# How Reproducible are QM/MM Simulations? Lessons from Computational Studies of the Covalent Inhibition of the SARS-CoV-2 Main Protease by Carmofur

Goran Giudetti[a], Igor Polyakov[b], Bella L. Grigorenko[b], Shirin Faraji[c], Alexander V. Nemukhin[b], and Anna I. Krylov[a,*]

[a] *Department of Chemistry, University of Southern California, Los Angeles, California, 90089-0482, United States*

[b] *Department of Chemistry, Lomonosov Moscow State University, Moscow, 119991, Russia*

[c] *Zernike Institute for Advanced Materials, University of Groningen, Groningen, 9747 AG, The Netherlands and*

* *Corresponding author: krylov@usc.edu*

This work explores what level of transparency in reporting the details is required for practical reproducibility of quantum mechanics/molecular mechanics (QM/MM) simulations. Using the reaction of an essential SARS-CoV-2 enzyme (the main protease) with a covalent inhibitor (carmofur) as a test case of chemical reactions in biomolecules, we carried out QM/MM calculations to determine the structures and energies of the reactants, the product, and the transition state/intermediate using analogous QM/MM models implemented in two software packages, NWChem and Q-Chem. Our main benchmarking goal was to reproduce the key energetics computed with the two packages. Our results indicate that quantitative agreement (within the numerical thresholds used in calculations) is difficult to achieve. We show that rather minor details of QM/MM simulations must be reported in order to ensure the reproducibility of the results and offer suggestions towards developing practical guidelines for the reporting results of biosimulations.

## I. INTRODUCTION

Computer simulations of structures and properties of biomolecules are now routinely used to aid biomedical studies, including characterization of prospective drug candidates and their interaction with pathogenes' enzymes. Among various simulation tools, quantum

mechanics/molecular mechanics (QM/MM) approaches[1,2] play an important role because they are able to describe making and breaking of chemical bonds. Computational search for efficient covalent inhibitors, which operate by binding covalently to the protein, relies on QM/MM as an essential tool.

The COVID-19 pandemic prompted massive research efforts to reveal the mechanisms of the action of SARS-CoV-2 enzymes at a molecular level, with an ultimate goal of designing drugs to fight the disease[3]. In the past two years, numerous computational papers describing various non-covalent and covalent inhibitors, which potentially can inactivate these enzymes were published[4–14].

Given the urgency and the significance of the subject, the question of reliability and reproducibility of the results of these and future simulations is of an utmost importance. Reproducibility of computational modeling of biological systems is not trivial because of the complexity of underlying theoretical models, of the computational protocols implementing these models, and of software stacks executing these protocols[15]. The standards of reporting the details of calculations developed for electronic structure calculations[16] are simply not sufficient in this context.

The question of reproducibility of research results is of course much broader than molecular simulations. A recent study[17] investigated the reproducibility of the computational results from a random sample of computational papers published in *Science* since 2011. The authors were able to reproduce the findings of only 44% of the studies and attributed the difficulties to a variety of problems, ranging from authors' desire to protect their data or software to the lack of standards and mechanisms for depositing digital artifacts, as well as the complexity of the data and protocols. Given that biomolecular simulations are much more complex than an average computational study (in terms of the protocols, the sheer size of the data and the codes), the problems of reproducibility are likely to be more severe, and would not be easily addressed by the proposed policies[18].

Well-justified by the urgency of the situation, a rapid pace of publications reporting computational studies related to COVID-19 calls for careful assessment of various aspects of QM/MM simulations. We do not imply that the software has not been properly tested or that the algorithms are not reliable; rather we point out that the results depend on numerous parameters hidden in the simulation setups.

The pitfalls in biomolecular QM/MM simulations are well-known: QM/MM is not a

black-box tool, the computational protocols are not standardized, the software is constantly evolving, and computational workflows are not fully automated[15]. The reproducibility of the results of QM/MM simulations is further hindered by the omission of details—often perceived by authors as minor, irrelevant, or trivial—in the published papers.

In this contribution, we explore what level of transparency in reporting the details is required for practical reproducibility of QM/MM simulations, with an aim to provide a guide, in the spirit of IUPAC guidelines[16], for future studies. We use the reaction of an essential SARS-CoV-2 enzyme, the main protease ($M^{Pro}$)[19], with a covalent inhibitor, carmofur,[20] as a test case of chemical reactions in biomolecules. The main protease $M^{Pro}$ is a cysteine protease; in SARS-CoV-2, it catalyzes the cleavage of the polyprotein of the virus into the working proteins, which is the key step in the virus replication in human cells[21]. Carmofur is a certified drug for other diseases[22]. Its presumed inhibitory activity towards $M^{Pro}$ is shown in Fig. 1. The deprotonated side chain of the cysteine residue Cys145 of $M^{Pro}$ , which is formed upon proton transfer to His41, reacts with the electrophilic carbon atom of the carmofur tail attached to the fluoro-uracil warhead. This results in a covalently bound adduct, thus blocking the function of the enzyme.

Computational characterization of such a reaction entails calculations of the energy profile along the reaction coordinate. These calculations can confirm (or dispute) proposed mechanisms and provide an insight into elementary steps involved. The comparison of the reaction energy profiles computed for different target molecules can then be used to evaluate their relative effectiveness in deactivating the enzyme. In this work, we carried out QM/MM calculations to determine the structures and energies of the reactants, the product, and the transition state/intermediate for the reaction of carmofur with $M^{Pro}$.

A reaction of $M^{Pro}$ with another covalent inhibitor (called N3) was recently investigated computationally by two expert groups[9,12] using QM/MM and molecular dynamics simulations with QM/MM potentials. Although the two papers report reaction energies within 3 kcal/mol from each other, the difference in the reaction barriers are much larger, up to 10 kcal/mol, illustrating an extent by which the results can be affected by different QM/MM-based schemes.

We employed two different software packages, NWChem[23] and Q-Chem[24,25], in order to utilize various technical innovations available in them. Using analogous QM/MM models within the two packages, we constructed segments of the potential energy profile for the

reaction in the enzyme's active site. Our main benchmarking goal was to reproduce the key energetics computed with the two packages—in the ideal case of perfectly reproducible protocols, the structures of the key stationary points along the reaction profile and the respective energetics computed by the two softwares should be identical, within small error bars consistent with numerical thresholds used in the calculations. Our results indicate that such agreement is difficult to achieve.

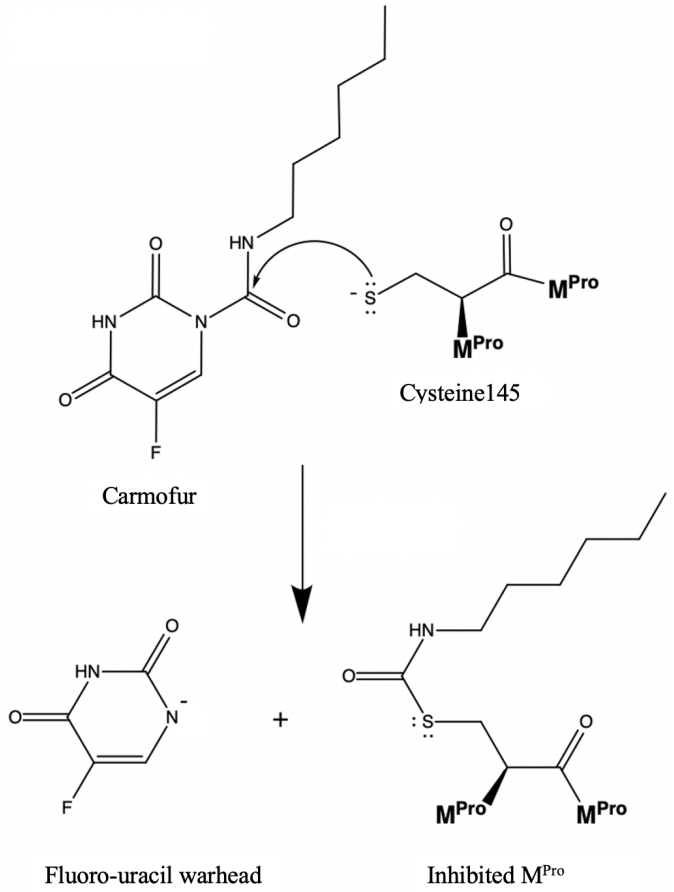


FIG. 1: The key step of the reaction between carmofur and $M^{Pro}$ —the nucleophilic attack of the thiolate of the catalytic cysteine.

QM/MM is a versatile approach[1,2] for multi-scale modeling, suitable for simulating chemical processes in complex environments, such as solutions, solids, interfaces, or proteins. The key idea is to partition the system into the important part (i.e., a subsystem where the chemical reaction occurs), to be treated quantum mechanically, and the environment, to be treated by less demanding methods, e.g., by classical forcefields. However, there is no unique recipe for how to break the system into the QM and MM parts, how to treat them, and

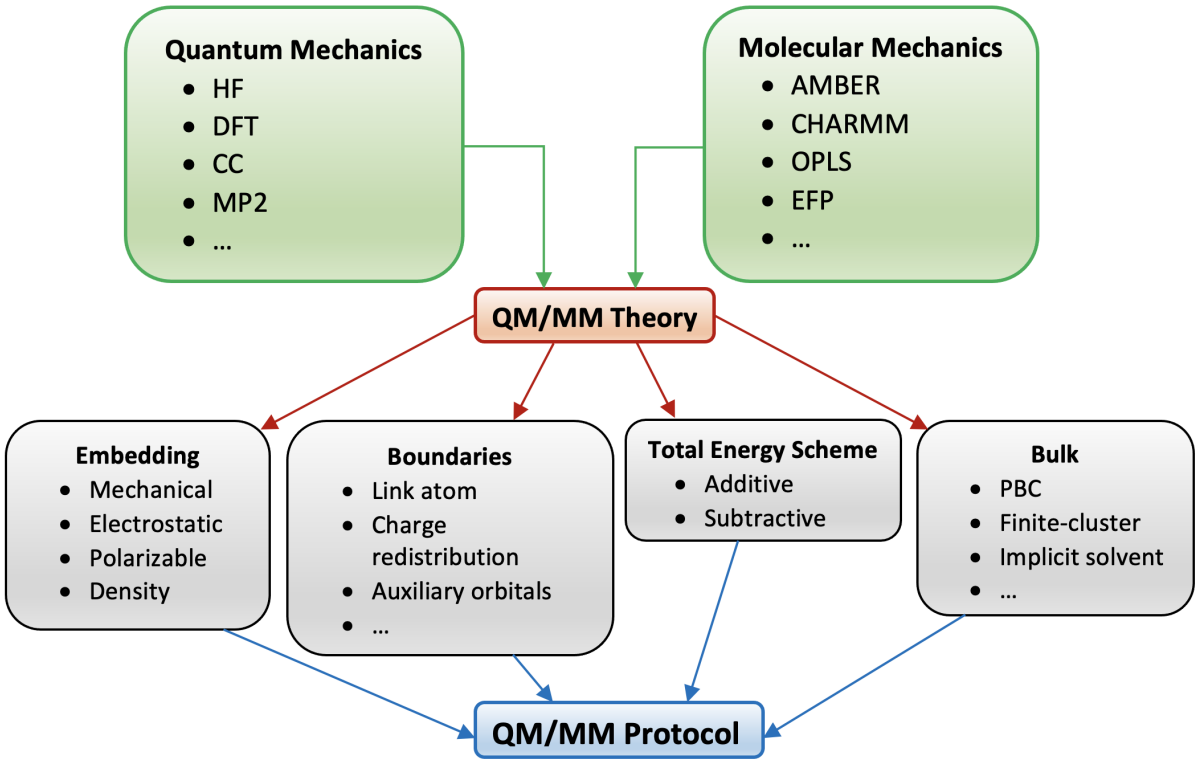


FIG. 2: Various components defining a QM/MM protocol.

how to describe their interaction. Even at a high level of generalization, as depicted in Fig. 2, QM/MM theory comprises multiple models and techniques. The ensuing computational protocols are far from being black box and tend to be system specific. To complicate the matters further, different software implementation of QM/MM models may lead to differences in computed properties obtained with seemingly identical protocols.

Once the QM and MM parts are defined, one needs to specify how to treat them (i.e., which level of theory for the QM and MM parts), their boundary (i.e., what to do with the broken bonds) and how to describe their interaction (i.e., embedding type). In this work, we describe the QM part by the density functional theory (DFT) using the PBE0-(D3)/6-31G* level of theory and the MM part by the AMBER99 forcefield. We saturate the broken bonds by hydrogen link atoms and use the electrostatic embedding QM/MM scheme, which is capable of accounting for changes in charge distributions in the course of a reaction. Specifying the above details defines the essential features of the QM/MM setup; however, as we show below, this alone is not sufficient for reproducibility of the results, especially between different software packages. Below we analyze the impact of other parameters on

the computed properties and quantify their effect by comparing the results computed with the two software packages. We consider:

- Details of treating the boundary between the QM and MM regions, specifically, charge redistribution schemes;
- The geometry optimization protocols, specifically, whether microiterations are used and how they are implemented;
- Versions of the AMBER99 forcefields and processing of the topology files;
- Details of the grid and dispersion correction used in the DFT calculations.

In addition to these details, which could, at least in principle, be specified by a precise description of the implementation and relevant input keywords/parameters, there is always a concern of the execution of the workflow. Numerous tasks involved in setting the calculations are not fully automated and often involve manual inspection of the structure (e.g., in order to assign proper protonation states of histidines and other titratable residues); in-house scripts are often used to convert topology files and coordinates from one software to another (e.g., to convert the results from initial equilibration procedure by classical molecular dynamics into inputs for QM/MM models). This creates additional challenges for reproducing the results of QM/MM simulations even within the same research group.

## II. DETAILS OF THE QM/MM SIMULATIONS

Figure 3 illustrates the treatment of covalent bonds at the QM/MM boundary by using the link atom approach. Here we used hydrogens as the link atoms to saturate the dangling bonds; however, this alone does not fully define the model. The exact placement of the link atoms and the definition of charges on the boundary region varies among different implementations[26,27].

We used two different QM/MM partitioning schemes— one with a moderate-size QM system (83 atoms) and one with a larger QM system (155 atoms). Figure 4 shows the moderate-size QM region, which is used in most of our QM/MM calculations. It comprises 83 atoms belonging to the carmofur molecule, side chain/backbone atoms from His41, Asn142, Gly143, Cys145, one water molecule, and five hydrogen link atoms. Figure 4 shows that five

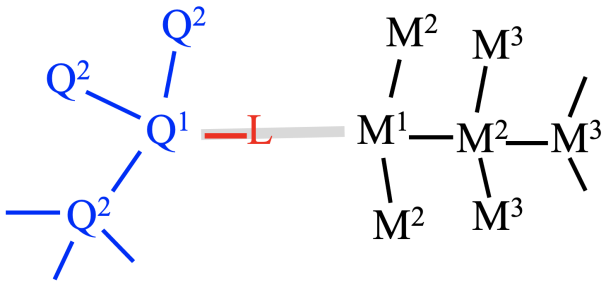


FIG. 3: Illustration of the link atoms approach. QM and MM atoms are marked as $Q^i$ and $M^i$, respectively. The covalent bond between $Q^1$ and $M^1$ is cut and saturated by the link atom L. To avoid overpolarization, the MM charges on the boundary ($M^1$ here) are set to zero; in some approaches, the excess charge is redistributed among the neighboring atoms to conserve the total charge.

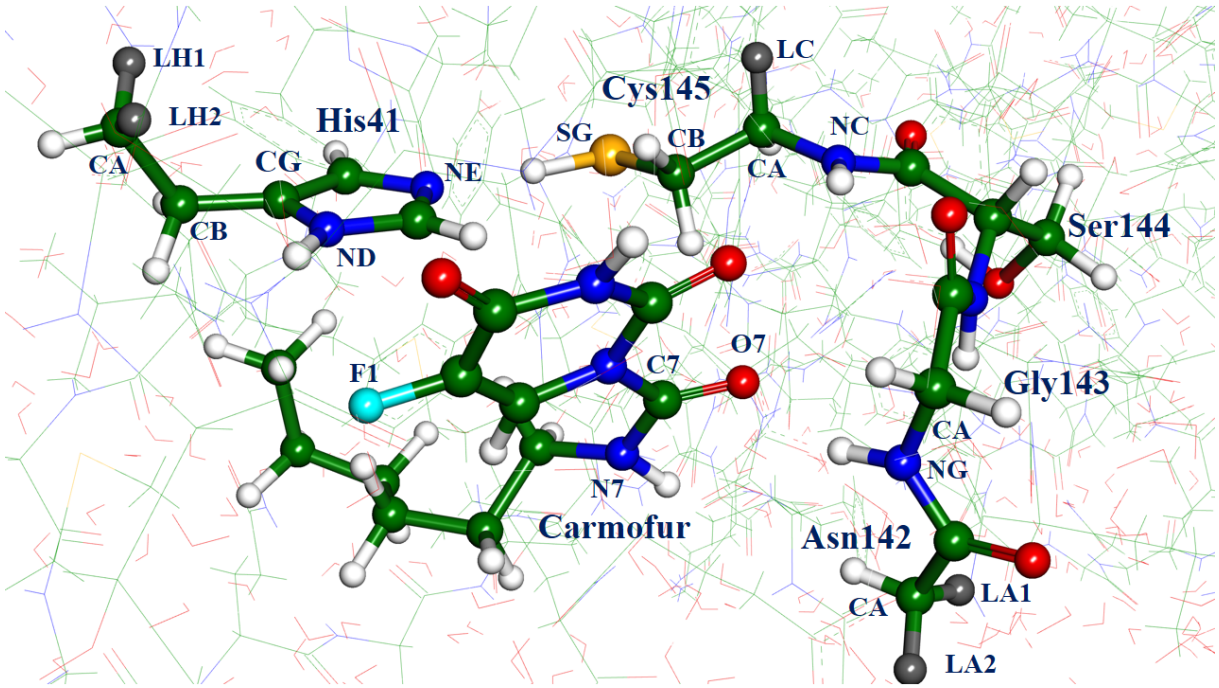


FIG. 4: Definition of the moderate-size QM region (83 atoms). Carbon atoms are colored green, oxygen – red, nitrogen – blue, hydrogen – white, fluorine – cyan, sulfur – orange. The hydrogen link atoms are shown in dark grey; they correspond to His41 (LH1, LH2), Asn142 (LA1, LA2) and Cys145 (LC). Designation of several atoms in the system, which are referenced below (e.g., C7, O7, N7 from carmofur, SG from Cys145, etc.) is specified.

covalent bonds are cut in this QM/MM partitioning (83 atoms in QM). Cutting the Cys145

chain entails adding the hydrogen link atom (LC) at the bond between CA(Cys145) (i.e., $Q^1$ in Fig. 3) and C(backbone) (i.e., $M^1$ in Fig. 3). Correspondingly, in Fig. 3 the $Q^2$ atoms are CB(Cys145), N(backbone), H; the $M^2$ atoms are O and N; the $M^3$ atoms are C and H. The large QM system (155 atoms) comprises the moderate-size QM system plus the side chain/backbone atoms from Thr25, Thr26, Leu27, Leu141, Asn142, Gly146, His164, Met165, Asp187, four water molecules, and 15 link atoms. The relevant details are given in the Supporting Information (SI).

When using electrostatic embedding, the QM region is polarized by the Coulomb potential due to the MM point charges. The charges are determined by the forcefield and the treatment of the boundary. Forcefields are generally well documented, although some variations exist among different software packages that do not share the same topology with the AMBER99 parameters. For example, NWChem and Q-Chem differ in description of C- and N-terminal groups, i.e., His41 is described by the HIP topology in NWChem and by the HID topology in Q-Chem.

In addition, the boundary treatments may also vary, which, as we show below, can lead to substantial discrepancies in QM energies. To avoid overpolarization of the QM region, the charges of the boundary atoms ($M^1$ in Fig. 3) are usually set to zero; however, the subsequent treatment varies—some implementations redistribute the charges of the boundary atoms among the neighboring atoms to preserve the total charge whereas others simply ignore it; the redistribution schemes can also vary[26].

The Q-Chem calculations were carried out using the *HLINK* option implemented in a developer's version of Q-Chem. The charge of the boundary atom ($M^1$ in Fig. 3) was set to 0 and its original forcefield charge was uniformly distributed among the neighboring MM atoms, e.g., 1/3 of the original charge on $M^1$ was added to the three $M^2$ atoms. This procedure is automated in the Q-Chem *HLINK* implementation and was tested by carrying out additional single-point calculation in which the QM energies were computed in the field of the manually prepared point charges. The NWChem calculations were carried out using the *"mm_charges exclude none"* option. It includes all MM point charges in the calculation except the ones located on the covalent QM/MM boundary, i.e., the charge of $M^1$ set to 0 and there is no charge redistribution over neighboring MM atoms.

To reduce cost of the MM force evaluation step, classical molecular dynamics simulations often use electrostatic cutoffs (10-14 Å) in the MM force evaluation step. In QM/MM

simulations, electrostatic cutoffs afford a speed-up in evaluating one-electron contributions to the Hamiltonian. However, such cutoffs can cause problems with optimization, e.g., convergence issues when charges cross the cutoff line. Therefore, in this work we did not apply cutoffs of the electrostatic contributions, i.e., the cutoff radius was larger than the system size. This detail is crucial for achieving convergence in the energy minimization procedure.

The results of the calculations can be also affected by the details of the geometry optimization algorithms (full QM/MM optimization versus micro-iterations). The QM energy includes the Coulomb interaction with the MM region. The MM energy includes forcefield interactions between the MM atoms and van der Waals interactions between the QM and MM parts. In the standard optimization step, the total energy and gradient include the electrostatic interaction between the MM charges and polarized electron density of the QM system.

To speed-up the calculations, NWChem affords a multi-region optimization procedure (called micro-iterations), such that at each optimization cycle the QM region is optimized for $M$ steps (10 in our calculations) with the MM region being frozen, followed by $N$ steps (300 in our case) of the optimization of the MM region with the QM region being frozen. In the MM micro-iterations, the QM/MM electrostatic interaction is described using two options: *"density espfit"* or *"density static"*. The first option approximates the electron density of the QM region with the point charges obtained in the end of the QM optimization cycle, whereas the second option uses exact frozen electron density of the QM region computed in the end of the QM optimization cycle[23]. As discussed below, the two schemes yielded slightly different structures.

The micro-iteration feature is not available in Q-Chem, such that the QM and MM regions are optimized together in each cycle. Q-Chem can only carry out full unconstrained optimizations because the current implementation of the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm[28] does not allow for the constrained geometry optimization or saddle point searches. Hence, we could only compare optimized structures of the reactants, products, and intermediates, but not of transition states.

We prepared the model system of the enzyme-carmofur complex starting from the coordinates of the heavy atoms provided by the protein data dank (PDB) structure 7BUY[20]. Because the PDB structure contains only the alyphatic carmofur tail covalently bound to

the catalytic Cys145 residue, the entire carmofur molecule with the fluoro-uracil warhead (see Fig. 1) was manually docked into the active site after careful inspection of the structure of the reaction product. Protons were added to the aminoacid residues according to their conventional states at neutral pH; i.e., all Arg and Lys residues were positively charged, Glu and Asp were negatively charged. The N-terminal Ser and C-terminal Gln residues were protonated, yielding positive and negative charges, respectively. The histidine residues were protonated according to the hydrogen-bond pattern implied by the heavy atoms positions, i.e., $N_\epsilon$-protonated His64, His163, His164, His172, His246, and $N_\delta$-protonated His41, His80. The protein was fully solvated by TIP3P water molecules. The final model system contained 1,250 water molecules, including those present in the crystal structure PDB 7BUY.

Following QM/MM partitioning, we optimized these model systems with NWChem as described below. To generate input files for Q-Chem calculations, we employed the Tinker package[29], which can read PDB files, recognize the names of the amino acid residues, and generate topology files suitable for Q-Chem by assigning the atom type labels according to the Tinker convention (also used by Q-Chem). Some atom labels used in NWChem were adjusted to match the convention of the AMBER99 forcefield as implemented in Tinker; we show an example (the serine side chain) in the SI. We also provide the inputs for QM/MM simulations for both packages.

We treated the QM part using the PBE0 functional[30] with the 6-31G* basis set and the MM parth using the AMBER99 forcefield[31]. We used default grids (SG-1) and the original variant of Grimme's D3 correction[32]. We note that the Q-Chem implementation uses by default slightly different damping functions than NWChem; however, the same damping functions can be deployed using appropriate keywords. The details of the grid, the exact variant of D3 correction, and relevant input keywords are given in the SI. We also investigated the effect of the functional choice on the reaction energetics and report additional results obtained with $\omega$B97X-D[33,34].

We consider the segment of the potential energy surface (PES) between the reactant (denoted below as REAC) and the product (denoted as PROD) of the reaction of $M^{Pro}$ with carmofur. PROD corresponds to the covalent complex between $M^{Pro}$ and the alyphatic tail of carmofur (whose structure can be compared to the crystallography data) and the separated fluoro-uracil warhead in the active site. REAC corresponds to the reactive conformation of the reactants, formed upon proton transfer within the catalytic dyad from the initially

neutral side chain of Cys145 to the initially neutral side chain of His41 (see the upper part of Fig. 1). We did not model the initial enzyme-substrate (ES) complexes with the neutral Cys145 and His41 species, because the initial step of proton transfer from Cys/His to $Cys^-/His^+$ in cysteine proteases is well studied[8,35,36] and is not considered to be critical in the reaction mechanism. To investigate the effect of the QM size, we computed reaction profile using the two QM-MM partitioning schemes described above (83 and 155 QM atoms).

We consider two different protocols for computing the reaction profile. In Protocol 1, we use the same structures (optimized with NWChem) to carry out single-point energy calculations with NWChem and Q-Chem. In Protocol 2, we compute reaction energy profile using structures optimized with respective packages, i.e., NWChem energetics is computed using NWChem-optimized structures and Q-Chem energetics is computed using Q-Chem-optimized structures (Q-Chem optimizations were carried out starting from the NWChem optimized structures). As discussed below, the agreement between the NWChem and Q-Chem optimized structures is reasonable but not perfect.

For the moderate-size QM, we located the minimum energy structures of REAC, PROD, and the respective transition state (TS) by optimizing the geometry with the *density espfit* option in NWChem. The REAC and PROD structures were re-optimized for the 155-atomic QM part with the *density espfit* and *density static* options. Using Protocol 2 (geometry optimization in each software package), we obtained the structures and energies of three minimum energy points, REAC, PROD, and the reaction intermediate (INT). Below we compare the total QM/MM energies and various individual contributions for each calculation. We emphasize that we compare the results obtained with NWChem and Q-Chem using precisely the same QM/MM partitioning, the same MM parameters, and the same QM level of theory. We note that often only these computational details are reported in the QM/MM studies.

## III. ANALYSIS OF QM/MM RESULTS: PROTOCOL 1

We begin by analyzing the three key points on the PES—REAC, TS, and PROD—located with NWChem using the (more economical) *density espfit* option with the moderate-size (83 atoms) QM part. In this calculation we were able to locate the true TS structure as the stationary point with a single imaginary frequency of 293i $cm^{-1}$.

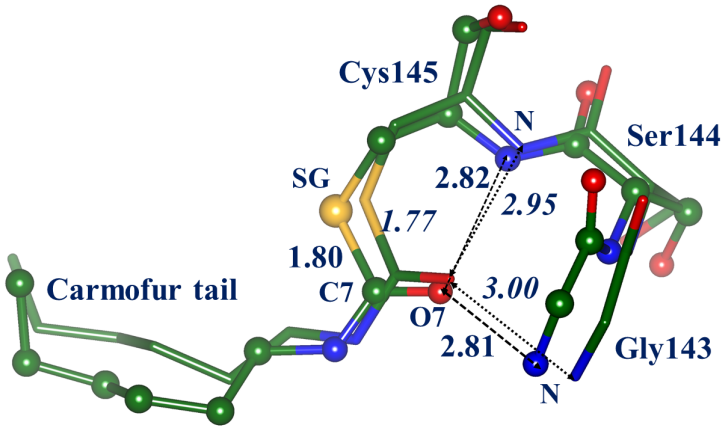


FIG. 5: Alignment of the PROD and X-ray structures. Distances are in Å. The values in italics refer to the crystal structure. Hydrogen atoms are omitted.

It is instructive to compare the results of QM/MM optimization with the only available piece of experimental information—the X-ray structure. Fig. 5 shows the structure of PROD superimposed over the crystal structure PDB 7BUY, focusing on the fragments that are important for this reaction. Fig. 5 shows the most important structural parameters, namely, the length of the formed covalent bond SG(Cys145)-C7(carmofur) and the parameters of the formed oxyanion hole (the distances between O7 atom of carmofur and the backbone nitrogen atoms). The X-ray and computed parameters agree reasonably well, especially taking into account that the PROD model system includes the leaving group (the fluorouracil warhead), which is absent in the X-ray structure. The magnitude of discrepancies is typical for QM/MM simulations.

Fig. 6 summarizes the results of calculations. The individual panels show the REAC, TS, and PROD structures, and the reaction energy profiles computed with NWChem and Q-Chem. The reaction energetics computed with the two softwares shows the discrepancies of 3 kcal/mol at the TS point and of 6 kcal/mol at the PROD point, which is discouraging, especially given that the calculations used the same QM/MM partitioning, same geometries, and the same QM and MM treatments.

Table I shows various energy contributions to the total QM/MM energy obtained with NWChem and Q-Chem. The total energy in the last row is what determines the reaction profile (energies of the TS and PROD relative to REAC; shown the left lower panel in Fig. 6); it is the sum of QM, QM/MM, and MM terms. To understand the sources of discrepan-

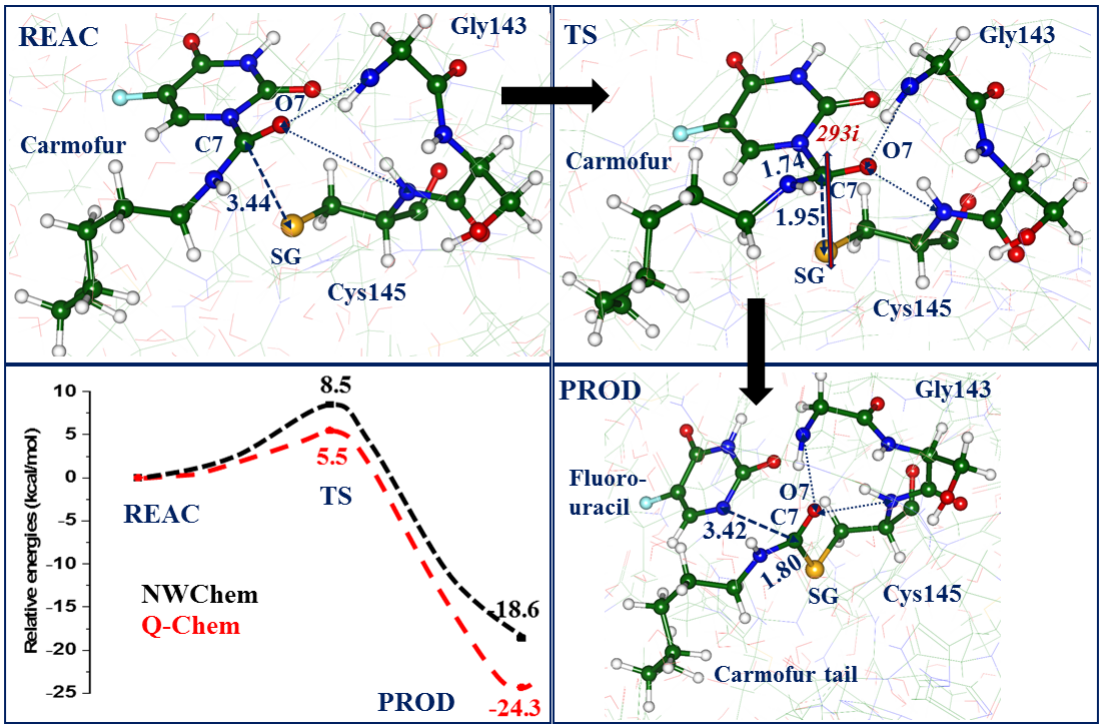


FIG. 6: Structures of REAC, TS, and PROD obtained with Protocol 1 and moderate-size (83 atoms) QM part and the corresponding total energy profile. Distances are in Å. The TS panel shows the vibrational mode with the imaginary frequency 293i $cm^{-1}$.

cies between the calculations performed with two software packages, we compare different contributions to the total energy. These terms are: "QM in gas phase" is the quantum mechanical energy of the isolated QM subsystem (no MM point charges; D3 correction is subtracted); "QM in MM charges" is the energy of the QM subsystem in the presence of the external MM charges from which the explicit charge-density and charge-nuclei contributions are subtracted[37]; "QM + QM/MM" is the full energy of the QM system in the presence of the MM charges (i.e., "QM in MM charges" plus the explicit charge-density and charge-nuclei contributions); and "MM" is the forcefield energy of the MM region.

We attribute small differences in the absolute QM energies ("QM in gas phase") to the differences in: (i) the precise positions of the link atoms and (ii) the parameters of the DFT grids (the grid parameters are given in the SI; they are similar for all elements except for sulfur). The differences in the absolute and relative QM energies are 0.0015 hartree and 0.3 kcal/mol, respectively. Thus, the single-point QM energies are consistent between the two packages.

The discrepancies in total energies between the two packages are due to the different ways of redistributing external charges on the QM boundary. The MM energies follow a similar pattern; relative energies are consistent, whereas differences in absolute energies are larger. The discrepancies in the MM energies are due to slight differences in topologies within the versions of the AMBER99 forcefield implemented in the two software packages.

This analysis of the discrepancies between individual terms allows us to attribute the discrepancies in the total energies to the "QM+QM/MM" term, i.e., the explicit electrostatic contribution (charge-nuclei and charge-density interaction), whereas the implicit electrostatic contribution (energy of the system polarized by the MM charges) that is accounted for in the "QM in MM charges" term is consistent in the two softwares.

Our results illustrate that small differences in the treatment of the QM/MM boundary can and ultimately do lead to substantial differences in the computed reaction energy profiles. The overall differences in the computed energetics for TS and PROD relative to REAC are around 3-5 kcal/mol. The magnitude of these discrepancies is disappointingly large compared to the desired accuracy of 1 kcal/mol; it also exceeds the errors due to approximations in the quantum-chemistry treatments. This qualifies the uncertainties in the QM/MM calculations, setting the bar for reproducibility of the QM/MM results between different software packages.

TABLE I: Energies for the REAC, TS, and PROD structures computed using the same structures.[a]

| Energy contribution | Software | Energy (a.u.) REAC | Energy of PROD relative to REAC (kcal/mol) | |
|---|---|---|---|---|
| | | | TS | PROD |
| QM in gas phase | NWChem | -2474.6196 | 15.6 | -5.2 |
| | Q-Chem | -2474.6211 | 15.8 | -4.9 |
| QM in MM charges | NWChem | -2474.5907 | 12.2 | -5.7 |
| | Q-Chem | -2474.6545 | 12.9 | -4.7 |
| QM + QM/MM | NWChem | -2475.1349 | 14.9 | -9.7 |
| | Q-Chem | -2475.9042 | 11.5 | -15.5 |
| MM | NWChem | -39.1656 | -6.3 | -8.9 |
| | Q-Chem | -39.2451 | -5.9 | -8.8 |
| Total energy QM + QM/MM + MM | NWChem | -2514.3005 | 8.5 | -18.6 |
| | Q-Chem | -2514.1494 | 5.5 | -24.3 |

[a] 83-atom QM subsystem (see Fig. 4); structures optimized with NWChem using the *density espfit* option. D3 correction is subtracted from the "QM in gas phase" but is included in all other terms. Definitions: "QM in gas phase" is the quantum-mechanical energy of the isolated QM subsystem (no MM point charges); "QM in MM charges" is the energy of the QM subsystem in the presence of the external MM charges without the explicit charge-density and charge-nuclei contributions; "QM + QM/MM" is the total QM energy in the presence of the MM charges (i.e., "QM in MM charges" energy plus the explicit charge-density and charge-nuclei contributions; "MM" is the forcefield energy of the MM region.

As the next step, we analyze the effect of increasing the size of the QM subsystem (up to 155 atoms, see Section II) and using different optimization protocols (*density espfit* versus *density static*). In these calculations, we only consider the REAC and PROD structures and their relative energies. The energies are collected in Table II.

First, we note that the structures computed using these two optimization options differ slightly. Fig. 7, which compares the two structures of REAC, shows that the discrepancies do not exceed 0.1 Å for the critical distances: i.e., the distance of the nucleophilic attack

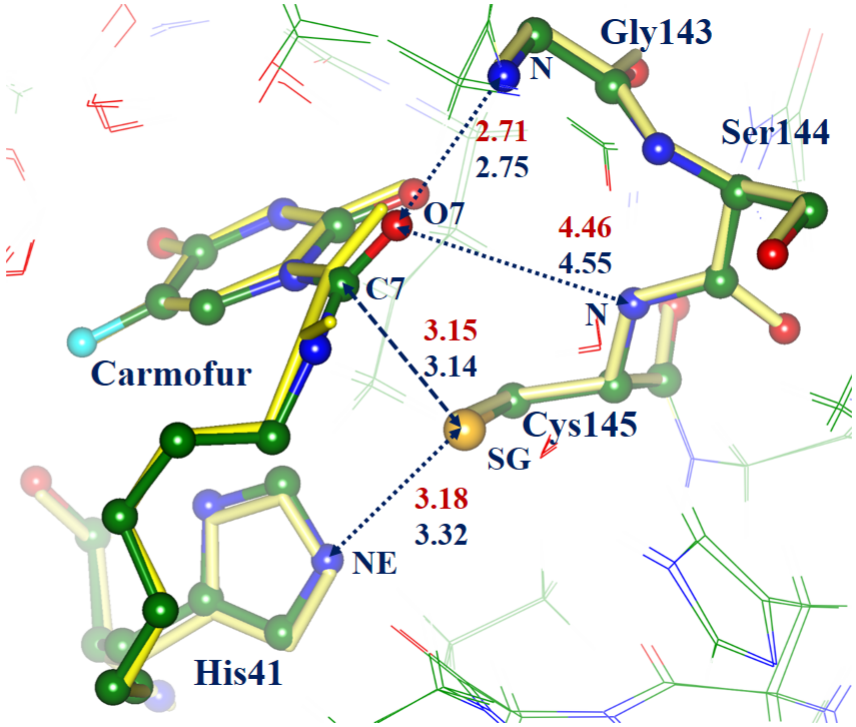


FIG. 7: Alignment of the REAC structures optimized with NWChem using the *density espfit* (colored balls and sticks, red values for distances) and *density static* (yellow sticks, dark blue values for distances) options. Hydrogen atoms are omitted. Distances are in Å.

(SG(Cys145)-C7(carmofur)), the distance between SG(Cys145) and NE(His145), and the distances describing the future oxyanion hole between O7 (carmofur) and nitrogen atoms in the Gly143-Cys145 chain.

However, these slight discrepancies in the optimized structures lead to discrepancies in the relative energies of REAC and PROD, up to 2-3 kcal/mol, as shown in Table II; cf the corresponding rows "NWChem *density espfit*" and "NWChem *density static*". Comparison of the respective rows in the "QM + QM/MM" section shows the discrepancies between the NWChem and Q-Chem energies increases up to 6 kcal/mol, giving rise to the same-magnitude discrepancies (5 kcal/mol) in the total energy as in the calculations with a moderate-size QM subsystem.

This is a counter-intuitive finding—we expected that increasing the QM subsystem would reduce the discrepancies due to slightly different treatment of the boundary as the boundary moves further way from the reaction center. However, larger QM subsystem resulted in a larger QM-MM boundary, which entailed cutting more covalent bonds and, consequently,

more link atoms and more points where the redistribution of boundary charges occur. It is indeed disappointing that increasing the QM does not improve the agreement between the two softwares.

TABLE II: Energies for the REAC and PROD structures.[a]

| Energy contribution | Software | Optimization option in NWChem | E (a.u.), REAC | Energy of PROD relative REAC (kcal/mol) |
|---|---|---|---|---|
| QM in gas phase | NWChem | density espfit | -4092.3434 | -3.7 |
| | Q-Chem | | -4092.3472 | -4.0 |
| | NWChem | density static | -4092.3439 | -5.2 |
| | Q-Chem | | -4092.3475 | -5.3 |
| QM in MM charges | NWChem | density espfit | -4091.8895 | -1.0 |
| | Q-Chem | | -4092.2708 | -1.0 |
| | NWChem | density static | -4091.8897 | -2.5 |
| | Q-Chem | | -4092.2712 | -2.5 |
| QM + QM/MM | NWChem | density espfit | -4093.1537 | -28.4 |
| | Q-Chem | | -4092.8357 | -34.1 |
| | NWChem | density static | -4093.1711 | -26.0 |
| | Q-Chem | | -4092.8553 | -31.5 |
| MM | NWChem | density espfit | -38.3278 | +5.9 |
| | Q-Chem | | -38.5575 | +6.1 |
| | NWChem | density static | -38.3284 | +5.4 |
| | Q-Chem | | -38.5543 | +5.9 |
| Total energy QM + QM/MM + MM | NWChem | density espfit | –4131.4814 | -22.5 |
| | Q-Chem | | -4131.3932 | -28.0 |
| | NWChem | density static | -4131.4995 | -20.5 |
| | Q-Chem | | -4131.4096 | -25.6 |

[a] 155-atom QM subsystem, structures optimized with NWChem using the *density espfit* and *density static* options. See footnotes in Table I for the definition of various energy terms.

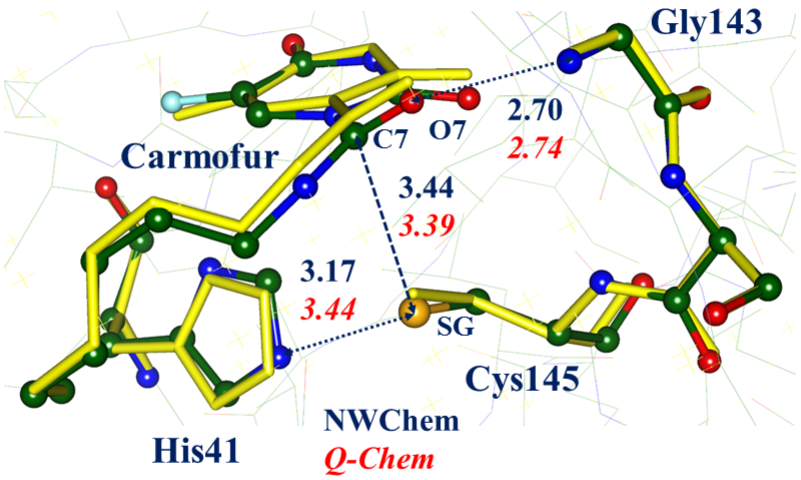


FIG. 8: Alignment of the REAC structures optimized by Q-Chem (balls and sticks colored by elements) and by NWChem (stick colored yellow). Distances are in Å. Hydrogen atoms are omitted.

## IV. EFFECT OF DIFFERENT STRUCTURES: PROTOCOL 2

Having established the effect of the structures computed using two different protocols in NWChem on the reaction profile, we now focus on the differences in the key energetics computed with MWChem and Q-Chem using the structures optimized with each software. Here we use moderate-size QM part (83 atoms). The structures were first optimized with NWChem (using *"density static"*) and then reoptimized with Q-Chem.

In these calculations, we located one more minimum energy point on the PES between REAC and PROD, which was overlooked in the *density espfit* calculations. This structure denoted as INT corresponds to the tetrahedral intermediate typical for the serine or cysteine protease catalysis. The energy of INT is slightly below the level of the previously located TS. Also, we could not locate the second transition state, separating INT and PROD. Multiple scans of the PES in the region around the first TS (separating REAC and INT) reveal very shallow landscape and allow us to estimate that the height of the second transition state is below 1 kcal/mol. Therefore, here we focus on the three stationary points, REAC, INT and

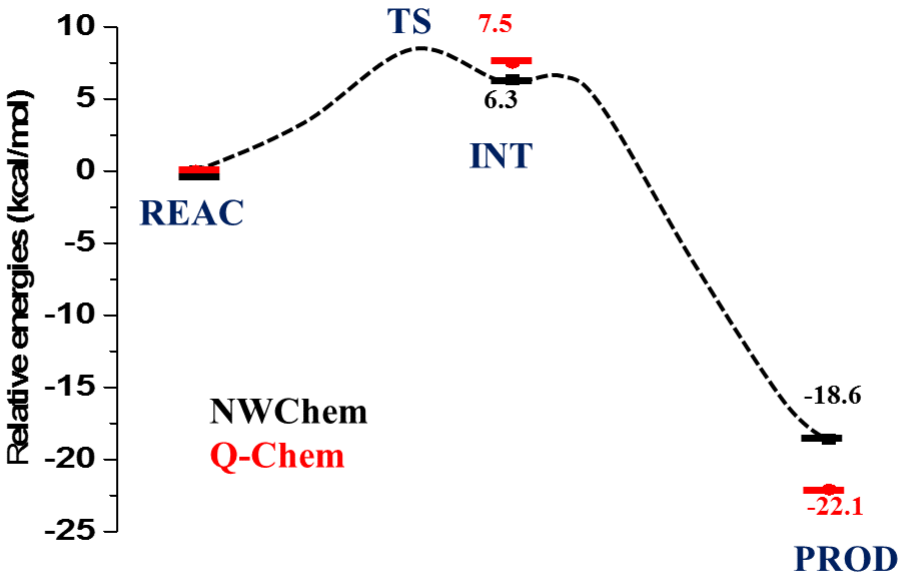


FIG. 9: The reaction energy profile showing the relative energies of REAC, INT and PROD located using Protocol 2.

PROD.

Fig. 8 shows superposition of the groups assigned to the QM part in the REAC structure obtained in two packages. The distances along the chemical bonds are practically identical in two optimized structures. There are slight discrepancies in intermolecular distances. For instance, the distance of the nucleophilic attack (i.e., the SG(Cys145)-C7(carmofur) distances) are 3.44 Å and 3.39 Å in the NWChem and Q-Chem optimized structures, respectively. These small differences can affect relative energies obtained with two packages. The results are summarized in Fig. 9 and in Table III. We observe discrepancies of 1-3.5 kcal/mol in relative energies computed using this protocol, which are slightly smaller than the discrepancies obtained with Protocol 1.

TABLE III: Energies for the REAC, INT, and PROD structures.[a]

| Energy contribution | Software | Energy (a.u.) REAC | Energy of PROD relative to REAC (kcal/mol) | |
|---|---|---|---|---|
| | | | INT | PROD |
| QM in | NWChem | -2474.5961 | 13.5 | -3.5 |
| MM charges | Q-Chem | -2474.6744 | 10.9 | -5.6 |
| QM | NWChem | -2475.1218 | 7.1 | -16.8 |
| + QM/MM | Q-Chem | -2474.8649 | 13.2 | -16.8 |
| MM | NWChem | -39.1798 | -0.8 | -1.8 |
| | Q-Chem | -39.2606 | -5.8 | -5.3 |
| Total energy | NWChem | -2514.3017 | 6.3 | -18.6 |
| QM(D3) + QM/MM + MM | Q-Chem | -2514.1255 | 7.5 | -22.1 |

[a] 83-atom QM subsystem optimized with NWChem *density static* and independently with Q-Chem. See footnotes in Table I for the definition of various energy terms.

## V. DISCUSSION

Our initial research plan was to validate the QM/MM protocols in the two software packages, NWChem and Q-Chem, using the $M^{Pro}$-carmofur reaction as a test case, and then follow with computational design of other prospective covalent inhibitors of $M^{Pro}$ by the joint efforts of our research groups located at different parts of the globe. However, our benchmarking calculations revealed numerous problems of QM/MM calculations, which we did not anticipate to encounter in such a mature field. Indeed, the QM-MM boundary issues have been extensively discussed in many studies going back to the late $20^{th}$ century[2,38–41]. Several software packages are used nowadays, almost routinely, to scan reaction energy profiles for enzyme catalysis. Most users commonly employ the default options of the QM/MM algorithms with confidence that these broadly used tools are robust and produce reliable results. However, our study documents that the codes hide some serious pitfalls related to the QM-MM boundary treatments.

The results discussed above show that small differences in protocols such as optimization algorithms can lead to slightly different stationary points, even when starting from the iden-

tical staring points. Moreover, small differences in implementations can lead to discrepancies of up to 5 kcal/mol in relative energies computed with two different software packages, even when using the identical structures.

Strictly speaking, these problems are not critical for the computational prediction of covalent inhibitors of enzymes. For the purpose of computational screening of prospective inhibitors, it is sufficient to estimate whether the binding energy of the complex is sufficiently large (i.e., greater than 15 kcal/mol) while the barrier for the rate-limiting step is not too high (e.g., less than 15 kcal/mol). This is clearly the case for the $M^{Pro}$-carmofur reaction, as we report here. Despite the noted quantitative discrepancies in calculation results, the overall picture that emerged from the QM/MM simulations is consistent in the sense that: (1) the reaction energy (the relative energy between PROD and REAC) is large enough (around 20 kcal/mol) to explain strong covalent binding of the carmofur tail by the protein; and (2) the energy barrier of the reaction is small enough (less than 10 kcal/mol) to explain efficient chemical reaction between carmofur and $M^{Pro}$.

The key features of the energy landscape are illustrated in Fig. 9 and in Tables I, II, and III. The structure called TS, lying within 8 kcal/mol above REAC, separates REAC from the reaction intermediate INT, the energy of which is slightly below the TS level. The shallow energy landscape around the TS-INT region should contain another low saddle between INT and PROD; however, locating all possible stationary points is not necessary (a more rigorous approach would be to compute free energy profiles by QM/MM-based molecular dynamics simulations[11,13,42–45], possibly augmented by machine learning methods[46,47]).

The results of the present simulations are consistent with the experimental observation that carmofur binds covalently to $M^{Pro}$ and can act as an efficient inhibitor. Following this initial study, we have already carried out series of QM/MM calculations aiming to identify novel covalent inhibitors of $M^{Pro}$. The results are presented in the companion paper, where we followed the lessons learned here and carefully reported all technical details of QM/MM calculations, as required for reproducibility.

Coming back to the benchmarking, we comment here on the choice of a quantum-chemistry method and the size of the QM part. Currently, there is no practical alternative to DFT due to large sizes of typical QM subsystems, which often go up to hundreds of atoms, and the need to perform numerous energy and gradient calculations in QM/MM optimizations or free-energy simulations. The hybrid functionals such as B3LYP or PBE0

are commonly used in such calculations[48–55]. However, modern range-separated functionals might offer better accuracy and reliability, especially because the QM subsystem undergoes massive charge redistribution in the course of the reaction, so that the energetics might be spoiled by self-interaction errors[56,57]. To look into this, we computed single point QM/MM energies with the $\omega$B97X-D3 functional[33,34] using Q-Chem. We obtained relative energies of +7.3 for INT and -22.2 kcal/mol for PROD, which are close to the +7.5 and -22.1 kcal/mol values obtained with PBE0-D3 (shown in Table III). These differences due to the choice of the functional are clearly smaller than the discrepancies due to different QM/MM implementations in different software packages.

The conventional wisdom is that using large QM subsystems generally improves the accuracy and the robustness of the results. The effect of the size of the QM region and other aspects of QM/MM simulations have been investigated by many researchers[58–62]. For example, Ochsenfeld and co-workers have shown[58,59] that reaction energetics can be reliably computed even with mechanical embedding, provided the QM system size is large enough. When using electrostatic embedding, they reported much faster convergence with respect to the QM system size: about 1,000 atoms for proton transfer in DNA[59] and 150-300 atoms for an isomerization reaction in a peptidic system[58]. We investigated the effect of increasing the QM part by comparing the results of the 83-atom-large and 155-atom-large QM subsystems. Contrary to our expectation that a larger QM region would make the results less sensitive to the details of the protocols and, therefore, improve the reproducibility, we found that the effect is more nuanced. It turns out that the larger QM subsystem requires cutting more covalent bonds, giving rise to an extended boundary. The presence of the extended boundary exacerbates the effect of small differences in the treatment of the QM/MM boundary between the two software packages. This finding poses a bigger question about multi-scale methods—can one expect the convergence of the results towards the exact answer (full QM treatment) as the QM subsystem size increases and—if yes—how smooth this convergence might be?

## VI. CONCLUSIONS

In this contribution we show that seemingly minor details of QM/MM simulations, such as the treatment of the MM atom types, placements of link atoms, and charges in the boundary region, should be reported in order to ensure the reproducibility of the results. Specifically, to

accurately reproduce the results obtained with the specific electrostatic embedding QM/MM approach and a specific software, one would need all details of the calculation, to a keyword in the input file, and the exact version of the code. The PDB structures of the stationary points are simply not sufficient. Even with such level of details, one can expect differences up to 5 kcal/mol between different software packages, because of the differences in implementation and the inability for a user to control every small detail of the algorithm, as many parameters are hard-coded and cannot be changed via input. The problem of reproducibility exists even within the same package and the same group. Presently, it is not clear what one can do, but the first step is to acknowledge this problem and this is what we attempted to do in this paper.

The QM treatment (with no embedding) is under control[16], provided that thresholds, cutoffs, damping functions, grids, etc are specified. For DFT, which is the standard choice for QM/MM calculations, the cutoff, grid parameters, and exact details for empirical dispersion corrections should be reported for quantitative reproducibility of the results. We encourage researchers to share more details, including but not limited to the actual grids used in numerical integration. Such details can be provided in the supplementary materials or uploaded to relevant databases (such as MolSSI's COVID hub).

Reproducibly of QM/MM results is challenging because one keyword can change the conclusions quantitatively. The basic idea of multi-scale modeling is great—it is well justified by physics and is practical. The basic idea of the electrostatic embedding scheme is also great and very useful. The scientific community uses extensively QM/MM-based techniques to describe chemical processes happening in complex environments. Presently, there is no alternative for modeling chemical transformations in complex biomolecular systems. The next step to maturity of the field is a standardization of the protocols and ways to store, access and analyze results obtained by different scientists with different software. Simply stating "QM/MM electrostatic embedding scheme" in a paper is clearly not sufficient, as the details of the scheme could be fine-tuned by several keywords in the software, and that could lead to quantitatively different results.

It is difficult to find a practical solution to the changes of the software—the codes are constantly evolving in order to adapt to new hardware, improve efficiency, or expand the functionality. Even if no bugs are introduced by the updates, the results produced by different versions of the same software can differ because defaults were changed.

Some proposals go as far as suggesting to use docker containers with snapshots of the exact software executables and even operational system used in a research project[17], however, we consider this to be impractical and burdensome to the researchers and the environment (e.g., the infrastructure for keeping such vast amount of data would have a significant carbon footprint). Equally impractical is providing all output files for a project because of their large sizes—tens of gigabytes for a single reaction profile or terabites for dynamics and free-energy simulations.

Instead, we suggest to find a reasonable compromise in reporting the details—clearly there is a vast space between not showing any details and having the full containers with the exact software, libraries, input and output files. Instead of aiming at the exact reproducibility we propose to attempt to quantify anticipated error bars due to software implementations, as we have done in this work in the context of energetics of an enzymatic reaction.

On a positive note, despite of these pitfalls, we emphasize the consistency of the qualitative conclusions based on the results obtained in the different parts of the globe and using different software packages applied to the important and urgent problem. We conclude that prediction of prospective covalent inhibitors for troublesome enzymes can be successfully accomplished by properly documented QM/MM modeling.

## VII. ACKNOWLEDGMENTS

Goran Giudetti and Igor Polyakov contributed equally. We thank Prof. John Herbert from the Ohio State University for the help with QM/MM optimizations using the L-BFGS algorithm (in Q-Chem), and Prof. Yihan Shao from the University of Oklahoma for providing the *HLINK* code (in Q-Chem). This work was supported by the U.S. National Science Foundation (No. CHE-1856342 to AIK) and by the Russian Science Foundation (grant No. 19-73-20032 to IVP, BLG and AVN). SF thanks the Innovational Research Incentives Scheme Vidi 2017 with project number 016.Vidi.189.044 financed by the Dutch Research Council (NWO).

We used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 (XSEDE resource: Comet; allocation ID: TG-CHE200131). We also acknowledge resources of MolSSI. IVP, BLG, and AVN also acknowledge the use of the equipment of the shared research facilities

of HPC computing resources at Lomonosov Moscow State University.

[1] A. Warshel and M. Levitt, Theoretical studies of enzymatic reactions: Dielectric electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, J. Mol. Biol. **103**, 227 (1976).

[2] H. M. Senn and W. Thiel, QM/MM methods for biomolecular systems, Angew. Chem., Int. Ed. **48**, 1198 (2009).

[3] Q. Chen, A. Allot, and Z. Lu, Keep up with the latest coronavirus research, Nature **579**, 193 (2020).

[4] V. Frecer and S. Miertus, Antiviral agents against COVID-19: structure-based design of specific peptidomimetic inhibitors of SARS-CoV-2 main protease, RSC Adv. **10**, 40244 (2020).

[5] V. Zarezade, H. Rezaei, G. Shakerinezhad, A. Safavi, Z. Nazeri, A. Veisi, O. Azadbakht, M. Hatami, M. Sabaghan, and Z. Shajirat, The identification of novel inhibitors of human angiotensin-converting enzyme 2 and main protease of SARS-CoV-2: A combination of in silico methods for treatment of COVID-19, J. Molec. Struct. **1237**, 130409 (2021).

[6] L. Riva, S. Yuan andd X. Yin, L. Martin-Sancho, N. Matsunaga, L. Pache, S. Burgstaller-Muehlbacher, P. D. De Jesus, P. Teriete, M. V. Hull, M. W. Chang, J. F.-W. Chan, J. Cao, V. K-M. Poon, K. M. Herbert, K. Cheng, T.-T. H. Nguyen, A. Rubanov, Y. Pu, C. Nguyen, A. Choi R. Rathnasinghe, M. Schotsaert, L. Miorin, M. Dejosez, T. P. Zwaka, K.-Y. Sit, L. Martinez-Sobrido, W.-C. Liu, K. M. White, M. E. Chapman, E. K. Lendy, R. J. Glynne, R. Albrecht, E. Ruppin, A. D. Mesecar, J. R. Johnson, C. Benner, R. Sun, P. G. Schultz, A. I. Su, A. García-Sastre, A. K. Chatterjee, K.-Y. Yuen, and S. K. Chanda, Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing, Nature **586**, 113 (2020).

[7] D. Mondal and A. Warshel, Exploring the mechanism of covalent inhibition: Simulating the binding free energy of $\alpha$-ketoamide inhibitors of the main protease of SARS-CoV-2, Biochemistry **48**, 4601 (2020).

[8] K. Świderek and V. Moliner, Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods, Chem. Sci. **11**, 10626 (2020).

[9] K. Arafet, N. Serrano-Aparicio, A. Lodola, A. J. Mulholland, F. V. González, K. Swiderek, and V. Moliner, Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity,

Chem. Sci. **12**, 1433 (2021).

[10] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I., and Tuñón, Unraveling the SARS-CoV-2 main protease mechanism using multiscale methods, ACS Cat. **10**, 12544 (2020).

[11] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, and Tuñón, Computational simulations on the binding and reactivity of a nitrile inhibitor of the SARS-CoV-2 main protease, Chem. Comm. **57**, 9096 (2021).

[12] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I., and Tuñón, A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design, Chem. Sci. **12**, 3489 (2021).

[13] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, and Tuñón, Inhibition mechanism of SARS-CoV-2 main protease with ketone-based inhibitors unveiled by multiscale simulations: Insights for improved designs, Angew. Chem., Int. Ed. **60**, 25933 (2021).

[14] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, and Tuñón, Multiscale simulations of SARS-CoV-2 3CL protease inhibition with aldehyde derivatives. Role of protein and inhibitor conformational changes in the reaction mechanism, ACS Cat. **11**, 4157 (2021).

[15] A. I. Krylov, T. Windus, T. Barnes, E. Marin-Rimoldi, J. Nash, B. Pritchard, D. Smith, D. Altarawy, P. Saxe, C. Clementi, T. D. Crawford, R. Harrison, S. Jha, V. Pande, and T. Head-Gordon, Computational chemistry software and its advancement: Three Grand Challenge cases for computational molecular science, J. Chem. Phys. **149**, 180901 (2018).

[16] J. E. Boggs, Guidelines for presentation of methodological choices in the publication of computatiional results. A. Ab initio electronic structure calculations, Pure & Appl. Chem. **70**, 1015 (1998).

[17] V. Stodden, J. Seiler, and Z. Ma, An empirical analysis of journal policy effectiveness for computational reproducibility, Proc. Nat. Acad. Sci. **115**, 2584 (2018).

[18] D. B. Allison, R. M. Shiffrin, and V. Stodden, Reproducibility of research: Issues and proposed remedies, Proc. Nat. Acad. Sci. **115**, 2561 (2018).

[19] S. Ullrich and C. Nitsche, The SARS-CoV-2 main protease as drug target, Bioorg. & Med. Chem. Lett. **30**, 127377 (2020).

[20] Z. Jin, Y. Zhao, Y. Sun, B. Zhang, H. Wang, Y. Wu, Y. Zhu, C. Zhu, T. Hu, X. Du, Y. Duan, J. Yu, X. Yang, X. Yang, K. Yang, X. Liu, L. W. Guddat, G. Xiao, L. Zhang, H. Yang, and Z. Rao, Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug

carmofur, Nat. Struct. & Mol. Bio. **27**, 529 (2020).

[21] W. Dai, B. Zhang, X. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jinand J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L. Zhang, Y. Xu, H. Yang, and H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, Science **368**, 1331 (2020).

[22] J. Sakamoto, C. Hamada, M. Rahman, S. Kodaira, K. Ito, H. Nakazato, Y. Ohashi, and M. Yasutomi, An individual patient data meta-analysis of adjuvant therapy with carmofur in patients with curatively resected colon cancer, Jap. J. of Clin. Onc. **35**, 536 (2005).

[23] E. Aprá, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. M. del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero de-la Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, Á. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, and R. J. Harrison, NWChem: Past, present, and future, J. Chem. Phys. **152**, 184102 (2020).

[24] A. I. Krylov and P. M. W. Gill, Q-Chem: An engine for innovation, WIREs: Comput. Mol. Sci. **3**, 317 (2013).

[25] E. Epifanovsky, T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison,

K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, A. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, A. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert, and A. I. Krylov, Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package, J. Chem. Phys. **155**, 084801 (2021).

[26] Y. Shao and J. Kong, YinYang atom: A simple combined ab initio quantum mechanical

molecular mechanical model, J. Phys. Chem. A **111**, 3661 (2007).

[27] T. Vreven and K. Morokuma, Chapter 3 Hybrid Methods: ONIOM(QM:MM) and QM/MM, volume 2, pages 35–51. 2006.

[28] D. C. Liu and J. Nocedal, On the limited memory BFGS method for large scale optimization, Mathematical Programming **45**, 503 (1989).

[29] J. W. Ponder, TINKER – Software Tools for Molecular Design, URL http://dasher.wustl.edu/tinker/ (accessed on April 23, 2017).

[30] C. Adamo and V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, J. Chem. Phys. **110**, 6158 (1999).

[31] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, Proteins-Struct. Funct. Bioinf. **65**, 712 (2006).

[32] S. Grimme, Accurate description of van der Waals complexes by density functional theory including empirical corrections, J. Comput. Chem. **25**, 1463 (2004).

[33] J.-D. Chai and M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, J. Chem. Phys. **128**, 084106 (2008).

[34] J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion interactions, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

[35] J. C. Powers, J. L. Asgian, Ö. D. Ekici, and K. E. James, Irreversible inhibitors of serine, cysteine, and threonine proteases, Chem. Rev. **102**, 4639 (2002).

[36] A. Paasche, T. Schirmeister, and B. Engels, Benchmark study for the cysteine–histidine proton transfer reaction in a protein environment: Gas phase, COSMO, QM/MM approaches, J. Chem. Theory Comput. **9**, 1765 (2013).

[37] The "QM in MM charges" energy is printed in the output by NWChem, but requires an additional calculation with Q-Chem, i.e., a single-point energy calculation of the QM region surrounded by the point charges that are used in *HLINK* QM/MM jobs. These point charges can be extracted from the output file of Q-Chem.

[38] M. Liu, Y. Wang, Y. Chen, M. J. Field, and J. Gao, QM/MM through the 1990s: The first twenty years of method development and applications, Isr. J. of Chem. **54**, 1250 (2014).

[39] B. Wang and D. G. Truhlar, Tuned and balanced redistributed charge scheme for combined quantum mechanical and molecular mechanical (QM/MM) methods and fragment methods:

Tuning based on the CM5 charge model, J. Chem. Theory Comput. **9**, 1036 (2013).

[40] A. Monari, J. L. Rivail, and X. Assfeld, Theoretical modeling of large molecular systems. Advances in the local self consistent field method for mixed quantum mechanics/molecular mechanics calculations, Acc. Chem. Res. **46**, 596 (2013).

[41] B.L. Grigorenko, A.V. Nemukhin, I.A. Topol, and S.K. Burt, Modeling of biomolecular systems with the quantum mechanical and molecular mechanical method based on the effective fragment potential technique: Proposal of flexible fragments, J. Phys. Chem. A **106**, 10663 (2002).

[42] A. Rizzi, P. Carloni, and M. Parrinello, Targeted free energy perturbation revisited: Accurate free energies from mapped reference potentials, J. Phys. Chem. Lett. **12**, 9449 (2021).

[43] M. G. Khrenova, B. L. Grigorenko, and A. V. Nemukhin, Molecular modeling reveals the mechanism of ran-ranGAP-catalyzed guanosine triphosphate hydrolysis without an arginine finger, ACS Cat. **11**, 8985 (2021).

[44] M. G. Khrenova, A. M. Kulakova, and A. V. Nemukhin, Light-induced change of arginine conformation modulates the rate of adenosine triphosphate to cyclic adenosine monophosphate conversion in the optogenetic system containing photoactivated adenylyl cyclase, J. Chem. Inf. and Mod. **61**, 1215 (2021).

[45] C. H. S. da Costa, V. Bonatto, A. M. dos Santos, J. Lameira, A. Leitão, and C. A. Montanari, Evaluating QM/MM free energy surfaces for ranking cysteine protease covalent inhibitors, J. Chem. Inf. and Mod. **60**, 880 (2020).

[46] L. Shen and W. Yang, Molecular dynamics simulations with quantum mechanics/molecular mechanics and adaptive neural networks, J. Chem. Theory Comput. **14**, 1442 (2018).

[47] X. Pan, J. Yang, R. Van, E. Epifanovsky, J. Ho, J. Huang, J. Pu, Y. Mei, K. Nam, and Y. Shao, Machine-learning-assisted free energy simulation of solution-phase and enzyme reactions, J. Chem. Theory Comput. **17**, 5745 (2021).

[48] K. B. Bravaya, M. G. Khrenova, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, Effect of protein environment on electronically excited and ionized states of the green fluorescent protein chromophore, J. Phys. Chem. B **115**, 8296 (2011).

[49] E. Epifanovsky, I. Polyakov, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, The effect of oxidation on the electronic structure of the green fluorescent protein chromophore, J. Chem. Phys. **132**, 115104 (2010).

[50] M. Khrenova, A. V. Nemukhin, B. L. Grigorenko, A. I. Krylov, and T. Domratcheva, Quantum

chemistry calculations provide support to the mechanism of the light-induced structural changes in the flavin-binding photoreceptor protein, J. Chem. Theory Comput. **6**, 2293 (2010).

[51] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, and A. I. Krylov, Turning on and off photoinduced electron transfer in fluorescent proteins by $\pi$-stacking, halide binding, and Tyr145 mutations, J. Am. Chem. Soc. **138**, 4807 (2016).

[52] T. Sen, A. V. Mamontova, A. V. Titelmayer, A. M. Shakhov, A. A. Astafiev, A. Acharya, K. A. Lukyanov, A. I. Krylov, and A. M. Bogdanov, Influence of the first chromophore-forming residue on photobleaching and oxidative photoconversion of EGFP and EYFP, Int. J. Mol. Sci. **20**, 5229 (2019).

[53] T. Sen, Y. Ma, I. Polyakov, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, Interplay between locally excited and charge transfer states governs the photoswitching mechanism in fluorescent protein Dreiklang, J. Phys. Chem. B **125**, 757 (2021).

[54] A. Barrozo, M. Y. El-Naggar, and A. I. Krylov, Distinct electron conductance regimes in bacterial decaheme cytochromes, Angew. Chem., Int. Ed. **57**, 6805 (2018).

[55] S. Xu, A. Barrozo, L. M. Tender, A. I. Krylov, and M. Y. El-Naggar, Multiheme cytochrome mediated redox conduction through Shewanella oneidensis MR-1 cells, J. Am. Chem. Soc. **140**, 10085 (2018).

[56] Y. Zhang and W. Yang, A challenge for density functionals: Self-interaction error increases for systems with a noninteger number of electrons, J. Chem. Phys. **109**, 2604 (1998).

[57] M. Lundber and P.E.M. Siegbahn, Quantifying the effects of the self-interaction error in DFT: When do the delocalized states appear?, J. Chem. Phys. **122**, 224103 (2005).

[58] C. V. Sumowski and C. Ochsenfeld, A convergence study of QM/MM isomerization energies with the selected size of the QM region for peptidic systems, J. Phys. Chem. A **113**, 11734 (2009).

[59] S. Roßbach and C. Ochsenfeld, Influence of coupling and embedding schemes on QM size convergence in QM/MM approaches for the example of a proton transfer in DNA, J. Chem. Theory Comput. **13**, 1102 (2017).

[60] H. J. Kulik, J. Zhang, J. P. Klinman, and T. J. Martínez, How large should the QM region be in QM/MM calculations? The case of catechol o-methyltransferase, J. Phys. Chem. B **120**, 11381 (2016).

[61] Q. Cui, T. Pal, and L. Xie, Biomolecular QM/MM simulations: What are some of the "burning issues"?, J. Phys. Chem. B **125**, 689 (2021).

[62] L. O. Jones, M. A. Mosquera, G. C. Schatz, and M. A. Ratner, Embedding methods for quantum chemistry: Applications from materials to life sciences, J. Am. Chem. Soc. **142**, 3281 (2020).

**TOC graphics:**