple protein conformations docking protocols instead represent a protein by an ensemble of conformations of similar energy. These conformations can be available from experiments or generated via computational techniques, such as Monte Carlo or Molecular Dynamics simulations. The idea is to try to take into account all these diverse conformations either by sequentially docking the ligand into each receptor structure[119] or by building up a single averaged grid or by constructing the best performing "chimera" grid.[119,125]

3.2.2 | HPC approaches

The accuracy and the performance of SBVS in large-scale drug design campaigns may profit greatly from HPC approaches. Indeed, the latter may increase (a) global optimization procedures, (b) the number of the generated molecular conformations, and (c) the mathematical complexity of the scoring function.[130] Moreover, HPC approaches can distribute efficiently the computational cost of the scoring functions, that is, a highly time-consuming step. This has in turn allowed to (c) improve the complexity (and the accuracy) of the scoring functions, without significantly compromising the speed of drug screening.[130]

Therefore, in the next sections, we will describe how several research groups have developed parallel versions of molecular docking software based on common HPC platforms: grid, multiprocessor, cluster/supercomputer, accelerator/heterogeneous system machine grid, and cloud.

3.2.3 | Examples

Multiprocessor computing (Section 2.1.2) with the implementation of multiple data stream (MIMD) systems,[131] are very suitable for docking software because of the inherent independence of the individual processes: Many time-consuming processes in several docking softwares are indeed already composed of independent parts, like for instance ligands or receptors preparations in AutoDock,[132] that can be run in parallel. Moreover, one can also exploit different parallelization paradigms for facilitating software development in a parallel environment, like distributed (e.g., MPI) and shared-memory approaches (e.g. OpenMPI; Section 2.2). For instance, a mixed parallel scheme that combines MPI and multithreading (Section 2.1.2) was implemented in the widely used AutoDock Vina (Vina) code.[128] This exploits the parallelism of shared memory hardware, such as multi-core CPU or multi-CPU workstations,[133] allowing to reach hyper linearity speedup on a multi-processor machine (or to be Embarrassingly Parallel [EP]).[134] Vina's amazing performance can be achieved thanks to the high independencies in each process.

The Glide,[135] LigandFit,[136] and FlexX[103] codes use Grid computing[137] (Section 2.2) to distribute docking jobs to multiple computers over a network. This shortens the time to solution of large-scale molecular docking that allows for introducing receptor flexibility[6,119] and improves the accuracy of the scoring function.

Cluster computing (Section 2.2) may be particularly efficient in high throughput and HPC-based docking, as usually, the number of ligands to be screened is much larger than the number of CPUs available on a single machine. Therefore, the ligands are divided into several packets equal to the number of CPUs. Each CPU docks different ligands on the receptor. After completing docking, the results are collected, and then a predefined percentage of the best ligands are selected for further analysis. However, several docking software packages like for instance, the MPI version of the DOCK code,[138] face the problem of overloading the main node: the allocation efficiency of the main node decreases with increasing the number of nodes. This issue may be overcome by the effective distributed virtual screening data management system (DVSDMS) introduced by Caflisch et al.,[139] where the data processing and work distribution are realized by an open-source structured query language database software MySQL,[140] and that can be freely connected with different molecular docking software.

Supercomputing clusters (Section 2.2) require some complexity in the programming paradigms employed. For instance, the Dock 6 code[141] required few adjustments to run on large-scale parallel computers such as the IBM System Blue Gene® and Blue Gene/L,[142]: The performance on every single node was optimized. Next, also the parallelization was optimized: the nodes were divided into the main processor, responsible for scheduling tasks, and the worker processors executing their respective molecular docking tasks (master-working scheme). Due to the overload of the host, this scheme cannot scale well with a large number of nodes.[143] Thus, work units are sent asynchronously to individual computing nodes on the large-scale parallel computers to be executed. However, sometimes the workload of each worker processor turns out not to be equally distributed.[144] To face this challenge, the VinaMPI code uses a distribution scheme in which tasks are evenly distributed to the workers based on the complexity of each task.[145]

Parallel optimization of molecular docking also exploits accelerators like GPUs (Section 2.2). The calculation of the non-bonded (such as electrostatics or van der Waals forces) interactions usually represents up to 80% of the total execution time. Using GPU, these calculations can speed by a factor of 100.[146–148] They are implemented in codes such as FlexScreen.[149]

The cloud arcitecture is more scalable, flexible, and cost-effective than several other HPC approaches.[150] A cloud platform[151] is available for the widely used molecular docking program, AutoDock as AutoDockcloud.[152] Also, wFReDoW, a Cloud-Based Web environment managed to handle the important challenge of virtual screening of millions of ligands using molecular docking simulations of a fully flexible receptor model.[153]

3.3 | HPC-based MD and docking: Two complementary approaches

Even with state-of-the-art scoring functions-based predictions and the strategies to introduce target flexibility, docking approaches present further limitations, as they do not account well for conformational entropy or solvation energy

contributions. This is detrimental for binding affinity prediction.[154–156] Integration of parallel docking programs with postprocessing approaches may partially alleviate these limits.

Top-ranking docking poses may be re-scored by applying Molecular Mechanics/Generalized Born Surface Area (MM/GBSA methods)[27] or Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA).[27] These approaches are cheaper and more approximate than all-atom MD in explicit solvent. In MM/GBSA or MM/PBSA, one performs MD simulations on each ligand–receptor complex and then uses the GB or the PB to calculate the electrostatic contribution and a surface area model to estimate the nonpolar contributions.[157] Entropic contributions can be roughly estimated by using a harmonic model in MM/PB. In any case, both MM/GBSA and MM/PBSA contain some drastic approximations that limit the accuracy of conformational entropies and hydration free energies. These render the value of such post-docking rescoring, in some cases, inconclusive,[158] although these limitations can be overcome by all-atom MD and enhanced sampling approaches (Section 3.1),[154,159] far more accurate (and expensive). In addition, by improving dramatically the estimation of binding affinities,[125] the number of false positives and false negatives can be largely lowered in a virtual screening campaign.[109] Further improvements are possible using QM-based methods.[160]

An interesting experiment was performed to quantify the advantage of using HPC in a MD-based virtual screening (MDVS).[161] The authors showed that without using HPC, MDVS for a 10 K compound library with tens of nanoseconds of force-field-based MD simulations requires years of computer time. In contrast, a state of the art HPC machine can be 600 times faster than an eight-core PC server is in screening a typical drug target (which can contain from 10 K up to 70 K atoms without solvent) and that also careful design of the GPU/CPU architecture can decrease the HPC costs.[161]

4 | CONCLUSION

CSBDD complements in a powerful synergy experimental drug discovery research. Here, we have highlighted some of the current advances in HPC to accelerate docking and MD algorithms. Code developers have ported the existing software to different HPC platforms and even designed novel parallel algorithms able to fully exploit the computing power of parallel computing, and, eventually to face the exascale challenge[162]—that is, the possibility of running 10^{18} FLOPS. Exascale machines could literally revolutionize CSBDD. To reach this goal, at least three major challenges/caveats need to be met: (a) usage of millions of cores requires a complete re-working of existing algorithms; (b) heterogeneous accelerated architectures require careful software design; (c) the codes should continue to work in case of hardware failure that is increasingly more likely for large parallel computers.[163]

In addition to HPC, CSBDD is currently greatly boosted by the explosion of Artificial Intelligence (AI) and in particular Machine Learning (ML) and Deep Learning (DL) methods. In this framework, in October 2018, the Defense Advanced Research Projects Agency (DARPA) announced the launch of the accelerated molecular discovery (AMD) program, which aims to develop new AI-based systematic approaches to accelerate the discovery and optimization of high-quality molecules, including drug molecules. Moreover, internationally renowned pharmaceutical companies such as Merck, Sanofi, Genentech, and Takeda, have launched relevant cooperation efforts with AI companies.[164] Finally, AlphaFold developed by the DeepMind team has shown extremely impressive results in protein structure prediction.[165] Given these premises, we expect that the combination of AI with HPC, that is, currently just at the very beginning, will revolutionize drug design and drug discovery in the next future.

ACKNOWLEDGMENTS

We gratefully acknowledge the funding by (a) the Deutsche Forschungsgemeinschaft DFG via the Research Unit FOR2518 “Functional Dynamics of Ion Channels and Transporters – DynIon”, project P6 (PC); (b) the Swiss National Science Foundation via individual grant No. 200020-185092 and the NCCR MUST and the Swiss National Supercomputing Centre CSCS for computing time (UR); (c) the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3, G.R., and P.C.). Giulia Rossetti and Paolo Carloni also acknowledge the Helmholtz European Partnering grant: “Innovative High Performance Computing Approaches for Molecular Neuromedicine”. Paolo Carloni acknowledges the funding by the Deutsche Forschungsgemeinschaft via FOR 2518 DynIon 2 project and that of BioExcel 2. Viacheslav Bolnykh and Ursula Rothlisberger thank the Swiss National Science Foundation for the support via individual grants and the NCCR MUST. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Viacheslav Bolnykh: Conceptualization; investigation; resources; software; validation; visualization; writing-original draft; writing-review & editing. **Giulia Rossetti:** Conceptualization; data curation; funding acquisition; resources; software; validation; visualization; writing-original draft; writing-review & editing. **Ursula Roethlisberger:** Conceptualization; data curation; investigation; methodology; project administration; writing-original draft; writing-review & editing. **Paolo Carloni:** Conceptualization; funding acquisition; methodology; project administration; resources; software; visualization; writing-original draft; writing-review & editing.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study

ORCID

Viacheslav Bolnykh https://orcid.org/0000-0002-7090-1993
Giulia Rossetti https://orcid.org/0000-0002-2032-4630
Ursula Rothlisberger https://orcid.org/0000-0002-1704-8591
Paolo Carloni https://orcid.org/0000-0002-9010-0149

RELATED WIRES ARTICLES

Recent advances in dynamic docking for drug discovery
Computational structure-based drug design: Predicting target flexibility
Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity
From machine learning to deep learning: Advances in scoring functions for protein-ligand docking
Receptor flexibility in small-molecule docking calculations

REFERENCES

1. Brogi. Computational approaches for drug discovery. *Molecules*. 2019;*24*(17):3061. https://doi.org/10.3390/molecules24173061.
2. Schneider G, Clark DE. Automated de novo drug design: are we nearly there yet? *Angew Chem Int Ed*. 2019;*58*(32):10792–803. https://doi.org/10.1002/anie.201814681.
3. Kiriiri GK, Njogu PM, Mwangi AN. Exploring different approaches to improve the success of drug discovery and development projects: a review. *Future J Pharm Sci*. 2020;*6*(1):1–12. https://doi.org/10.1186/s43094-020-00047-9.
4. Albanese SK, Chodera JD, Volkamer A, Keng S, Abel R, Wang L. Is structure-based drug design ready for selectivity optimization? *J Chem Inf Model*. 2020;*60*(12):6211–27. https://doi.org/10.1021/acs.jcim.0c00815.
5. Batool M, Ahmad B, Choi S. A structure-based drug discovery paradigm. *Int J Mol Sci*. 2019;*20*(11):1–18. https://doi.org/10.3390/ijms20112783.
6. Iglesias J, Saen-oon S, Soliva R, Guallar V. Computational structure-based drug design: predicting target flexibility. *WIREs Comput Mol Sci*. 2018;*8*(5):e1367. https://doi.org/10.1002/wcms.1367.
7. Verlinde CL, Hol WG. Structure-based drug design: progress, results and challenges. *Structure*. 1994;*2*(7):577–87. https://doi.org/10.1016/S0969-2126(00)00060-5.
8. Clark DE. What has computer-aided molecular design ever done for drug discovery? *Expert Opin Drug Discov*. 2006;*1*(2):103–10. https://doi.org/10.1517/17460441.1.2.103.
9. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*. 2004;*3*(11):935–49. https://doi.org/10.1038/nrd1549.
10. Talele T, Khedkar S, Rigby A. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr Top Med Chem*. 2010;*10*(1):127–41. https://doi.org/10.2174/156802610790232251.
11. Wang X, Song K, Li L, Chen L. Structure-based drug design strategies and challenges. *Curr Top Med Chem*. 2018;*18*(12):998–1006. https://doi.org/10.2174/1568026618666180813152921.
12. Huggins DJ, Biggin PC, Dämgen MA, Essex JW, Harris SA, Henchman RH, et al. Biomolecular simulations: from dynamics and mechanisms to computational assays of biological activity. *WIREs Comput Mol Sci*. 2019;*9*(3):e1393. https://doi.org/10.1002/wcms.1393.
13. Jagger BR, Kochanek SE, Haldar S, Amaro RE, Mulholland AJ. Multiscale simulation approaches to modeling drug–protein binding. *Curr Opin Struct Biol*. 2020;*61*:213–21. https://doi.org/10.1016/j.sbi.2020.01.014.
14. Moroni E, Paladino A, Colombo G. The dynamics of drug discovery. *Curr Top Med Chem*. 2015;*15*(20):2043–55. https://doi.org/10.2174/1568026615666150519102950.

15. Cerqueira NMFSA, Gesto D, Oliveira EF, Santos-Martins D, Brás NF, Sousa SF, et al. Receptor-based virtual screening protocol for drug discovery. *Arch Biochem Biophys*. 2015;*582*:56–67. https://doi.org/10.1016/j.abb.2015.05.011.
16. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc*. 2016;*11*(5):905–19. https://doi.org/10.1038/nprot.2016.051.
17. Budzik B, Garzya V, Shi D, Walker G, Woolley-Roberts M, Pardoe J, et al. Novel N-substituted benzimidazolones as potent, selective, CNS-penetrant, and orally active M1 mAChR agonists. *ACS Med Chem Lett*. 2010;*1*(6):244–8. https://doi.org/10.1021/ml100105x.
18. Li N, Wang F, Niu S, Cao J, Wu K, Li Y, et al. Discovery of novel inhibitors of *Streptococcus pneumoniae* based on the virtual screening with the homology-modeled structure of histidine kinase (VicK). *BMC Microbiol*. 2009;*9*(1):129. https://doi.org/10.1186/1471-2180-9-129.
19. Lu IL, Huang CF, Peng YH, Lin YT, Hsieh HP, Chen CT, et al. Structure-based drug design of a novel family of PPARγ partial agonists: virtual screening, X-ray crystallography, and in vitro/in vivo biological activities. *J Med Chem*. 2006;*49*(9):2703–12. https://doi.org/10.1021/jm051129s.
20. Ruiz FM, Gil-Redondo R, Morreale A, Ortiz ÁR, Fábrega C, Bravo J. Structure-based discovery of novel non-nucleosidic DNA alkyltransferase inhibitors: virtual screening and in vitro and in vivo activities. *J Chem Inf Model*. 2008;*48*(4):844–54. https://doi.org/10.1021/ci700447r.
21. Simmons KJ, Chopra I, Fishwick CWG. Structure-based discovery of antibacterial drugs. *Nat Rev Microbiol*. 2010;*8*(7):501–10. https://doi.org/10.1038/nrmicro2349.
22. Triballeau N, Van Name E, Laslier G, Cai D, Paillard G, Sorensen PW, et al. High-potency olfactory receptor agonists discovered by virtual high-throughput screening: molecular probes for receptor structure and olfactory function. *Neuron*. 2008;*60*(5):767–74. https://doi.org/10.1016/j.neuron.2008.11.014.
23. Singh J, Chuaqui CE, Boriack-Sjodin PA, Lee WC, Pontz T, Corbley MJ, et al. Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFβ receptor kinase (TβRI). *Bioorg Med Chem Lett*. 2003;*13*(24):4355–9. https://doi.org/10.1016/j.bmcl.2003.09.028.
24. Shekhar C. In Silico pharmacology: computer-aided methods could transform drug development. *Chem Biol*. 2008;*15*(5):413–4. https://doi.org/10.1016/j.chembiol.2008.05.001.
25. Sawyer JS, Anderson BD, Beight DW, Campbell RM, Jones ML, Herron DK, et al. Synthesis and activity of new aryl- and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor-β type I receptor kinase domain. *J Med Chem*. 2003;*46*(19):3953–6. https://doi.org/10.1021/jm0205705.
26. Ayton GS, Noid WG, Voth GA. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr Opin Struct Biol*. 2007;*17*(2):192–8. https://doi.org/10.1016/j.sbi.2007.03.004.
27. Leach A. *Molecular Modelling: Principles and Applications*. Edinburgh: Longman; 1996.
28. Hospital A, Goñi JR, Orozco M, Gelpí JL. Molecular dynamics simulations: advances and applications. *Adv Appl Bioinf Chem*. 2015;*8*(1):37–47. https://doi.org/10.2147/AABC.S70333.
29. Kuzmanic A, Sutto L, Saladino G, Nebreda AR, Gervasio FL, Orozco M. Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. *Elife*. 2017;*6*:1–20. https://doi.org/10.7554/eLife.22175.
30. Shirts MR, Mobley DL, Brown SP. Free-energy calculations in structure-based drug design. *Drug Design*. Vol 1. Cambridge: Cambridge University Press; 2010. pp. 61–86. https://doi.org/10.1017/cbo9780511730412.007.
31. Song LF, Lee TS, Zhu C, York DM, Merz KM. Using AMBER18 for relative free energy calculations. *J Chem Inf Model*. 2019;*59*(7):3128–35. https://doi.org/10.1021/acs.jcim.9b00105.
32. Tsai CJ, Nussinov R. A unified view of "how Allostery works.". *PLoS Comput Biol*. 2014;*10*(2):e1003394. https://doi.org/10.1371/journal.pcbi.1003394.
33. Bruce NJ, Ganotra GK, Kokh DB, Sadiq SK, Wade RC. New approaches for computing ligand–receptor binding kinetics. *Curr Opin Struct Biol*. 2018;*49*:1–10. https://doi.org/10.1016/j.sbi.2017.10.001.
34. Capelli R, Lyu W, Bolnykh V, Meloni S, Olsen JMH, Rothlisberger U, et al. Accuracy of molecular simulation-based predictions of koffValues: a Metadynamics study. *J Phys Chem Lett*. 2020;*11*(15):6373–81. https://doi.org/10.1021/acs.jpclett.0c00999.
35. Plattner N, Noé F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun*. 2015;*6*:1–10. https://doi.org/10.1038/ncomms8653.
36. Durrant JD, McCammon JA. Computer-aided drug-discovery techniques that account for receptor flexibility. *Curr Opin Pharmacol*. 2010;*10*(6):770–4. https://doi.org/10.1016/j.coph.2010.09.001.
37. Aranda J, Orozco M. RNA-dependent RNA polymerase from SARS-CoV-2. Mechanism of reaction and inhibition by Remdesivir. *BioRxiv*. 2020;*2020*:163592. https://doi.org/10.1101/2020.06.21.163592.
38. Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, et al. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Central Science*. 2020;*6*(10):1722–34. https://doi.org/10.1021/acscentsci.0c01056.
39. Oliveira ASF, Ibarra AA, Bermudez I, Casalino L, Gaieb Z, Shoemark DK, et al. Simulations support the interaction of the SARS-CoV-2 spike protein with nicotinic acetylcholine receptors and suggest subtype specificity. *bioRxiv*. 2020;1–14. https://doi.org/10.1101/2020.07.16.206680.
40. Shaw DE. (2020). D. E. Shaw Research Download Simulation Data for SARS-CoV-2 Proteins Registration. https://www.deshawresearch.com/downloads/download_trajectory_sarscov2.cgi/

41. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of molecular dynamics and related methods in drug discovery. *J Med Chem*. 2016; *59*(9):4035–61. https://doi.org/10.1021/acs.jmedchem.5b01684.
42. Herbert C, Schieborr U, Saxena K, Juraszek J, De Smet F, Alcouffe C, et al. Molecular mechanism of SSR128129E, an extracellularly acting, small-molecule, allosteric inhibitor of fgf receptor signaling. *Cancer Cell*. 2013;*23*(4):489–501. https://doi.org/10.1016/j.ccr.2013.02.018.
43. Liu T, Lu D, Zhang H, Zheng M, Yang H, Xu Y, et al. Applying high-performance computing in drug discovery and molecular simulation. *Natl Sci Rev*. 2016;*3*(1):49–63. https://doi.org/10.1093/nsr/nww003.
44. Comitani F, Gervasio FL. Exploring cryptic pockets formation in targets of pharmaceutical interest with SWISH. *J Chem Theory Comput*. 2018;*14*(6):3321–31. https://doi.org/10.1021/acs.jctc.8b00263.
45. Feinstein WP, Brylinski M. Accelerated structural bioinformatics for drug discovery. *High Performance Parallelism Pearls: Multicore Many-core Program Approach*. 2015;*2*:55–72. https://doi.org/10.1016/B978-0-12-803819-2.00012-4.
46. Samdani A, Vetrivel U. POAP: a GNU parallel based multithreaded pipeline of open babel and AutoDock suite for boosted high throughput virtual screening. *Comput Biol Chem*. 2018;*74*:39–48. https://doi.org/10.1016/j.compbiolchem.2018.02.012.
47. Warshel A, Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol*. 1976;*103*(2):227–49. https://doi.org/10.1016/0022-2836(76)90311-9.
48. Lyu W, Arnesano F, Carloni P, Natile G, Rossetti G. Effect of in vivo post-translational modifications of the HMGB1 protein upon binding to platinated DNA: a molecular simulation study. *Nucleic Acids Res*. 2018;*46*(22):11687–97. https://doi.org/10.1093/nar/gky1082.
49. Riccardi L, Genna V, De Vivo M. Metal–ligand interactions in drug design. *Nat Rev Chem*. 2018;*2*(7):100–12. https://doi.org/10.1038/s41570-018-0018-6.
50. Schramm VL. Transition states, analogues, and drug development. *ACS Chem Biol*. 2013;*8*(1):71–81. https://doi.org/10.1021/cb300631k.
51. Bolnykh V, Olsen JMH, Meloni S, Bircher MP, Ippoliti E, Carloni P, et al. Extreme scalability of DFT-based QM/MM MD simulations using MiMiC. *J Chem Theory Comput*. 2019;*15*(10):5601–13. https://doi.org/10.1021/acs.jctc.9b00424.
52. Carbó R, Chakraborty T. *Theoretical and quantum chemistry at the Dawn of the 21st century*. New York: Apple Academic Press Incorporated; 2018.
53. Patterson D, John LH. *Computer organization and design: the hardware/software interface*. Vol 1. San Mateo, CA: M Organ Kaufmann Publishers; 2005. p. 998.
54. Vanderbauwhede W, Benkrid K. *High-performance computing using FPGAs*. Vol 3. New York: Springer; 2013.
55. MPI Standard. (2019). https://www.mpi-forum.org/docs/mpi-3.1/mpi31-report.pdf
56. Shainer G. *It is about latency*. *HPCWire*. San Diego: Tabor Communications Inc.; 2019. https://www.hpcwire.com/2019/12/02/it-is-about-latency/
57. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A second generation force Field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*. 1996;*118*(9):2309–9.
58. Larsson P, Hess B, Lindahl E. Algorithm improvements for molecular dynamics simulations. *WIREs Comput Mol Sci*. 2011;*1*(1):93–108. https://doi.org/10.1002/wcms.3.
59. Tenorio Y, Hernandez-Santoyo A, Altuzar V, Vivanco-Cid H, Mendoza-Barrera C. Protein-protein and protein-ligand docking. In: Ogawa T, editor. *Protein Engineering—Technology and Application*. London: InTech; 2013. p. 187. https://doi.org/10.5772/56376.
60. Tuckerman M. *Statistical mechanics: theory and molecular simulation*. Oxford: Oxford University Press; 2010.
61. Åqvist J, Warshel A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem Rev*. 1993;*93*(7):2523–44. https://doi.org/10.1021/cr00023a010.
62. Brunk E, Rothlisberger U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological Systems in Ground and Electronically Excited States. *Chem Rev*. 2015;*115*(12):6217–63. https://doi.org/10.1021/cr500628b.
63. Grillo IB, Bachega JFR, Timmers LFSM, Caceres RA, de Souza ON, Field MJ, et al. Theoretical characterization of the shikimate 5-dehydrogenase reaction from *Mycobacterium tuberculosis* by hybrid QC/MM simulations and quantum chemical descriptors. *J Mol Model*. 2020;*26*(11):1–12. https://doi.org/10.1007/s00894-020-04536-9.
64. Peraro MD, Ruggerone P, Raugei S, Gervasio FL, Carloni P. Investigating biological systems using first principles car–Parrinello molecular dynamics simulations. *Curr Opin Struct Biol*. 2007;*17*(2):149–56. https://doi.org/10.1016/j.sbi.2007.03.018.
65. Raich L, Nin-Hill A, Ardèvol A, Rovira C. Enzymatic cleavage of Glycosidic bonds: strategies on how to set up and control a QM/MM Metadynamics simulation. *Methods Enzymol*. 2016;*577*:159–83. https://doi.org/10.1016/bs.mie.2016.05.015.
66. Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew Chem Int Ed*. 2009;*48*(7):1198–229. https://doi.org/10.1002/anie.200802019.
67. Albini A, Monti S. Photophysics and photochemistry of fluoroquinolones. *Chem Soc Rev*. 2003;*32*(4):238–50. https://doi.org/10.1039/b209220b.
68. Stern RS, Kleinerman RA, Parrish JA, Fitzpatrick TB, Bleich HL. Phototoxic reactions to photoactive drugs in patients treated with PUVA. *Arch Dermatol*. 1980;*116*(11):1269–71. https://doi.org/10.1001/archderm.1980.01640350059015.
69. Taura J, Nolen EG, Cabré G, Hernando J, Squarcialupi L, López-Cano M, et al. Remote control of movement disorders using a photoactive adenosine a 2A receptor antagonist. *J Control Release*. 2018;*283*:135–42. https://doi.org/10.1016/j.jconrel.2018.05.033.
70. Calandrini V, Arnesano F, Galliani A, Nguyen TH, Ippoliti E, Carloni P, et al. Platination of the copper transporter ATP7A involved in anticancer drug resistance. *Dalton Trans*. 2014;*43*(31):12085–94. https://doi.org/10.1039/C4DT01339E.

71. Genna V, Vidossich P, Ippoliti E, Carloni P, De Vivo M. A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. *J Am Chem Soc*. 2016;*138*(44):14592–8. https://doi.org/10.1021/jacs.6b05475.
72. Park H, Brothers EN, Merz KM. Hybrid QM/MM and DFT investigations of the catalytic mechanism and inhibition of the Dinuclear zinc Metallo-β-lactamase CcrA from *Bacteroides fragilis*. *J Am Chem Soc*. 2005;*127*(12):4232–41. https://doi.org/10.1021/ja042607b.
73. Villar-Piqué A, Lopes da Fonseca T, Sant'Anna R, Szegö ÉM, Fonseca-Ornelas L, Pinho R, et al. Environmental and genetic factors support the dissociation between α-synuclein aggregation and toxicity. *Proc Natl Acad Sci USA*. 2016;*113*(42):E6506–15. https://doi.org/10.1073/pnas.1606791113.
74. Lopes PEM, Guvench O, MacKerell AD Jr. Current status of protein force fields for molecular dynamics simulations. *Methods Mol Biol*. 2015;*1215*:47–71. https://doi.org/10.1007/978-1-4939-1465-4_3.
75. Song LF, Sengupta A, Merz KM. Thermodynamics of transition metal ion binding to proteins. *J Am Chem Soc*. 2020;*142*(13):6365–74. https://doi.org/10.1021/jacs.0c01329.
76. van Duin ACT, Dasgupta S, Lorant F, Goddard WA. ReaxFF: a reactive force Field for hydrocarbons. *Chem A Eur J*. 2001;*105*(41):9396–409. https://doi.org/10.1021/jp004368u.
77. Pople JA, Beveridge DL. *Approximate molecular orbital theory*. New York: McGraw-Hill; 1970.
78. Kamerlin SCL, Warshel A. The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. *Faraday Discuss*. 2010;*145*:71–106. https://doi.org/10.1039/b907354j.
79. Minenkov Y, Sharapa DI, Cavallo L. Application of Semiempirical methods to transition metal complexes: fast results but hard-to-predict accuracy. *J Chem Theory Comput*. 2018;*14*(7):3428–39. https://doi.org/10.1021/acs.jctc.8b00018.
80. Vujović M, Huynh M, Steiner S, Garcia-Fernandez P, Elstner M, Cui Q, et al. Exploring the applicability of density functional tight binding to transition metal ions. Parameterization for nickel with the spin-polarized DFTB3 model. *J Comput Chem*. 2019;*40*(2):400–13. https://doi.org/10.1002/jcc.25614.
81. Shaw D, Grossman JP, Bank J, Batson B, Butts J, Chao J, et al. (2015). Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. International Conference for High Performance Computing, Networking, Storage and Analysis, SC, 2015, 41–53. https://doi.org/10.1109/SC.2014.9
82. Abraham MJ, van der Spoel D, Lindahl E, Hess B, The GROMACS Development Team. *GROMACS User Manual version 2019*. GROMACS; 2019.
83. Páll S, Zhmurov A, Bauer P, Abraham M, Lundborg M, Gray A, et al. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J Chem Phys*. 2020;*153*(13):134110. https://doi.org/10.1063/5.0018516.
84. Marinari E, Parisi G. Simulated tempering: a new Monte Carlo scheme. *Eur Phys Lett*. 1992;*19*(6):451–8. https://doi.org/10.1209/0295-5075/19/6/002.
85. Geyer CJ. Markov chain Monte Carlo maximum likelihood, computing science and statistics. *Proc 23rd Symp Interface*. 1991;*1*:156–63.
86. Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci USA*. 2002;*99*(20):12562–6. https://doi.org/10.1073/pnas.202427399.
87. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comput Phys*. 1977;*23*(2):187–99. https://doi.org/10.1016/0021-9991(77)90121-8.
88. Babin V, Roland C, Sagui C. Adaptively biased molecular dynamics for free energy calculations. *J Chem Phys*. 2008;*128*(13):134101. https://doi.org/10.1063/1.2844595.
89. Bonati L, Rizzi V, Parrinello M. Data-driven collective variables for enhanced sampling. *J Phys Chem Lett*. 2020;*11*(8):2998–3004. https://doi.org/10.1021/acs.jpclett.0c00535.
90. Noé F, Olsson S, Köhler J, Wu H. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science*. 2019;*365*(6457):eaaw1147. https://doi.org/10.1126/science.aaw1147.
91. Roberts K, Sebsebie R, Curotto E. A rare event sampling method for diffusion Monte Carlo using smart darting. *J Chem Phys*. 2012;*136*(7):074104. https://doi.org/10.1063/1.3685453.
92. Sbailò L, Dibak M, Noé F. Neural Mode Jump Monte Carlo. *ArXiv*. 2019;1–17. http://arxiv.org/abs/1912.05216.
93. Zhang J, Yang YI, Noé F. Targeted adversarial learning optimized sampling. *J Phys Chem Lett*. 2019;*10*(19):5791–7. https://doi.org/10.1021/acs.jpclett.9b02173.
94. Olsen JMH, Bolnykh V, Meloni S, Ippoliti E, Bircher MP, Carloni P, et al. MiMiC: a novel framework for multiscale modeling in computational chemistry. *J Chem Theory Comput*. 2019;*15*(6):3810–23. https://doi.org/10.1021/acs.jctc.9b00093.
95. Hutter J, Alavi A, Deutsch T, Bernasconi M, Goedecker S, Marx D, et al. (2018). CPMD copyright IBM Corp 1990-2019, MPI für Festkörperforschung Stuttgart 1997–2012. http://www.cpmd.org/
96. Cooley JW, Tukey JW. An algorithm for the machine calculation of complex Fourier series. *Math Comput*. 1965;*19*:297–301.
97. Weber V, Bekas C, Laino T, Curioni A, Bertsch A, Futral S. (2014). Shedding light on lithium/air batteries using millions of threads on the BG/Q supercomputer BT. 2014 IEEE 28th International Parallel and Distributed Processing Symposium. https://doi.org/10.1109/ipdps.2014.81
98. Browning NJ (2019). Applications of Artificial Intelligence to Computational Chemistry [EPFL PP - Lausanne]. https://doi.org/10.5075/epfl-thesis-9645
99. Chiariello MG, Bolnykh V, Ippoliti E, Meloni S, Olsen JMH, Beck T, et al. Molecular basis of CLC Antiporter inhibition by fluoride. *J Am Chem Soc*. 2020;*142*(16):7254–8. https://doi.org/10.1021/jacs.9b13588.

100. Maia EHB, Assis LC, de Oliveira TA, da Silva AM, Taranto AG. Structure-based virtual screening: from classical to artificial intelligence. *Front Chem*. 2020;*8*:343. https://doi.org/10.3389/fchem.2020.00343.
101. Li J, Fu A, Zhang L. An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdiscipl Sci Comput Life Sci*. 2019;*11*(2):320–8. https://doi.org/10.1007/s12539-019-00327-w.
102. Shen C, Ding J, Wang Z, Cao D, Ding X, Hou T. From machine learning to deep learning: advances in scoring functions for protein–ligand docking. *WIREs Comput Mol Sci*. 2020;*10*(1):1–23. https://doi.org/10.1002/wcms.1429.
103. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*. 1996;*261*(3):470–89. https://doi.org/10.1006/jmbi.1996.0477.
104. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev*. 2017;*9*(2):91–102. https://doi.org/10.1007/s12551-016-0247-1.
105. Pinzi L, Rastelli G. Molecular docking: shifting paradigms in drug discovery. *Int J Mol Sci*. 2019;*20*(18):1–23. https://doi.org/10.3390/ijms20184331.
106. Luksch T, Chan N-S, Brass S, Sotriffer CA, Klebe G, Diederich WE. Computer-aided design and synthesis of Nonpeptidic Plasmepsin II and IV inhibitors. *ChemMedChem*. 2008;*3*(9):1323–36. https://doi.org/10.1002/cmdc.200700270.
107. Chen D, Misra M, Sower L, Peterson JW, Kellogg GE, Schein CH. Novel inhibitors of anthrax edema factor. *Bioorg Med Chem*. 2008;*16*(15):7225–33. https://doi.org/10.1016/j.bmc.2008.06.036.
108. Bui NK, Turk S, Buckenmaier S, Stevenson-Jones F, Zeuch B, Gobec S, et al. Development of screening assays and discovery of initial inhibitors of pneumococcal peptidoglycan deacetylase PgdA. *Biochem Pharmacol*. 2011;*82*(1):43–52. https://doi.org/10.1016/j.bcp.2011.03.028.
109. Sotriffer CA. Molecular dynamics simulations in drug design. *Encyclopedic reference of genomics and proteomics in molecular medicine*. Berlin, Heidelberg: Springer; 2006. p. 1153–60. https://doi.org/10.1007/3-540-29623-9_0820.
110. Shah F, Mukherjee P, Gut J, Legac J, Rosenthal PJ, Tekwani BL, et al. Identification of novel malarial cysteine protease inhibitors using structure-based virtual screening of a focused cysteine protease inhibitor library. *J Chem Inf Model*. 2011;*51*(4):852–64. https://doi.org/10.1021/ci200029y.
111. Caporuscio F, Rastelli G, Imbriano C, Del Rio A. Structure-based design of potent aromatase inhibitors by high-throughput docking. *J Med Chem*. 2011;*54*(12):4006–17. https://doi.org/10.1021/jm2000689.
112. Yang Z, Wolf IM, Chen H, Periyasamy S, Chen Z, Yong W, et al. FK506-binding protein 52 is essential to uterine reproductive physiology controlled by the progesterone receptor a isoform. *Mol Endocrinol*. 2006;*20*(11):2682–94. https://doi.org/10.1210/me.2006-0024.
113. Cerchietti LC, Hatzi K, Caldas-Lopes E, Yang SN, Figueroa ME, Morin RD, et al. BCL6 repression of EP300 in human diffuse large B cell lymphoma cells provides a basis for rational combinatorial therapy. *J Clin Investig*. 2010;*120*(12):4569–82. https://doi.org/10.1172/JCI42869.
114. Jo S-H, Yang C, Miao Q, Marzec M, Wasik MA, Lu P, et al. Peroxisome proliferator-activated receptor γ promotes lymphocyte survival through its actions on cellular metabolic activities. *J Immunol*. 2006;*177*(6):3737–45. https://doi.org/10.4049/jimmunol.177.6.3737.
115. Segers K, Sperandio O, Sack M, Fischer R, Miteva MA, Rosing J, et al. Design of protein-membrane interaction inhibitors by virtual ligand screening, proof of concept with the C2 domain of factor V. *Proc Natl Acad Sci USA*. 2007;*104*(31):12697–702. https://doi.org/10.1073/pnas.0701051104.
116. Vidal D, Blobel J, Pérez Y, Thormann M, Pons M. Structure-based discovery of new small molecule inhibitors of low molecular weight protein tyrosine phosphatase. *Eur J Med Chem*. 2007;*42*(8):1102–8. https://doi.org/10.1016/j.ejmech.2007.01.017.
117. Mezei M. A new method for mapping macromolecular topography. *J Mol Graph Model*. 2003;*21*(5):463–72. https://doi.org/10.1016/S1093-3263(02)00203-6.
118. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol*. 2011;*9*(1):71. https://doi.org/10.1186/1741-7007-9-71.
119. Kokh DB, Wade RC, Wenzel W. Receptor flexibility in small-molecule docking calculations. *WIREs Comput Mol Sci*. 2011;*1*(2):298–314. https://doi.org/10.1002/wcms.29.
120. Jiang F, Kim SH. "Soft docking": matching of molecular surface cubes. *J Mol Biol*. 1991;*219*(1):79–102. https://doi.org/10.1016/0022-2836(91)90859-5.
121. Leach AR. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol*. 1994;*235*(1):345–56. https://doi.org/10.1016/S0022-2836(05)80038-5.
122. Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles. *Protein Sci*. 1998;*7*(4):938–50. https://doi.org/10.1002/pro.5560070411.
123. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem*. 2006;*49*(2):534–53. https://doi.org/10.1021/jm050540c.
124. Morris GM, Ruth H, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. Software news and updates AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*. 2009;*30*(16):2785–91. https://doi.org/10.1002/jcc.21256.
125. Salmaso V, Moro S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. *Front Pharmacol*. 2018;*9*:923. https://doi.org/10.3389/fphar.2018.00923.
126. Anderson AC, O'Neil RH, Surti TS, Stroud RM. Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking. *Chem Biol*. 2001;*8*(5):445–57. https://doi.org/10.1016/S1074-5521(01)00023-0.

127. Ravindranath PA, Forli S, Goodsell DS, Olson AJ, Sanner MF. AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS Comput Biol*. 2015;*11*(12):e1004586. https://doi.org/10.1371/journal.pcbi.1004586.
128. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2009;*31*(2):455–461. https://doi.org/10.1002/jcc.21334.
129. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*. 1997;*267*(3):727–48. https://doi.org/10.1006/jmbi.1996.0897.
130. Dong D, Xu Z, Zhong W, Peng S. Parallelization of molecular docking: a review. *Curr Top Med Chem*. 2018;*18*(12):1015–28. https://doi.org/10.2174/1568026618666180821145215.
131. Hatcher PJ, Lapadula AJ, Jones RR, Quinn MJ, Seevers BK, Anderson RJ. Data-parallel programming on MIMD computers. *IEEE Trans Parallel Distrib Syst*. 1991;*2*(3):377–83. https://doi.org/10.1109/71.86112.
132. Goodsell DS, Lauble H, Stout CD, Olson AJ. Automated docking in crystallography: analysis of the substrates of aconitase. *Proteins Struct Funct Bioinf*. 1993;*17*(1):1–10. https://doi.org/10.1002/prot.340170104.
133. Zhang X, Wong SE, Lightstone FC. Message passing interface and multithreading hybrid for parallel molecular docking of large databases on petascale high performance computing machines. *J Comput Chem*. 2013;*34*(11):915–27. https://doi.org/10.1002/jcc.23214.
134. Hilman M, Suhartanto H, Yanuar A. (2013). Performance Analysis of Embarassingly Parallel Application on Cluster Computer Environment: A Case Study of Virtual Screening with Autodock Vina 1.1 on Hastinapura Cluster. http://arxiv.org/abs/1305.3123
135. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*. 2004;*47*(7):1750–9. https://doi.org/10.1021/jm030644s.
136. Venkatachalam CM, Jiang X, Oldfield T, Waldman M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graph Model*. 2003;*21*(4):289–307. https://doi.org/10.1016/S1093-3263(02)00164-X.
137. Foster I. The anatomy of the grid: enabling scalable virtual organizations. *Lect Notes Comput Sci*. 2001;*2150*:1–4. https://doi.org/10.1007/3-540-44681-8_1.
138. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, et al. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des*. 2006;*20*(10–11):601–19. https://doi.org/10.1007/s10822-006-9060-4.
139. Zhou T, Caflisch A. Data management system for distributed virtual screening. *J Chem Inf Model*. 2009;*49*(1):145–52. https://doi.org/10.1021/ci800295q.
140. MySQL: The World's Most Popular Open Source Database. (2020). https://chnm.gmu.edu/digitalhistory/links/cached/appendix/a5_mysql.htm
141. Peters A, Lundberg M, Lang P, Sosa CP. *High throughput computing validation for drug discovery using the DOCK program on a massively parallel system*. New York: IBM; 2007.
142. Allsopp N, Tabary A, Follows J, Vezolle P, Hennecke M, Reddy H, et al. (2005). Unfolding IBM eServer Blue Gene Solution.
143. Thorsen O, Smith B, Sosa CP, Jiang K, Lin H, Peters A., Feng WC (2007). Parallel genomic sequence-search on a massively parallel system. Proceedings of the Conference on 2007 Computing Frontiers, 59–68. https://doi.org/10.1145/1242531.1242542
144. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. How does a drug molecule find its target binding site? *J Am Chem Soc*. 2011;*133*(24):9181–3. https://doi.org/10.1021/ja202726y.
145. Ellingson SR, Smith JC, Baudry J. VinaMPI: facilitating multiple receptor high-throughput virtual docking on high-performance computers. *J Comput Chem*. 2013;*34*(25):2212–21. https://doi.org/10.1002/jcc.23367.
146. Guerrero GD, Pérez-Sánchez H, Wenzel W, Cecilia JM, García JM. Effective parallelization of non-bonded interactions kernel for virtual screening on GPUs. *Adv Intell Soft Comput*. 2011;*93*:63–9. https://doi.org/10.1007/978-3-642-19914-1_9.
147. Harvey MJ, De Fabritiis G. An implementation of the smooth particle mesh Ewald method on GPU hardware. *J Chem Theory Comput*. 2009;*5*(9):2371–7. https://doi.org/10.1021/ct900275y.
148. Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K. Accelerating molecular modeling applications with graphics processors. *J Comput Chem*. 2007;*28*(16):2618–40. https://doi.org/10.1002/jcc.20829.
149. Kokh DB, Wenzel W. Flexible side chain models improve enrichment rates in in silico screening. *J Med Chem*. 2008;*51*(19):5919–31. https://doi.org/10.1021/jm800217k.
150. Hanbury A, Müller H, Langs G, Weber MA, Menze BH, Fernandez TS. Bringing the algorithms to the data: cloud-based benchmarking for medical image analysis. *Lect Notes Comput Sci*. 2012;*7488*:24–9. https://doi.org/10.1007/978-3-642-33247-0_3.
151. Armbrust M, Fox A, Griffith R, Joseph AD, Katz RH, Konwinski A, et al. (2009). Above the Clouds: A Berkeley View of Cloud Computing. http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html
152. Ellingson SR, Baudry J. High-throughput virtual molecular docking with AutoDockCloud. *Concurr Comput Pract Exp*. 2014;*26*(4):907–16. https://doi.org/10.1002/cpe.2926.
153. De Paris R, Frantz FA, Norberto De Souza O, Ruiz DDA. WFReDoW: a cloud-based web environment to handle molecular docking simulations of a fully flexible receptor model. *Biomed Res Int*. 2013;*2013*:12. https://doi.org/10.1155/2013/469363.
154. Cournia Z, Allen B, Sherman W. Relative binding free energy calculations in drug discovery: recent advances and practical considerations. *J Chem Inform Model*. 2017;*57*(12):2911–37. https://doi.org/10.1021/acs.jcim.7b00564.
155. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct*. 2007;*36*:21–42. https://doi.org/10.1146/annurev.biophys.36.040306.132550.
156. Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. *Sci Rep*. 2017;*7*(1):1–10. https://doi.org/10.1038/srep46710.

157. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov*. 2015;*10*(5):449–61. https://doi.org/10.1517/17460441.2015.1032936.
158. Cournia Z, Allen BK, Beuming T, Pearlman DA, Radak BK, Sherman W. Rigorous free energy simulations in virtual screening. *J Chem Inform Model*. 2020;*60*(9):4153–69. https://doi.org/10.1021/acs.jcim.0c00116.
159. Lee TS, Allen BK, Giese TJ, Guo Z, Li P, Lin C, et al. Alchemical binding free energy calculations in AMBER20: advances and best practices for drug discovery. *J Chem Inf Model*. 2020;*60*(11):5595–623. https://doi.org/10.1021/acs.jcim.0c00613.
160. Cavasotto CN, Adler NS, Aucar MG. Quantum chemical approaches in structure-based virtual screening and Lead optimization. *Front Chem*. 2018;*6*(MAY):188. https://doi.org/10.3389/fchem.2018.00188.
161. Ge H, Wang Y, Li C, Chen N, Xie Y, Xu M, et al. Molecular dynamics-based virtual screening: accelerating the drug discovery process by high-performance computing. *J Chem Inf Model*. 2013;*53*(10):2757–64. https://doi.org/10.1021/ci400391s.
162. Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E. *Tackling Exascale software challenges in molecular dynamics simulations with GROMACS BT. Lecture Notes in Computer Science*. New York: Springer International Publishing; 2015.p. 3–27. https://doi.org/10.1007/978-3-319-15976-8_1.
163. Dauwe D, Pasricha S, Maciejewski AA, Siegel HJ. (2016). A performance and energy comparison of fault tolerance techniques for exascale computing systems. 2016 IEEE International Conference on Computer and Information Technology (CIT), 436–443. https://doi.org/10.1109/CIT.2016.44
164. Smalley E. AI-powered drug discovery captures pharma interest. *Nat Biotechnol*. 2017;*35*(7):604–5. https://doi.org/10.1038/nbt0717-604.
165. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020;*577*(7792):706–10. https://doi.org/10.1038/s41586-019-1923-7.

How to cite this article: Bolnykh V, Rossetti G, Rothlisberger U, Carloni P. Expanding the boundaries of ligand–target modeling by exascale calculations. *WIREs Comput Mol Sci*. 2021;e1535. https://doi.org/10.1002/wcms.1535