

Tuning proton transfer thermodynamics in SARS-Cov-2 main protease: implications for catalysis and inhibitor design

Laura Zanetti-Polzi,^{*,†} Micholas Dean Smith,[‡] Chris Chipot,^{¶,§} James C. Gumbart,^{||} Diane L. Lynch,^{||} Anna Pavlova,^{||} Jeremy C. Smith,^{⊥,‡} and Isabella Daidone^{*,#}

[†]*Center S3, CNR Institute of Nanoscience, Via Campi 213/A, I-41125 Modena, Italy*

[‡]*Department of Biochemistry, Molecular and Cellular Biology, The University of Tennessee, Knoxville. 309 Ken and Blaire Mossman Bldg. 1311 Cumberland Avenue, Knoxville, TN 37996, United States*

[¶]*UMR 7019, Universite de Lorraine, Laboratoire International Associe CNRS*

[§]*University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL, 61801, United States*

^{||}*School of Physics, Georgia Institute of Technology, Atlanta GA 30332, United States*

[⊥]*UT/ORNL Center for Molecular Biophysics, Biosciences Division, Oak Ridge National Laboratory, TN 37831, United States*

[#]*Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio, I-67010 L'Aquila, Italy*

E-mail: laura.zanettipolzi@nano.cnr.it; isabella.daidone@univaq.it

Abstract

The catalytic reaction in SARS-CoV-2 main protease is activated by a proton transfer (PT) from Cys145 to His41. The same PT reaction is likely also required for the covalent binding of some classes of inhibitors. Here we use a hybrid quantum/classical approach to investigate the PT thermodynamics in the apo state and in the presence of two covalent inhibitors, N3 and the α -ketoamide **13b**. We show that in the presence of the inhibitors the free energy cost to reach the charge-separated state of the active-site dyad is lower, with N3 inducing the most significant reduction in the free energy change upon PT. We also show that a few key sites (including specific water molecules) are able to significantly enhance or reduce the thermodynamic feasibility of the PT reaction, with selective desolvation of the active site playing a crucial role. In general, the approach presented is a cost-effective procedure to identify the enzyme regions that control the activation of the catalytic reaction and, more specifically, is also useful to guide the screening and design of potential covalent inhibitors.

Keywords

Free energy, Proton transfer, Quantum mechanics/molecular mechanics, SARS-Cov-2 main protease, Thermodynamics

Introduction

The rapid and broad spread of the pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to an urgent need for effective therapeutics. One of the most promising targets for drug development among coronaviruses is the main protease (M^{pro}), as this protein plays a key role in viral replication and transcription. SARS-CoV-2 M^{pro} cleaves the virus non-structural polyprotein at 11 sites, acting on the sequence Leu/Phe-Gln*Ser/Ala/Gly (where * marks the cleavage site).¹ Inhibition of its cleaving activity would

therefore block the viral replication cycle. In addition, the recognition sequence of SARS-CoV-2 M^{pro} is different from that of all human proteases and thus inhibitors of its activity are less likely to be toxic.² Furthermore, the structure of M^{pro} and of its catalytic pocket is very similar among the coronaviruses family, suggesting that broad-spectrum antiviral drugs might be obtained by targeting this enzyme.¹

SARS-CoV-2 M^{pro} is a three-domain cysteine protease. Domains I (residues 8-101) and II (residues 102-184) are arranged in an antiparallel β -barrel structure, whereas domain III (residues 201-303) contains five α -helices arranged in a globular cluster.³ Domain III has been suggested to be essential in the proteolytic activity by keeping domain II and the long loop connecting domains II and III (residues 185-200) in the proper orientation, and/or by orienting the N-terminal residues that are essential for the dimerization.⁴ Dimerization of the enzyme was itself shown to be essential for catalytic activity by maintaining the proper shape of the pocket of the substrate-binding site.²

Similarly to other cysteine proteases, SARS-CoV-2 M^{pro} features a cysteine-histidine catalytic dyad (Cys145/His41) (see Figure 1). Protein hydrolysis is mediated by the catalytic Cys145 via a nucleophilic attack on the carbonyl carbon of a susceptible peptide bond. It is widely accepted that the imidazole of His41 is the base of the proton transfer (PT) reaction that itself leads to a highly reactive zwitterionic couple (Cys145⁻/His41⁺) which reacts with the substrate.¹ However, the PT mechanism in the apo enzyme and in the presence of the functional substrate or inhibitors is still under debate.⁵⁻⁹ For a comprehensive understanding of the catalytic activity of SARS-CoV-2 M^{pro}, it is crucial to achieve a clear picture of the chemical, structural and dynamical features of the active site and its neighborhood in the apo state. This point is also relevant for the design and screening of potential inhibitors of SARS-CoV-2 M^{pro}. The covalent binding of an inhibitor to the thiol of Cys145 requires the deprotonated cysteine, and therefore the protonation state of the residues of the catalytic dyad in the apo state plays a crucial role in effective inhibition. The propensity of catalytic cysteines to be deprotonated is in fact considered the prime determinant of their reactivity

toward inhibitors.¹⁰

The present work focuses on the investigation of the thermodynamics of the protonation states of the catalytic dyad and of the PT reaction in SARS-CoV-2 M^{pro} by means of MD simulations and the perturbed matrix method (PMM), a hybrid quantum/classical approach for the investigation of chemical processes in complex systems.^{11,12} To determine the protonation state of the catalytic dyad in the apo state, we calculate by means of the perturbative MD-PMM approach the energy variation upon PT and, from that, the corresponding free energy change. We also calculate with the same approach the energy change (and related free energy change) for the tautomerization reaction (His41E \rightleftharpoons His41D). In its neutral state, His41 can be in fact protonated either at its ϵ or at its δ nitrogen, possibly leading to a different structural and dynamical behavior of the catalytic site and its neighborhood. We therefore investigate three possible protonation states: the neutral dyad Cys145H-His41E (i.e., His41 protonated at N $_{\epsilon}$), the neutral dyad Cys145H-His41D (i.e., His41 protonated at N $_{\delta}$) and the zwitterionic dyad Cys145⁻-His41H⁺. The energy variation upon PT is also calculated in the presence of inhibitors previously shown to covalently bind to the SARS-CoV-2 M^{pro} catalytic cysteine: N3 and α -ketoamide **13b**.^{2,3}

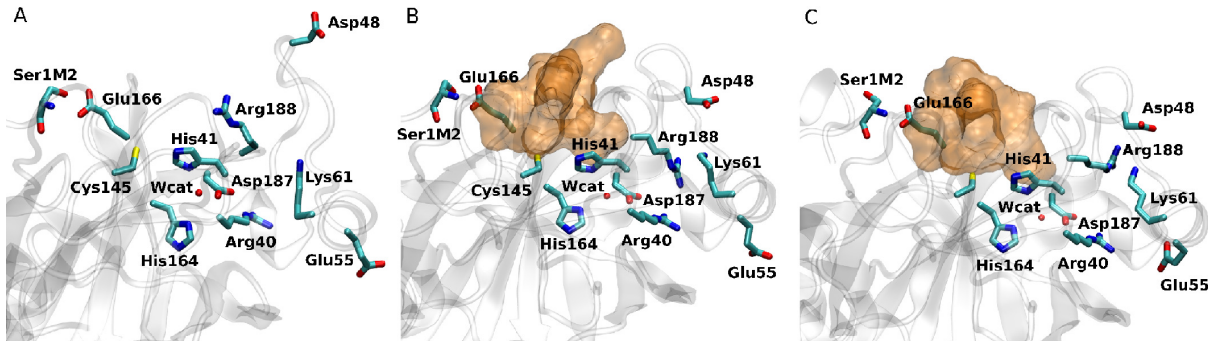


Figure 1: Representative structure of the binding site of SARS-CoV-2 M^{pro} and its neighborhood in the apo state (A) and in the presence of inhibitors N3 (B) and α -ketoamide **13b** (C). The residues of the catalytic dyad (Cys145 and His41), the catalytic water molecule (Wcat) and some key residues surrounding the binding site are highlighted in licorice.

The hybrid quantum/classical approach we apply here, the MD-PMM approach,^{11,12} is based on the joint use of classical MD simulations and quantum mechanical (QM) calcula-

tions. As commonly done in hybrid multiscale approaches,^{13–16} also for the investigation of enzyme catalysis,^{5,7,17–21} the portion of the system in which the chemical event takes place is treated quantum mechanically (the quantum center, QC) while the rest of the system is treated classically and atomistically and exerts an electrostatic perturbation on the QC electronic states. However, the main difference with other hybrid methods is that in the MD-PMM the whole system configurational space (including the QC) is sampled by fully classical MD simulations. The electrostatic perturbation of the environment on the electronic properties of the QC is included *a posteriori*: the electronic properties of the isolated QC (unperturbed properties) are calculated quantum-mechanically in vacuum (i.e., in the gas phase) and then, for each configuration generated by all-atom classical MD simulations of the whole system, the electrostatic effect of the instantaneous atomistic configurations of the environment is included as a perturbing term to the quantum properties of the isolated QC, providing the properties of the QC embedded in the perturbing environment. To include the perturbation of the environment on the QC quantum states, the electrostatic potential and electric field that each atom of the environment exerts on the QC is evaluated at each frame of the MD simulation, and their effect on the QC quantum properties (e.g., energies) are calculated. In the present case, the energy change upon PT is calculated at each frame of the MD simulation, and used to compute the free energy change associated to the PT reaction. Free energy calculations are therefore based on a sampling that is much more extensive than the one that can be typically obtained with standard QM/MM approaches. In addition, the MD-PMM approach also allows us to treat explicitly and dynamically the coupling of the quantum observables with the external environment, and in particular with the collective motions of the system. This is a very relevant point in charge transfer processes, which have been shown to strongly depend on the structural fluctuations of the environment.^{22,23} More details on the MD-PMM approach can be found in the Supporting Information (SI).

The MD-PMM approach has been successfully used to model electron transfer reactions and to calculate reduction potentials in proteins,^{24–32} and, more recently, to evaluate the pK_a

of amino acids and to compute PT reactions free energies.^{33,34} The investigation of charge transfer processes in complex bio-molecular systems with the MD-PMM approach has been also used to identify the protein/enzyme sites that are able to modulate the charge transfer energetics,^{28,29,34} explaining at the molecular level the effect of point mutations and also providing hints for new possible mutations. With the same approach, we focus here on the identification of the enzyme regions that can be targeted to inhibit its catalytic activity with particular attention to the role played by the solvent.

Results

Proton transfer thermodynamics

To investigate the active site PT reaction in SARS-CoV-2 M^{pro} apo state, three MD simulations are performed. Two of these simulations feature both Cys145 and His41 in their neutral state (the reactant ensemble, Cys145H + His41): one with His41 protonated at N_ε (His41E) and one with His41 protonated at N_δ (His41D). The third MD simulation is carried out with Cys145 and His41 forming an ionic couple (the product ensemble, Cys145⁻ + His41H⁺). More details on the MD simulations can be found in the SI.

To apply the above mentioned MD-PMM approach, we select the side chains of Cys145 and His41 as QCs. In the MD simulations in the apo state reactant ensemble, Cys145 and His41 sample a highly variable range of relative conformations, with an average distance between the sulfur of Cys145 and the ϵ and δ nitrogen of His41 of ≈ 0.38 nm and ≈ 0.50 nm, respectively (see Figure S1 in the SI). In addition, the side-chain proton of Cys145 only seldom points toward the ϵ nitrogen of His41 (see Figure S2 in the SI). In the absence of a direct hydrogen bond (HB) between the catalytic dyad residues, Cys145 and His41 are treated as separate QCs in the MD-PMM (see The Perturbed Matrix Method section in the SI). In a small number of configurations in the MD trajectory with His41D, the relative position of Cys145 and His41 is however compatible with a direct HB between the sulfur

and the ε nitrogen involving the proton to be exchanged (i.e., the sulfur proton). In these configurations the residues of the catalytic dyad cannot be treated as separate QCs. In the MD simulations in the apo-state product ensemble (i.e., the ionic couple) the relative configurations of Cys145 and His41 compatible with a direct S-N $_{\varepsilon}$ HB are much more frequent (see Figure S3). For the configurations in which this HB is present, Cys145 and His41 are treated as a unique QC both in the reactant and in the product ensemble. More details on the application of the MD-PMM can be found in the SI. Therefore, QM calculations (see SI) are performed on Cys145 and His41 both as separate groups and as a unique QC in both the reactant and product state, as defined by the following PT reaction:



Then, by applying the MD-PMM approach, we compute at each MD frame for each simulation ensemble the time evolution of the energy change upon PT that provides the reaction free energy ΔG^0 (see Eq. 6 in the SI).

The computed PT reaction free energies reported in Table 1 clearly show that in the apo state both residues of the catalytic dyad are neutral. The zwitterionic couple is in fact at a higher energy for both His41E and His41D in the reactant state (by 34 and 37 kJ/mol for His41D and His41E, respectively, see Table 1). Previous findings obtained with a different computational approach indicate, in agreement with the present results, that the zwitterionic couple lies at a higher energy.^{7,9} However, the free energy cost for the formation of the ionic couple reported in these works is ≈ 12 kJ/mol, i.e., lower than the present one by 22-25 kJ/mol considering the His41D and His41E states, respectively. Our estimate of a higher free energy change upon PT can be attributed to the fact that, in contrast with the aforementioned studies, in our free energy calculations all the configurations of the MD simulations are used. Therefore, the active site dyad configurations in which there is no direct HB between Cys145 and His41 and in which the two residues are relatively far apart are also included. These configurations, wherein the PT is discouraged, represent the majority

of the sampled configurations in the reactant state and thus contribute to a higher free energy change upon PT. Nevertheless, we cannot rule out that the approximations of the MD-PMM approach can lead to an overestimation of the PT free energy. We also note that in a very recent work³⁵ the rate constants for the hydrolysis of various substrates were measured for SARS-CoV-2 M^{pro}, providing higher activation free energies with respect to the ones measured for analogous processes in SARS-CoV M^{pro}. While these higher activation free energies can stem from the different substrates used and/or to different experimental procedures, they are also compatible with a higher PT energy in SARS-CoV-2 M^{pro} with respect to SARS-CoV M^{pro}, possibly also in the apo state. In another very recent work,³⁶ it was suggested, on the basis of constant-pH MD simulations, that Cys44 can be deprotonated in the apo state of SARS-CoV-2 M^{pro}. As this residue is located only ≈ 6.5 Å away from His41, in its deprotonated state Cys44 could relevantly affect the PT energy, favouring the dyad charge separated state. This would provide a lower PT free energy compared to the one computed here using protonated Cys44.

The PT reaction free energy was also calculated within the linear response approximation, i.e. by assuming a Gaussian distribution for the energy change upon PT and considering the average of the mean values of the PT energy obtained in the two ensembles along the MD trajectories.^{37,38} The results are in qualitative agreement with those obtained by explicitly calculating the reaction free energy ΔG^0 (see Eq. 6 in the SI) providing estimates for the free energy change upon PT of 40 and 49 kJ/mol for His41D and His41E, respectively, with a standard error of ≈ 5 kJ/mol. The full calculation of ΔG^0 , being based on an exact relation (see Eq. 6 in the SI), provides a more accurate result than the one that can be obtained within the linear response approximation. However, the latter is less affected by inaccuracies due to finite-sampling issues. The qualitative agreement between the results obtained with the two approaches enhances therefore the reliability of the computed estimates.

The above results also show that the His41E and His41D reactant states are essentially isoenergetic, with His41E slightly more stable with respect to His41D (by ≈ 3 kJ/mol).

This small free energy difference upon tautomerization, which is in agreement with previous relative free energy calculations with the free energy perturbation (FEP) method,³⁹ suggests that in the apo state at physiological temperatures both protonation states may be accessible. However, the His41D state appears to be the most probable reactant state for the PT reaction on the basis of structural considerations. In the crystal structure and in the MD simulations in the reactant state with both His41D and His41E, the N_ϵ of His41 is in fact at a lower distance from the sulfur of Cys145 with respect to the N_δ (see Figure S1). The proximity of the sulfur and N_ϵ would strongly favor a direct PT from Cys145 to His41 from a kinetic point of view.

Table 1: Calculated free energy difference ΔG^0 in kJ/mol for the proton transfer reaction in Eq. 1 in the apo state and in the presence of inhibitors N3 and **13b** and for the tautomerization reaction of His41 in the apo state. The standard error for the computed values is ≈ 6 kJ/mol and is obtained using three subtrajectories for each MD simulation. Details on the ΔG^0 calculation can be found in the SI.

	ΔG^0
$(\text{Cys145H} + \text{His41E} \rightleftharpoons \text{Cys145}^- + \text{His41H}^+)_{\text{apo}}$	37
$(\text{Cys145H} + \text{His41D} \rightleftharpoons \text{Cys145}^- + \text{His41H}^+)_{\text{apo}}$	34
$(\text{His41E} \rightleftharpoons \text{His41D})_{\text{apo}}$	3
$(\text{Cys145H} + \text{His41D} \rightleftharpoons \text{Cys145}^- + \text{His41H}^+)_{\text{N3}}$	20
$(\text{Cys145H} + \text{His41E} \rightleftharpoons \text{Cys145}^- + \text{His41H}^+)_{\text{13b}}$	31

A number of previously and newly designed inhibitors were recently tested against SARS-CoV-2 M^{pro}. Among others, a previously designed Michael acceptor inhibitor (N3) was shown to exhibit potent inhibition for SARS-CoV-2 M^{pro}. The crystal structure of the protein in complex with N3 reveals that the C_β atom of the inhibitor vinyl group is covalently bound to the sulfur of Cys145 of the protein catalytic dyad. Besides Michael acceptors, α -ketoamides were also previously designed as broad-spectrum inhibitors of the main proteases of beta-coronaviruses. Recently, the crystal structure of the complex between **13b** α -ketoamide and SARS-CoV-2 M^{pro} showed that through the nucleophilic attack of the catalytic Cys145 onto the α -keto group of the inhibitor, a thiohemiketal is formed in a reversible reaction.

We thus investigate the active site PT reaction in the non-covalent complex between SARS-CoV-2 M^{pro} and the two inhibitors N3 and **13b** α -ketoamide. In order to compute the energy variation upon PT in the presence of the N3 and **13b** inhibitors inside the substrate binding pocket, we simulate the complex between the non-covalently bound inhibitors and the enzyme. A recent computational work addressing the role of histidine protonation in influencing the structural stability of SARS-CoV-2 M^{pro},³⁹ showed that in the presence of the inhibitor N3 the most stable state is the one with His41D, while in the presence of **13b** the most stable state is the His41E one. Therefore, the computation of the PT energy in the presence of N3 focuses on the His41D reactant state. Two MD simulations are performed for the non-covalent complex between SARS-CoV-2 M^{pro} and N3: one in the reactant ensemble (Cys145H + His41D) and one in the product ensemble (Cys145⁻ + His41H⁺). The computation of the PT energy in the presence of **13b**, instead, focuses on the His41E state. Two MD simulations are also carried out for the non-covalent complex between SARS-CoV-2 M^{pro} and **13b**: one in the reactant ensemble (Cys145H + His41E) and two in the product ensemble (Cys145⁻ + His41H⁺).

From our calculations (see Methods section in the SI) we obtain that, in the presence of the inhibitor N3 in the active-site, the PT reaction free energy is significantly lower, changing from $\Delta G^0 = 34 \pm 6$ kJ/mol to $\Delta G^0 = 20 \pm 6$ kJ/mol (see Table 1). These results clearly show that the zwitterionic couple is at a lower energy in the presence of the inhibitor N3 than in the apo state. Therefore, the PT reaction is more thermodynamically favoured in the presence of N3 than in the apo state.

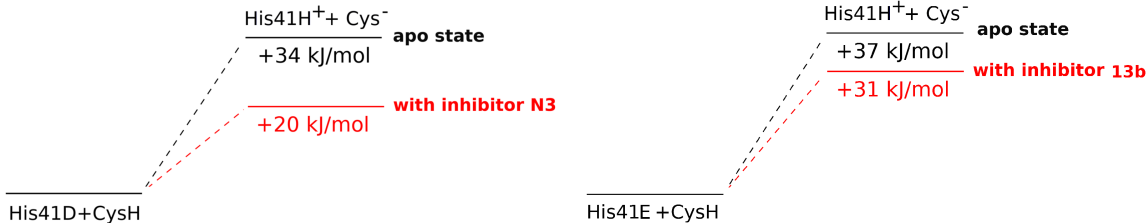
The free energy difference upon PT estimated here is in agreement with previous recent computational works^{21,40} showing that in the presence of N3 the ionic couple is at a higher energy with respect to the reactant state. These two studies report a positive free energy difference upon PT, yet the two estimates are remarkably different: in the work of Arafet et al. a free energy change of ≈ 5.5 kJ/mol upon PT is reported while in the work of Ramos et al. the ionic couple is ≈ 43 kJ/mol higher in energy than the neutral one. Nevertheless, these

two computational investigations report similar values for the free energy barrier separating the ionic couple from the covalent complex ($\approx 41/44$ kJ/mol) and for the overall free energy difference between the covalent and non-covalent complex ($\approx -75/-64$ kJ/mol).

Experimentally, the inactivation rate constant of SARS-CoV-2 M^{pro} by N3 could not be measured because the process is too fast. For the very similar protease of SARS-CoV, the inactivation rate constant is slower and could be measured providing an activation free energy of ≈ 90 kJ/mol. This experimental estimate of the activation free energy (that is given by the sum of the free energy change upon PT and the subsequent free energy barrier for the formation of the covalent complex) agrees well with the activation free energy obtained using the PT free energy calculated by Ramos et al. (i.e. $43 + 44 = 87$ kJ/mol). On the other hand, the activation free energy estimated by Arafet et al. ($\approx 5 + 41 = 46$ kJ/mol) appears to be underestimated because it would require an unlikely fast inactivation rate constant (i.e. 8 orders of magnitude faster with respect to SARS-CoV M^{pro}). Our calculated PT free energy provides an activation free energy of $\approx 60-65$ kJ/mol (i.e. $\approx 20 + 41/44$ kJ/mol, where $41/44$ kJ/mol are the estimates provided by Arafet et al. and Ramos et al.) which is lower than the one of SARS-CoV M^{pro} and, hence, lends support of a faster reaction expected in SARS-CoV-2 with respect to SARS-CoV M^{pro}. Our estimate of the PT free energy in the presence of N3, i.e. 20 kJ/mol, thus reconciles the previous estimates.^{21,40}

In contrast with Ramos et al., we also find that the PT reaction is thermodynamically favoured in the presence of N3 with respect to the apo state. A reduction of the free energy cost in forming the ionic couple was suggested for SARS-CoV in the presence of the functional substrate.⁴¹ We believe that the decrease of the PT free energy in the presence of N3 is in agreement with the experimental observation of the covalent binding of N3 to the catalytic cysteine, which requires a deprotonated Cys145. As a matter of fact, the inability to efficiently promote the PT reaction has been suggested to determine the low inhibition potencies of known inhibitors.⁴¹

The calculation of the free energy change upon PT in the presence of **13b** shows a



Scheme 1: Representation of the free energy change upon PT in the apo state and in the presence of inhibitors N3 and **13b**.

modest reduction of the PT reaction free energy with respect to the apo state: $\Delta G^0 = 37 \pm 6$ kJ/mol in the His41E apo state and $\Delta G^0 = 31 \pm 6$ kJ/mol in the His41E state in the presence of **13b** (see Table 1). The computed value for the free energy difference upon PT is in good agreement with a previous computational estimate for the **13b**-SARS-CoV-2 M^{pro} non-covalent complex (≈ 30 kJ/mol).⁹ In the same work the free energy barrier separating the ionic couple from the covalent complex is estimated to be ≈ 50 kJ/mol, i.e. 5-10 kJ/mol higher than the one reported for N3, which, put together with the free energy change upon PT, leads to an activation free energy of ≈ 80 kJ/mol, i.e. 15-20 kJ/mol higher for **13b** than for N3.

Contribution of protein residues and water molecules to the proton transfer energy

To gain insights into the molecular mechanisms that tune the energetics of the catalytic site in the reactant state, we analyzed the electrostatic potential contribution to the PT energy in the apo and inhibitor-bound states. The most important contribution to the PT energy fluctuations around the unperturbed ground state energy difference along the MD simulations is given by the electrostatic potential felt by the QCs²⁹ (see The Perturbed Matrix Method section in the SI). We therefore analyzed the contribution of each protein residue to the electrostatic potential in order to understand which protein regions contribute

the most to the PT energy. This analysis reveals which are the enzyme sites that control, *via* a dynamic electrostatic effect, the catalytic activity that is triggered by the PT reaction in the binding site.

In what follows we report the results obtained for the apo state with His41D and the changes induced by the presence of N3, while the results for the apo state with His41E and the changes occurring upon binding of **13b** are relegated to the SI, as the case of N3 is more informative on the protein and solvent contributions to the PT energy change upon binding. This is because the lowering of the free energy cost in forming the ionic couple is larger upon binding to N3 (from 34 kJ/mol in the apo state to 20 kJ/mol in the bound state) than to **13b** (from 37 kJ/mol in the apo state to 31 kJ/mol in the bound state).

In Figure 2A we report qV , i.e., the contribution to the PT energy due to the electrostatic potential, of each residue in the MD simulation of the apo state with His41D. In the figure the residues with a negative contribution exert an electrostatic effect that favors the PT reaction, while the opposite is true for the residues with a positive contribution. The residues with the highest (> 20 kJ/mol) positive or negative qV are labelled in the figure, being those that most contribute to determine the PT energy.

The residues that most relevantly contribute to the PT energy are Arg40, Glu166 and Arg188 (positive contributions disfavouring PT) and Asp187 (negative contribution favouring PT). These are charged residues in the vicinity of the catalytic dyad that, besides being energetically relevant, are also structurally relevant. Arg40 and Asp187 are bound in a salt bridge linking domains I and II of the protein. Glu166 forms a salt bridge with the amino group of the N-terminus residue of the other monomer (Ser1) and is involved in the dimerization.² In addition, the native loop containing Asp187 and Arg188 was shown to be essential for the proteolytic activity of a very similar protease from another coronavirus.⁴² The contribution of Ser1 of the other monomer, related to the one of Glu166 (see above), can be also seen.

Three additional minor contributions can also be observed in Figure 2A: Asp48, Glu55

and Lys61. The contribution of these charged residues, that are located on the protein surface in the His41 loop not far from the catalytic dyad cleft, shows that the catalytic activity might be modulated by allosteric control by acting on regions outside the active site cleft. A possible inhibition strategy could be to target one of the identified regions in order to promote an allosteric modulation of the PT energetics.

More in general, the present approach can be used to suggest promising candidates to enhance or inhibit the catalytic activity of proteases by point mutations without altering the structural and chemical properties of the active site cleft. In fact, one of the residues located outside the active site pocket and identified as key site for tuning the PT energy could be mutated from a charged to a neutral residue to induce a non-negligible change in the PT energy favoring/disfavoring the PT reaction.

Concerning the contribution of the solvent it can be seen from Figure 2A that the water molecules, considered collectively, do not contribute to the PT energy change. However, the analysis of the separate contributions arising from the water molecules closest to the catalytic dyad, reveals interesting features (see Figure 2B). To identify these water molecules, we firstly analyzed the available crystal structures and then the MD simulation.

In the crystal structure of SARS-CoV-2 M^{pro} (pdb 6wqf) a highly buried water molecule is present, which is packed in a tight HB network involving His41 N δ , His41 backbone NH group, the side chain oxygens of Asp187 and His164 N δ (see Figure 2). This HB network is conserved along other available crystal structures (e.g., 6y2e, 7bqy, 6y2g, 6yb7) and the buried water molecule, Wcat, has been suggested to play an active role in the catalytic PT reaction.⁴² Another well conserved water molecule, Wdyad, can be found in several crystal structures (see for example pdbs 6wqf and 6y2e) and is located within the catalytic dyad, almost bridging the sulfur atom of Cys145 and the N δ atom of His41 (see Scheme 2). This water molecule was suggested to play a role in the proton transfer from His41 to the inhibitor after the formation of the covalent bond between the inhibitor and the thiolate of Cys145.⁴⁰ The analysis of the occupancy of water molecules close to the catalytic dyad (i.e.

the water molecules residing within 0.5 nm of any atom of the dyad) along the apo-state simulation reveals that, besides the two aforementioned conserved water molecules, there is a high occupancy of small water wires of 3-4 molecules starting from Wdyad (see Scheme 2). Therefore, the contributions of Wcat, Wdyad and the small water wires identified were analyzed separately in the PT energetics analysis (see Figure 2B). It can be seen that there is an opposite effect between Wcat (negative contribution favoring the PT reaction), and Wdyad and the small water wires (positive contribution disfavoring the PT reaction) that almost counterbalance giving rise to an almost nil total contribution of the solvent.

The analysis of the residues and water molecules that contribute the most to the PT energy was also performed for the MD simulation in the reactant ensemble in the presence of N3, along with the direct contribution of N3. The results, reported in Figure 2C, show that in the presence of the inhibitor the residues that most significantly contribute to the PT energy are those already identified in the apo state. In Figure 2E the difference $\Delta(qV)$ between the single residue contribution obtained from the MD in the presence of N3 and that obtained in the apo ($\Delta(qV) = qV(N3) - qV(apo)$) is also reported. This difference highlights the protein regions that more relevantly contribute to the variation of the energy change upon PT in the presence of N3: the residues with a high negative contribution in Figure 2E are those that contribute to lowering the PT energy in the presence of the inhibitor with respect to the apo state.

The most relevant (positive and negative) contributions are exerted by Glu166, Arg188, N3 and the solvent. Concerning Glu166 and Arg188, the analysis of the crystal structure of SARS-CoV-2 M^{pro} bound to N3 highlights that the inhibitor backbone forms an antiparallel sheet with residues 164-168 on one side and with residues 189-191 on the other side. The local interactions of N3 with protein sites different from the catalytic site thus favours the PT reaction. The presence of N3 reduces the electrostatic effect of Arg188 and Glu166 that disfavors the PT in the apo state, both by screening the catalytic dyad from their charge and by sterically pushing away their side chains (see Scheme 2). This reduced electrostatic

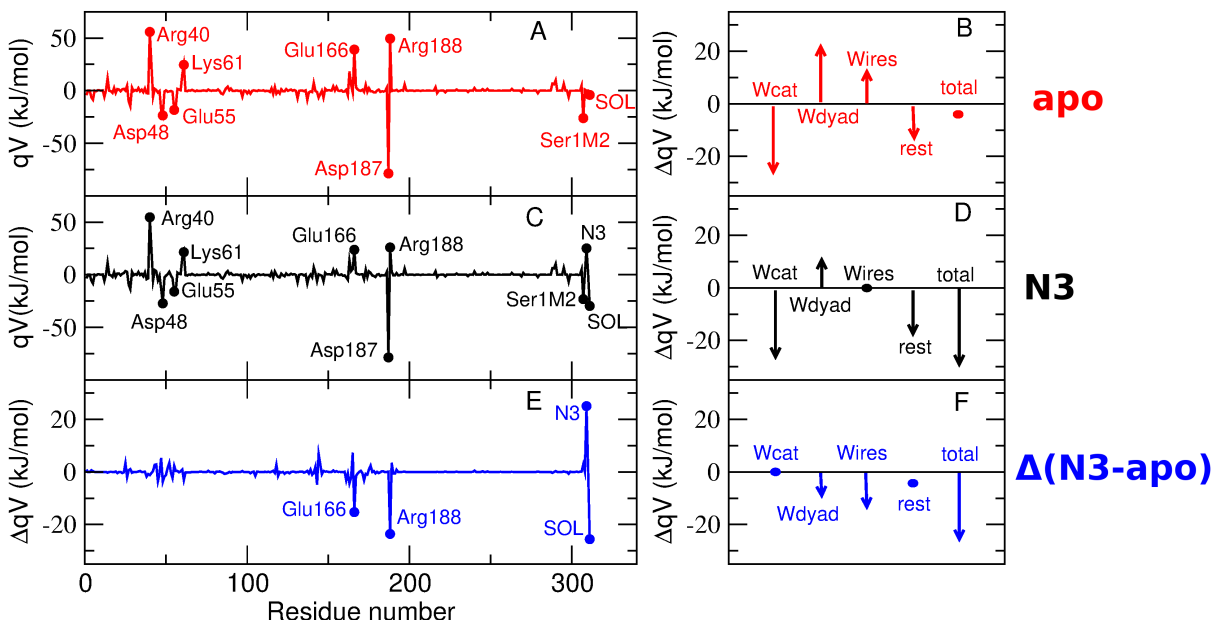


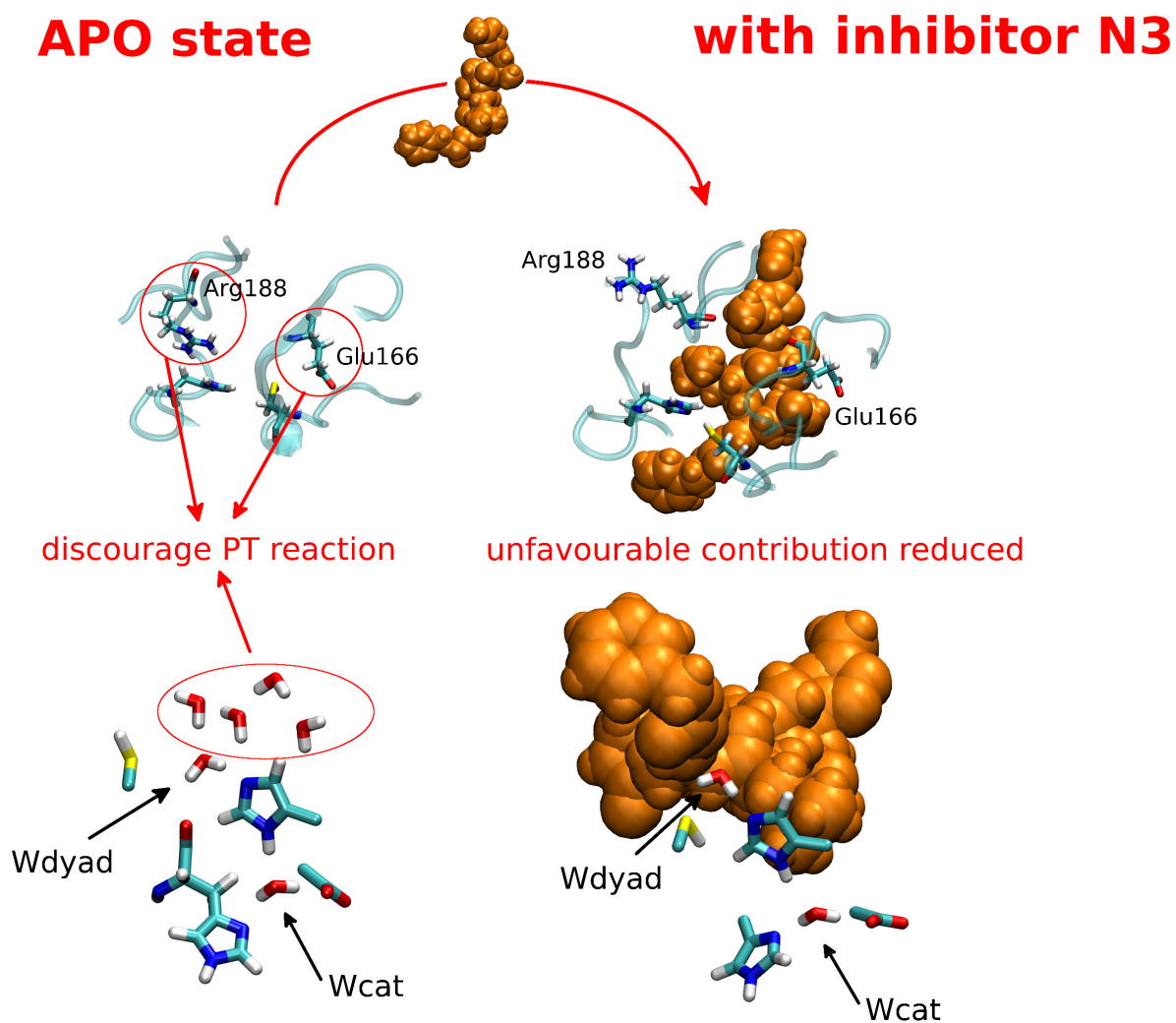
Figure 2: A and C: qV is plotted for each protein residue and all the water molecules as an additional virtual residue SOL for the apo state (A) and in the presence of the inhibitor N3 (C). qV is the mean value along the MD trajectories of the contribution due to the electrostatic potential to the PT energy. The residues featuring an absolute value of qV higher than 20 kJ/mol are labelled in the figure. The residues with a negative contribution exert an electrostatic effect that favors the PT reaction, while the opposite is true for the residues with a positive contribution. The contribution of the solvent in the apo state is labelled for the sake of comparison between the panels, but is approximately null. The contributions of the residues of the catalytic dyad (His41 and Cys145) are not included in the plot. The contribution of the residues of the second monomer (except Ser1M2) are not shown being negligible. E: $\Delta(qV) = qV(N3) - qV(apo)$ is plotted for each protein residue and SOL. The residues featuring an absolute value of qV higher than 10 kJ/mol are labelled. The contributions of the residues of the catalytic dyad (His41 and Cys145) are not included in the plot. The residues with a negative contribution are those that contribute to lower the PT energy in the presence of the inhibitor with respect to the apo state while the opposite is true for the residues with a positive contribution. B, D and F: dissection of the contribution of the solvent (SOL): the contribution of Wcat, Wdyad, the molecules forming the wires and the rest of the water molecules are reported together with the total solvent contribution for the apo state (B), in the presence of N3 (D) and for the difference $\Delta(qV) = qV(N3) - qV(apo)$ (F).

effect contributes to the observed lowering of the PT energy.

Concerning the direct effect of the presence of N3, it can be seen that the electrostatic contribution of N3 in Figure 2C is positive and therefore does not favor the lowering of the PT energy. It should be however noted that this contribution only takes into account the direct electrostatic effect of N3, neglecting any possible higher-level interaction between the inhibitor and the catalytic dyad. The positive contribution of N3 is counterbalanced by a negative contribution of the solvent, which implies that the desolvation of the active site occurring upon binding of N3 gives a favourable contribution to the PT process. Although this result might sound counterintuitive (i.e. that PT is favoured by desolvation), it is very well explained by the detailed analysis of the contributions of the water molecules residing closest to the catalytic dyad, as described hereafter.

In the N3-bound state the contribution of Wcat is essentially the same as in the apo state (see Figure 2, panels B and D), consistent with the fact that in the MD simulation in the presence of the inhibitor the same HB pattern for Wcat is observed as in the MD simulation of the apo state (see Scheme 2). Instead, significant differences are observed for Wdyad and its water wires, which show an overall less unfavourable contribution than in the apo state. This is due to the fact that when N3 is bound, while Wdyad is maintained (although with a slightly different orientation with respect to the catalytic dyad), the water wires are expelled (see Figure 2, panels D and F, and Scheme 2 for representative configurations showing the different arrangement of the water molecules). Hence, the rearrangement of Wdyad and the release of the "unfavourable" water wires upon binding of N3 are responsible for the lowering of the PT energy upon binding of the inhibitor (i.e. the negative peak in Figure 2E arising from the solvent).

As a last remark, in the presence of N3 the feasibility of the PT reaction is also enhanced by a better relative position of the two reaction partners (Cys145 and His41). In fact, the distance between the sulfur atom of Cys145 and the N ϵ atom of His41 is lower and less variable in the MD simulation with N3 than in the apo-state simulation (see Figure S4).



Scheme 2: Representation of the mechanism that determines a lower PT free energy in the presence of inhibitor N3

Discussion

Some classes of covalent inhibitors of SARS-CoV-2 M^{pro} require the deprotonation of the sulfhydryl group of Cys145 and the most viable route for this is the PT reaction from Cys145 to His41, which is proposed to be the physiological reaction in the presence of the substrate. Therefore, the knowledge of the thermodynamic feasibility of this PT reaction, and of the protein regions that are responsible for its enhancement or reduction, can guide the design of the recognition motives of an inhibitor to the catalytic site of the enzyme with the aim of promoting the PT reaction, and thus the subsequent formation of a covalent bond. In addition, previous computational works, as well as the present work, showed that the PT reaction free energy accounts for 1/3-1/2 of the total activation free energy for the formation of the covalent complex and thus the PT relevantly contributes to the rate determining step.

Here we investigated the thermodynamics of the PT reaction within the catalytic dyad of SARS-CoV-2 M^{pro} and identified which protein regions and solvent molecules have an important role in modulating the energy change upon PT. Our results support the hypothesis that in the apo enzyme the most stable state of the catalytic dyad is that in which both residues are neutral, and suggest that for the neutral His41 the protonation at N_ε is slightly favoured with respect to the protonation at N_δ (the two tautomers can however easily interconvert, given that the tautomerization free energy is very small, i.e. 3 kJ/mol). We further showed that the presence of a number of key groups (Arg40, Glu166, Asp187, Arg188 and water molecules) within, or in the vicinity of, the catalytic site are able to significantly enhance or reduce the thermodynamic feasibility of the PT reaction in the apo state.

To demonstrate that the residues and water molecules identified here can be reliably used for the screening of new inhibitors, we turned as test cases to previously designed inhibitors, i.e. the Michael acceptor inhibitor N3 and the α -ketoamide **13b**, which covalently bind to the Cys145 sulfur and exhibit potent inhibition for SARS-CoV-2 M^{pro}. The most significant difference in the PT reaction free energy is observed for N3, upon the binding of which the PT free energy change decreases significantly (by ≈ 15 kJ/mol) with respect

to the His41D apo enzyme. The observed free energy decrease does not arise from a direct electrostatic effect of the inhibitor, but rather from local conformational changes induced by the non-covalent binding of the inhibitor. In fact, the presence of N3 not only favors a lower average Cys145S-His41N ϵ distance, but also induces a favourable conformational change of two residues, Glu166 and Arg188, which were found to disfavor the PT in the apo state. Moreover, and more surprisingly, a strong contribution is found to arise from "selective" desolvation of the active site: upon binding of N3 a few water molecules that showed a disfavourable contribution to the PT free energy in the active site of the apo enzyme are expelled providing a net favourable contribution to the PT reaction.

The approach presented in this work is envisioned to be of significant aid in the design of new inhibitors by means of computer simulations. In particular, our approach can help identify compounds that can promote the catalytic PT reaction and, therefore, be good candidates as covalent inhibitors. Knowledge of the key residues and water molecules that tune the catalytic PT reaction can in fact guide the analysis of the interactions between the recognition moieties of a candidate compound and the different sub-sites of the binding pocket of the protein, which is a crucial step in the design and screening of potential inhibitors. In addition, identification of specific conserved water molecules able to affect the PT energetics could be used in docking procedures to select which water molecules should be explicitly considered.

Design of analogues of known covalently-bound inhibitors (e.g. analogues of N3 or of the **13b** α -ketoamide) commonly relies on the geometrical and structural analysis of X-ray or MD-derived structures of the reactant complex for modulating the recognition portion. Here, we propose a procedure based on the calculation of the molecular determinants of the proton transfer thermodynamics to complement the information that can be currently obtained from the available computational approaches for the screening of covalently-bound candidates.^{43–46} As a more general application, this approach is a cost-effective procedure to reveal key sites in enzymes to be mutated, or to be targeted with ligands, in order to

enhance or reduce the catalytic activity.

Acknowledgement

This work used the Hive cluster, which is supported by the US National Science Foundation under grant number 1828187 and is managed by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology. JCG acknowledges support from the US National Institutes of Health (R01-AI148740).

Supporting Information Available

The following files are available free of charge.

- details on the MD-PMM approach
- details on the quantum mechanical calculations
- details on the molecular dynamics simulations
- additional structural analyses
- the His41E apo state in comparison with the **13b**-bound state

References

- (1) Su, H.; Su, H.; Zhou, F.; Huang, Z.; Ma, X.; Natarajan, K.; Zhang, M.; Huang, Y. Molecular Insights into Small Molecule Drug Discovery for SARS-CoV-2. *Angew. Chem. Int. Ed.* **2020**, doi:10.1002/anie.202008835.
- (2) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* **2020**, *368*, 409–412.

- (3) Jin, Z. et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293.
- (4) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13190–13195.
- (5) Swiderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chem. Sci.* **2020**, *11*, 10626–10630.
- (6) Kneller, D. W.; Phillips, G.; Weiss, K. L.; Pant, S.; Zhang, Q.; O'Neill, H. M.; Coates, L.; Kovalevsky, A. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J. Biol. Chem.* **2020**, *295*, 17365–17373.
- (7) Ramos-Guzman, C. A.; Ruiz-Pernia, J. J.; Tunon, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, *10*, 12544–12554.
- (8) Hoffman, R. L. et al. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J. Med. Chem.* **2020**, *63*, 12725–12747.
- (9) Mondal, D.; Warshel, A. Exploring the Mechanism of Covalent Inhibition: Simulating the Binding Free Energy of α -Ketoamide Inhibitors of the Main Protease of SARS-CoV-2. *Biochem.* **2020**, *59*, 4601–4608.
- (10) Awoonor-Williams, E.; Rowley, C. N. How reactive are druggable cysteines in protein kinases? *J. Chem. Inf. Model.* **2018**, *58*, 1935–1946.
- (11) Aschi, M.; Spezia, R.; Di Nola, A.; Amadei, A. A first principles method to model perturbed electronic wavefunctions: the effect of an external electric field. *Chem. Phys. Lett.* **2001**, *344*, 374–380.

- (12) Zanetti-Polzi, L.; Del Galdo, S.; Daidone, I.; D'Abramo, M.; Barone, V.; Aschi, M.; Amadei, A. Extending the perturbed matrix method beyond the dipolar approximation: comparison of different levels of theory. *Phys. Chem. Chem. Phys.* **2018**, *20*, 24369–24378.
- (13) Vreven, T.; Morokuma, K. Chapter 3 Hybrid Methods: ONIOM(QM:MM) and QM/MM. *Ann. Rep. Comp. Chem.* **2006**, *2*, 35–51.
- (14) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.
- (15) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- (16) Liu, M.; Wang, Y.; Chen, Y.; Field, M. J.; Gao, J. QM/MM through the 1990s: the first twenty years of method development and applications. *Isr. J. Chem.* **2014**, *54*, 1250–1263.
- (17) Gao, J.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Ann. Rev. Phys. Chem.* **2002**, *53*, 467–505.
- (18) Sgrignani, J.; Cavalli, A.; Colombo, G.; Magistrato, A. Enzymatic and inhibition mechanism of human aromatase (CYP19A1) enzyme. A computational perspective from QM/MM and classical molecular dynamics simulations. *Mini Rev. Med. Chem.* **2016**, *16*, 1112–1124.
- (19) Faraji, S.; Zhong, D.; Dreuw, A. Characterization of the Intermediate in and Identification of the Repair Mechanism of (6-4) Photolesions by Photolyases. *Angew. Chem. Int. Ed.* **2016**, *55*, 5175–5178.
- (20) Rinaldi, S.; Van der Kamp, M. W.; Ranaghan, K. E.; Mulholland, A. J.; Colombo, G.

- Understanding complex mechanisms of enzyme reactivity: the case of Limonene-1, 2-epoxide hydrolases. *Acs Catal.* **2018**, *8*, 5698–5707.
- (21) Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A.; González, F. V.; Swiderek, K.; Moliner, V. 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. *ChemRxiv* **2020**, doi:10.26434/chemrxiv.12941819.v1.
- (22) Kundu, M.; He, T.-F.; Lu, Y.; Wang, L.; Zhong, D. Short-Range Electron Transfer in Reduced Flavodoxin: Ultrafast Nonequilibrium Dynamics Coupled with Protein Fluctuations. *J. Phys. Chem. Lett.* **2018**, *9*, 2782–2790.
- (23) Lu, Y.; Zhong, D. Nonequilibrium dynamics of photoinduced forward and backward electron transfer reactions. *J. Chem. Phys.* **2020**, *152*, 065102.
- (24) Amadei, A.; Daidone, I.; Aschi, M. A general theoretical model for electron transfer reactions in complex systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1360–13770.
- (25) Bortolotti, C. A.; Amadei, A.; Aschi, M.; Borsari, M.; Corni, S.; Sola, M.; Daidone, I. The Reversible Opening of Water Channels in Cytochrome c Modulates the Heme Iron Reduction Potential. *J. Am. Chem. Soc.* **2012**, *134*, 13670–13678.
- (26) Zanetti-Polzi, L.; Daidone, I.; Bortolotti, C. A.; Corni, S. Surface Packing Determines the Redox Potential Shift of Cytochrome c Adsorbed on Gold. *J. Am. Chem. Soc.* **2014**, *136*, 12929–12937.
- (27) Daidone, I.; Amadei, A.; Zaccanti, F.; Borsari, M.; Bortolotti, C. A. How the Reorganization Free Energy Affects the Reduction Potential of Structurally Homologous Cytochromes. *J. Phys. Chem. Lett.* **2014**, *5*, 1534–1540.

- (28) Zanetti-Polzi, L.; Bortolotti, C. A.; Daidone, I.; Aschi, M.; Amadei, A.; Corni, S. A few key residues determine the high redox potential shift in azurin mutants. *Org. Biomol. Chem.* **2015**, *13*, 11003–11013.
- (29) Zanetti-Polzi, L.; Corni, S.; Daidone, I.; Amadei, A. Extending the essential dynamics analysis to investigate molecular properties: application to the redox potential of proteins. *Phys. Chem. Chem. Phys.* **2016**, *18*, 18450–18459.
- (30) Paltrinieri, L.; Di Rocco, G.; Battistuzzi, G.; Borsari, M.; Sola, M.; Ranieri, A.; Zanetti-Polzi, L.; Daidone, I.; Bortolotti, C. A. Computational evidence support the hypothesis of neuroglobin also acting as an electron transfer species. *J. Biol. Inorg. Chem.* **2017**, 1–9.
- (31) Zanetti-Polzi, L.; Aschi, M.; Amadei, A.; Daidone, I. Alternative electron-transfer channels ensure ultrafast deactivation of light-induced excited states in Riboflavin binding protein. *J. Phys. Chem. Lett.* **2017**, *8*, 3321–3327.
- (32) Zanetti-Polzi, L.; Daidone, I.; Corni, S. Evidence of a Thermodynamic Ramp for Hole Hopping to Protect a Redox Enzyme from Oxidative Damage. *J. Phys. Chem. Lett.* **2019**, *10*, 1450–1456.
- (33) Zanetti-Polzi, L.; Daidone, I.; Amadei, A. Fully Atomistic Multiscale Approach for pKa Prediction. *J. Phys. Chem. B* **2020**, *124*, 4712–4722.
- (34) Zanetti-Polzi, L.; Aschi, M.; Daidone, I. Cooperative protein–solvent tuning of proton transfer energetics: carbonic anhydrase as a case study. *Phys. Chem. Chem. Phys.* **2020**, *22*, 19975–19981.
- (35) Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.; Pawlik, B.; Wang, X.; Jochmans, D.; Neyts, J.; Młynarski, W.; Hilgenfeld, R.; Drag, M. SARS-CoV-2 M pro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* **2020**, 1–7.

- (36) Verma, N.; Henderson, J. A.; Shen, J. Proton-Coupled Conformational Activation of SARS Coronavirus Main Proteases and Opportunity for Designing Small-Molecule Broad-Spectrum Targeted Covalent Inhibitors. *J. Am. Chem. Soc.* **2020**, *142*, 21883–21890.
- (37) Muegge, I.; Qi, P.; Wand, A.; Chu, Z.; Warshel, A. The reorganization energy of cytochrome c revisited. *J. Phys. Chem. B* **1997**, *101*, 825–836.
- (38) Amadei, A.; Daidone, I.; Bortolotti, C. A. A general statistical mechanical approach for modeling redox thermodynamics: the reaction and reorganization free energies. *RSC Adv.* **2013**, *3*, 19657–19665.
- (39) Pavlova, A. et al. Inhibitor binding influences the protonation states of histidines in SARS-CoV-2 main protease. *bioRxiv* **2020**, doi: 10.1101/2020.09.07.286344.
- (40) Ramos-Guzman, C. A.; Ruiz-Pernia, J. J.; Tunon, I. A Microscopic Description of SARS-CoV-2 Main Protease Inhibition with Michael Acceptors. Strategies for Improving Inhibitors Design. *chemRxiv* **2020**, doi: 10.26434/chemrxiv.12895064.v1.
- (41) Paasche, A.; Zipper, A.; Schafer, S.; Ziebuhr, J.; Schirmeister, T.; Engels, B. Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His dyad of the SARS-CoV main protease. *Biochem.* **2014**, *53*, 5930–5946.
- (42) Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **2003**, *300*, 1763–1767.
- (43) Sørensen, J.; Demir, Ö.; Swift, R. V.; Feher, V. A.; Amaro, R. E. *Molecular Modeling of Proteins*; Springer, 2015; pp 445–469.
- (44) Awoonor-Williams, E.; Walsh, A. G.; Rowley, C. N. Modeling covalent-modifier drugs. *Biochim. Biophys. Acta Proteins Proteom.* **2017**, *1865*, 1664–1675.

- (45) Zaidman, D.; Gehrtz, P.; Filep, M.; Fearon, D.; Prilusky, J.; Duberstein, S.; Cohen, G.; Owen, D.; Resnick, E.; Strain-Damerell, C.; Lukacik, P.; Barr, H.; Walsh, M. A.; von Delft, F.; London, N. An automatic pipeline for the design of irreversible derivatives identifies a potent SARS-CoV-2 Mpro inhibitor. *bioRxiv* **2020**, doi: 10.1101/2020.09.21.299776.
- (46) Jagger, B. R.; Kochanek, S. E.; Haldar, S.; Amaro, R. E.; Mulholland, A. J. Multiscale simulation approaches to modeling drug–protein binding. *Curr. Opin. Struct. Biol.* **2020**, *61*, 213 – 221.