



HAL
open science

Protein - ligand interactions from a quantum fragmentation perspective: the case of the SARS-CoV-2 main protease interacting with α -ketoamide inhibitors

Luigi Genovese, William Dawson, Takahito Nakajima, Viviana Cristiglio,
Valérie Vallet, Michel Masella

► To cite this version:

Luigi Genovese, William Dawson, Takahito Nakajima, Viviana Cristiglio, Valérie Vallet, et al.. Protein - ligand interactions from a quantum fragmentation perspective: the case of the SARS-CoV-2 main protease interacting with α -ketoamide inhibitors. *The Journal of Chemical Physics*, 2023, 158 (21), pp.214121. 10.1063/5.0148064 . hal-02933298

HAL Id: hal-02933298

<https://hal.science/hal-02933298v1>

Submitted on 15 Jul 2024




HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE | JUNE 05 2023

Protein–ligand interactions from a quantum fragmentation perspective: The case of the SARS-CoV-2 main protease interacting with α -ketoamide inhibitors

Special Collection: [High Performance Computing in Chemical Physics](#)

Luigi Genovese ; William Dawson ; Takahito Nakajima ; Viviana Cristiglio ; Valérie Vallet ; Michel Masella 



J. Chem. Phys. 158, 214121 (2023)

<https://doi.org/10.1063/5.0148064>

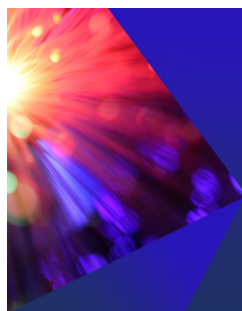


View
Online



Export
Citation

15 July 2024 14:17:00



The Journal of Chemical Physics
2024 Emerging Investigators
Special Collection

Submit Today

Protein–ligand interactions from a quantum fragmentation perspective: The case of the SARS-CoV-2 main protease interacting with α -ketoamide inhibitors

Cite as: J. Chem. Phys. 158, 214121 (2023); doi: 10.1063/5.0148064

Submitted: 28 February 2023 • Accepted: 12 May 2023 •

Published Online: 5 June 2023



View Online



Export Citation



CrossMark

Luigi Genovese,^{1,a)} William Dawson,² Takahito Nakajima,² Viviana Cristiglio,³ Valérie Vallet,⁴ and Michel Masella^{5,b)}

AFFILIATIONS

¹Univ. Grenoble Alpes, CEA, IRIG-MEM-L_Sim, 38000 Grenoble, France

²RIKEN Center for Computational Science, Kobe, Japan

³Institut Laue Langevin, 71 Av. des Martyrs, 38000 Grenoble, France

⁴Université de Lille, CNRS, UMR 8523 - PhLAM - Physique des Lasers Atomes et Molécules, F-59000 Lille, France

⁵Laboratoire de Biologie Structurale et Radiobiologie, Service de Bioénergétique, Biologie Structurale et Mécanismes, Institut Joliot, CEA Saclay, F-91191 Gif sur Yvette Cedex, France

Note: This paper is part of the JCP Special Topic on High Performance Computing in Chemical Physics.

^{a)}Author to whom correspondence should be addressed: luigi.genovese@cea.fr

^{b)}E-mail: michel.masella@cea.fr

ABSTRACT

We present a hybrid, multi-method, computational scheme for protein/ligand systems well suited to be used on modern and forthcoming massively parallel computing systems. The scheme relies on a multi-scale polarizable molecular modeling, approach to perform molecular dynamics simulations, and on an efficient Density Functional Theory (DFT) linear scaling method to post-process simulation snapshots. We use this scheme to investigate recent α -ketoamide inhibitors targeting the main protease of the SARS-CoV-2 virus. We assessed the reliability and the coherence of the hybrid scheme, in particular, by checking the ability of MM and DFT to reproduce results from high-end *ab initio* computations regarding such inhibitors. The DFT approach enables an *a posteriori* fragmentation of the system and an investigation into the strength of interaction among identified fragment pairs. We show the necessity of accounting for a large set of plausible protease/inhibitor conformations to generate reliable interaction data. Finally, we point out ways to further improve α -ketoamide inhibitors to more strongly interact with particular protease domains neighboring the active site.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0148064>

I. INTRODUCTION

To study molecular systems at the atomic scale, Quantum Mechanical (QM) methods are the most accurate theoretical approaches. Presently, the availability of efficient numerical schemes and ongoing increases in available computational resources are enabling one to perform *full* QM (QMF), computations of complete microscopic systems at the 10k atoms scale and above.^{1,2} This is in comparison to commonly used hybrid approaches that

combine QM schemes with Molecular Mechanics (MM) methods (see, among others, Refs. 3–13). A number of challenges exist when employing QM/MM;¹⁴ one benefit of using QMF is there is no need to arbitrarily define a frontier between the QM and MM regions. A further interesting feature of QMF methods is their ability to decompose a molecular system into chemical fragments. Then, using energy-decomposition schemes, it is possible to construct maps of these fragments that effectively describe the strength of the local microscopic interactions.^{15–18} QMF schemes have the benefit not

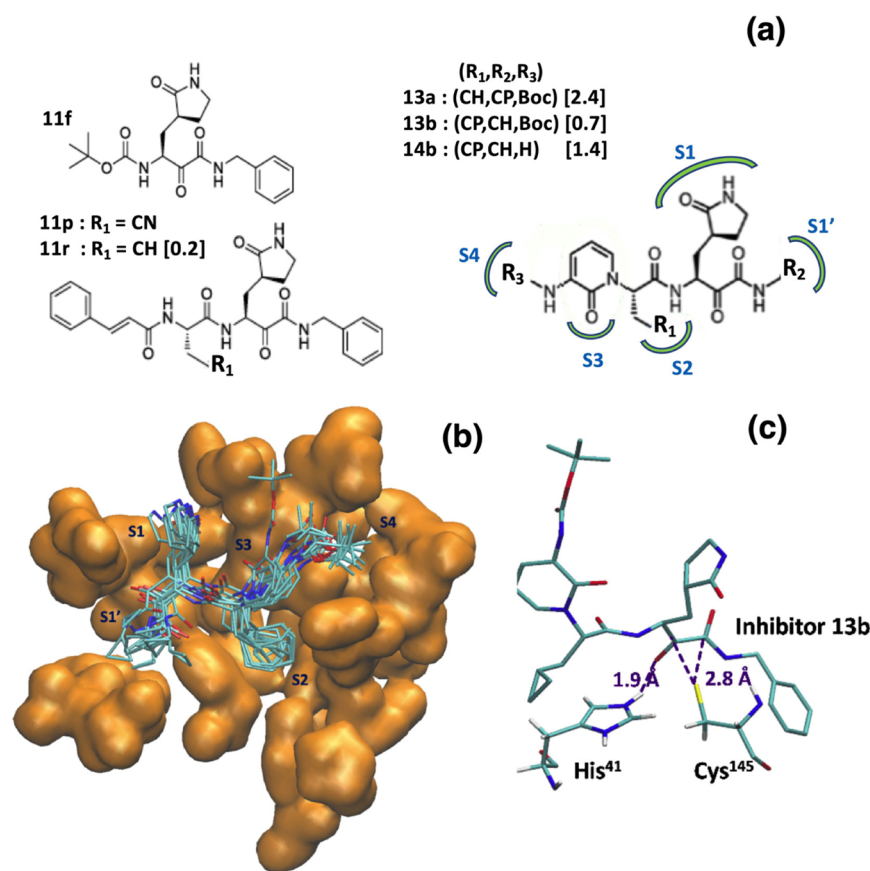
only of opening up the treatment of new categories of systems^{2,19} but are also promising theoretical tools from which to build a new generation of docking methods that will help speed up the development/optimization process of new drugs,^{20,21} for instance. We note here that QMF schemes have already been used to investigate the binding modes of drugs targeting the main protease M^{Pro} of the SARS-CoV-2 virus and of its variants.^{22–26}

Despite their efficiency, we cannot rely on QMF schemes to perform the intensive sampling of the conformational space of large systems that is needed to compute reliable averaged quantities (such as interaction energies). Most of the available QMF studies take into account only a handful of conformations of the systems investigated, see Refs. 27 and 28, for instance. To remediate the drawbacks arising from using limited conformational datasets, we may investigate systems using Molecular Dynamics (MD) techniques based on efficient MM approaches; a QMF scheme can then be used to post-process a large enough snapshot set extracted along the MD trajectories (see, among others, Refs. 29–32). We denote this kind of sequentially coupled approach as MM-QMF hereafter.

MD/MM schemes are now particularly efficient and long-time simulations, from the 0.1–1 μ s scale, have been repeatedly reported. Regarding M^{Pro} /inhibitor complexes, we may quote here Refs. 6 and 33–36, and even MD runs up to the 10 μ s scale of the M^{Pro} dimer are available.³⁷ However, the reliability of the snapshot sets

extracted along MD trajectories does not only depend on the MD run duration but also on the accuracy of the MM approach. Up until now, investigations of protein/ligand complexes mostly use standard pairwise MM approaches. These MM approaches neglect all the dynamical/cooperative effects, such as microscopic polarization (or they account for them according to a pairwise mean field approach) and their accuracy is still largely debated, in particular, to model polyelectrolyte systems.^{38–41}

Our aim is to show the ability of an original polarizable MM-QMF scheme to study large-scale sets of protein/ligand conformations using massively parallel computational resources and to discuss the origin of the microscopic local interactions stabilizing those complexes from reliable mean values of electronic density-based quantities. Moreover, we will also show how modern computational resources help in assessing new force field approaches built from purely quantum *ab initio* data, which is also pivotal to addressing challenging new chemical systems for which experimental data are rare or unavailable. Regarding the MM approach, we consider a multi-scale scheme relying on the polarizable *ab initio*-based force field TCPEp⁴² to model the protein/ligand complexes and on a polarizable coarse-grained approach^{43,44} to account for the solvent. This multi-scale MM approach is efficient enough to readily generate 100 ns scale MD simulations of multiple hydrated systems. To post-process the large set of complex conformations (for the present



study 100 snapshots) regularly extracted along MD trajectories, we consider our own QMF scheme based on Density Functional Theory (DFT) and implemented in the Daubechies wavelets basis set^{45,46} as well as its features for defining meaningful chemical fragments within a system from which to compute local bond orders among fragments.^{15,16}

Here, we focus our study on complexes corresponding to the SARS-CoV-2 main protease M^{pro} interacting with a new family of peptidomimetic α -ketoamide molecules. α -ketoamides are an example of potent (and *a priori* nontoxic⁴⁷) inhibitors that are able to target the main protease of a wide group of coronaviruses.⁴⁸ In particular, they are able to covalently bond (based on a reversible mechanism^{49,50}) to the M^{pro} catalytic domain, yielding the inhibition of biological activity.^{51–55} In Fig. 1, we summarize the inhibitors considered in the present study. The first four (namely, **13a**, **13b**, **14b**, and **11r**) are potent M^{pro} inhibitors,⁵¹ whereas the last two (**11p** and **11f**) are characterized by weak or no inhibitory potency.⁴⁸ Note that the lack of inhibitory potency does not prejudice the ability of an α -ketoamide to interact with the M^{pro} active site or to form a stable complex, as shown by crystallographic data in the particular case of **11f** (see Ref. 48 and crystallographic structures of **11f** interacting with the main protease of the human coronavirus HCoV-NL63, PDB label 5NHO⁵⁶). With the exception of **11f**, the inhibitors investigated using the MM approach will be further studied using the QMF scheme in order to compare the role of inhibitor side chains on the M^{pro} /inhibitor interactions.

II. METHODS

A. The multi-scale MM polarizable approach

The proteases are modeled using an updated version of the polarizable *all atom* force field TCPEp that relies on an induced dipole moment polarization approach and whose parameters are assigned to reproduce *ab initio* quantum data.⁴² Most of the force field parameters used to model the α -ketoamide inhibitors correspond to those that we use to model proteins. For instance, repulsion, dispersion, and polarizability parameters to model the interactions between a thiolate and a ketoamide moiety NH–OCCO correspond to those modeling the interactions between thiolate and the amide moiety NH–CO in the latest force field version. However, we assigned a reduced set of parameters to model the specific standard 1–4 dihedral and improper torsional energy terms from quantum *ab initio* data computed at the MP2(FC)/aug-cc-pVDZ level of theory (see Figs. S1–S3).

All the inhibitor Coulombic charges and all the non-bonded energy term parameters regarding inhibitor acetonitrile moieties were specifically assigned for the present study according to the protocol detailed in former studies (see Ref. 57, for instance). Note here the Coulombic charges of the sp^3 carbons connecting the inhibitor moieties are set to values zeroing the inhibitor total charge. The Coulombic charges of inhibitors **13b** and **11p** are reported in Fig. 2(a). A novelty of the polarizable TCPEp force field is to only account for the electric field components generated by charged

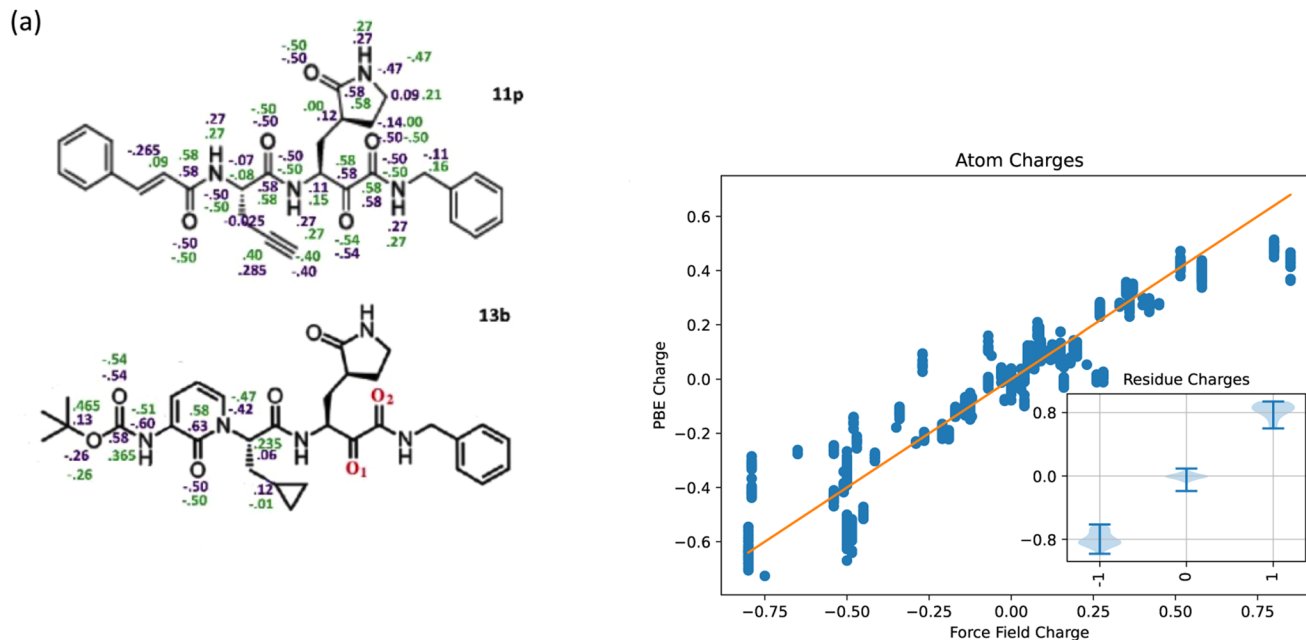


FIG. 2. (a) Atomic charges to compute the Coulombic term (violet) and the electric fields acting on polarizable atoms (green) for inhibitors **11p** and **13b** (charges not shown for **11p** correspond to their values for standard peptide/protein systems, whereas the charges already defined for **11p** are not shown for **13b**). The definitions of α -ketoamide oxygens O_1 and O_2 are shown in red. (b) QMF atomic charges as a function of MM ones for the M^{pro} monomer in its crystallographic x-ray geometry (PDB label 6Y7M). Solid line: the linear correlation fit (the regression coefficient is 0.91, the slope is 0.8, and the charge at the origin is 0.0e). The comparison of the aggregated charge on the residues is also shown in the inset.

and dipolar chemical groups. The particular case of α -ketoamide inhibitors leads us to define a specific set of charges for the latter “connecting” carbons so that the sum of the charges generating the inhibitor electric field is again zero.

We account for the solvent water using a polarizable coarse-grained approach well suited to be used in conjunction with all atom polarizable force fields based on the induced dipole moment approach to model solutes.^{43,44} The water molecule is modeled as a single polarizable particle (its polarizability obeys a Clausius–Mossotti relation) and electrostatic effects among solvent particles are neglected. The interparticle interactions are modeled using a Lennard-Jones-like term to which is added an energy term that is a function of the water local density. The solute/solvent parameters are assigned to reproduce the hydration structural properties (such as mean water–solute atom distances and mean hydration numbers) and hydration Gibbs free energies at ambient conditions of a training set of small molecular and ionic entities (such as methanol and the methyl ammonium ion).

As with all force field approaches, the TCPEp and the solvent coarse-grained parameter sets have been regularly updated since their original versions.^{42,44} To assess the quality of the latest version of the our MM multi-scale approach, we plot in Fig. S4 the potential of mean surfaces of the dipeptide alanine in gas phase and in liquid water, and in Fig. S5, we plot the temporal evolution at the μ s scale of the Root Mean Square Deviation (RMSD) of the protein 1PGB⁵⁸ backbone along eight distinct MD simulations in aqueous phase performed according to the protocol detailed below and using our latest parameter sets. The geometry of 1PGB is not restrained to its crystallographic starting conformation. The RMSDs are stable along long time periods (up to 900 ns) and they fluctuate within 1.1–2.2 Å. To further assess the ability of our MM approach to model molecular assemblies, we show in Fig. S5 a superposition of the starting and final conformations of an 11-residue cleavage site peptide within the active site from a 100 ns MD simulation using our MM approach (here the protease backbone is restrained to its starting conformation). The protease/peptide starting structure corresponds to the final one⁵⁹ of a MD simulation in water performed using the GROMACS code (version 2019.2⁶⁰) and the non-polarizable AMBER99SB-ILDN force field.⁶¹ Along our MD simulation, the peptide backbone RMSD is included within 0.8 and 1.5 Å, a range of values that corresponds to the peptide RMSD fluctuations along the GROMACS/AMBER trajectory.

B. Molecular dynamics protocol

All MD simulations were performed at ambient conditions using the code POLARIS(MD).⁶² The Newtonian equations of motion are solved using a multiple time step algorithm with two time steps: 0.25 fs for intra-solute stretching, bending, and improper torsional energy terms and 2 fs for all the other interactions.⁶³ All the covalent X–H bonds and H–X–H angles are constrained to their force field reference values by means of the iterative RATTLE procedure (convergence criterion of 10^{-5} Å). The system temperature and pressure in NVT and NPT simulations are monitored by means of the Langevin dynamics approach⁶⁴ and the Nosé–Hoover barostat,⁶⁵ respectively. Solute and solvent induced dipole moments are solved iteratively until the mean difference in these dipoles between two successive iterations is less than 10^{-6} D.

Our study investigates the local interactions among the monomeric form of the SARS-CoV-2 main protease and α -ketoamide inhibitors. Protease dimerization is *a priori* needed for catalytic activity as local dimer interactions stabilize the catalytic pocket shape.⁶⁶ Moreover, recent MD simulation studies have revealed significant structural differences between the M^{Pro} monomeric and dimeric forms that would *a priori* favor catalytic efficiency in the dimer state.⁶⁷ As we only consider monomeric protease systems, we systematically restrain the positions of the protease backbone carbons (α , C_α) that lie at a distance larger than 5 Å from any inhibitor atom within the starting conformations by means of the following term:

$$U^{\text{dir}}(C_\alpha) = \sum_{C_\alpha} k_c [1 - \mathbf{u}_i \cdot \mathbf{u}_i^0]^2. \quad (1)$$

Here, the sum runs over the restrained C_α 's, and \mathbf{u}_i is the unit vector connecting the protease center of mass (COM) to its i th C_α (\mathbf{u}_i^0 is that vector in the starting structures). We set the constant k_c to $20 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. Note that most of the C_α 's of the two flexible loops close to the protease catalytic site are restrained along our simulations.⁶⁸

C. QMF treatment with density functional theory

Kohn–Sham Density Functional Theory (DFT) is implemented using a Daubechies wavelets formalism^{45,46} in the BigDFT code.⁶⁹ We use the PBE functional⁷⁰ with D3 corrections⁷¹ (PBE + D3) and Hartwigsen–Goedecker–Hutter pseudopotentials.⁷² For systems with a suitable electronic structure, the one-body density matrix can be represented in terms of a set of support functions ϕ_α , which can be optimized while preserving their strict locality (their domain is within a predefined localized space region). This nearsightedness principle, yielding quantum DFT approaches in which the computational cost scales linearly with the number of atoms, is also at the core of similar DFT approaches available in the ONETEP⁷³ and CONQUEST⁷⁴ codes.

Besides computational efficiency, linear scaling DFT approaches are also at the core of original analysis tools, such as the Complexity Reduction Framework (CRF) procedure.^{15,16} The details of the CRF procedure for application to protein systems are available in Refs. 59 and 75. Briefly, the CRF procedure relies on the identification of independent fragments from the analysis of the system density matrix F expressed in terms of the support functions. For determining fragments, it provides an index $\Pi^f = \frac{1}{q_f} \text{Tr}((F^f)^2 - F^f)$ (termed the purity index, q_f the number of electrons) that can be interpreted as the *level of confidence* within which a given fragment can be considered as an independent unit of the system. Once a fragmentation is provided that meets the desired level of confidence, we may define quantities, such as the inter-Fragment Bond Order (FBO), that measures the amount of the electronic density shared by two fragments (strength of the chemical bond). Fragments may also be used as a basis set on which to project the total interaction energy of a molecular assembly. The sum of interfragment energies ΔU^f may be used to estimate total interaction energies and to analyze individual fragment contributions to assembly stabilization.

CRF fragments do not necessarily meet standard groups/moieties based on chemical intuition. As an illustration, let us consider M^{Pro} in its apo form and in its crystallographic geometry (from PDB ID 6LU7) to which we added hydrogens

(see details below), and we analyze the purity values of suitable portions of the protein. We plot in Fig. 3, the distributions of the purity index values of each residue type present in M^{Pro} (also split in its side chain and its backbone unit $NHCOC_\alpha$) as a CRF

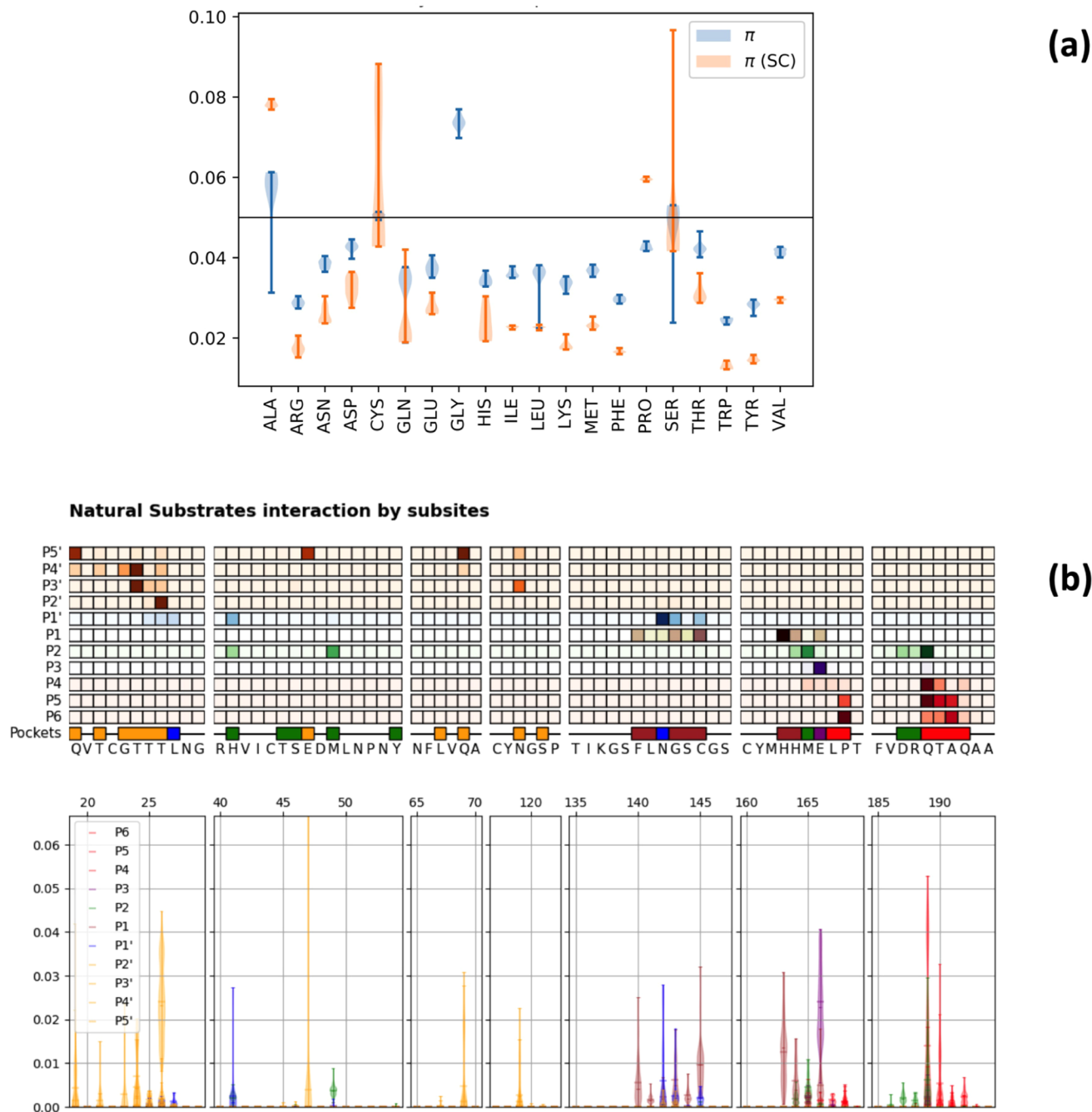


FIG. 3. (a) Purity index values required to get the equivalence between CRF fragments and residues for M^{Pro} in its crystallographic geometry 6LU7. Blue and orange violin plots: purity index distributions for the full residues (π) and for their side chains [$\pi(SC)$], respectively. (b) Identification of pocket leading fragments within the active site from FBO data between M^{Pro} and the subsites (labeled P and P') of two of its natural substrates.⁵⁹ The violin plots show the FBO distributions for each M^{Pro} residues with the substrate subsites among the MD snapshots post-processed (see main text for details). The boxes on top of the M^{Pro} sequence represent the FBO averages (darker colors correspond to higher FBO values). Boxes are colored according to the substrate subsites. On the bottom row, we represent the final set of leading fragments.

fragment. From this figure, it appears that particular residues, such as arginines, tryptophans, and tyrosines, correspond to pure fragments (they are identified as fragments even with purity index values as low as about 1%) whereas glycines and cysteines (in particular, that of the M^{pro} active site) are identified as fragments only for large purity index values (>5%). The high purity values needed to identify the latter two residue types as fragments is a marker of a (relatively) more diffuse electronic density around them. To meet lower purity values around those regions, one should merge such subunits into larger fragments that include some of their neighboring residues to reach lower purity values. From this case-study, the purity index should be set to about 8% to interpret all the M^{pro} residues as fragments. That is still an acceptable level of confidence in value, which will be considered for the present study.

Within this setup, we further analyzed the interaction between M^{pro} and its natural substrates following the guidelines of the study in Ref. 59 to identify leading amino acid fragments interacting with the substrates subsites. Such an approach can be thought as an unbiased, *ab initio* identification of the enzymatic pockets that only relies on chemical interaction among the system's moieties. The leading fragments are summarized in Table I and are obtained from the representations of the interaction in Fig. 2. Three fragment sets correspond to the usual M^{pro} pockets S1, S2, and S4.⁴⁸ We thus denote those fragments accordingly to the latter pockets. We also identify a leading fragment set (denoted S3) composed of a single residue (namely, Glu¹⁶⁶), and two distinct sets (denoted S1' and S2'–S5'), where the latter encompasses the interaction residues with the last part of the substrates. When discussing FBO data below, we denote as a “pocket” each set of leading fragments even if they include a single residue (S3).

D. Standard quantum *ab initio* computations

To assess the reliability of our QMF approach and to build up an accurate polarizable force field to model M^{pro} /inhibitor complexes, we performed standard MP2 computations, with the frozen core approximation MP2(FC), using the GAUSSIAN09 package.⁷⁶ We also performed PNO-L-CCSD(T)^{77,78} computations (using the MOLPRO code⁷⁹) of a set of **13a** conformations using the aug-cc-pVTZ basis set,⁸⁰ with density-fitting for all integrals using the aug-cc-pVTZ/JKFIT⁸¹ and aug-cc-pVTZ/MP2FIT basis sets.⁸² For the DFT calculations, we have employed a grid spacing of 0.4 bohr for the underlying wavelet sampling of the support functions in all the data presented in this study.

TABLE I. Leading fragments of the M^{pro} pockets (the fragments are here all residues).

Pocket	Leading fragments (residues) for M^{pro}
S1	140, 141, 143, 144, 145, 163, 164
S2	41, 45, 46, 49, 54, 165, 187, 188
S3	166
S4	167, 168, 189, 190, 191, 192
S1'	27, 142
S2'–S5'	19, 21, 23, 24, 25, 26, 47, 67, 69, 119, 121

E. Starting structures of protease/inhibitor complexes

For inhibitors **11f**, **13a**, and **13b**, their x-ray crystallographic 3D structures interacting with the HCov-NL63, SARS-CoV, and SARS-CoV-2 main proteases, respectively, have been reported.^{56,83,84} As the latter two proteases share 96% of homology, we assume the binding mode of **13a** *wrt* SARS-CoV M^{pro} to be the same as for SARS-CoV-2. Moreover, because of the similarity of the binding modes of **11f**, **13a**, and **13b** with reference to the proteases of HCov-NL63, SARS-CoV, and SARS-CoV-2,^{48,51} respectively, we also assume the binding mode of inhibitors **14b** and **11r**, **11p** and **11f** correspond to that of **13b** with reference to SARS-CoV-2.

We built the starting structures of all complexes from the crystallographic x-ray PDB structure 6Y2G⁸⁴ (i.e., **13b** interacting with SARS-CoV-2 M^{pro}). Each complex is made from a single protease monomer, and we model the covalent bond between the inhibitor and the M^{pro} Cys¹⁴⁵ cysteine as an ionic bond: the M^{pro} catalytic dyad His⁴¹/Cys¹⁴⁵ is taken under its ionic form (the side chains are thus an imidazolium and thiolate group, respectively). In correspondence to a previous analysis of 91 M^{pro} x-ray crystallographic structures,⁶ we set His⁴¹ to be in an ϵ -rotameric state, i.e., the His⁴¹ side chain nitrogen NE2 (PDB labeling) points toward the inhibitor ketoamide oxygen O₁ [see Fig. 2(a)].

We docked the **13a**, **14b**, **11r**, **11p**, and **11f** inhibitors to best fit the **13b** conformation within the active site in the original PDB structure 6Y2G. For all inhibitors, their dihedral angle OCCO is set to 180° (that angle is termed below as ω_{OCCO}). The six resulting complex structures were first optimized in the gas phase by performing 100 steps of standard conjugate gradient without imposing any constraints on the complex geometry. The optimized complexes were then embedded in cubic boxes filled with polarizable pseudo particles (PPP) modeling water. These particles (about 57 000) are set on the nodes of a grid whose dimension is 3 Å (and particles located at less than 3 Å from any non-hydrogen atom are removed). The resulting hydrated protease/inhibitor structures were then relaxed by performing a series of eight independent preliminary 10 ns scale MD simulations per complex in the NPT ensemble. Each simulation corresponds to a different set of starting atomic velocities and the protease C _{α} positions are restrained by means of the energy term U^{dir} [see relation (1)]. For each complex, we selected a final simulation snapshot from one of these MD runs along which the inhibitor maintains an interaction pattern that agrees with those of X-ray experimental structures 6Y7M and 6Y2G. These final structures (protease/inhibitor and solvent particles) were again optimized as above to yield the starting simulation structures for a final series of MD simulations performed in the NVT ensemble and at the 100 ns scale (one per protease/inhibitor complex).

Our assumption to model the M^{pro} /inhibitor binding by considering a zwitterionic (activated) form of the cysteine/histidine catalytic dyad may be interpreted as modeling an ionic pre-reactive state before the formation of a covalent bond between M^{pro} and the inhibitors. This state has been taken into account in recent QM/MM studies investigating the proteolysis reaction mechanism³ as well as its covalent bonding reaction mechanism with aldehyde derivatives^{6,7} and α -ketoamide inhibitors.^{23,85} A recent experimental neutron crystallographic study⁸⁶ (supported by large scale

quantum MD simulations⁸⁷) suggests the catalytic dyad to already be under an ionic form in the M^{Pro} apo form as well as the histidine His¹⁶⁴ that appears to be hydrogen bonded to the catalytic His⁴¹ in all the reported experimental structures (see Ref. 6 and references therein). We investigated the effect of the protonation state of His¹⁶⁴ on the structural stability of the active site for our six M^{Pro} /inhibitor complexes using our MD/MM protocol. Our simulations systematically show the protonation of His¹⁶⁴ to fully disrupt the original interaction pattern reported in the PDB experimental structures 6Y2G and 6Y7M, regardless of the inhibitor. Hence, we consider His¹⁶⁴ as neutral and we set it in a δ -rotameric state in our starting structures to favor a strong hydrogen bond between it and the His⁴¹ imidazolium group. Note, in recent x-ray/neutron studies, Kneller *et al.* show the protonation state of His¹⁶⁴ to be *a priori* ligand dependent.^{88,89}

III. ACCURACY AND COHERENCE OF THE MM-QMF APPROACH

To assess the quality of both our MM and QMF approaches to model α -ketoamide inhibitors, we performed a 40 ns MD simulation of **13a** in the aqueous phase, from which we extracted 300 inhibitor snapshots regularly spaced along the last 30 ns segment. From these snapshots, we extracted a subset of 30 inhibitor conformations (from the least to the most stable MM ones) for re-investigated at the quantum *ab initio* PNO-L-CCSD(T) level of theory. The MM, QMF, and *ab initio* intramolecular inhibitor total energies all agree (see Fig. 4). The agreement is particularly good between QMF and *ab initio* data, whereas our force field agrees with *ab initio* data only on average.

To assess the consistency of post-processing M^{Pro} /inhibitor snapshots from our MD/MM simulations using our QMF scheme, we extracted 100 M^{Pro} /inhibitor structures along a subset of the twelve preliminary NPT simulations performed to relax the hydrated M^{Pro} /inhibitor complexes. Along the full NPT simulation set, inhibitors **11r**, **13a**, **13b**, and **14b** maintain their starting binding pattern within the active site. Regarding **11r** and **11f**, they are both rapidly ejected away from M^{Pro} binding site along 4 and 6 simulations, respectively, but they still lie close to the active site. The above (post-processed) simulation subset mostly comprises **11r** and **11f** trajectories corresponding to different M^{Pro} /inhibitor binding modes and the snapshots are extracted along the final MD segments on which the M^{Pro} /inhibitor binding pattern is stable.

For each extracted snapshot set, we computed the M^{Pro} /inhibitor mean interaction energies $\Delta\bar{U}$ from our MM and QMF approaches (the details of these energies are summarized in Tables S2–S4 of the supplementary material). The $\Delta\bar{U}$ s are computed as the difference between the M^{Pro} /inhibitor complex energy and the energies of M^{Pro} and the inhibitor separately in their complex poses. Both the MM and QMF $\Delta\bar{U}$ values are linearly correlated [see Fig. 4(b)]. However, the MM energies are more stable than the QMF ones by about 20%.

To investigate the origin of this discrepancy, we compared both the MM and QMF (PBE + D3 corrections) interaction energies with reference to their *ab initio* CCSD(T)/CBS values using the 36 non uracil and non ethyne dimers of the S66 benchmark database.⁹⁰ All these energy datasets are all linearly correlated with each other (the Pearson correlation coefficient, P_r , is at least 0.83). Note here that the D3 and MM dispersion components of the S66 dimer energies are

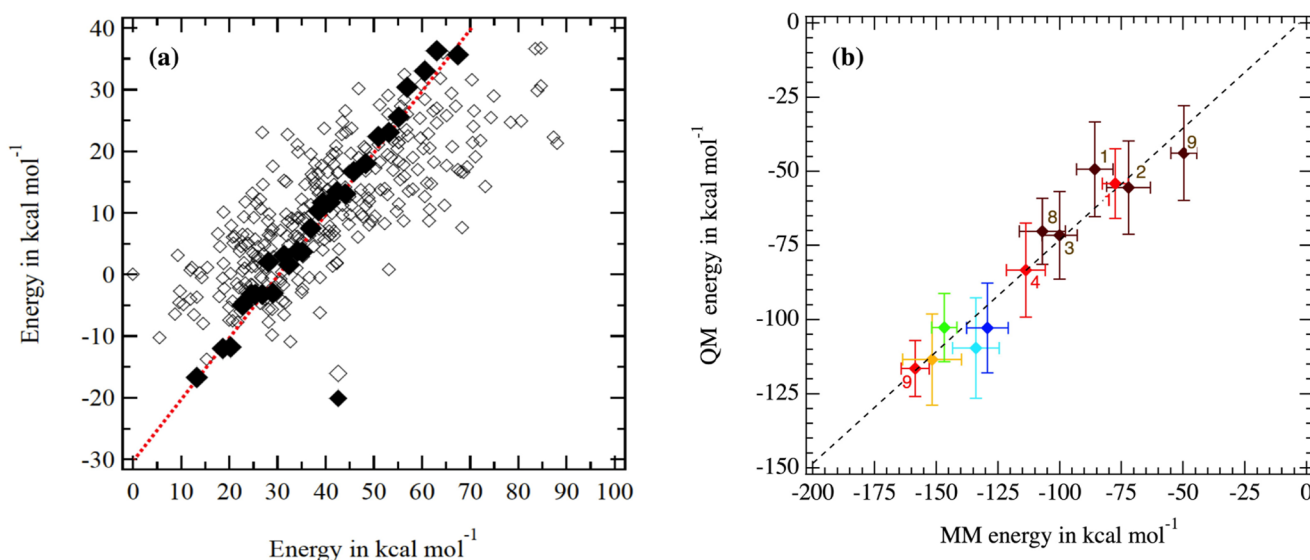


FIG. 4. (a) Comparison of the inhibitor **13a** intramolecular energy estimated from our MM approach (empty symbols) and *ab initio* PNO-L-CCSD(T) computations (full symbols) to QMF (PBE + D3) data for snapshots extracted along a MD/MM simulation of the inhibitor in water. The energies are shifted to be zero for the first snapshot (MM and CCSD data) and for the most stable one (QMF). The dashed red line corresponds to the linear function $y = x + b$ whose parameter b is set to minimize the difference between MM and QMF energies. (b) Mean QMF M^{Pro} /inhibitor interaction energies $\Delta\bar{U}$ as a function of their MM counterparts (brown **11f**, red **11p**, orange **13a**, green **11r**, light blue **13b**, dark blue **14b**). The error bars correspond to the standard deviations of these mean energy values. For **11f** and **11p** data, the numbers shown corresponds to the simulation labels (see Figs. S8 and S9 of the supplementary material).

linearly correlated ($P_r = 0.92$) and their magnitudes are close (within less than $0.5 \text{ kcal mol}^{-1}$ on average; see Fig. S6 of the supplementary material). We also investigated the properties of a trimer modeling the local interactions between the M^{Pro} catalytic pair (in its ionic form) and the inhibitor ketoamide moiety (in a geometry corresponding to that observed along our MD runs) by means of our MM, QMF and *ab initio* MP2(FC)/aug-cc-pVDZ computations. All the trimer interaction energies agree within $\pm 5 \text{ kcal mol}^{-1}$ (see Table S1 of the supplementary material). Hence, the discrepancy between our MM and QMF approaches regarding the M^{Pro} /inhibitor energies $\Delta\bar{U}$ s does not arise from drawbacks in modeling local M^{Pro} /inhibitor interactions.

We also computed the Mulliken charges corresponding to M^{Pro} from our QMF approach. The formalism of the BigDFT code, employed in this study, is based on a set of localized basis functions that are expressed in Daubechies wavelets and optimized on-site to suitably express the electronic structure of the system, being at the same time close to orthogonality. Thanks to this quasi-orthogonality, we have proven in Ref. 15 that Mulliken and Loewdin population analysis are almost identical in this basis set. For these reasons, they provide an effective representation for the density matrix of the system and enable precise and chemically sound results, especially when projected on the fragment with low purity index (the residues in this case). Because of the fragmentation process, those charges are reliable quantities from which to compute fragment electrostatic properties (such as dipole moment and quadrupole tensor). The charges are here systematically lower in magnitude compared to our MM charges by about 20% [see Fig. 2(b)]. This suggests that the discrepancies between MM and QMF M^{Pro} /inhibitor interaction energies that arise are more likely from under-polarization of the M^{Pro} chemical bonds using the PBE functional, which yields an electrostatic potential within the active site weaker by 20% for QMF than MM. A similar PBE bond under-polarization compared to a polarizable force field has already been reported for liquid water.²

Both our MM and QMF approaches provide a similar description of the microscopic interactions between the inhibitors and

M^{Pro} . This description is in line with that expected from higher levels of quantum *ab initio* theory. In particular, the set of M^{Pro} /inhibitor conformations (and their relative weight) sampled along our MD/MM simulations should thus be close to that expected along a MD/QMF simulation. That supports the reliability of the data resulting from post-processing MM snapshots using QMF.

IV. INHIBITORS WITHIN M^{Pro} ACTIVE SITE FROM MM SIMULATIONS

We discuss here data computed along the final 100 ns of the NVT simulations. Regarding inhibitor α -ketoamide moieties, their final ω_{OCCO} values remain close to their starting value (about 160° – 180°) for **13b**, **14b**, **11r**, and **11p**; for **11f** and **13a**, they evolve toward a value $\leq 15^\circ$, see Table II. In Fig. 5, we plot the *ab initio* energy from MP2(FC)/aug-cc-pVDZ computations as a function of ω_{OCCO} for the model molecule $\text{CH}_3\text{-NH-OCCO-CH}_3$. Two snapshots that show the detail of the inhibitor interactions with the protease catalytic dyad from simulations along which ω_{OCCO} evolves toward a weak value ($\leq 30^\circ$) are also provided in Fig. 5.

According to the *ab initio* data regarding $\text{CH}_3\text{-NH-OCCO-CH}_3$, conformations corresponding to ω_{OCCO} values $\leq 60^\circ$ are less stable than those for which $\omega_{\text{OCCO}} = 160^\circ$ – 180° by about 10 kcal mol^{-1} . In addition to strong local interactions between ketoamide oxygens and the imidazolium hydrogen HE2 (PDB labeling) of His¹⁴⁵, inhibitor conformations within the active site corresponding to weak ω_{OCCO} values are also stabilized by hydrogen bonds with the lateral chains or the backbone residues such as His¹⁶³/Asn¹⁴², see Fig. 5. Experimentally, all the reported crystallographic ketoamide inhibitor/protease structures correspond to ω_{OCCO} values of about 180° .^{48,51} However, the inhibitors are covalently bonded to the thiolate sulfur of Cys¹⁴⁵ yielding a less hindered rotation around the ketoamide OC–CO bond than in free inhibitors. As we model the interaction between the inhibitors and the catalytic dyad as being purely ionic, our results regarding ω_{OCCO} values $\leq 90^\circ$ does not prejudice the inhibitor conformational behavior once covalently bonded to M^{Pro} .

TABLE II. Inhibitor moiety spatial volumes values (\bar{V}_m) computed along MD simulations (in \AA^3), and ω_{OCCO} mean values along the last 10 ns simulation segments for the ketoamide moiety $M_{\text{ketoamide}}$ (the OCCO RMSD values are $\pm 20^\circ$).

	13b	13a	14b	11r	11p	11f
\bar{V}_m data and mean OCCO dihedral angle values						
$M_{\text{ketoamide}}$ (deg)	0.10–180	0.18–15	0.04–160	0.04–180	0.02–180	0.09–0
M_{amide}	0.04	0.25	0.09	0.04	0.02	0.17
M_{S1}	0.05	0.59	0.34	0.04	0.05	0.39
$M_{S1'}$	0.05	...	0.05	0.07	0.13	0.14
M_{S2}	0.05	0.15	0.11	0.06	0.06	...
M_{S4}	0.10	0.36	0.22	0.04	0.03	...
Moiety definition						
$M_{\text{ketoamide}}$	OCCO–NH backbone					
M_{amide}	OC–NH backbone ending group					
M_{S1}	Lactam side chain					
$M_{S1'}$	Phenyl	Cyclo-propyl	Phenyl	Phenyl	Phenyl	Phenyl
M_{S2}	Cyclo-propyl	Cyclo-hexyl	Cyclo-propyl	Cyclo-hexyl	Acetonitril	...
M_{S4}	Pyridone + Boc	Pyridone + Boc	Pyridone + Boc	Cinnamoyl	Cinnamoyl	Boc

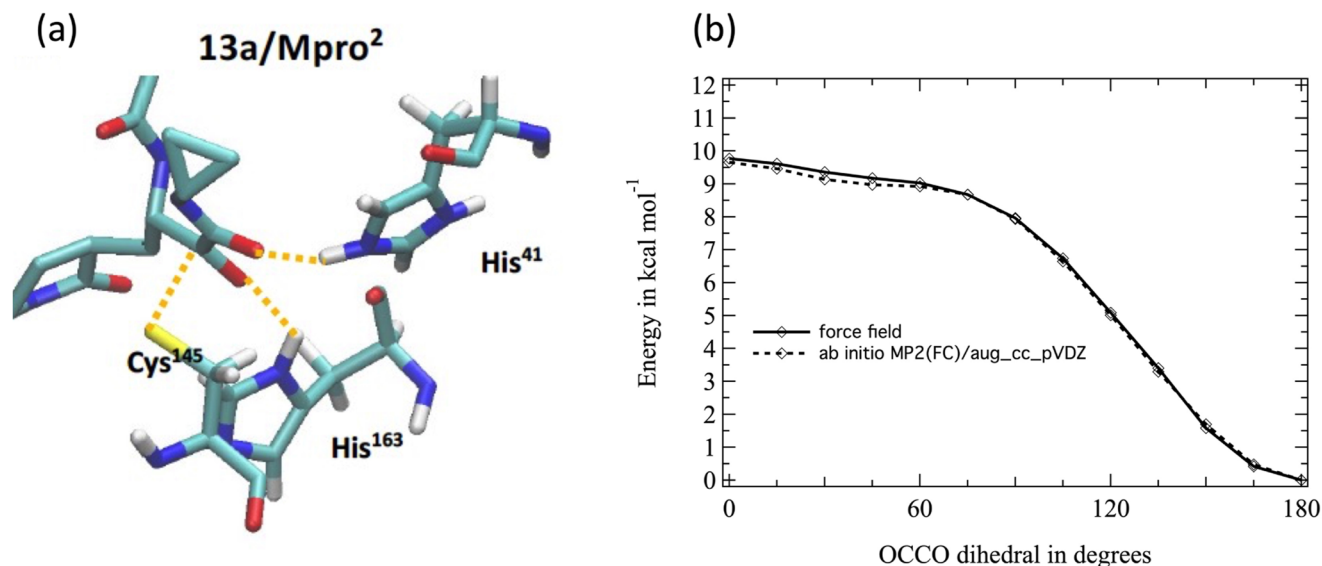


FIG. 5. (a) Details of the local interactions between **13a** and the M^{Pro} catalytic residues His⁴¹ and Cys¹⁴⁵. (b) Energy profiles with regard to ω_{OCCO} for the model molecule $\text{CH}_3\text{-NH-OCO-CH}_3$ from *ab initio* MP2(FC)/aug-cc-pVDZ computations (dashed line) and from our force field (solid line). The mean difference between force field and *ab initio* energies is here $0.07 \text{ kcal mol}^{-1}$ (the corresponding mean standard deviation is $0.1 \text{ kcal mol}^{-1}$), and the root mean square deviation between both sets of energy values is $0.12 \text{ kcal mol}^{-1}$.

Regardless of the inhibitor OCCO conformation, both the local interaction **LI**₁ between the His⁴¹ imidazolium group and the inhibitor ketoamide oxygens and the interaction **LI**₂ between the thiolate sulfur of Cys¹⁴⁵ and the ketoamide carbons are preserved along all the simulations, as shown by the plots of Fig. 6, with the exception of interaction **LI**₁ in the particular case of **11f** (the smallest inhibitor that demonstrates no inhibitory potency). For instance, the distance R_1 between the imidazolium hydrogen HE2 of His⁴¹ and the ketoamide oxygen O₁ with which it interacts the most closely in the starting complex structure is about $1.8(2) \text{ \AA}$ along all our simulations. The distances R_2 between the thiolate sulfur of Cys¹⁴⁵ and the two ketoamide sp² carbons are all about $3.0(2) \text{ \AA}$ (with the exception of **11f**), with a dissymmetry of the order of 0.2 \AA . All these data show the robustness of the ionic interaction pattern between the inhibitor ketoamide group and the lateral chains of His⁴¹ and Cys¹⁴⁵, regardless of the inhibitor.

Regarding the inhibitor side chains, they all maintain their starting position and preserve their interactions with individual M^{Pro} pockets except for **14b** and **11p**. For **14b**, its γ -actam group moves to mainly interact with the solvent whereas its terminal phenyl group moves from its starting position within pocket S4 to an intermediate position in between pockets S1 and S4. Regarding the terminal phenyl group of **11p**, it evolves from pocket S1' toward pocket S2 (see Fig. 7). For the latter three groups, their structural transition within the active site is achieved within the first 10 ns segment of the MD simulations.

To further investigate the inhibitor behavior within the active site, we compute the spatial volumes \bar{V}_m explored by the center of masses (COMs) of the inhibitor main moieties, such as the γ -lactam side chain and the ketoamide “backbone” moieties. These groups are here defined from common chemical arguments and not from the

QMF procedure detailed in Sec. II C. They are denoted M_{S_m} where S_m is the largest pocket to which the moiety is the closest to in the simulation starting structures (see Fig. 1). This arbitrary definition is made for readability purpose (as well as the assignment of the ending cyclopropyl group of **13a** to its $M_{\text{ketoamide}}$ moiety). Besides the M_{S_m} moieties, we define two inhibitor “backbone” moieties that are denoted $M_{\text{ketoamide}}$ and M_{amide} , respectively. The volumes \bar{V}_m are computed from the RMSDs (termed δx , δy , and δz) of the moiety COM Cartesian positions along the simulations according to

$$\bar{V}_m = \frac{4}{3} \pi \delta x \delta y \delta z. \quad (2)$$

For this calculation, the M^{Pro}/inhibitor complex structures along a MD trajectory are projected on to orthogonal axis frameworks for which the RMSD (computed from the starting simulation structures) of the backbone C _{α} positions is the weakest. The moiety definition and the \bar{V}_m values are summarized in Table II.

Estimates of mean position uncertainties in experimental crystallographic structures range from 0.2 to 0.4 \AA for x-ray resolutions of 1.6 and 1.7 \AA , respectively.^{91,92} That corresponds to an equivalent spherical volume \bar{V} of at most 0.08 \AA^3 . This value is an upper bound of most of our \bar{V}_m data (see Table II) and, in particular, for **13b** from which we built all the M^{Pro}/inhibitor starting structures (the resolution of the experimental x-ray crystallographic complex structure 6Y2G for inhibitor **13b** is 2.2 \AA). Even if our \bar{V}_m values for **13a** are larger than \bar{V} , they all show the structural stability of the inhibitors within the M^{Pro} pockets along our 100 ns simulations. Note, however, the length of our simulations prevents us from forming conclusions about the thermodynamic nature of the

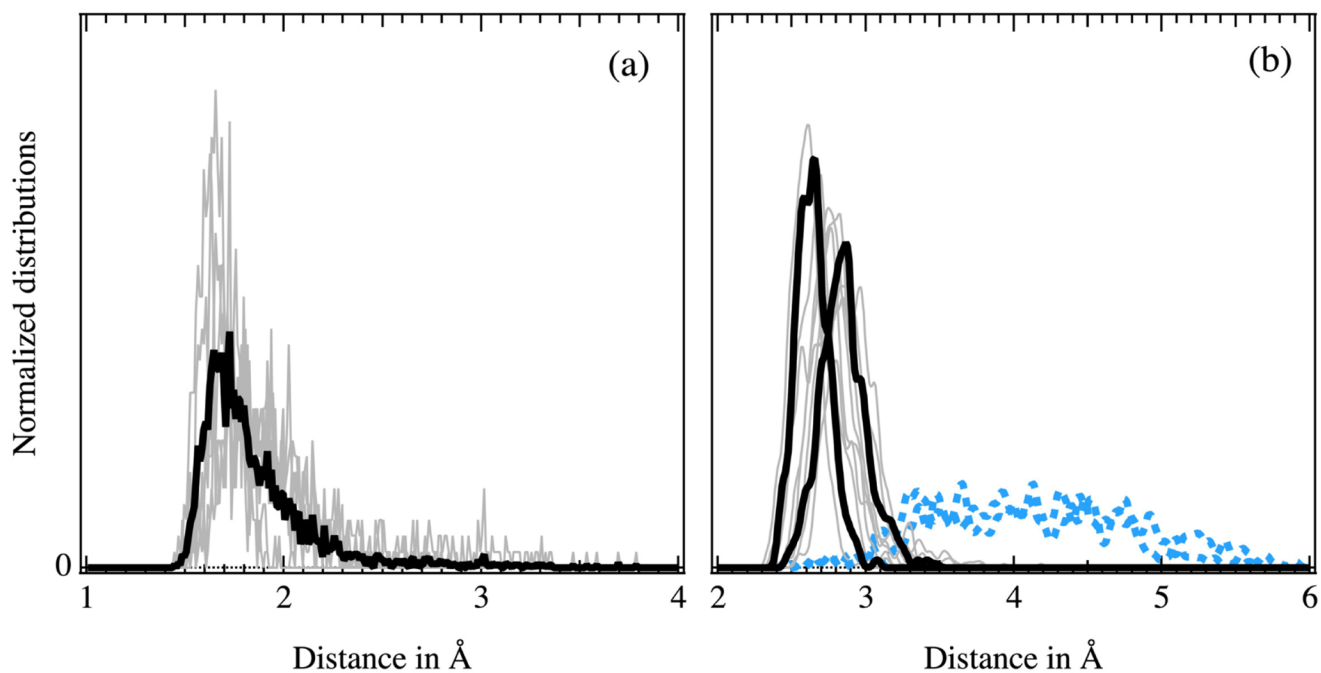


FIG. 6. Geometrical details of the ionic interactions between the M^{Pro} catalytic dyad and the inhibitor ketoamide moiety. (a) Normalized distributions of the distance R_1 between the hydrogen HE2 (PDB labeling) of His⁴¹ and the ketoamide oxygen O₁ [see Fig. 2(a) for definition]. (b) Normalized distributions of the distances R_2 between the thiolate sulfur of Cys⁴¹ and the two ketoamide sp² carbons. Light gray and bold black lines: distributions for individual inhibitor and averaged over all inhibitors, respectively, with the exception of **11f** whose specific R_2 distributions are shown in blue dashed lines.

M^{Pro} /inhibitor binding patterns here investigated (i.e., true minima or transient states).

Regarding the particular case of inhibitors **11f** and **11p**, our simulation data support the experimental evidence that the lack of inhibitory potency (as experimentally reported for those inhibitors) does not necessarily correspond to an inability of an α -ketoamide inhibitor to form relatively stable complexes with the main proteases of coronaviruses.⁴⁸ Moreover and contrary to recent non-polarizable force field simulations,^{93–95} our simulations also show the stability of the ionic interaction between the inhibitor ketoamide moieties and the activated (charged) residues of the catalytic dyad. This supports the existence of an equilibrium between a state corresponding to α -ketoamide inhibitors covalently bonded to M^{Pro} and a pre-reactive “ionic” state. We may also note here recent MD simulations showing the presence of the α -ketoamide **13b** within the M^{Pro} active site to stabilize the catalytic dyad in its ionic form by a few kcal mol⁻¹.⁹⁶

V. QM FRAGMENTS AND M^{Pro} /INHIBITORS INTERACTIONS

Our QMF approach enables us to quantify the strength of interactions among fragments of a molecular system. By selecting a target region it is possible to identify sets of fragments that share a non-negligible fraction of the electronic density; in particular, we identify fragment pairs for which the (symmetrized) FBO value is larger than 0.007 atomic units. Note FBO values for standard hydrogen bonds are included within 0.5×10^{-2} , and 1×10^{-2} , as discussed in Ref. 59.

Here, we present such FBO data for the five inhibitors that are able to interact with the overall large portion of the active site (i.e., all inhibitors except of **11f**). Our aim is to compare the role of inhibitor side chains on the M^{Pro} /inhibitor complex stability. All the mean FBO values discussed below correspond to averages performed on sets of 200 M^{Pro} /inhibitor complex snapshots regularly extracted along the 100 ns NVT simulations discussed above, after discarding the first 10 ns.

A. Inhibitor QM fragments

Besides the nature of the atoms and the conformation of a molecular system, the results of the CRF fragmentation process mainly depend on the magnitude of the purity index. Here, we use the purity index defined in Sec. II C (8%) to automatically generate a set of stable QM fragments for M^{Pro} along our MD trajectories. Those fragments generally correspond to single residues (with the exceptions corresponding to those identified using crystallographic structures, see Sec. II C).

As expected from the overall weak fluctuations of the inhibitor moiety positions within the active site, the inhibitor QM fragments are also consistent within the entire extracted snapshot sets. Those fragments are shown in Fig. 8. Interestingly, they correspond to standard (intuitive) chemical groups (like phenyl). Of course, the final fragmentation is not unique, and slightly altering the purity index or generation procedure may lead to alternative fragment sets. This arbitrariness explains why a methyl connecting two “chemical”

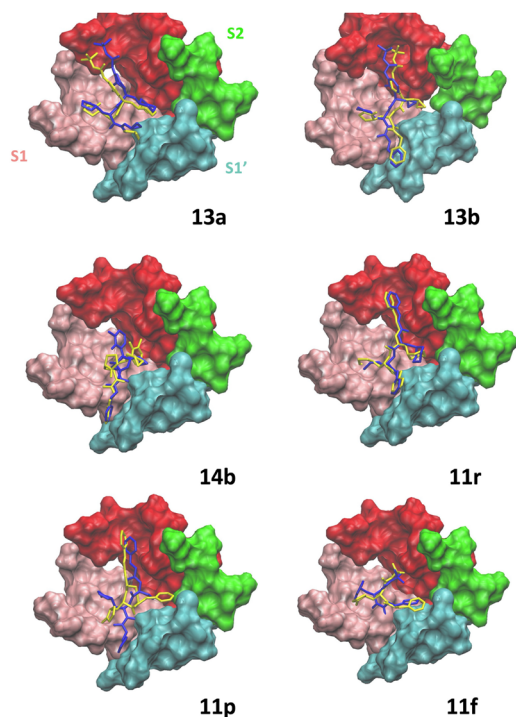


FIG. 7. Inhibitor conformations within M^{Pro} pockets. The inhibitors are shown in a licorice representation (yellow: starting structure, blue: final MD/MM structure). The $S2'-S5'$ pockets are represented together as pocket $S1'$ in for readability purpose.

groups may belong to one of these groups for a particular inhibitor and to a second group for another as observed for the ending phenyl group of **11p** and **11p** connected to ketoamide moiety, for instance. Thus, the interaction of an inhibitor within the active site may be interpreted as arising from a set of temporally stable fragments that can be used as two “fragment basis sets” (one for the protease and one for the inhibitor) from which to decompose the global M^{Pro} /inhibitor interaction into a sum of almost “pure” pairwise interactions (within the confidence level defined by the purity index).

B. Protease/inhibitor interactions

To quantify M^{Pro} /inhibitor interactions from the protease point of view, we sum up the mean FBO values for each protease residue with the inhibitor fragments. The magnitudes of the resulting $\sigma(\text{FBO})_{\text{res}}$ data are shown in Fig. 9(a). As expected from the local nature of the FBO quantity, only a reduced set of residues strongly interact with the inhibitor: their $\sigma(\text{FBO})$ values range from 0.01 up to about $0.2e$ in the particular case of the catalytic residue Cys¹⁴⁵.

FBO data also allow one to investigate M^{Pro} /inhibitor interactions from the protease pocket point of view: for that purpose we sum for each inhibitor the $\sigma(\text{FBO})_{\text{res}}$ data corresponding to the pocket leading residues that we have defined in Table I. Those new values are denoted $\sigma(\text{FBO})_{\text{pocket}}$ and their schematic representation provides a simple way to readily compare the strength

of protease pocket/inhibitor interactions [see Fig. 9(b)]. Note here that $\sigma(\text{FBO})_{\text{pocket}}$ data correspond (with the exception of pocket S3) to super-fragments and thus to a lower purity index. Stated otherwise, the pocket data are better defined QM quasi-observables. As expected, the $\sigma(\text{FBO})_{\text{pocket}}$ data corresponding to the “catalytic pocket” S1 are the strongest because of our choice to model the catalytic dyad as charged residues. For other pockets, the magnitude of the M^{Pro} /inhibitor interactions decreases in the series $S2 > S4 \approx S3 > S2'-S5' > S1'$. The magnitude of each set of $\sigma(\text{FBO})_{\text{pocket}}$ data is remarkably close, regardless of the inhibitor. These data clearly show that the inhibitor side chains to interact more strongly with pocket S2 and they also show the pivotal role of residue Glu¹⁶⁶ (“pocket” S3) to anchor the inhibitors within the active site; because of the anionic head of Glu¹⁶⁶, the magnitude of $\sigma(\text{FBO})_{S3}$ value is close to that of pocket S4 that comprises six leading residues.

The spread among the $\sigma(\text{FBO})_{\text{pocket}}$ values is large within each MD snapshot set: it amounts to about 20% of the mean $\sigma(\text{FBO})_{\text{pocket}}$ values. FBO data, as with the Natural Bond Orders (NBOs),⁹⁷ are mostly sensitive to inter-atomic distances (see, among others, Ref. 98). As discussed above, the volumes \bar{V}_m explored by inhibitor side chain COMs along MD simulations agree with those computed from the atomic position uncertainty in experimental crystallographic x-ray structures. In Fig. 9(c), we compare the $\sigma(\text{FBO})_{\text{pocket}}$ data for **13b** averaged over MD snapshots to those computed from a single crystallographic geometry (PDB structure 6Y2G); we find that the “crystallographic” FBO points are all included within the range of MD snapshot-based values, including the $\sigma(\text{FBO})_{\text{res}}$ residue data (see also Fig. 10). The present spread in $\sigma(\text{FBO})_{\text{pocket}}$ values from MD snapshots can serve as an estimate of the uncertainty regarding FBO data (and other related quantities, such as NBO) computed from a single experimental geometry. On one hand, this means that such single structure-based FBO data may be used to get quick insight into the local interaction networks stabilizing a molecular assembly; on the other hand, the inclusion of statistical distribution of these data thanks to the molecular modeling enables to further discuss the relative weight of interfragment interactions. For instance, for **13b** the single structure-based FBO data suggest the strength of the interactions between pockets S2, S3, S4, and $S2'-S5'$ are much closer than averaged data from MD simulations.

For the present M^{Pro} /inhibitor complexes, we have seen that the short-range interaction components that are related to the chemical bond (including D3 dispersion energies) represent the most important contribution to the ligand-enzyme interaction (see Fig. S13), and correlate with the magnitudes of the corresponding FBO data (see Fig. 9). In particular, the FBO uncertainty when using a single geometry, about 20%, is still valid for chemical interaction energy components [see Fig. 9(c)].

The FBO data suggest that inhibitor interactions with pocket S4 are not energetically optimal: the $\sigma(\text{FBO})_{S4}$ values are about twice smaller than $\sigma(\text{FBO})_{S2}$ (and even negligible for **14b**) and of the same order of magnitude as the “single fragment” pocket S3 (pocket S4 comprises six leading fragments). Such consideration are based on interactions that are not necessarily sterically obvious, as they are not directly based on inter-fragment distances, and yet enable to distinguish the relative weight of interaction of vicinal residues. As MD snapshot-based FBO values are in line with experimental single structure-based ones for **13b**, this suggests the latter result to not arise from a drawback tied to our MM approach and/or to our

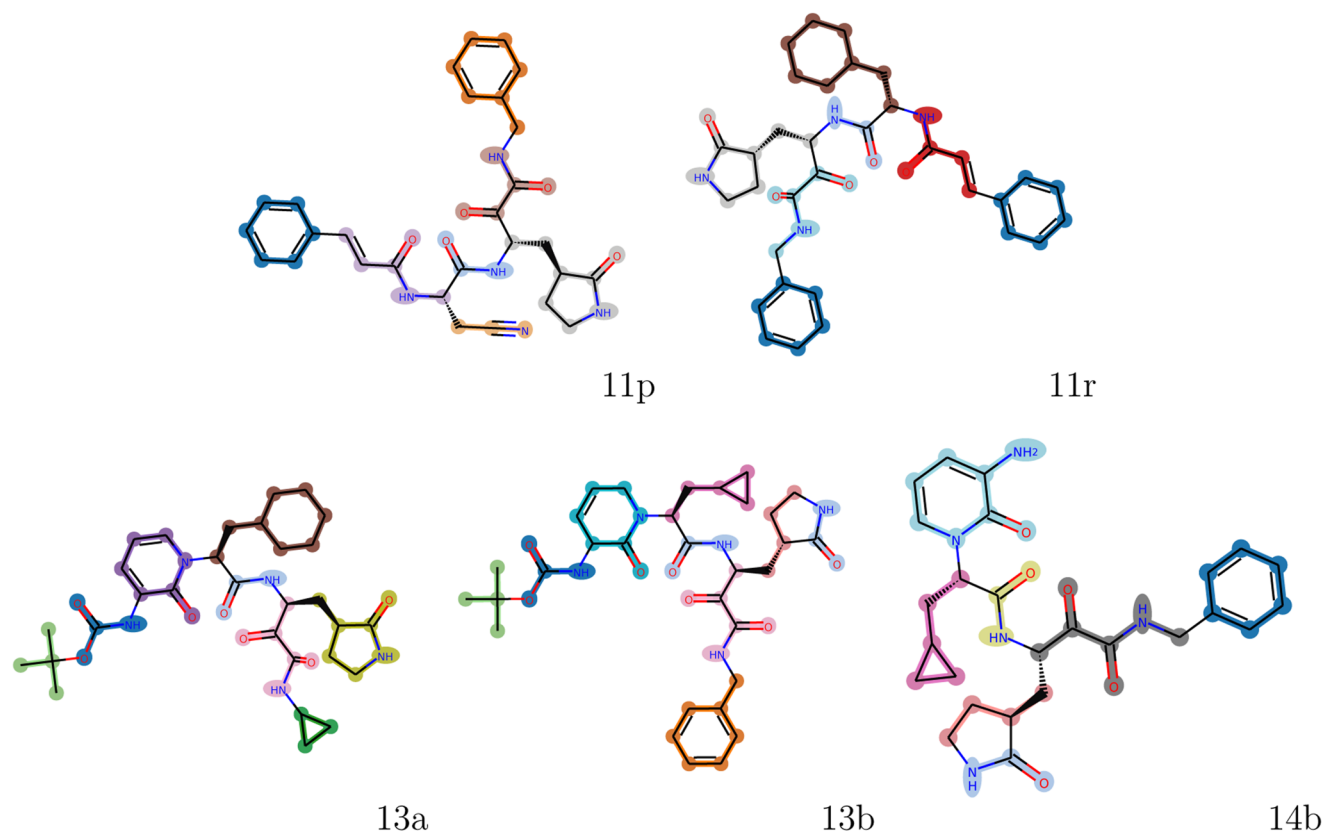


FIG. 8. QM fragmentation (color coded) of the inhibitors emerging from the processing of the MD trajectories using a purity index of 8%.

MD protocol. Hence, we may conclude that the five α -ketoamide inhibitors here investigated may be further improved to strengthen local interactions with pocket S4 following the method proposed in Ref. 99, for instance. The same conclusion may be drawn for inhibitor interactions with pockets S1' and S2'-S5', though the weakness of those interactions was expected from the fact that the drug protrude less in the direction of these pockets (see Fig. 1).

VI. CONCLUSION AND PERSPECTIVE

Efficient DFT linear scaling approaches, such as our own implemented in BigDFT, are promising theoretical tools to investigate large molecular assemblies such as protein/ligand complexes. They are also well suited to be used on modern massively parallel computing systems. These DFT methods remain, however, computationally demanding. It is unrealistic, if not unnecessary, to expect DFT calculations to replace commonly used force field-based methods in view of the need for exhaustive sampling of the configuration space. However, they do allow one to post-process large enough sets of molecular conformations extracted from force field-based simulations using original theoretical tools, such as our complexity reduction framework. The CRF method yields a fragmentation

of a molecular system and measures of interaction between pairs of identified fragments.

Here, we have considered a hybrid MM/QMF computational scheme, which sequentially couples a multi-scale polarizable MM approach together with our DFT linear scaling and post-processing methods to investigate a set of five new α -ketoamide inhibitors targeting the main protease M^{Pro} of the SARS-CoV-2 virus. The MM approach relies on an *all atom ab initio*-based force field to handle M^{Pro} and the inhibitors and on a coarse-grained scheme to account for the solvent. The reliability and the coherence of that hybrid scheme were assessed by cross-checking the predictions of the polarizable MM and DFT approaches regarding mean interaction energies of a large set of M^{Pro} /inhibitor complex snapshots extracted along MD simulations, as well as by assessing the reliability of both methods to reproduce high-end quantum *ab initio* computations for small molecular aggregates and a set of α -ketoamide inhibitor conformations.

Using the multi-scale polarizable MM approach, we performed MD simulations of the M^{Pro} /inhibitor complexes in aqueous phase at the 100 ns scale. Along those simulations, the inhibitor conformations and interaction patterns within the protease active site remain stable. Moreover, the fluctuations of the center of mass positions of the main inhibitor moieties measured along the simulations are

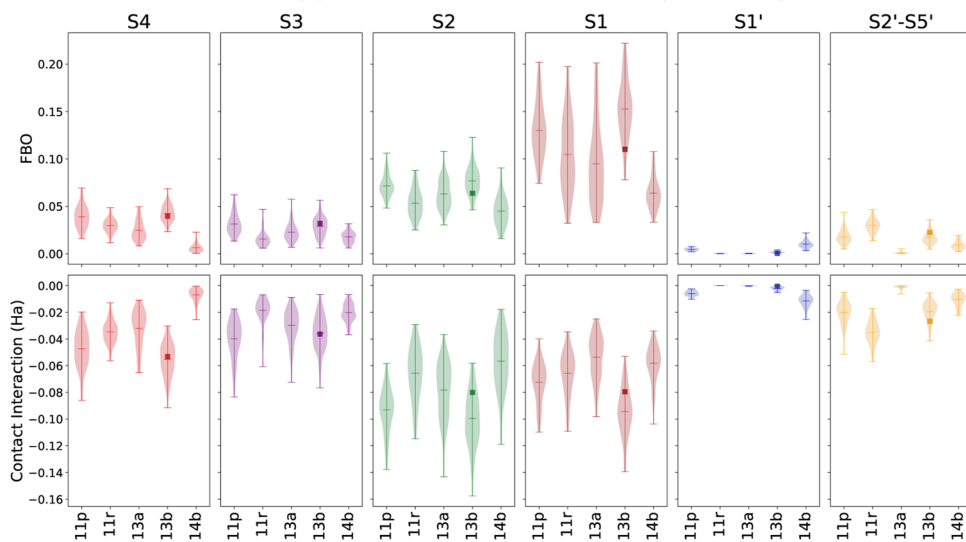
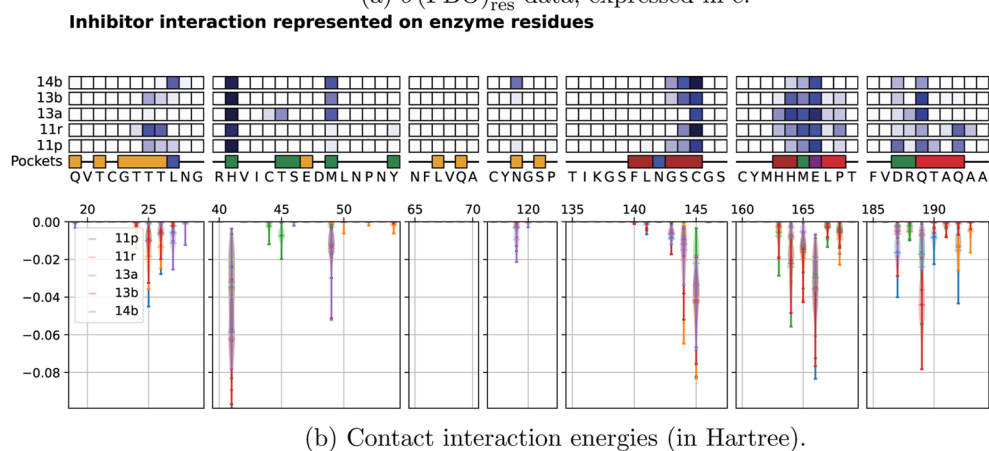
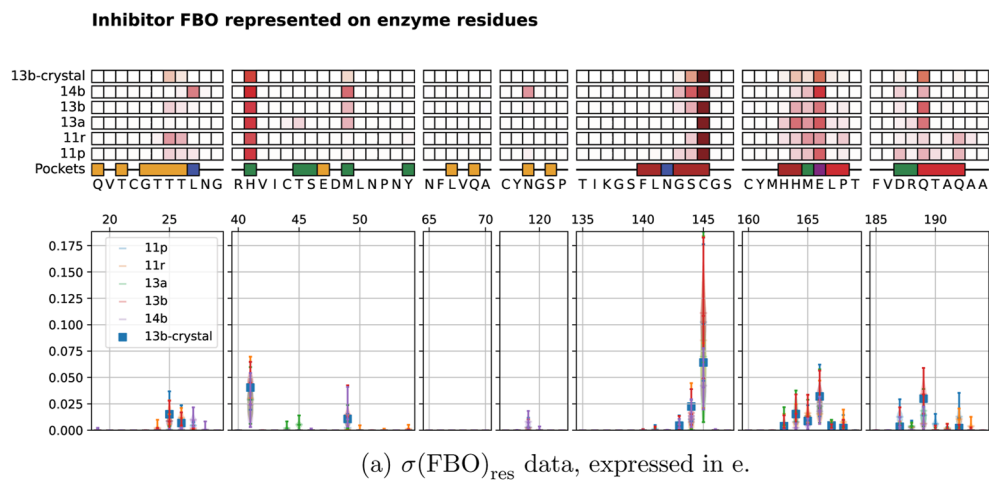


FIG. 9. Interaction of the inhibitors projected on enzyme residues (pockets).

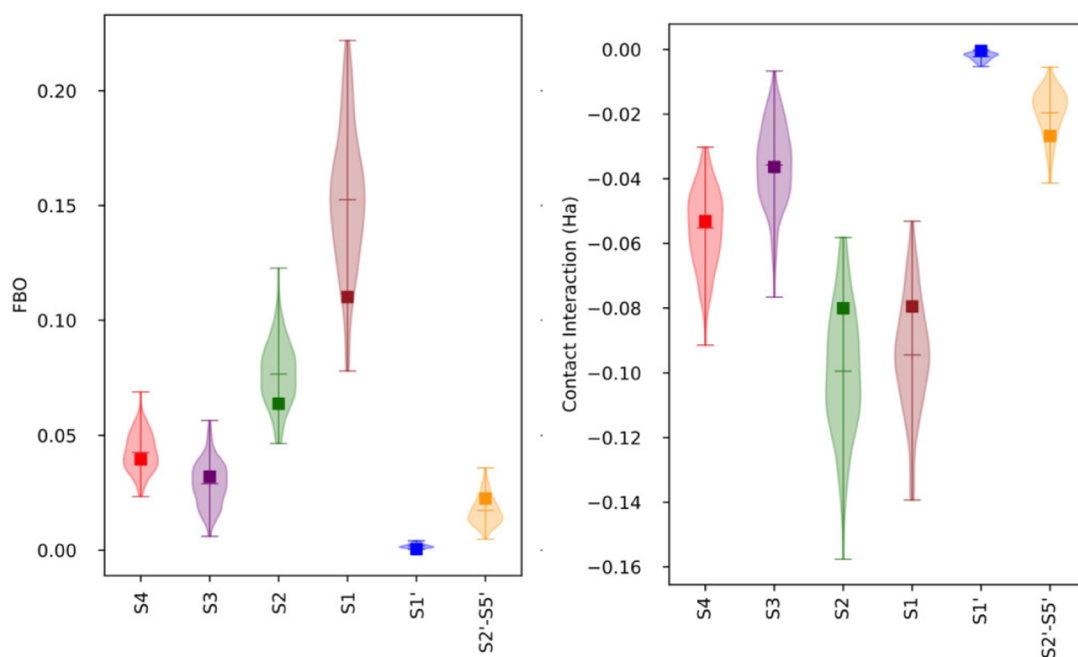


FIG. 10. $\sigma(\text{FBO})_{\text{pocket}}$ (left) and corresponding contact energy (right) data for **13b**. The full square symbols within the violin plots correspond to data computed from the original experimental x-ray crystallographic geometry 6Y2G.

in line with the atomic position uncertainties in crystallographic x-ray structures.^{91,92} Even though we do not demonstrate here the need of using a polarizable force field to simulate protein/ligand complexes, we nevertheless show the ability of polarizable force fields whose parameters are only assigned from quantum *ab initio* results (without considering any experimental data) to provide an overall reliable description of such molecular assemblies. This paves the road toward computationally assessing the potential of new classes of drugs ahead of any experiment.

By post-processing sets of 100 snapshots of the M^{Pro} /inhibitor complexes extracted along our MD simulations, we show the existence of quantum-based fragments temporally stable along the MD simulations (here quantum means the CRF fragmentation procedure). The M^{Pro} fragments are mainly standard residues with very few exceptions, whereas they correspond to standard chemical groups for inhibitors. From inspections of mean FBO data corresponding to M^{Pro} /inhibitor fragment pairs, we show inhibitor side chains to more strongly interact with the M^{Pro} pocket S2 (and with residue Glu¹⁶⁶) rather than pocket S4, regardless of the inhibitor. Even though the same analysis could be carried out on single experimentally resolved crystallographic structures, with the present study we also provide variabilities and uncertainties of the derived quantities, which strengthen our findings. Our present conclusion *a priori* contrasts with a recent large-scale MD simulations study³³ that shows the opposite. Note that

besides simulation artifacts underlined by the authors of the latter study, like the accuracy of their scoring functions and of their pairwise force fields, the set of drugs that were then investigated is *a priori* not able to specifically target pocket S2. Our study also yields an estimate of the uncertainty in FBO data when taking into account a single experimental x-ray structure instead of a large set of plausible geometries for a molecular assembly: that uncertainty amounts to 20%, a value that may be considered for any kind of electronic density analysis scheme (like the NBO one,⁹⁷ for instance). Taking into account such a large uncertainty should yield to mitigate the conclusions drawn as discussing very weak M^{Pro} /ligand interactions from electronic density arguments as in Ref. 100.

As a perspective, we may mention the interesting feature of FBO data to provide useful synthetic graph views of the local interaction networks occurring among the components of a molecular assembly. In Fig. 11, we plot such interaction graphs for the five largest α -ketoamide inhibitors. Theoretical FBO-based interaction graphs can be used to develop new cost functions for docking approaches, for instance. To achieve the present study, all the QMF computations needed the use of about 1% of the total year resource of available ten petaflop computational systems. With the multiple exaflop computational systems that are expected to land within a few years, such large scale QMF computational studies can be routinely performed in order to investigate new classes of original drugs.

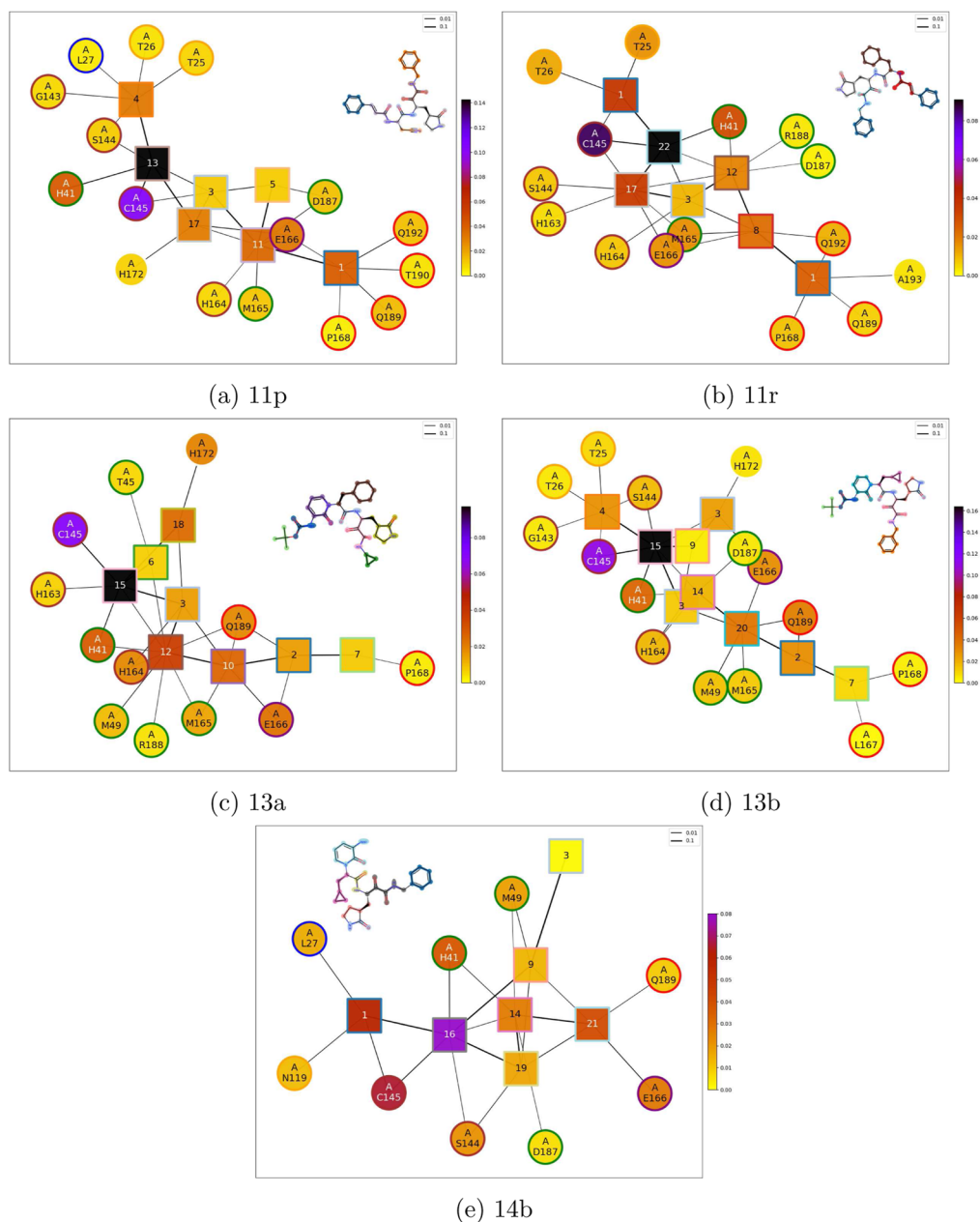


FIG. 11. M^{Pro} /inhibitor interaction networks from FBO data. Inhibitor QM fragments are shown by squares and the M^{Pro} ones (mostly protease residues) are shown by circles. The lines interconnect M^{Pro} /inhibitor fragment pairs whose FBO value is larger than $7 \times 10^{-3}e$. The colors in the inner part of the node, related to the side colorbar, indicate the total value of the fragment FBO with all the connected fragment of the protein. The node edges are associated with the pocket colors and to the inhibitor fragment for the enzyme and the inhibitor nodes, respectively. For the case of the ligand, the fragments are defined in the molecule drawing next to the graph. Node numbers indicate the fragment id as per the QMF fragment library definition that is provided in the supplementary material: (a) 11p, (b) 11r, (c) 13a, (d) 13b, and (e) 14b.

SUPPLEMENTARY MATERIAL

See the supplementary material for more detailed discussions regarding the accuracy of the force field and of the multi-scale

MM approach to model proteins, as well as of the agreement between our force field and PBE computations to model the microscopic interactions occurring in our protease/ligand systems. The final structures of the M^{Pro} /inhibitor complexes along with

all our MD simulations (in PDB format) are freely available at <http://biodev.cea.fr/polaris/download.html>.

ACKNOWLEDGMENTS

M.M. thanks Emmanuel Oseret (Exascale Computing Research Laboratory, a joint INTEL/CEA/UVSQ/GENCI laboratory) for his help in optimizing the code POLARIS(MD). This work was granted access to the TGCC HPC resources under the Allocation 2019–2020 (No. A0070307078) and the Grand Challenge Allocation (No. GC0429) made by GENCI. L.G., T.N., and W.D. gratefully acknowledge the joint CEA-RIKEN collaboration action. We also acknowledge support from the MaX EU Center of Excellence. We are also grateful to the financial support by the Labex CaPPA (Contract No. ANR-11-LABX-0005-01) and I-SITE ULNE project OVERSEE (Contract No. ANR-16-IDEX-0004), and by the CPER CLIMIBIO (European Regional Development Fund, Hauts de France council, French Ministry of Higher Education and Research).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Luigi Genovese: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **William Dawson:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Takahito Nakajima:** Methodology (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal). **Viviana Cristiglio:** Conceptualization (equal); Investigation (equal); Methodology (equal); Supervision (supporting); Validation (equal); Writing – review & editing (equal). **Valérie Vallet:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Michel Masella:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- W. Dawson, A. Degomme, M. Stella, T. Nakajima, L. E. Ratcliff, and L. Genovese, “Density functional theory calculations of large systems: Interplay between fragments, observables, and computational complexity,” *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **12**, e1574 (2022).
- L. E. Ratcliff, S. Mohr, G. Huhs, T. Deutsch, M. Masella, and L. Genovese, “Challenges in large scale quantum mechanical calculations,” *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **7**, e1290 (2017).
- K. Świderek and V. Moliner, “Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 M^{pro} by QM/MM computational methods,” *Chem. Sci.* **11**, 10626–10630 (2020).
- K. Arafet, N. Serrano-Aparicio, A. Lodola, A. J. Mulholland, F. V. González, K. Świderek, and V. Moliner, “Mechanism of inhibition of SARS-CoV-2 M^{pro} by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity,” *Chem. Sci.* **12**, 1433–1444 (2021).
- E. Awoonor-Williams and A. A.-A. Abu-Saleh, “Covalent and non-covalent binding free energy calculations for peptidomimetic inhibitors of SARS-CoV-2 main protease,” *Phys. Chem. Chem. Phys.* **23**, 6746–6757 (2021).
- C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, and I. 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 Catal.* **11**, 4157–4168 (2021).
- C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, and I. Tuñón, “A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design,” *Chem. Sci.* **12**, 3489–3496 (2021).
- H. S. Fernandes, S. F. Sousa, and N. M. F. S. A. Cerqueira, “New insights into the catalytic mechanism of the SARS-CoV-2 main protease: An ONIOM QM/MM approach,” *Mol. Diversity* **26**, 1373–1381 (2022).
- J. Nochebuena and G. A. Cisneros, “Polarizable MD and QM/MM investigation of acrylamide-based leads to target the main protease of SARS-CoV-2,” *J. Chem. Phys.* **157**, 185101 (2022).
- K. Gao, R. Wang, J. Chen, L. Cheng, J. Frishcosy, Y. Huzumi, Y. Qiu, T. Schluckbier, X. Wei, and G.-W. Wei, “Methodology-centered review of molecular modeling, simulation, and prediction of SARS-CoV-2,” *Chem. Rev.* **122**, 11287–11368 (2022).
- S. Martí, K. Arafet, A. Lodola, A. J. Mulholland, K. Świderek, and V. Moliner, “Impact of warhead modulations on the covalent inhibition of SARS-CoV-2 M^{pro} explored by QM/MM simulations,” *ACS Catal.* **12**, 698–708 (2022).
- N. Ye, Z. Yang, and Y. Liu, “Applications of density functional theory in COVID-19 drug modeling,” *Drug Discovery Today* **27**, 1411–1419 (2022).
- H. Sharma, B. Raju, G. Narendra, M. Motiwale, B. Sharma, H. Verma, and O. Silakari, “QM/MM studies on enzyme catalysis and insight into designing of new inhibitors by ONIOM approach: Recent update,” *ChemistrySelect* **8**, e202203319 (2023).
- G. Giudetti, I. Polyakov, B. L. Grigorenko, S. Faraji, A. V. Nemukhin, and A. I. Krylov, “How reproducible are QM/MM simulations? Lessons from computational studies of the covalent inhibition of the SARS-CoV-2 main protease by carmofur,” *J. Chem. Theory Comput.* **18**, 5056–5067 (2022).
- S. Mohr, M. Masella, L. E. Ratcliff, and L. Genovese, “Complexity reduction in large quantum systems: Fragment identification and population analysis via a local optimized minimal basis,” *J. Chem. Theory Comput.* **13**, 4079–4088 (2017).
- W. Dawson, S. Mohr, L. E. Ratcliff, T. Nakajima, and L. Genovese, “Complexity reduction in density functional theory calculations of large systems: System partitioning and fragment embedding,” *J. Chem. Theory Comput.* **16**, 2952 (2020).
- D. G. Fedorov and K. Kitaura, “Pair interaction energy decomposition analysis,” *J. Comput. Chem.* **28**, 222–237 (2007).
- M. J. S. Phipps, T. Fox, C. S. Tautermann, and C.-K. Skylaris, “Intuitive density functional theory-based energy decomposition analysis for protein–ligand interactions,” *J. Chem. Theory Comput.* **13**, 1837–1850 (2017).
- M. Zaccaria, W. Dawson, V. Cristiglio, M. Reverberi, L. E. Ratcliff, T. Nakajima, L. Genovese, and B. Momeni, “Designing a bioremediator: Mechanistic models guide cellular and molecular specialization,” *Curr. Opin. Biotechnol.* **62**, 98–105 (2020).
- M. A. Llanos, M. E. Gantner, S. Rodriguez, L. N. Alberca, C. L. Bellera, A. Talevi, and L. Gavernet, “Strengths and weaknesses of docking simulations in the

- SARS-CoV-2 era: The main protease (Mpro) case study,” *J. Chem. Inf. Model.* **61**, 3758–3770 (2021).
- ²¹E. N. Muratov, R. Amaro, C. H. Andrade, N. Brown, S. Ekins, D. Fourches, O. Isayev, D. Kozakov, J. L. Medina-Franco, K. M. Merz, T. I. Oprea, V. Poroikov, G. Schneider, M. H. Todd, A. Varnek, D. A. Winkler, A. V. Zakharov, A. Cherkasov, and A. Tropsha, “A critical overview of computational approaches employed for COVID-19 drug discovery,” *Chem. Soc. Rev.* **50**, 9121–9151 (2021).
- ²²R. Hatada, K. Okuwaki, Y. Mochizuki, Y. Handa, K. Fukuzawa, Y. Komeiji, Y. Okiyama, and S. Tanaka, “Fragment molecular orbital based interaction analyses on COVID-19 main protease - inhibitor N3 complex (PDB ID: 6LU7),” *J. Chem. Inf. Model.* **60**, 3593–3602 (2020).
- ²³S. Banerjee, “An insight into the interaction between α -ketoamide-based inhibitor and coronavirus main protease: A detailed *in silico* study,” *Biophys. Chem.* **269**, 106510 (2021).
- ²⁴K. Fukuzawa, K. Kato, C. Watanabe, Y. Kawashima, Y. Handa, A. Yamamoto, K. Watanabe, T. Ohyama, K. Kamisaka, D. Takaya, and T. Honma, “Special features of COVID-19 in the FMOB: Fragment molecular orbital calculations and interaction energy analysis of SARS-CoV-2-related proteins,” *J. Chem. Inf. Model.* **61**, 4594–4612 (2021).
- ²⁵T. Ishikawa, H. Ozono, K. Akisawa, R. Hatada, K. Okuwaki, and Y. Mochizuki, “Interaction analysis on the SARS-CoV-2 spike protein receptor binding domain using visualization of the interfacial electrostatic complementarity,” *J. Chem. Phys. Lett.* **12**, 11267–11272 (2021).
- ²⁶S. Hwang, S.-H. Baek, and D. Park, “Interaction analysis of the spike protein of delta and omicron variants of SARS-CoV-2 with hACE2 and eight monoclonal antibodies using the fragment molecular orbital method,” *J. Chem. Inf. Model.* **62**, 1771–1782 (2022).
- ²⁷K. Watanabe, C. Watanabe, T. Honma, Y.-S. Tian, Y. Kawashima, N. Kawashita, T. Takagi, and K. Fukuzawa, “Intermolecular interaction analyses on SARS-CoV-2 spike protein receptor binding domain and human angiotensin-converting enzyme 2 receptor-blocking antibody/peptide using fragment molecular orbital calculation,” *J. Chem. Phys. Lett.* **12**, 4059–4066 (2021).
- ²⁸C. Watanabe, Y. Okiyama, S. Tanaka, K. Fukuzawa, and T. Honma, “Molecular recognition of SARS-CoV-2 spike glycoprotein: Quantum chemical hot spot and epitope analyses,” *Chem. Sci.* **12**, 4722–4739 (2021).
- ²⁹S. J. Fox, C. Pittock, C. S. Tautermann, T. Fox, C. Christ, N. O. J. Malcolm, J. W. Essex, and C.-K. Skylaris, “Free energies of binding from large-scale first-principles quantum mechanical calculations: Application to ligand hydration energies,” *J. Phys. Chem. B* **117**, 9478–9485 (2013).
- ³⁰S. Tokutomi, K. Shimamura, K. Fukuzawa, and S. Tanaka, “Machine learning prediction of inter-fragment interaction energies between ligand and amino acid residues on the fragment molecular orbital calculations for Janus kinase – inhibitor complex,” *Chem. Phys. Lett.* **757**, 137883 (2020).
- ³¹S. Tanaka, S. Tokutomi, R. Hatada, K. Okuwaki, K. Akisawa, K. Fukuzawa, Y. Komeiji, Y. Okiyama, and Y. Mochizuki, “Dynamic cooperativity of ligand–residue interactions evaluated with the fragment molecular orbital method,” *J. Phys. Chem. B* **125**, 6501–6512 (2021).
- ³²K. Takaba, C. Watanabe, A. Tokuhisa, Y. Akinaga, B. Ma, R. Kanada, M. Araki, Y. Okuno, Y. Kawashima, H. Moriwaki, N. Kawashita, T. Honma, K. Fukuzawa, and S. Tanaka, “Protein-ligand binding affinity prediction of cyclin-dependent kinase-2 inhibitors by dynamically averaged fragment molecular orbital-based interaction energy,” *J. Comput. Chem.* **43**, 1362 (2022).
- ³³T. Huynh, H. Wang, and B. Luan, “*In silico* exploration of the molecular mechanism of clinically oriented drugs for possibly inhibiting SARS-CoV-2’s main protease,” *J. Phys. Chem. Lett.* **11**, 4413–4420 (2020).
- ³⁴S. A. Khan, K. Zia, S. Ashraf, R. Uddin, and Z. Ul-Haq, “Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach,” *J. Biomol. Struct. Dyn.* **39**, 2607 (2020).
- ³⁵D. Gentile, V. Patamia, A. Scala, M. T. Sciortino, A. Piperno, and A. Rescifina, “Putative inhibitors of SARS-CoV-2 main protease from a library of marine natural products: A virtual screening and molecular modeling study,” *Mar. Drugs* **18**, 225 (2020).
- ³⁶S. Kumar, P. P. Sharma, U. Shankar, D. Kumar, S. K. Joshi, L. Pena, R. Durvasula, A. Kumar, P. Kempaiah, Poonam, and B. Rathi, “Discovery of new hydroxyethylamine analogs against 3CL^{pro} protein target of SARS-CoV-2: Molecular docking, molecular dynamics simulation, and structure-activity relationship studies,” *J. Chem. Inf. Model.* **60**, 5754–5770 (2020).
- ³⁷D. E. Shaw Research, Molecular Dynamics Simulations Related to SARS-CoV-2, D. E. Shaw Research Technical Data, 2020; http://www.deshawresearch.com/resources_sarscov2.html.
- ³⁸J. A. Lemkul and A. D. MacKerell, “Polarizable force field for DNA based on the classical Drude oscillator: I. Refinement using quantum mechanical base stacking and conformational energetics,” *J. Chem. Theory Comput.* **13**, 2053–2071 (2017).
- ³⁹V. S. Inakollu, D. P. Geerke, C. N. Rowley, and H. Yu, “Polarisable force fields: What do they add in biomolecular simulations?,” *Curr. Opin. Struct. Biol.* **61**, 182–190 (2020).
- ⁴⁰A. N. Singh and A. Yethiraj, “Driving force for the complexation of charged polypeptides,” *J. Phys. Chem. B* **124**, 1285–1292 (2020).
- ⁴¹M. Masella, A. Crudu, and F. Léonforté, “Hybrid polarizable simulations of a conventional hydrophobic polyelectrolyte. Toward a theoretical tool for green science innovation,” *J. Chem. Phys.* **155**, 114903 (2021).
- ⁴²M. Masella and P. Cuniasse, “A many-body model to study proteins. I. Applications to ML_n^{m+} complexes, $M^{m+} = Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}$, and Zn^{2+} , $L = H_2O, CH_3OH, HCONH_2$, $n = 1–6$, and to small hydrogen bonded systems,” *J. Chem. Phys.* **119**, 1866–1873 (2003).
- ⁴³M. Masella, D. Borgis, and P. Cuniasse, “Combining a polarizable force-field and a coarse-grained polarizable solvent model: Application to long dynamics simulations of bovine pancreatic trypsin inhibitor,” *J. Comput. Chem.* **29**, 1707–1724 (2008).
- ⁴⁴M. Masella, D. Borgis, and P. Cuniasse, “Combining a polarizable force-field and a coarse-grained polarizable solvent model. II. Accounting for hydrophobic effects,” *J. Comput. Chem.* **32**, 2664–2678 (2011).
- ⁴⁵L. E. Ratcliff, W. Dawson, G. Fiscaro, D. Caliste, S. Mohr, A. Degomme, B. Videau, V. Cristiglio, M. Stella, M. D’Alessandro, S. Goedecker, T. Nakajima, T. Deutsch, and L. Genovese, “Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations,” *J. Chem. Phys.* **152**, 194110 (2020).
- ⁴⁶L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, and R. Schneider, “Daubechies wavelets as a basis set for density functional pseudopotential calculations,” *J. Chem. Phys.* **129**, 014109 (2008).
- ⁴⁷M. Robello, E. Barresi, E. Baglini, S. Salerno, S. Taliani, and F. D. Settimo, “The alpha keto amide moiety as a privileged motif in medicinal chemistry: Current insights and emerging opportunities,” *J. Med. Chem.* **64**, 3508–3545 (2021).
- ⁴⁸L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E. J. Snijder, H. Liu, and R. Hilgenfeld, “ α -ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: Structure-based design, synthesis, and activity assessment,” *J. Med. Chem.* **63**, 4562–4578 (2020).
- ⁴⁹T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, and S.-H. Jung, “An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy,” *J. Med. Chem.* **59**, 6595–6628 (2016).
- ⁵⁰A. Tichá, S. Stanchev, K. R. Vinothkumar, D. C. Mikles, P. Pächl, J. Began, J. Škerle, K. Švehlová, M. T. N. Nguyen, S. H. L. Verhelst, D. A. Johnson, D. A. Bachovchin, M. Lepšík, P. Majer, and K. Strisovsky, “General and modular strategy for designing potent, selective, and pharmacologically compliant inhibitors of rhomboid proteases,” *Cell Chem. Biol.* **24**, 1523–1536.e4 (2017).
- ⁵¹L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerherring, S. Becker, K. Rox, and R. Hilgenfeld, “Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors,” *Science* **368**, 409–412 (2020).
- ⁵²Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, and H. Yang, “Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors,” *Nature* **582**, 289–293 (2020).
- ⁵³Z. 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. Biol.* **27**, 529–532 (2020).
- ⁵⁴W. Dai, B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, 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.-K. Zhang, Y. Xu, H. Yang, and H. Liu, "Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease," *Science* **368**, 1331–1335 (2020).
- ⁵⁵S. VanPatten, M. He, A. Altiti, K. F. Cheng, M. H. Ghanem, and Y. Al-Abed, "Evidence supporting the use of peptides and peptidomimetics as potential SARS-CoV-2 (COVID-19) therapeutics," *Future Med. Chem.* **12**, 1647–1656 (2020).
- ⁵⁶L. Zhang, D. Lin, and R. Hilgenfeld, "Structure of human coronavirus NL63 main protease in complex with the alpha-ketoamide tert-butyl ((S)-4-(benzylamino)-3,4-dioxo-1-((S)-2-oxopyrrolidin-3-yl)butan-2-yl)carbamate (tert-butyl -GlnLactam-CO-CO-NH-benzyl)," <http://doi.org/10.2210/pdb6fv2/pdb> (2018).
- ⁵⁷C. Houriez, V. Vallet, F. Réal, M. Meot-Ner (Mautner), and M. Masella, "Organic ion association in aqueous phase and *ab initio*-based force fields: The case of carboxylate/ammonium salts," *J. Chem. Phys.* **147**, 161720 (2017).
- ⁵⁸T. Gallagher, P. Alexander, P. Bryan, and G. L. Gilliland, "Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR," *Biochemistry* **33**, 4721–4729 (1994).
- ⁵⁹H. T. H. Chan, M. A. Moesser, R. K. Walters, T. R. Malla, R. M. Twidale, T. John, H. M. Deeks, T. Johnston-Wood, V. Mikhailov, R. B. Sessions, W. Dawson, E. Salah, P. Lukacik, C. Strain-Damerell, C. D. Owen, T. Nakajima, K. Świderek, A. Lodola, V. Moliner, D. R. Glowacki, J. Spencer, M. A. Walsh, C. J. Schofield, L. Genovese, D. K. Shoemark, A. J. Mulholland, F. Duarte, and G. M. Morris, "Discovery of SARS-CoV-2 M^{pro} peptide inhibitors from modelling substrate and ligand binding," *Chem. Sci.* **12**, 13686–13703 (2021).
- ⁶⁰M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX* **1–2**, 19–25 (2015).
- ⁶¹K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. Klepeis, R. Dror, and D. Shaw, "Improved side-chain torsion potentials for the amber ff99SB protein force field," *Proteins* **78**, 1950–1958 (2010).
- ⁶²See <http://biodev.cea.fr/polaris/> for information about the POLARIS code.
- ⁶³M. Masella, "The multiple time step r-RESPA procedure and polarizable potentials based on induced dipole moments," *Mol. Phys.* **104**, 415–428 (2006).
- ⁶⁴E. Cancès, F. Legoll, and G. Stoltz, "Theoretical and numerical comparison of some sampling methods for molecular dynamics," *ESAIM: M2AN* **41**, 351–389 (2007).
- ⁶⁵G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Mol. Phys.* **87**, 1117–1157 (1996).
- ⁶⁶K. Anand, G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld, "Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain," *EMBO J.* **21**, 3213–3224 (2002).
- ⁶⁷D. Suárez and N. Díaz, "SARS-CoV-2 main protease: A molecular dynamics study," *J. Chem. Inf. Model.* **60**, 5815–5831 (2020).
- ⁶⁸M. Bzówka, K. Mitusińska, A. Raczyńska, A. Samol, J. A. Tuszyński, and A. Góra, "Structural and evolutionary analysis indicate that the SARS-CoV-2 M^{pro} is a challenging target for small-molecule inhibitor design," *Int. J. Mol. Sci.* **21**, 3099 (2020).
- ⁶⁹See <http://www.bigdft.org> for information about the BigDFT code.
- ⁷⁰J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- ⁷¹S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *J. Comput. Chem.* **32**, 1456–1465 (2011).
- ⁷²A. Willand, Y. O. Kvashnin, L. Genovese, A. Vázquez-Mayagoitia, A. K. Deb, A. Sadeghi, T. Deutsch, and S. Goedecker, "Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations," *J. Chem. Phys.* **138**, 104109 (2013).
- ⁷³C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, "Introducing ONETEP: Linear-scaling density functional simulations on parallel computers," *J. Chem. Phys.* **122**, 084119 (2005).
- ⁷⁴D. R. Bowler and T. Miyazaki, "Calculations for millions of atoms with density functional theory: Linear scaling shows its potential," *J. Phys.: Condens. Matter* **22**, 074207 (2010).
- ⁷⁵M. Zaccaria, L. Genovese, W. Dawson, V. Cristiglio, T. Nakajima, W. Johnson, M. Farzan, and B. Momeni, "Probing the mutational landscape of the SARS-CoV-2 spike protein via quantum mechanical modeling of crystallographic structures," *PNAS Nexus* **1**, pgac180 (2022).
- ⁷⁶M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian *et al.*, Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford, CT, 2009.
- ⁷⁷M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, "Scalable electron correlation methods. 3. Efficient and accurate parallel local coupled cluster with pair natural orbitals (PNO-LCCSD)," *J. Chem. Theory Comput.* **13**, 3650–3675 (2017).
- ⁷⁸Q. Ma and H.-J. Werner, "Scalable electron correlation methods. 5. Parallel perturbative triples correction for explicitly correlated local coupled cluster with pair natural orbitals," *J. Chem. Theory Comput.* **14**, 198–215 (2018).
- ⁷⁹H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györfy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, MOLPRO, version 2019.2, a package of *ab initio* programs, 2019, see <https://www.molpro.net>.
- ⁸⁰R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," *J. Chem. Phys.* **96**, 6796–6806 (1992).
- ⁸¹F. Weigend, "A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency," *Phys. Chem. Chem. Phys.* **4**, 4285–4291 (2002).
- ⁸²F. Weigend, A. Köhn, and C. Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," *J. Chem. Phys.* **116**, 3175–3183 (2002).
- ⁸³L. Zhang, D. Lin, and R. Hilgenfeld, "Crystal structure of the complex resulting from the reaction between the SARS-CoV main protease and tert-butyl (1-((S)-3-cyclohexyl-1-(((S)-4-(cyclopropylamino)-3,4-dioxo-1-((S)-2-oxopyrrolidin-3-yl)butan-2-yl)amino)-1-oxopropan-2-yl)-2-oxo-1,2-dihydropyridin-3-yl)carbamate," <https://doi.org/10.2210/pdb6Y2F/pdb> (2020).
- ⁸⁴L. Zhang, D. Lin, X. Sun, and R. Hilgenfeld, "Crystal structure (orthorhombic form) of the complex resulting from the reaction between SARS-CoV-2 (2019-nCoV) main protease and tert-butyl (1-((S)-1-(((S)-4-(benzylamino)-3,4-dioxo-1-((S)-2-oxopyrrolidin-3-yl)butan-2-yl)amino)-3-cyclopropyl-1-oxopropan-2-yl)-2-oxo-1,2-dihydropyridin-3-yl)carbamate (alpha-ketoamide 13b)," <https://doi.org/10.2210/pdb6Y2G/pdb> (2020).
- ⁸⁵D. Mondal and A. Warshel, "Exploring the mechanism of covalent inhibition: Simulating the binding free energy of α -ketoamide inhibitors of the main protease of SARS-CoV-2," *Biochemistry* **59**, 4601–4608 (2020).
- ⁸⁶D. W. Kneller, G. Phillips, K. L. Weiss, S. Pant, Q. Zhang, H. M. O'Neill, L. Coates, and A. Kovalevsky, "Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography," *J. Mol. Biol.* **295**, 17365–17373 (2020).
- ⁸⁷J. Ono, U. Koshimizu, Y. Fukunishi, and H. Nakai, "Multiple protonation states in ligand-free SARS-CoV-2 main protease revealed by large-scale quantum molecular dynamics simulations," *Chem. Phys. Lett.* **794**, 139489 (2022).
- ⁸⁸D. W. Kneller, H. Li, S. Galanie, G. Phillips, A. Labbé, K. L. Weiss, Q. Zhang, M. A. Arnould, A. Clyde, H. Ma, A. Ramanathan, C. B. Jonsson, M. S. Head, L. Coates, J. M. Louis, P. V. Bonnesen, and A. Kovalevsky, "Structural, electronic, and electrostatic determinants for inhibitor binding to subsites S1 and S2 in SARS-CoV-2 main protease," *J. Med. Chem.* **64**, 17366–17383 (2021).
- ⁸⁹D. W. Kneller, G. Phillips, K. L. Weiss, Q. Zhang, L. Coates, and A. Kovalevsky, "Direct observation of protonation state modulation in SARS-CoV-2 main

protease upon inhibitor binding with neutron crystallography," *J. Med. Chem.* **64**, 4991–5000 (2021).

⁹⁰J. Řezáč, K. E. Riley, and P. Hobza, "S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures," *J. Chem. Theory Comput.* **7**, 2427–2438 (2011).

⁹¹A. A. Rashin, A. H. L. Rashin, and R. L. Jernigan, "Protein flexibility: Coordinate uncertainties and interpretation of structural differences," *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **65**, 1140–1161 (2009).

⁹²N. G. Ramos, G. F. Sarmanho, F. de Sá Ribeiro, V. de Souza, and L. M. T. Lima, "The reproducible normality of the crystallographic B-factor," *Anal. Biochem.* **645**, 114594 (2022).

⁹³A. A.-A. Abu-Saleh, I. E. Awad, A. Yadav, and R. A. Poirier, "Discovery of potent inhibitors for SARS-CoV-2's main protease by ligand-based/structure-based virtual screening, MD simulations, and binding energy calculations," *Phys. Chem. Chem. Phys.* **22**, 23099–23106 (2020).

⁹⁴N. Díaz and D. Suárez, "Influence of charge configuration on substrate binding to SARS-CoV-2 main protease," *Chem. Commun.* **57**, 5314–5317 (2021).

⁹⁵A. Pavlova, D. L. Lynch, I. Daidone, L. Zanetti-Polzi, M. D. Smith, C. Chipot, D. W. Kneller, A. Kovalevsky, L. Coates, A. A. Golosov, C. J. Dickson, C. Velez-Vega, J. S. Duca, J. V. Vermaas, Y. T. Pang, A. Acharya, J. M. Parks, J. C. Smith, and J. C. Gumbart, "Inhibitor binding influences the protonation

states of histidines in SARS-CoV-2 main protease," *Chem. Sci.* **12**, 1513–1527 (2021).

⁹⁶L. Zanetti-Polzi, M. D. Smith, C. Chipot, J. C. Gumbart, D. L. Lynch, A. Pavlova, J. C. Smith, and I. Daidone, "Tuning proton transfer thermodynamics in SARS-CoV-2 main protease: Implications for catalysis and inhibitor design," *J. Chem. Phys. Lett.* **12**, 4195–4202 (2021).

⁹⁷A. E. Reed, F. Weinhold, L. A. Curtiss, and D. J. Pochatko, "Natural bond orbital analysis of molecular interactions: Theoretical studies of binary complexes of HF, H₂O, NH₃, N₂, O₂, F₂, CO, and CO₂ with HF, H₂O, and NH₃," *J. Chem. Phys.* **84**, 5687–5705 (1986).

⁹⁸C. Houriez, M. Masella, and N. Ferré, "Structural and atoms-in-molecules analysis of hydrogen-bond network around nitroxides in liquid water," *J. Chem. Phys.* **133**, 124508 (2010).

⁹⁹Y. Sun, B. Zhao, Y. Wang, Z. Chen, H. Zhang, L. Qu, Y. Zhao, and J. Song, "Optimization of potential non-covalent inhibitors for the SARS-CoV-2 main protease inspected by a descriptor of the subpocket occupancy," *Phys. Chem. Chem. Phys.* **24**, 29940–29951 (2022).

¹⁰⁰P. García-Gutiérrez, R. A. Zubillaga, I. A. Ibarra, A. Martínez, R. Vargas, and J. Garza, "Non-conventional interactions of N3 inhibitor with the main protease of SARS-CoV and SARS-CoV-2," *Comput. Struct. Biotechnol. J.* **19**, 4669–4675 (2021).