Drug Design in the Exascale Era: A Perspective from Massively Parallel QM/MM Simulations

Bharath Raghavan,[†,‡] Mirko Paulikat,[†] Katya Ahmad,[†] Lara Callea,[¶] Andrea Rizzi,[†,§] Emiliano Ippoliti,[†] Davide Mandelli,[*,†] Laura Bonati,[¶] Marco De Vivo,[§] and Paolo Carloni[*,†,‡]

†*Computational Biomedicine, Institute of Advanced Simulations IAS-5/Institute for Neuroscience and Medicine INM-9, Forschungszentrum Jülich GmbH, Jülich 52428, Germany*

‡*Department of Physics and Universitätsklinikum, RWTH Aachen University, Aachen 52074, Germany*

¶*Department of Earth and Environmental Sciences, University of Milano-Bicocca, Piazza della Scienza 1, 20126 Milan, Italy*

§*Atomistic Simulations, Italian Institute of Technology, Genova 16163, Italy*

E-mail: d.mandelli@fz-juelich.de; p.carloni@fz-juelich.de

Abstract

The initial phases of drug discovery - in silico drug design - could benefit from first principle Quantum Mechanics / Molecular Mechanics (QM/MM) molecular dynamics (MD) simulations in explicit solvent, yet many applications are currently limited by the short time scales that this approach can cover. Developing scalable first principle QM/MM MD interfaces fully exploiting current exascale machines - so far an unmet and crucial goal - will help overcome this problem, opening the way to the study of

the thermodynamics and kinetics of ligand binding to protein with first principle accuracy. Here, taking two relevant case studies involving the interactions of ligands with rather large enzymes, we showcase the use of our recently developed massively scalable MiMiC QM/MM framework (currently using DFT to describe the QM region) to investigate reactions and ligand binding in enzymes of pharmacological relevance. We also demonstrate for the first time strong scaling of MiMiC QM/MM MD simulations with parallel efficiency above 70% with over 40,000 cores. Thus, among many others, the MiMiC interface represents a promising candidate towards exascale applications by combining machine learning with statistical mechanics based algorithms tailored for exascale supercomputers.

Introduction

In the last decade, first-principle Quantum Mechanics/Molecular Mechanics (QM/MM) molecular dynamics (MD) simulations in explicit solvent proved to be a powerful tool to investigate biochemical processes where the electronic degrees of freedom play a major role.[1,2] In this approach, the region of interest (e.g. the active site of an enzyme) is treated at the QM level, while the rest is described by classical force fields.[3] The choice of the QM level of theory is generally a compromise between the accuracy required by the problem at hand and the associated computational burden. In this respect, non-empirical density functional theory (DFT) is a rather general (and relatively accurate) approach, and it comes at a far minor computational cost than wavefunction-based methods including electronic correlation.[4] As such, DFT-based QM/MM MD simulations are nowadays the method of choice in many state of the art *in-silico* studies of biochemical processes, including enzymatic reactions,[5–16] transition metals binding to proteins,[17–20] proton transfer[21–24] and photophysical processes.[25–28] Applications to drug design, on the other hand, have not been sufficiently explored, apart from notable exceptions.[29] *Static* DFT QM/MM calculations have been already shown to be very useful, by explaining drug action and by informing about routes

for structure based drug design.[30–34] Accessing the *dynamics* of enzymatic reactions at the DFT level will push the boundaries of pharmacological applications beyond the current state of the art,[12] extending in particular the domain of application of QM/MM approaches to most metalloenzymes (more than 30% of all proteins[35]). Long timescale DFT QM/MM MD can help in describing the flexibility and the dynamics of complex enzymes, which may be crucial for their function,[36–40] and in predicting accurate catalytic rates (k_{cat}) and transition states. The latter represent essential knowledge for the design of transition state analogs,[41,42] widely considered to be superior to substrate analogs.[43] First principle QM/MM MD simulations can also serve as a stepping stone towards accurate predictions of ligand binding free energies[44–46] and residence times (k_{off}^{-1}),[47,48] very important parameters to assess drug efficiency.[49–51] However, DFT QM/MM MD comes at a much larger computational cost than static and semi-empirical calculations. As a result, the accessible timescales currently reach few hundreds of ps in state of the art DFT QM/MM simulations including $\sim 10^2$ QM atoms,[52–54] severely limiting the statistical accuracy. This is the main bottleneck hindering the widespread utilization of this method for pharmacology, in both academia and industry.

The current exascale revolution in high performance computing presents an exciting opportunity for the DFT QM/MM community to transcend these limitations.[55] Reaching the exascale requires DFT QM/MM interfaces to scale effectively and take maximum advantage of the large number of networked CPU and GPU cores provided by modern supercomputers. Despite the many efficient DFT QM/MM software available,[56–67] to the best of our knowledge, scarce information can be found in the literature regarding their strong scaling in pure QM/MM MD applications. In this respect, the Multiscale Modeling in Computational Chemistry (MiMiC) QM/MM framework[68,69] that couples CPMD[70] (QM) and GROMACS[71] (MM), represents a notable exception. MiMiC has been recently developed within a European collaboration, including some of the authors. As we demonstrate in this work, the current version of MiMiC can scale over tens of thousands of processes in a single QM/MM MD run of a large enzymes at the B3LYP level of theory. As such, we believe that it is well

posed to break the limits of currently achievable timescales in applications to pharmacology. Recent trends in computational chemistry let us envision that this will occur via a clever combination with novel statistical mechanics based algorithms and machine learning techniques. So far, QM codes have not been able to scale efficiently on GPU-equipped distributed architectures. Machine learning methods have already shown to make excellent utilization of GPU resources, and could be excellent candidates to push DFT QM/MM MD into the exascale regime.[72,73]

Here, after summarizing some salient aspects of the MiMiC QM/MM interface and demonstrating its scalability, we present applications of the code to systems of pharmacological relevance, from enzymatic reactions for the prediction of the transition state, to inhibitor-enzyme binding towards the investigation of k_{off} values. We close by giving our perspective about QM/MM MD simulations for drug design in the exascale era.

The MiMiC framework

The MiMiC framework provides a general platform that enables the implementation of multiscale simulation methods through coupling of multiple external programs.[68] Since its inception, MiMiC has been designed for massively parallel applications. With this in mind, a multiple-programs multiple-data model has been adopted, where the external programs are allowed to run simultaneously on independent computing resources while exploiting their existing parallelization strategies. Specifically, MiMiC consists of two libraries: (i) the main MiMiC library,[74] which provides optimized routines for fast computation of the interactions between different subsystems and (ii) the MiMiC communication library (MCL),[75] a lightweight communication library that is used to exchange information between the main MiMiC library and the external programs. Adding a new external program to the MiMiC framework requires a relatively small effort that consists in implementing an MCL-based interface and, if needed, extending the main MiMiC library to support the computation of

new interaction terms. Preparing input files for MiMiC-based QM/MM simulations is made easy by the MiMiCPy python library,[76] which can be straightforwardly extended to deal with file formats of any MM and QM code. Overall, these features make MiMiC a highly flexible and efficient framework for multiscale simulations.

MiMiC allows performing QM/MM simulations at the DFT level of theory within an electrostatic embedding scheme[77] via coupling to the CPMD[78] and GROMACS[71,79] codes serving as the MM and QM subprograms, respectively. Thanks to CPMD's very efficient use of standard CPU nodes, this implementation has already displayed strong scaling well beyond ten thousand cores while maintaining an overall parallel efficiency above 70% in a single QM/MM MD simulation of an antiporter protein embedded in a solvated lipid bilayer.[69] More recently, MiMiC-QM/MM simulations have been used to investigate the thermodynamics of transport processes in membrane channels and transporters,[21,22] demonstrating the possibility of routine sub-ns QM/MM MD runs of rather large systems.

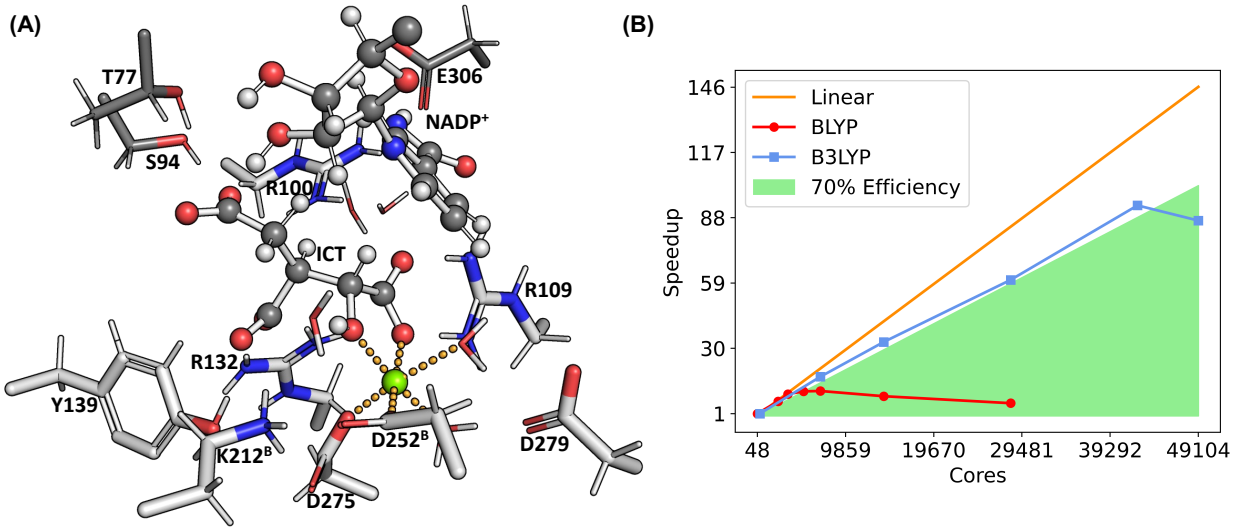


Figure 1: (A) Representation of the Michaelis complex of the IDH1 active site from classical MD simulations. ICT, part of the $NADP^+$ pictured and all residues in light gray are placed in the QM region in our MiMiC-QM/MM simulations. (B) Strong scaling of MiMiC-based DFT QM/MM MD simulations of IDH1 as a function of the number of cores assigned to CPMD. In all simulations, we assigned one node (48 cores) to GROMACS. The speedup is provided in terms of the CPU time required for one MD step, normalized with respect to the reference runs on one (BLYP) and seven nodes (B3LYP). All simulations have been performed on the JUWELS cluster.[80]

Here, for the first time, we demonstrate extreme scalability of MiMiC-QM/MM MD for the investigation of enzymatic reactions, considering the case study of the human Isocitrate Dehydrogenase-1 (IDH1). The solvated protein consists of 130,828 atoms in total, 142 atoms from the active site were assigned to the QM region with box size 46.0 a.u. $\times$ 46.0 a.u. $\times$ 46.0 a.u. Figure 1 shows the Michaelis complex of the enzyme, as obtained from preliminary classical MD simulations, together with the results of strong scaling benchmarks performed using the BLYP[81] and B3LYP[82] functionals. Using the former, CPMD (coupled to MiMiC) scales efficiently up to 5,184 cores, achieving a performance of 5.4 ps/day. Running at this configuration would require around 0.02 Mcore-hours/ps. The scaling obtained at the B3LYP[82] level is remarkable, with an efficiency $> 70\%$ up to 42,336 cores (882 JUWELS nodes), and achieving a performance of 0.4 ps/day. Running at this configuration would require around 2.5 Mcore-h/ps. The maximum amount of 1024 nodes can be requested for a single run on JUWELS. Thus, 882 nodes effectively corresponds to the largest amount of resources that can be still distributed efficiently to keep load balancing in CPMD for our system. Indeed, bad load balancing explains the sudden drop in performance observed when using 49,104 cores, corresponding to 1023 nodes (see Table 1 and related discussion in the Supporting Information for more details). This indicates that the overall performance could be potentially improved, given enough resources (like in exascale machines).

We also performed scaling benchmarks for the inhibitor-enzyme complex formed by the p38α enzyme and the ligand **2g**, which served as our second case study in this work. This system was smaller, with only the 46 atoms of the ligand included in the QM region. MiMiC QM/MM MD at the BLYP level achieved performance of $\approx$ 21 ps/day on 384 cores of the JUWELS cluster module. Running at this configuration would require around 439 core-hours/ps. Benchmarks at the B3LYP level, on the other hand, showed parallel efficiency above 70% up to 12,288 cores with a performance of $\approx$ 4.8 ps/day (see Supporting Information for details). Running at this configuration would require around 0.06 Mcore-hours/ps.

Investigation of an enzymatic reaction: the case of human Isocitrate Dehydrogenase-1

The $NADP^+$-dependent IDH1 enzyme catalyzes the conversion of isocitrate (ICT) to α-ketoglutarate (α-KG) in the Krebs cycle[83] . The enzyme is a homodimer and it requires both the $NADP^+$ cofactor and a Mg(II) ion:[84,85]

$$\mathrm{ICT} + \mathrm{NADP}^+ \xrightarrow[\mathrm{IDH1}]{\mathrm{Mg}^{2+}} \alpha\mathrm{KG} + \mathrm{NADPH}$$

Mutations at the Arg132 position in the active site impart the ability to convert αKG to 2-hydroxyglutarate (2-HG),[86] an oncometabolite that promotes stemness in human cells and inhibits DNA demethylases.[87,88] Such variants are involved in the progress of low-grade glioma, glioblastoma, and acute myeloid leukaemia (AML).[89] Describing wild-type and variant IDH1 reaction mechanisms may help design transition-state analogs that act as selective inhibitors of mutant IDHI and are able to interfere with such diseases.

The reaction of wild-type IDH1 has been proposed to occur in a multi-step way.[90] The first step comprises two sub-steps, corresponding to the deprotonation of the C_α hydroxyl of ICT to Oxalosuccinate (OXS) initiated by a base and followed by reduction of $NADP^+$ to NADPH by accepting the C_α hydride of ICT (see figure 2C). Notably, the base has not yet been definitively identified. This step is followed by the loss of C_β carboxylate of OXS to give enolate, with the protonation of this enolate resulting in α-ketoglutarate. The X-ray structure of the protein in complex with ICT and $NADP^+$ (figure 2A) shows that each of the two monomers consists of a large domain, a small domain and a clasp domain. Two active sites include residues from both monomers[1], held together in the dimer by the clasp domain.[83] The α-carboxylate group of ICT forms a direct H-bond with Arg100 and Arg109. $Lys212^B$, Arg132 and Tyr139 interact with the β-carboxylate group of ICT through H-bonds. Thr77 (through a water molecule), Ser94 and the $NADP^+$ ribose interact with the γ-carboxylate of

[1]Here, residues from the second subunit are labelled by the superscript B, while those from the first subunit are left unmarked

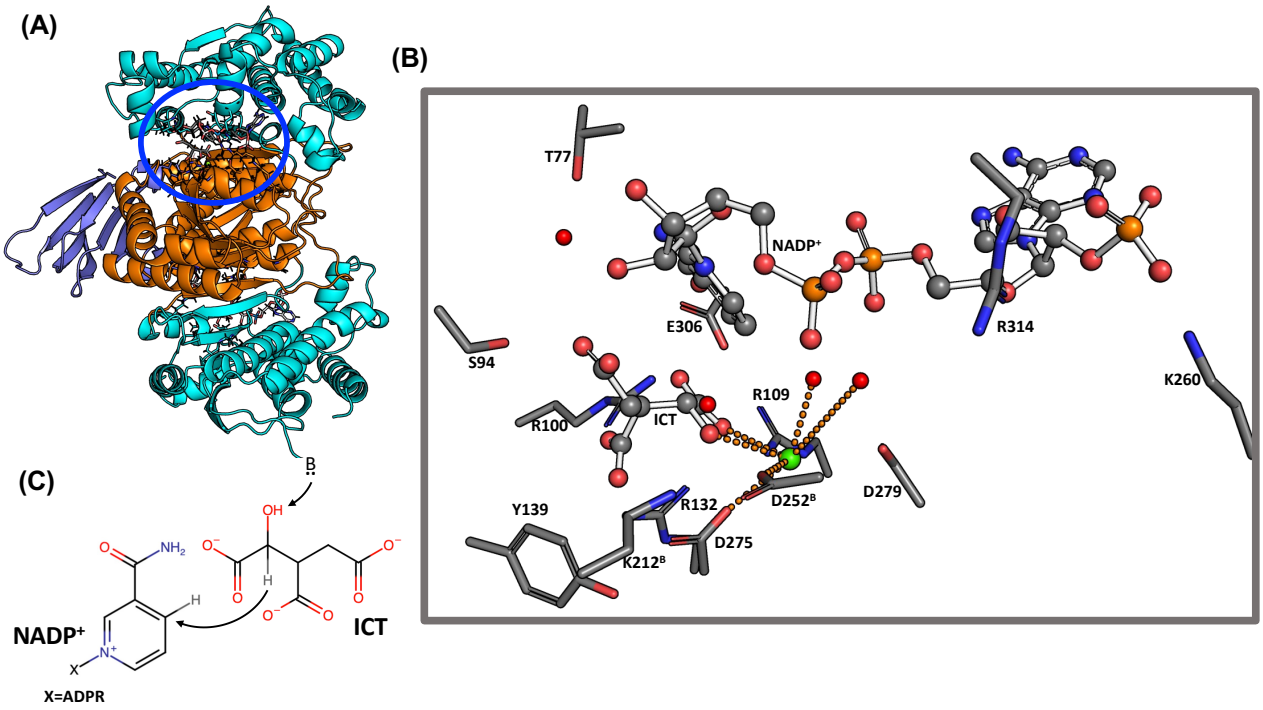


Figure 2: (A) Cartoon representation of the IDH1 enzyme with ICT and $NADP^+$. (B) Representation of the IDH1 active site from the X-ray structure. ICT and $NADP^+$ are shown in ball-and-sticks representation, while the protein residues are shown as sticks. Crystal waters are shown as red sphere. The Mg^{2+} ion (shown in green) coordination interactions are shown as orange dotted lines. (C) Proposed first step of the IDH1 enzymatic catalysis.[90]

ICT, while Glu306 forms an H-bond with the $NADP^+$ ring. These interactions anchors the $NADP^+$ nicotinamide ring close to ICT. The phosphate group of the ribose ring carrying the adenine moiety is held in the active site by interactions with Arg314 and Lys260. The Mg^{2+} ion coordination polyhedron consists of the α-carboxylate group of ICT, the α-alcohol of ICT, Asp275, $Asp252^B$, and two water molecules. A third water molecule forms an H-bond with $Asp252^B$ and with the α-alcohol of ICT. Because of this interaction, Hurley et al.[91] suggested that $Asp252^B$ is the base in the first step of the catalysis. Grodsky et al. proposed instead that this role was taken by Asp279, based on the finding that the activity of IDH1 with $Asp252^B$ mutated to Asn is similar to that of the wild-type.[92] Later, studies showing that IDH1 with $Lys212^B$ mutated to Arg, Gln, and Tyr exhibited lower activity, allowed for the suggestion that $Lys212^B$ in its deprotonated could be the key basic residue.[93] Classical MD

of the protein in which Lys212^{B} was either protonated or deprotonated, along with static QM/MM calculations corroborated this suggestion, showing that the activation free energy of the $NADP^+$ reduction step is larger when Asp279 is the initiator base (21.4 kcal/mol) than when deprotonated Lys212^{B} is the base (13.4 kcalmol) [2].[94] This step contributes significantly to the determination of the rate of reaction, and the latter pathway agrees fairly well with the experimentally observed k_{cat} value of $\approx$ 16 kcal/mol.[95]

Here, we apply classical and MiMiC-QM/MM MD to study the conversion of ICT to OXS, and the subsequent reduction of $NADP^+$ by IDH1 with Lys212^{B} in its protonated configuration [3].

Classical MD. Figure 1A shows the structure of the Michaelis complex of IDH1 with protonated Lys212^{B} as obtained from our simulations: the network of interaction involving the Mg^{2+} ion, ICT, $NADP^+$ and the protein residues in the active site are qualitatively very similar to the X-ray structure described previously. This includes the water molecule forming an H-bond with Asp252^{B} and the α-alcohol of ICT. A significant difference is that Arg100 has moved away from the α-carboxylate of ICT, establishing a water-mediated interaction, in agreement with similar observations by Neves et al.[94] Thr77, on the other hand, moves closer to and interacts directly with the γ-carboxylate of ICT compared to the crystal structure. Asp252^{B} is well positioned to abstract a proton from the C_α hydroxyl of ICT through an H-bonded water molecule (see figure 3A). This allows us to suggest that Asp252^{B} is a potential candidate base. Asp279, on the other hand, interacts with Mg^{2+} through one of the water molecules coordinating with the metal ion. This mediated interaction moves the residue farther away from the ICT alcohol. Thus, based on our model, we conclude that Asp279 is not a likely candidate for acting as a basis in the first step of the reaction.

MiMiC-QM/MM MD. The free energy associated with the conversion of ICT to OXS mediated by Asp252^{B}, followed by the reduction of the $NADP^+$ ring is investigated via

[2]These calculations used a two-layered ONIOM model at the B3LYP/6-31G(d) level of theory with entropic effects included via harmonic approximation

[3]Details of the simulation setups and additional analysis are reported in the online Supporting Information

thermodynamic integration.[96,97] The conversion of ICT to OXS mediated by the Asp252^{B}-water pair is described using as collective variable (CV) the difference $CV_1 = d_{O_h-H_h} - d_{H_h-O_W}$ between the distances of the proton from the two relevant oxygen atoms, while the reduction of $NADP^+$ to NADPH is described using the difference $CV_2 = d_{C_\alpha-H_\alpha} - d_{H_\alpha-C_N}$ (see figure 3A). The obtained free energy profiles along CV_1 and CV_2 are reported in figure 3B and C, where the insets show representative starting, transition state and final configurations. Simulations were performed at the BLYP level. A cumulative 39 ps of MD were performed, which were obtained in the span of one week (see Supporting Information for more details on the steps used in thermodynamic integration).

To investigate the nature of the bond breaking formation in the QM region during the course of the reaction (see figures 3D–F) we make use of the Wannier center analysis. In the reactant state ($CV_1 \approx -0.08$ Å), the C_α-O_h bond length is equal to 1.6 Å, with a Wannier center located at ≈ 1.0 Å from C_α, indicating a single bond character. At $CV_1 = 0$ Å, close to the putative transition state, the water molecule exists as a hydronium ion stabilized by Asp252^{B}. In this configuration, O_W interacts with H_h, while one of the hydrogen atoms bound to O_W interacts with the Asp252^{B} side chain. The Wannier center along the O_h-H_h bond is located farther away from H_h than in the reactant state by ≈ 0.2 Å, indicating an increasingly higher polar character of the bond and the transfer of a proton to O_W. This Wannier center is more closely associated with O_h, indicating a developing negative charge on it. In the final product ($CV_1 \approx 0.08$), the C_α-O_h bond length decreases to ≈ 1.3 Å, and the Wannier center along the bond is ≈ 0.8 Å away from C_α. Furthermore, Asp252^{B} is protonated and the ICT C_α hydroxyl group is deprotonated with a negatively charged O_h, due to the extra third Wannier center associated with it.

Starting from the product state of the first sub-step, we calculated the free energy change with increasing CV_2. In the reactant state ($CV_1 \approx -0.2$ Å), the Wannier center along the C_α-H_α bond is ≈ 0.7 Å from C_α and ≈ 3.5 Å from C_N of the $NADP^+$ ring. At the transition state ($CV_2 \approx 0$ Å), the hydride transfer of H_α to C_N takes place. The Wannier center along

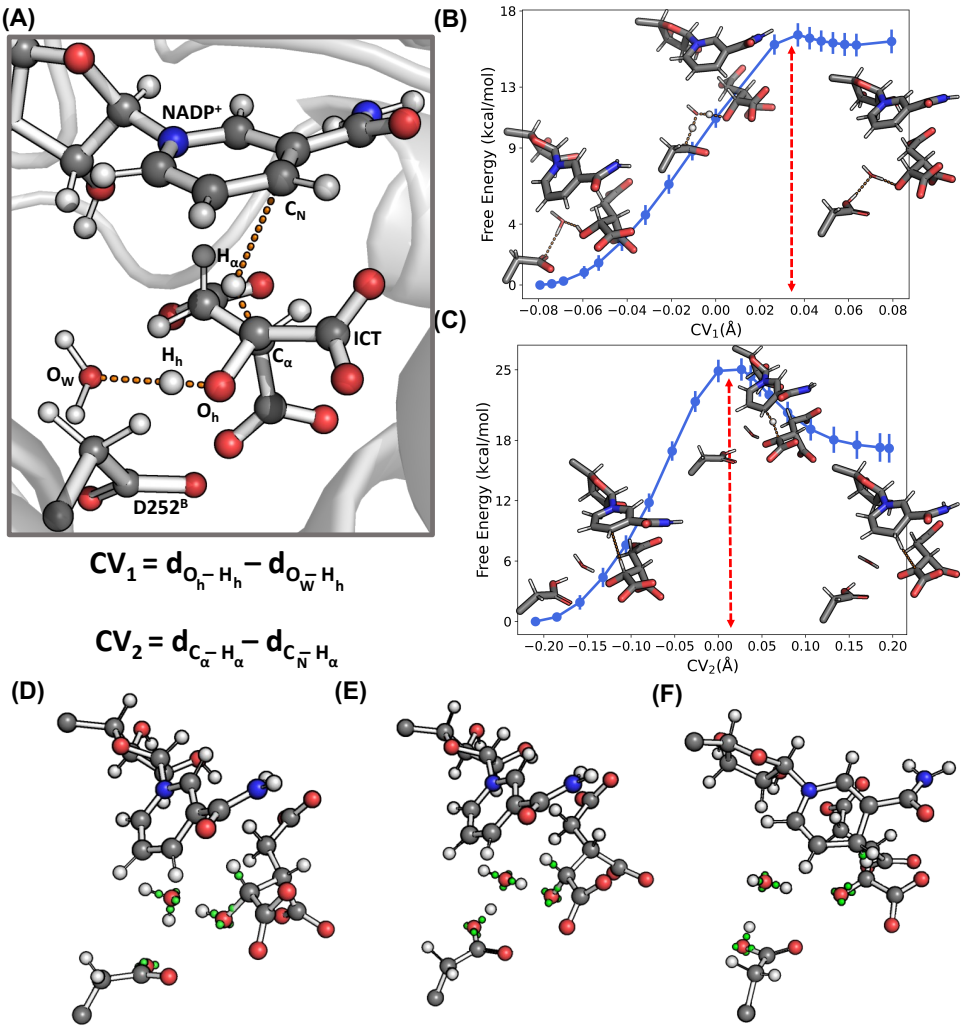


Figure 3: (A) Definition of the collective variables $CV_{1,2}$ used for thermodynamic integration. (B) Free energy of the ICT to OXS conversion with respect to CV_1. (C) Free energy of the $NADP^+$ reduction with respect to CV_2. Wannier centers depicted in green for select bonds at (D) $CV_1 \approx -0.08$ Å (E) $CV_1 \approx 0.08$ Å or $CV_2 = 0$ Å (F) $CV_2 \approx 0.2$ Å.

the C_α-H_α bond is now $\approx$ 1.3 Å from C_α and $\approx$ 1.5 Å from C_N. Furthermore, the third Wannier center associated with O_h from the product of the previous step, has now moved closer to C_α (from $\approx$ 1.5 Å to $\approx$ 1.1 Å) and more along the C_α-O_h bond. This, together with the fact that the C_α-O_h bond length reduces to 1.3 Å, indicates the emergence of a partial double bond character along the C_α-O_h bond. At the product ($CV_2 \approx 0.2$ Å), this extra Wannier center moves to $\approx$ 0.8 Å from C_α. This results in two Wannier centers along the C_α-O_h bond, and indicates the establishment of a full double bond, i.e., the formation of a

ketone. The Wannier center along the C_α-H_α bond moves $\approx$ 3.4 Å away from C_α, with this Wannier center falling along the newly formed C_N-H_α bond. The hydride transfer of H_α to the $NADP^+$ ring is complete.

Table 1: Free energies (in kcal/mol) associated with the first step of the IDH1 catalysis, for various base residues as initiators of the reaction. The Helmholtz free energy for $Asp252^B$ as base is from this work, while the Gibbs free energies for the pathways with $Lys212^B$ and Asp279 as base are from ref. [94]

	$Lys212^B$	Asp279	$Asp252^B$
Deprotonation of ICT	1.5	12.2	16.6 ($\pm$ 0.7)
Reduction of $NADP^+$	13.4	21.4	24.0 ($\pm$ 1.6)

The free energy barriers obtained for the two steps are $\approx$ 16 and $\approx$ 24 kcal/mol, respectively (Table 1). These values are not too dissimilar from those of the Asp279 pathway and both are significantly higher than the relevant barrier of 13.4 kcal/mol of the $Lys212^B$ pathway. Overall, our results thus support the conclusion of Ref. [94] indicating the deprotonated $Lys212^B$ as the residue that is more likely acting as the base in the first step of the catalytic process.

Investigation of drug/enzyme interactions: the case of 2g binding to p38α mitogen-activated protein kinase

An accurate description of ligand/enzyme interactions is mandatory to obtain quantitative insights that can guide drug screening and drug design. Indeed, biomolecular force fields-based estimates of the drug's residence time – a key parameter to assess a drug's efficacy [49–51] – show a large degree of variations, also depending on the enhanced sampling technique adopted. [48] Furthermore, static DFT QM/MM calculations directly suggested that limitations of current force fields, which cannot describe charge redistribution of the ligand during the unbinding processes, can contribute to this uncertainty. [47] DFT QM/MM *molecular dynamics* could be an excellent tool to include these variable charge distributions as such

effects are inherently incorporated in this first-principle MD and exascale computers, combined with the power of parallel programming, may help overcome the timescale limitation that currently hampers such applications. Since currently a fully QM/MM MD investigation of residence times is out of reach, as a first step towards this very ambitous goal, here we investigate for the first time substrate binding in a pharmacologically relevant enzyme by DFT QM/MM MD using MiMiC to analyse in details the most important interactions and the dynamics of the bound state. We focus on the p38α enzyme, a member of the mitogen-activated protein kinase (MAPK) family[98] in complex with the ligand **2g**. This is a serine/threonine kinase that controls cytokine biosynthesis, and it is involved in the initiation of chronic inflammation processes, development of cancer, heart disease, and many other diseases.[99–102] It adopts a typical kinase fold, including N-terminal lobe and C-terminal lobe that are connected via a hinge region (see Figure 4). The catalytic site of the protein is placed between the two lobes, where ATP molecules can bind. The binding of **2g** (Figure 4) is studied based on the X-ray structure complex with its close analogue **2a** (Figure 4, PDB code: 3FLN) [4].[103] The solvated **2g**/p38α complex was obtained by 500 ns-long MD followed by 100 ps-long QM/MM MD simulation at 300 K at the BLYP level.

In the X-ray structure, the phenoxy moiety forms hydrophobic interactions with the residues Ala51, Val38, Leu86, Leu104, Ile84, Thr106, and Leu167 (see Table 1 in Supporting Information). The pyrimidine N3 atom and the amino group interact instead with Met190 backbone unit (see Table 1 in Supporting Information). The rest of the molecule is solvent-exposed.

In the simulations the mode of binding of the ligand is the same (Figure 4 and Table 1 in Supporting Information). However, the pyridone oxygen atom interacts at times with Lys53 ($d(\mathrm{O_{2g}} \cdots \mathrm{H_{Lys53}}) \approx 3$ Å) because of a water-induced interruption of the Lys53-Glu71 salt bridge (Figure 4), and the tetrahydropyranyl oxygen atom forms a water mediated H-bond with Asp112.[5] This decisive role of water has also been observed in a recent MD study of

[4]The ligand names are adopted from Ref. 103. The IUPAC names are given in Supporting Information.
[5]These residues play no role for **2a** binding in the X-ray structure of the **2a**/p38α complex

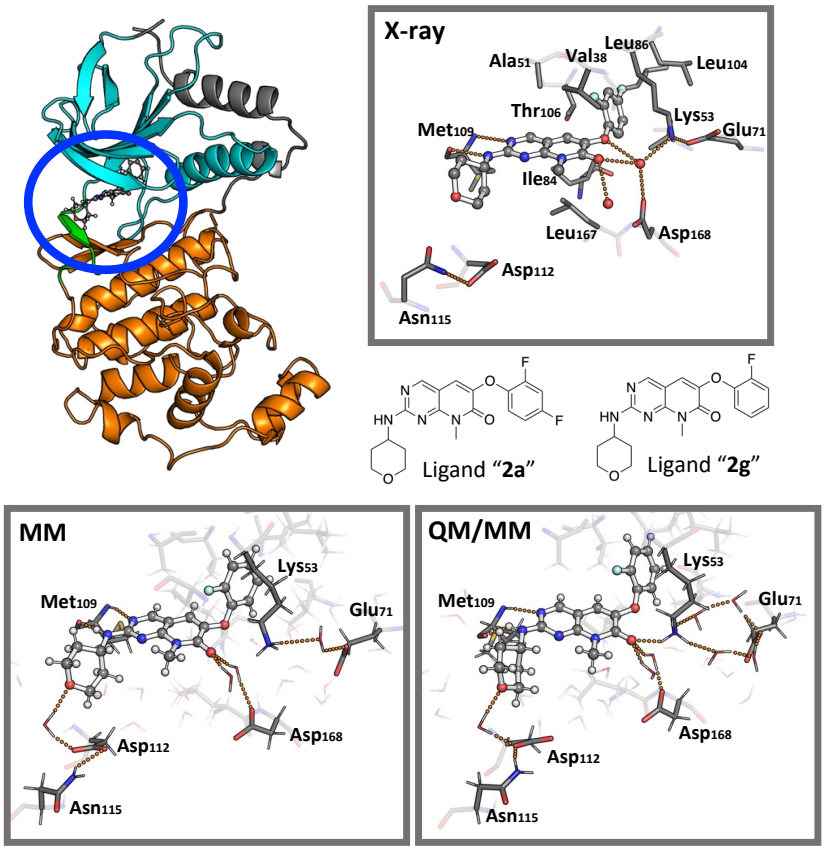


Figure 4: Comparison of X-ray, MM and MiMiC structures. Upper left panel: Cartoon representation of the p38α MAPK enzyme in complex with a ligand. The N-lobe (cyan) and the C-lobe (orange) of the enzyme are connected via a hinge region (green). The ligand binding pocket is located in between the lobes (blue ellipsis). Upper right panel: Representation of the enzyme binding pocket from the X-ray structure (PDB code: 3FLN).[103] The ligand **2a** is shown in ball-and-sticks representation, while the protein residues are shown as sticks. Crystal water molecules are shown as red sphere. H-bond interactions are shown as orange dotted lines. The structural formula of ligand **2a** (X-ray) and **2g** (MM and MiMiC simulations) are shown at the bottom of that panel. Lower left panel: Representative snapshot of the binding pocket from the classical MD simulation. Lower right panel: Representative snapshot of the binding pocket from MiMiC simulations.

the p38α MAPK enzyme in complex with different ligands.[104] In our MiMiC simulations, the first H-bond emerging from the MD simulations becomes persistent ($d(\mathrm{O}_{2\mathrm{g}} \cdots \mathrm{H}_{\mathrm{Lys53}} \approx$ 2.0$\pm$0.2 Å (see Figure 4), while the second is retained, although the water molecule mediating the interaction is exchanged within the solvent. The second coordinated water molecule retains its position during the whole simulation and mediates the interaction between the

pyridone oxygen atom and the Asp168 residue.

The effect of electronic polarization of the ligand is investigated here in terms of ligand's difference electronic density upon passing from vacuum to the enzyme-bound state.[47] The change in the atomic partial charges $\Delta Q(i)$ is then derived from integration around each atom (Figure 5).

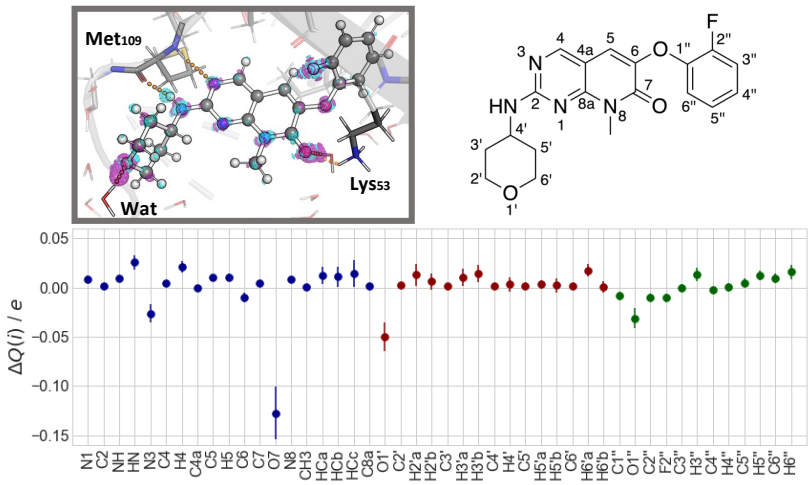


Figure 5: Electronic polarization analysis of the ligand upon passing from the vacuum to the enzyme-bound state. Top left panel: Difference density map for a representative QM/MM snapshot. The difference density is shown as isomesh with a contour level of 0.002 and -0.002 e Å^{-3} for increased (magenta) and decreased (cyan) electronic densities, respectively. Top right panel: Structural formula of ligand **2g** indicating the atom numbering. Bottom panel: Averaged change in the atomic partial charges $\Delta Q(i)$ for each ligand's atom. The data are grouped into the pyridopyrimidone (blue), tetrahydropyranyl (red) and phenoxy (green) moieties.

As expected, the polarization effects are more pronounced for the atomic species of the ligand which are involved in hydrogen bonds with the environment. The electronic density increased for the hydrogen bond acceptors of **2g**, while the hydrogen bond donor shows a decrease in the electronic density. On average, the total amount of redistributed charge within the ligand is 0.59 ± 0.05 e. The largest polarization effect is observed for the pyridone-O7 atom (-0.13 ± 0.03 e), possibly because of its strong interactions with the positively charged Lys53 residue and two water molecules. The tetrahydropyranyl-O1' atom shows a decrease of -0.05 ± 0.02 e from the interactions with a water molecule, while the hydrogen

bond interactions with the Met109 residue causes a charge shift of 0.03 ± 0.01 e and -0.03 ± 0.01 e for the amino HN and pyrimidine-N3 atoms, respectively.

To investigate the effect of the enzyme environment on the ligand dynamics, we calculate the ligand's infrared (IR) spectrum from the trajectory.[105] Comparison is made with the spectrum of the ligand in the gas phase from a normal mode analysis at the BLYP-D/def2-TZVP level of theory.[106–108] Since difference in the basis sets, plane waves and Gaussian-type orbitals, are known to have a minor impact on the harmonic frequencies,[109] the effect of ligand binding to the enzyme environment can be reasonably estimated from this comparison (Figure 6).

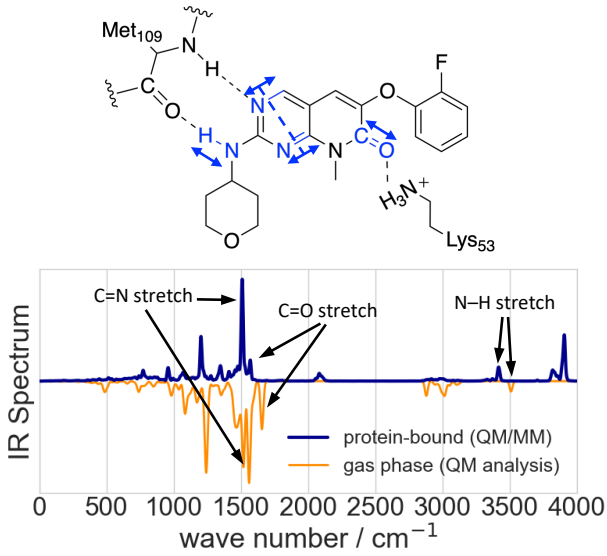


Figure 6: Vibrational analysis of the enzyme-bound ligand. Top panel: The scheme indicates important interactions of the ligand with the Met109 and Lys53 residues of p38α and vibrational modes of the ligand which are influenced by these interactions. Bottom panel: The IR spectrum of **2g** bound to the p38α MAPK enzyme is shown in blue. The spectrum was obtained from a 14.52 ps-long QM/MM MD simulation and the intensity normalized to its strongest absorption band. It is compared to a normalized spectrum of **2g** (orange), which is determined from normal mode analysis in gas phase.

The fundamental mode with the highest frequency is observed at $\approx$ 3500 cm^{-1}. It corresponds to the N–H stretching mode of the amino group of the ligand. A large shift of about -90 cm^{-1} is observed when passing from the vacuum to the enzyme-bound state,

because the strong hydrogen bond with the carbonyl oxygen atom of residue Met109 weakens the N−H bond. The symmetric C=N stretching vibration within the pyrimidine moiety of **2g** is the most intense IR band and appears at $\approx$ 1560 cm^{-1} in the gas phase spectrum. This band is shifted by about -50 cm^{-1} in the enzyme-bound state because of the H-bond between the Met109 residue and the pyrimidine-N3 atom. The latter exhibits more structural flexibility that results in a less pronounced shift (see Table 1 in Supporting Information). Finally, the C=O stretching vibration at $\approx$ 1655 cm^{-1} experiences a shift of about -85 cm^{-1} upon binding to the enzyme. This is caused by the interactions with water molecules and the Lys53 residue. Thus, Met109 and Lys53 play a key role in ligand dynamics.[6].

In conclusion, our QM/MM MD simulations properly describe the dynamic impact of the enzyme environment on the ligand's electronic structure and its internal dynamics - a prerequesite towards a balanced description of the unbinding process and, in turn, the accurate prediction of the ligand's residence time.

Conclusions and Outlook

The MiMiC framework was built with two main goals in mind: to be able to use different QM and MM codes with great ease and to scale as well as possible.[68,69] As showcased here and in previous work,[69] MiMiC scales up to thousands of standard CPU cores and allows running up to several ps/day in a single QM/MM MD run. In particular, the extreme scalability at the B3LYP level indicates viability for accurate description of enzymatic reactions when large computational resources are provided. Besides highlighting the efficient use of computational resources by the chosen QM layer (CPMD), these performances further demonstrate the effectiveness of a loose-coupling, multiple-program multiple-data paradigm for the development of extremely scalable first principle QM/MM interfaces.

As modern architectures make extensive use of heterogeneous nodes that combine mul-

[6]The simulation-based spectra calculations allows the observation of overtone and combination bands, which cannot be detected in the harmonic spectrum, as for example the spectral features at $\approx$ 3800 cm^{-1}.

ticore CPUs with GPUs,[55,110,111] achieving exascale will require coupling GPU-ready MM and QM software able to scale on many ($\approx 10^{2|3}$) such nodes. While a plethora of classical MD codes already exist that fully exploit GPUs,[112–115] including GROMACS[71] used in MiMiC, full implementation for these architectures is still an ongoing process for DFT codes,[61,63,65,66,116] except for the TeraChem proprietary software.[56,57] This is arguably the main reason why serious endeavours to port first principle QM/MM MD interfaces to GPUs are appearing only now in the literature.[117,118]

Strong scaling on heterogeneous nodes is actually the major challenge for molecular simulation. In force field based MD simulations, this is related to the relatively fixed size of the biological systems of interest[71] and the intrinsic seriality of the time evolution integration algorithms. Attempts to overcome these limitations have leveraged on statistical mechanics-based ensemble methods,[119] path sampling[120] and path-integral-like approaches,[121] often combined with machine learning (ML) techniques.[122] In DFT-based MD, only very recently scalability over thousands of GPUs has been achieved exploiting innovative linear scaling approaches and sparse algebra methods within an extended tight-binding scheme.[64] These observations indicate the necessity to develop innovative algorithms and statistical mechanics based methods beyond standard MD approaches as a route towards exascale DFT QM/MM MD, an idea already explored in the context of semiempirical QM/MM simulations.[113]

As a very flexible multiscale framework, MiMiC is an excellent candidate to bring DFT QM/MM MD simulations to the exascale by coupling codes running on GPUs and exploiting massively parallel free energy methods. Massively parallel, pharmacologically-oriented applications are envisaged in a not-too-far future.

Because of the cost associated with exascale calculations, we expect DFT QM/MM MD calculations to tremendously profit from the diffusion of ML techniques in molecular simulations.[123] Indeed, hybrid ML/MM models enable the simulation of biological systems using an ML representation of a quantum mechanical potential at near QM/MM accuracy and at a fraction of the computational cost.[44,124–127] These ML models work natively on GPUs,

and because they normally rely on local interactions alone, they can be exceptionally scalable on distributed architectures.[72,73] Furthermore, their training requires datasets generated through many single-point QM(/MM) calculations that are expensive but embarrassingly parallelizable. Finally, the recent introduction of ML-accelerated perturbative techniques provides an efficient and highly parallelizable way of recovering the accuracy of QM/MM potentials from simulations using cheaper methods (such as force fields or even ML/MM models) at the cost of only a few single-point energy and force QM/MM calculations.[46,128,129] These methods, in combination with enhanced sampling approaches,[130] promise to enable the QM/MM prediction of fundamental biophysical quantities such as drug-protein binding free energies or full free energy surfaces.

It is thus our hope that exascale DFT QM/MM MD simulations, combined with the power of ML approaches, will lead to a paradigm shift by bringing DFT-based QM/MM MD to the realm of drug discovery.

Acknowledgement

Discussions with Maria João Ramos are gratefully acknowledged. AR, BR, DM, MDV, PC thank the Helmholtz European Partnering program ("Innovative high-performance computing approaches for molecular neuromedicine") for funding. PC and MP thank the Human Brain Project (EU Horizon 2020) for funding. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS[80] at Jülich Supercomputing Centre (JSC). BR, EI, MDV, PC also acknowledge RWTH Aachen University for providing computer resources under project rwth0596. BR acknowledges useful discussions with Nitin Malapally.

Supporting Information Available

Computational details regarding the setup of the calculations; additional analyses of the classical MD simulations; QM/MM MD benchmark calculations for the **2g**/p38α system; X-ray, classical MD, and QM/MM MD structural data for the **2g**/p38α system.

References

(1) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chemical Reviews* **2015**, *115*, 6217–6263.

(2) Van Der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52*, 2708–2728.

(3) Hartsough, D. S.; Merz, K. M. J. Dynamic Force Field Models: Molecular Dynamics Simulations of Human Carbonic Anhydrase II Using a Quantum Mechanical/Molecular Mechanical Coupled Potential. *The Journal of Physical Chemistry* **1995**, *99*, 11266–11275.

(4) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie (International ed. in English)* **2009**, *48*, 1198–1229.

(5) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **1976**, *103*, 227–249.

(6) Friesner, R. A.; Guallar, V. AB INITIO QUANTUM CHEMICAL AND MIXED QUANTUM MECHANICS/MOLECULAR MECHANICS (QM/MM) METHODS

FOR STUDYING ENZYMATIC CATALYSIS. *Annual Review of Physical Chemistry* **2005**, *56*, 389–427.

(7) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and Free Energies of Enzymatic Reactions. *Chemical Reviews* **2006**, *106*, 3188–3209.

(8) Senn, H. M.; Thiel, W. QM/MM studies of enzymes. *Current Opinion in Chemical Biology* **2007**, *11*, 182–187.

(9) Ramos, M. J.; Fernandes, P. A. Computational enzymatic catalysis. *Accounts of Chemical Research* **2008**, *41*, 689–698.

(10) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728.

(11) Carvalho, A. T.; Barrozo, A.; Doron, D.; Kilshtain, A. V.; Major, D. T.; Kamerlin, S. C. L. Challenges in computational studies of enzyme structure, function and dynamics. *Journal of Molecular Graphics and Modelling* **2014**, *54*, 62–79.

(12) Merz, K. M. Using quantum mechanical approaches to study biological systems. *Accounts of Chemical Research* **2014**, *47*, 2804–2811.

(13) Das, C. K.; Nair, N. N. Elucidating the Molecular Basis of Avibactam-Mediated Inhibition of Class A β-Lactamases. *Chemistry – A European Journal* **2020**, *26*, 9639–9651.

(14) Prieß, M.; Göddeke, H.; Groenhof, G.; Schäfer, L. V. Molecular Mechanism of ATP Hydrolysis in an ABC Transporter. *ACS Central Science* **2018**, *4*, 1334–1343.

(15) Neves, R. P. P.; Fernandes, P. A.; Ramos, M. J. Role of Enzyme and Active Site Conformational Dynamics in the Catalysis by α-Amylase Explored with QM/MM Molecular Dynamics. *Journal of Chemical Information and Modeling* **2022**, *62*, 3638–3650.

(16) Streltsov, V. A. et al. Discovery of processive catalysis by an exo-hydrolase with a pocket-shaped active site. *Nature Communications* **2019**, *10*, 2222.

(17) Lukoyanov, D. A.; Yang, Z.-Y.; Dean, D. R.; Seefeldt, L. C.; Raugei, S.; Hoffman, B. M. Electron Redistribution within the Nitrogenase Active Site FeMo-Cofactor During Reductive Elimination of H_2 to Achieve N≡N Triple-Bond Activation. *Journal of the American Chemical Society* **2020**, *142*, 21679–21690.

(18) Calandrini, V.; Rossetti, G.; Arnesano, F.; Natile, G.; Carloni, P. Computational metallomics of the anticancer drug cisplatin. *Journal of Inorganic Biochemistry* **2015**, *153*, 231–238.

(19) Ho, M. H.; De Vivo, M.; Peraro, M. D.; Klein, M. L. Unraveling the catalytic pathway of metalloenzyme farnesyltransferase through QM/MM computation. *Journal of Chemical Theory and Computation* **2009**, *5*, 1657–1666.

(20) Wu, R.; Hu, P.; Wang, S.; Cao, Z.; Zhang, Y. Flexibility of catalytic zinc coordination in thermolysin and HDAC8: A Born-Oppenheimer ab initio QM/MM molecular dynamics study. *Journal of Chemical Theory and Computation* **2010**, *6*, 337–343.

(21) Chiariello, M. G.; Bolnykh, V.; Ippoliti, E.; Meloni, S.; Olsen, J. M. H.; Beck, T.; Rothlisberger, U.; Fahlke, C.; Carloni, P. Molecular Basis of CLC Antiporter Inhibition by Fluoride. *Journal of the American Chemical Society* **2020**, *142*, 7254–7258.

(22) Chiariello, M. G.; Alfonso-Prieto, M.; Ippoliti, E.; Fahlke, C.; Carloni, P. Mechanisms Underlying Proton Release in CLC-type F − /H + Antiporters. *The Journal of Physical Chemistry Letters* **2021**, *12*, 4415–4420.

(23) Paulikat, M.; Aranda, J.; Ippoliti, E.; Orozco, M.; Carloni, P. Proton Transfers to DNA in Native Electrospray Ionization Mass Spectrometry: A Quantum Mechanics/Molecular Mechanics Study. *The Journal of Physical Chemistry Letters* **2022**, *13*, 12004–12010.

(24) Tripathi, R.; Forbert, H.; Marx, D. Settling the Long-Standing Debate on the Proton Storage Site of the Prototype Light-Driven Proton Pump Bacteriorhodopsin. *The Journal of Physical Chemistry B* **2019**, *123*, 9598–9608.

(25) Valsson, O.; Campomanes, P.; Tavernelli, I.; Rothlisberger, U.; Filippi, C. Rhodopsin absorption from first principles: Bypassing common pitfalls. *Journal of Chemical Theory and Computation* **2013**, *9*, 2441–2454.

(26) Woiczikowski, P. B.; Steinbrecher, T.; Kubař, T.; Elstner, M. Nonadiabatic QM/MM simulations of fast charge transfer in Escherichia coli DNA photolyase. *Journal of Physical Chemistry B* **2011**, *115*, 9846–9863.

(27) Guglielmi, M.; Doemer, M.; Tavernelli, I.; Rothlisberger, U. Photodynamics of Lys+-Trp protein motifs: Hydrogen bonds ensure photostability. *Faraday Discussions* **2013**, *163*, 189–203.

(28) Murphy, A. R.; Hix, M. A.; Walker, A. Exploring the effects of mutagenesis on FusionRed using excited state QM/MM dynamics and classical force field simulations. *ChemBioChem* **2023**,

(29) Hitzenberger, M.; Schuster, D.; Hofer, T. S. The Binding Mode of the Sonic Hedgehog Inhibitor Robotnikinin, a Combined Docking and QM/MM MD Study. *Frontiers in Chemistry* **2017**, *5*.

(30) Spiegel, K.; Magistrato, A. Modeling anticancer drug–DNA interactions via mixed QM/MM molecular dynamics simulations. *Organic & Biomolecular Chemistry* **2006**, *4*, 2507–2517.

(31) Palermo, G.; Branduardi, D.; Masetti, M.; Lodola, A.; Mor, M.; Piomelli, D.; Cavalli, A.; De Vivo, M. Covalent Inhibitors of Fatty Acid Amide Hydrolase: A Rationale for the Activity of Piperidine and Piperazine Aryl Ureas. *Journal of Medicinal Chemistry* **2011**, *54*, 6612–6623.

(32) Callegari, D.; Ranaghan, K. E.; Woods, C. J.; Minari, R.; Tiseo, M.; Mor, M.; Mulholland, A. J.; Lodola, A. L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib. *Chemical Science* **2018**, *9*, 2740–2749.

(33) Zueva, I. V.; Lushchekina, S. V.; Pottie, I. R.; Darvesh, S.; Masson, P. 1-(3-Tert-Butylphenyl)-2,2,2-Trifluoroethanone as a Potent Transition-State Analogue Slow-Binding Inhibitor of Human Acetylcholinesterase: Kinetic, MD and QM/MM Studies. *Biomolecules 2020, Vol. 10, Page 1608* **2020**, *10*, 1608.

(34) Taguchi, M.; Oyama, R.; Kaneso, M.; Hayashi, S. Hybrid QM/MM Free-Energy Evaluation of Drug-Resistant Mutational Effect on the Binding of an Inhibitor Indinavir to HIV-1 Protease. *Journal of Chemical Information and Modeling* **2022**, *62*, 1328–1344.

(35) Cohen, S. M. A Bioinorganic Approach to Fragment-Based Drug Discovery Targeting Metalloenzymes. *Accounts of Chemical Research* **2017**, *50*, 2007–2016.

(36) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochemistry* **2011**, *50*, 10422–10430.

(37) Schwartz, S. D.; Schramm, V. L. Enzymatic transition states and dynamic motion in barrier crossing. *Nature Chemical Biology* **2009**, *5*, 551–558.

(38) Richard, J. P. Protein Flexibility and Stiffness Enable Efficient Enzymatic Catalysis. *Journal of the American Chemical Society* **2019**, *141*, 3320–3331.

(39) Cascella, M.; Micheletti, C.; Rothlisberger, U.; Carloni, P. Evolutionary conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *Journal of the American Chemical Society* **2005**, *127*, 3734–3742.

(40) Carnevale, V.; Raugei, S.; Micheletti, C.; Carloni, P. Convergent Dynamics in the

Protease Enzymatic Superfamily. *Journal of the American Chemical Society* **2006**, *128*, 9766–9772.

(41) Amaro, R. E.; Mulholland, A. J. Multiscale methods in drug design bridge chemical and biological complexity in the search for cures. *Nature Reviews Chemistry 2018 2:4* **2018**, *2*, 1–12.

(42) De Vivo, M. Bridging quantum mechanics and structure-based drug design. *Frontiers in Bioscience* **2011**, *16*, 1619.

(43) Evans, G. B.; Schramm, V. L.; Tyler, P. C. The transition to magic bullets – transition state analogue drug design. *MedChemComm* **2018**, *9*, 1983–1993.

(44) Rufa, D. A.; Bruce Macdonald, H. E.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D. Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning / molecular mechanics potentials. *bioRxiv* **2020**, *0*, 2020.07.29.227959.

(45) Hudson, P. S.; Aviat, F.; Meana-Pañeda, R.; Warrensford, L.; Pollard, B. C.; Prasad, S.; Jones, M. R.; Woodcock, H. L.; Brooks, B. R. Obtaining QM/MM binding free energies in the SAMPL8 drugs of abuse challenge: indirect approaches. *Journal of Computer-Aided Molecular Design* **2022**, *36*, 263–277.

(46) Rizzi, A.; Carloni, P.; Parrinello, M. Multimap targeted free energy estimation. **2023**,

(47) Capelli, R.; Lyu, W.; Bolnykh, V.; Meloni, S.; Olsen, J. M. H.; Rothlisberger, U.; Parrinello, M.; Carloni, P. Accuracy of Molecular Simulation-Based Predictions of k off Values: A Metadynamics Study. *The Journal of Physical Chemistry Letters* **2020**, *11*, 6373–6381.

(48) Ahmad, K.; Rizzi, A.; Capelli, R.; Mandelli, D.; Lyu, W.; Carloni, P. Enhanced-

Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. *Frontiers in Molecular Biosciences* **2022**, *9*, 1–17.

(49) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug–target residence time and its implications for lead optimization. *Nature Reviews Drug Discovery* **2006**, *5*, 730–739.

(50) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular determinants of drug–receptor binding kinetics. *Drug Discovery Today* **2013**, *18*, 667–673.

(51) Copeland, R. A. The drug–target residence time model: a 10-year retrospective. *Nature Reviews Drug Discovery* **2016**, *15*, 87–95.

(52) Vennelakanti, V.; Nazemi, A.; Mehmood, R.; Steeves, A. H.; Kulik, H. J. Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins. 2022.

(53) Casalino, L.; Nierzwicki, Ł.; Jinek, M.; Palermo, G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by ¡i¿Ab Initio¡/i¿ Molecular Dynamics. *ACS Catalysis* **2020**, *10*, 13596–13605.

(54) Nierzwicki, Ł.; East, K. W.; Binz, J. M.; Hsu, R. V.; Ahsan, M.; Arantes, P. R.; Skeens, E.; Pacesa, M.; Jinek, M.; Lisi, G. P.; Palermo, G. Principles of target DNA cleavage and the role of Mg2+ in the catalysis of CRISPR–Cas9. *Nature Catalysis* **2022**, *5*, 912–922.

(55) Ariana Remmel Welcome to exascale computing. *C&EN Global Enterprise* **2022**, *100*, 29–33.

(56) Ufimtsev, I. S.; Martínez, T. J. Graphical Processing Units for Quantum Chemistry. *Computing in Science & Engineering* **2008**, *10*, 26–34.

(57) Snyder, J. W.; Fales, B. S.; Hohenstein, E. G.; Levine, B. G.; Martínez, T. J. A direct-compatible formulation of the coupled perturbed complete active space self-consistent

field equations on graphical processing units. *The Journal of Chemical Physics* **2017**, *146*, 174113.

(58) Gonze, X. et al. Recent developments in the ABINIT software package. *Computer Physics Communications* **2016**, *205*, 106–131.

(59) Kalinowski, J.; Wennmohs, F.; Neese, F. Arbitrary Angular Momentum Electron Repulsion Integrals with Graphical Processing Units: Application to the Resolution of Identity Hartree–Fock Method. *Journal of Chemical Theory and Computation* **2017**, *13*, 3160–3170.

(60) García, A. et al. Siesta : Recent developments and applications. *The Journal of Chemical Physics* **2020**, *152*, 204108.

(61) Kowalski, K. et al. From NWChem to NWChemEx: Evolving with the Computational Chemistry Landscape. *Chemical Reviews* **2021**, *121*, 4962–4998.

(62) Barca, G. M. J. et al. Recent developments in the general atomic and molecular electronic structure system. *The Journal of Chemical Physics* **2020**, *152*, 154102.

(63) Kühne, T. D. et al. CP2K: An electronic structure and molecular dynamics software package -Quickstep: Efficient and accurate electronic structure calculations. *Journal of Chemical Physics* **2020**, *152*.

(64) Schade, R.; Kenter, T.; Elgabarty, H.; Lass, M.; Schütt, O.; Lazzaro, A.; Pabst, H.; Mohr, S.; Hutter, J.; Kühne, T. D.; Plessl, C. Towards electronic structure-based ab-initio molecular dynamics simulations with hundreds of millions of atoms. *Parallel Computing* **2022**, *111*, 102920.

(65) Giannozzi, P.; Baseggio, O.; Bonfà, P.; Brunato, D.; Car, R.; Carnimeo, I.; Cavazzoni, C.; de Gironcoli, S.; Delugas, P.; Ferrari Ruffino, F.; Ferretti, A.; Marzari, N.;

Timrov, I.; Urru, A.; Baroni, S. Quantum ESPRESSO toward the exascale. *The Journal of Chemical Physics* **2020**, *152*, 154105.

(66) Manathunga, M.; Shajan, A.; Giese, T. J.; Cruzeiro, V. W. D.; Smith, J.; Miao, Y.; He, X.; Ayers, K.; Brothers, E.; Götz, A. W.; Merz, K. M. QUICK-22.03, University of California San Diego, CA and Michigan State University, East Lansing, MI, 2022.

(67) `http://www.mimic-project.org/`.

(68) Olsen, J. M. H.; Bolnykh, V.; Meloni, S.; Ippoliti, E.; Bircher, M. P.; Carloni, P.; Rothlisberger, U. MiMiC: A Novel Framework for Multiscale Modeling in Computational Chemistry. *Journal of Chemical Theory and Computation* **2019**, *15*, 3810–3823.

(69) Bolnykh, V.; Olsen, J. M. H.; Meloni, S.; Bircher, M. P.; Ippoliti, E.; Carloni, P.; Rothlisberger, U. Extreme Scalability of DFT-Based QM/MM MD Simulations Using MiMiC. *Journal of Chemical Theory and Computation* **2019**, *15*, 5601–5613.

(70) Hutter, J.; Alavi, A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. CPMD, Copyright IBM Corp 1990-2022, Copyright MPI für Festkörperforschung Stuttgart 1997-2001. `http://www.cpmd.org/`.

(71) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. Solving Software Challenges for Exascale. Cham, 2015; pp 3–27.

(72) Lu, D.; Wang, H.; Chen, M.; Lin, L.; Car, R.; Weinan, E.; Jia, W.; Zhang, L. 86 PFLOPS Deep Potential Molecular Dynamics simulation of 100 million atoms with ab initio accuracy. *Computer Physics Communications* **2021**, *259*, 107624.

(73) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications* **2023**, *14*, 579.

(74) Olsen, J. M. H.; Bolnykh, V.; Meloni, S.; Ippoliti, E.; Carloni, P.; Rothlisberger, U. MiMiC: A Framework for Multiscale Modeling in Computational Chemistry (v0.2.0). 2022.

(75) Bolnykh, V.; Olsen, J. M. H.; Meloni, S.; Ippoliti, E.; Carloni, P.; Rothlisberger, U. MiMiC Communication Library (v2.0.1). 2022.

(76) Raghavan, B.; Schackert, F. K.; Levy, A.; Johnson, S. K.; Ippoliti, E.; Mandelli, D.; Olsen, J. M. H.; Rothlisberger, U.; Carloni, P. MiMiCPy: An Efficient Toolkit for MiMiC-Based QM/MM Simulations. *Journal of Chemical Information and Modeling* **2023**,

(77) Laio, A.; VandeVondele, J.; Rothlisberger, U. A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *The Journal of Chemical Physics* **2002**, *116*, 6941.

(78) Hutter, J.; Alavi, A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. CPMD 4.3. `https://github.com/CPMD-code/CPMD`.

(79) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(80) Alvarez, D. JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre. *Journal of large-scale research facilities JLSRF* **2021**, *7*.

(81) Becke, A. D. Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *The Journal of Chemical Physics* **1992**, *96*, 2155–2160.

(82) Becke, A. D. Density-functional thermochemistry. II. The effect of the Perdew–Wang

generalized-gradient correlation correction. *The Journal of Chemical Physics* **1992**, *97*, 9173–9177.

(83) Xu, X.; Zhao, J.; Xu, Z.; Peng, B.; Huang, Q.; Arnold, E.; Ding, J. Structures of human cytosolic NADP-dependent isocitrate dehydrogenase reveal a novel self-regulatory mechanism of activity. *The Journal of biological chemistry* **2004**, *279*, 33946–33957.

(84) Gonçalves, S.; Miller, S. P.; Carrondo, M. A.; Dean, A. M.; Matias, P. M. Induced fit and the catalytic mechanism of isocitrate dehydrogenase. *Biochemistry* **2012**, *51*, 7098–7115, Paper for PDB ID: 4AJ3.

(85) Quartararo, C. E.; Hazra, S.; Hadi, T.; Blanchard, J. S. Structural, kinetic and chemical mechanism of isocitrate dehydrogenase-1 from mycobacterium tuberculosis. *Biochemistry* **2013**, *52*, 1765–1775.

(86) Dang, L. et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **2009**, *462*, 739–744.

(87) Wu, H.; Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **2014**, *156*, 45–68.

(88) Guo, J. U.; Su, Y.; Zhong, C.; Ming, G. L.; Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **2011**, *145*, 423–434.

(89) Horbinski, C. What do we know about IDH1/2 mutations so far, and how do we use it? *Acta neuropathologica* **2013**, *125*, 621–636.

(90) Bolduc, J. M.; Dyer, D. H.; Scott, W. G.; Singer, P.; Sweet, R. M.; Koshland, D. E.; Stoddard, B. L. Mutagenesis and laue structures of enzyme intermediates: Isocitrate dehydrogenase. *Science* **1995**, *268*, 1312–1318.

(91) Hurley, J. H.; Dean, A. M.; Koshland, D. E.; Stroud, R. M. Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and NADP+ complexes. *Biochemistry* **1991**, *30*, 8671–8678.

(92) Grodsky, N. B.; Soundar, S.; Colman, R. F. Evaluation by site-directed mutagenesis of aspartic acid residues in the metal site of pig heart NADP-dependent isocitrate dehydrogenase. *Biochemistry* **2000**, *39*, 2193–2200.

(93) Kim, T. K.; Lee, P.; Colman, R. F. Critical role of Lys212 and Tyr140 in porcine NADP-dependent isocitrate dehydrogenase. *The Journal of biological chemistry* **2003**, *278*, 49323–49331.

(94) Neves, R. P.; Fernandes, P. A.; Ramos, M. J. Unveiling the Catalytic Mechanism of NADP+-Dependent Isocitrate Dehydrogenase with QM/MM Calculations. *ACS Catalysis* **2016**, *6*, 357–368.

(95) Rendina, A. R.; Pietrak, B.; Smallwood, A.; Zhao, H.; Qi, H.; Quinn, C.; Adams, N. D.; Concha, N.; Duraiswami, C.; Thrall, S. H.; Sweitzer, S.; Schwartz, B. Mutant IDH1 Enhances the Production of 2-Hydroxyglutarate Due to Its Kinetic Mechanism. *Biochemistry* **2013**, *52*, 4563–4577.

(96) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chemical Physics Letters* **1989**, *156*, 472–477.

(97) Carloni, P.; Sprik, M.; Andreoni, W. Key Steps of the cis-Platin-DNA Interaction: Density Functional Theory-Based Molecular Dynamics Simulations. *Journal of Physical Chemistry B* **2000**, *104*, 823–835.

(98) Wang, Z.; Harkins, P. C.; Ulevitch, R. J.; Han, J.; Cobb, M. H.; Goldsmith, E. J. The structure of mitogen-activated protein kinase p38 at 2.1Åresolution. *Proceedings of the National Academy of Sciences* **1997**, *94*, 2327–2332.

(99) Han, J.; Lee, J.-D.; Bibbs, L.; Ulevitch, R. J. A MAP Kinase Targeted by Endotoxin and Hyperosmolarity in Mammalian Cells. *Science* **1994**, *265*, 808–811.

(100) Lee, J. C. et al. A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature* **1994**, *372*, 739–746.

(101) Rouse, J.; Cohen, P.; Trigon, S.; Morange, M.; Alonso-Llamazares, A.; Zamanillo, D.; Hunt, T.; Nebreda, A. R. A novel kinase cascade triggered by stress and heat shock that stimulates MAPKAP kinase-2 and phosphorylation of the small heat shock proteins. *Cell* **1994**, *78*, 1027–1037.

(102) Cuenda, A.; Rousseau, S. p38 MAP-Kinases pathway regulation, function and role in human diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2007**, *1773*, 1358–1375.

(103) Goldstein, D. M. et al. Discovery of 6-(2,4-Difluorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl)propylamino]-8-methyl-8 *H* -pyrido[2,3-*d*]pyrimidin-7-one (Pamapimod) and 6-(2,4-Difluorophenoxy)-8-methyl-2-(tetrahydro-2 *H* -pyran-4-ylamino)pyrido[2,3- *d*]pyrimidin-7(8 *H*)-one (R1487) as Orally Bioavailable and Highly Selective Inhibitors of p38α Mitogen-Activated Protein Kinase. *Journal of Medicinal Chemistry* **2011**, *54*, 2255–2265.

(104) Pantsar, T.; Kaiser, P. D.; Kudolo, M.; Forster, M.; Rothbauer, U.; Laufer, S. A. Decisive role of water and protein dynamics in residence time of p38α MAP kinase inhibitors. *Nature Communications* **2022**, *13*, 569.

(105) Brehm, M.; Thomas, M.; Gehrke, S.; Kirchner, B. TRAVIS—A free analyzer for trajectories from molecular simulation. *The Journal of Chemical Physics* **2020**, *152*, 164105.

(106) Becke, A. D. Density-functional thermochemistry. II. The effect of the Perdew–Wang

generalized-gradient correlation correction. *The Journal of Chemical Physics* **1998**, *97*, 9173.

(107) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010**, *132*.

(108) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297.

(109) Tosoni, S.; Tuma, C.; Sauer, J.; Civalleri, B.; Ugliengo, P. A comparison between plane wave and Gaussian-type orbital basis sets for hydrogen bonded systems: Formic acid as a test case. *The Journal of Chemical Physics* **2007**, *127*, 154102.

(110) Schneider, D. The Exascale Era is Upon Us: The Frontier supercomputer may be the first to reach 1,000,000,000,000,000,000 operations per second. *IEEE Spectrum* **2022**, *59*, 34–35.

(111) Q. Choi, C. The Beating Heart of the World's First Exascale Supercomputer. *IEEE Spectrum* **2022**,

(112) Brooks, B. R. et al. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009**, *30*, 1545–1614.

(113) Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics* **2020**, *153*, 044130.

(114) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* **2012**, *8*, 1542–1555.

(115) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, *9*, 3878–3888.

(116) Gavini, V. et al. Roadmap on Electronic Structure Codes in the Exascale Era. **2022**,

(117) Manathunga, M.; Aktulga, H. M.; Götz, A. W.; Merz, K. M. Quantum Mechanics/Molecular Mechanics Simulations on NVIDIA and AMD Graphics Processing Units. *Journal of Chemical Information and Modeling* **2023**, *63*, 711–717.

(118) Cruzeiro, V. W. D.; Wang, Y.; Pieri, E.; Hohenstein, E. G.; Martínez, T. J. TeraChem protocol buffers (TCPB): Accelerating QM and QM/MM simulations with a client–server model. *The Journal of Chemical Physics* **2023**, *158*, 044801.

(119) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, *314*, 141–151.

(120) Pronk, S.; Pouya, I.; Lundborg, M.; Rotskoff, G.; Wesén, B.; Kasson, P. M.; Lindahl, E. Molecular Simulation Workflows as Parallel Algorithms: The Execution Engine of Copernicus, a Distributed High-Performance Computing Platform. *Journal of Chemical Theory and Computation* **2015**, *11*, 2600–2608.

(121) Mandelli, D.; Hirshberg, B.; Parrinello, M. Metadynamics of Paths. *Physical Review Letters* **2020**, *125*, 026001.

(122) Martin, W.; Sheynkman, G.; Lightstone, F. C.; Nussinov, R.; Cheng, F. Interpretable artificial intelligence and exascale molecular dynamics simulations to reveal kinetics: Applications to Alzheimer's disease. *Current Opinion in Structural Biology* **2022**, *72*, 103–113.

(123) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine learning for molecular simulation. *Annual review of physical chemistry* **2020**, *71*, 361–390.

(124) Lahey, S.-L. J.; Rowley, C. N. Simulating protein–ligand binding with neural network potentials. *Chemical Science* **2020**, *11*, 2362–2368.

(125) Pan, X.; Yang, J.; Van, R.; Epifanovsky, E.; Ho, J.; Huang, J.; Pu, J.; Mei, Y.; Nam, K.; Shao, Y. Machine-learning-assisted free energy simulation of solution-phase and enzyme reactions. *Journal of Chemical Theory and Computation* **2021**, *17*, 5745–5758.

(126) Xu, M.; Zhu, T.; Zhang, J. Z. Automatically constructed neural network potentials for molecular dynamics simulation of zinc proteins. *Frontiers in Chemistry* **2021**, *9*, 692200.

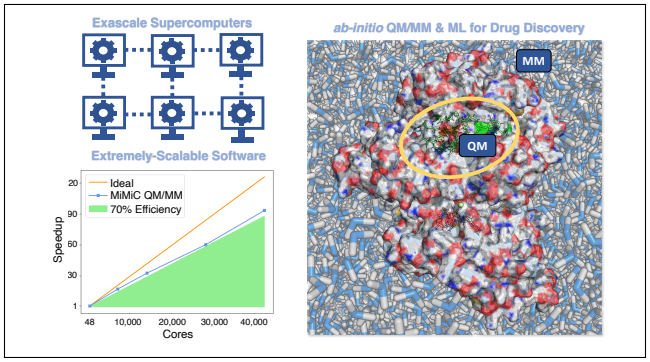
(127) Galvelis, R.; Varela-Rial, A.; Doerr, S.; Fino, R.; Eastman, P.; Markland, T. E.; Chodera, J. D.; De Fabritiis, G. NNP/MM: Fast molecular dynamics simulations with machine learning potentials and molecular mechanics. *arXiv preprint arXiv:2201.08110* **2022**,

(128) Wirnsberger, P.; Ballard, A. J.; Papamakarios, G.; Abercrombie, S.; Racanière, S.; Pritzel, A.; Jimenez Rezende, D.; Blundell, C. Targeted free energy estimation via learned mappings. *The Journal of Chemical Physics* **2020**, *153*, 144112.

(129) Rizzi, A.; Carloni, P.; Parrinello, M. Targeted Free Energy Perturbation Revisited: Accurate Free Energies from Mapped Reference Potentials. *The Journal of Physical Chemistry Letters* **2021**, *12*, 9449–9454.

(130) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living J. Comp. Mol. Sci.* **2022**, *4*, 1583.

TOC Graphic


Exascale Supercomputers
Extremely-Scalable Software
Ideal
MiMiC QM/MM
70% Efficiency
Speedup
Cores
48
10,000
20,000
30,000
40,000
ab-initio QM/MM & ML for Drug Discovery
MM
QM