

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 a covalent inhibitor, N3. We show that in the apo state a neutral catalytic dyad is favored whereas in the presence of N3 the PT reaction becomes thermodynamically favorable. We also show that a few key sites (including a water molecule) are able to significantly enhance or reduce the thermodynamic feasibility of the PT reaction. The approach presented is a general and cost-effective procedure to identify the enzyme regions that control the activation of the catalytic reaction. It 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, it is still debated whether the PT and the formation of the thiolate anion are induced by the presence of the functional substrate or occur in the apo state.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 an inhibitor (N3), which was previously shown to covalently bind to the SARS-CoV-2 M^{pro} catalytic cysteine.¹³

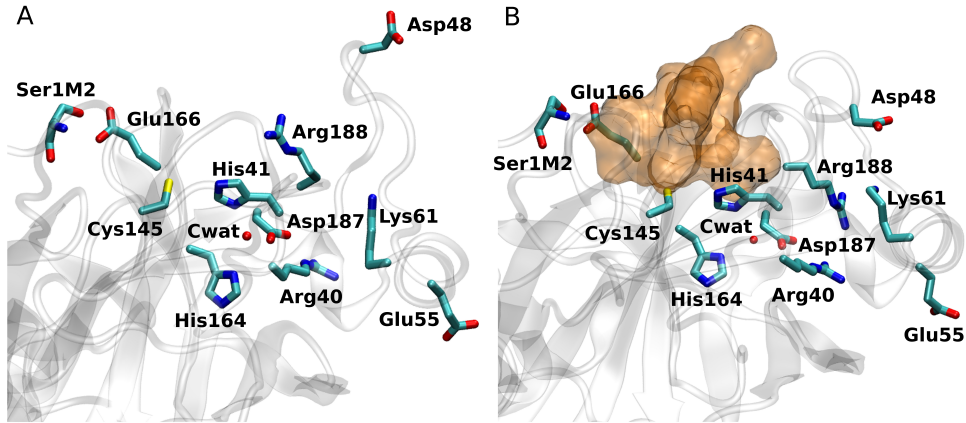


Figure 1: Representative structure of the binding site of SARS-CoV-2 M^{pro} and its neighborhood in the apo state (A) and in the presence of the inhibitor N3 (B). The residues of the catalytic dyad (Cys145 and His41), the catalytic water molecule (Cwat) and some key residues surrounding the binding 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) calculations. As commonly done in hybrid multiscale approaches,^{14–17} also for the investigation of

enzyme catalysis,^{6,8,18-22} 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.^{23,24} More details on the MD-PMM approach can be found in the Supporting Information (SI).

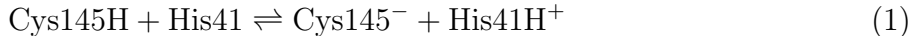
The MD-PMM approach has been successfully used to model electron transfer reactions and to calculate reduction potentials in proteins,²⁵⁻³³ and, more recently, to evaluate the pK_a of amino acids and to compute PT reactions free energies.^{34,35} The investigation of charge

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,^{29,30,35} 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.

Results

Proton transfer thermodynamics

To investigate the active site PT reaction in SARS-CoV-2 M^{Pro}, we select the side chains of Cys145 and His41 as QCs and perform quantum mechanical calculations on these QCs in both the reactant and product state, as defined by the following PT reaction:



Three MD simulations are performed. Two of these are performed in 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 performed in the product ensemble (Cys145⁻ + His41H⁺). More details on the quantum chemical calculations and MD simulations can be found in the SI.

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, according to our results, 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 31 and 36 kJ/mol for His41D and His41E, respectively, see Table 1). Previous

findings obtained with a different computational approach analyzed only the reactant state with His41D and showed, in agreement with the present results, that the zwitterionic couple is at higher energy.^{6,8}

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.^{36,37} 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 a free energy change upon PT of 41 and 45 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 ≈ 5 kJ/mol with the full free energy calculation and by ≈ 4 kJ/mol within the linear response approximation). This small free energy difference upon tautomerization, which is in agreement with previous relative free energy calculations with the free energy perturbation (FEP) method,³⁸ suggests that in the apo state at physiological temperatures both protonation states may be accessible. This also implies that the thermodynamic cost to reach the charge-separated state, which has a higher free energy than the neutral dyad couple, is slightly lower in the reactant state with His41D.

Beside the slightly lower free energy change upon PT, the His41D state appears the most probable reactant state for the PT reaction also 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 and the representative snapshots Figure 2). 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 inhibitor N3 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 can be found in the SI.

	ΔG^0
$(\text{Cys145H} + \text{His41E} \rightleftharpoons \text{Cys145}^- + \text{His41H}^+)_{\text{apo}}$	36
$(\text{Cys145H} + \text{His41D} \rightleftharpoons \text{Cys145}^- + \text{His41H}^+)_{\text{apo}}$	31
$(\text{Cys145H} + \text{His41D} \rightleftharpoons \text{Cys145}^- + \text{His41H}^+)_{\text{N3}}$	-2
$(\text{His41E} \rightleftharpoons \text{His41D})_{\text{apo}}$	5

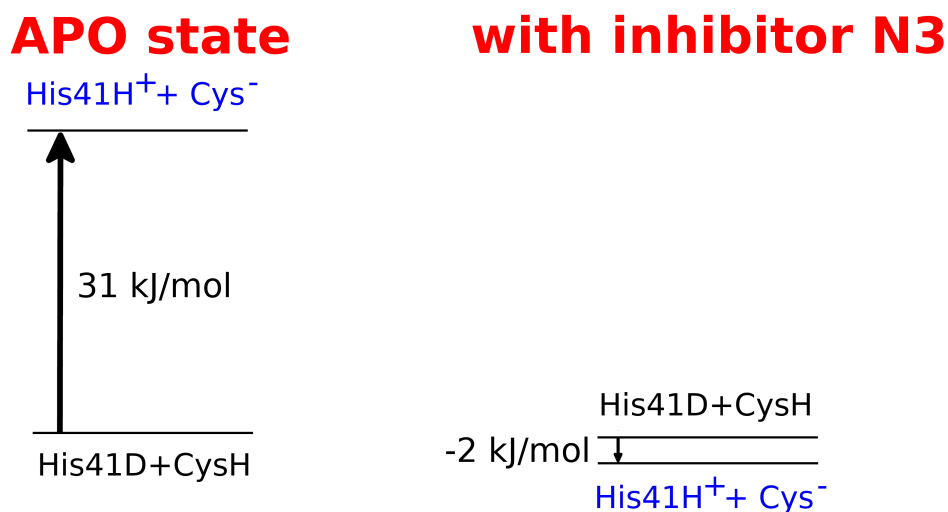
It was recently shown that a previously designed Michael acceptor inhibitor (N3) exhibits potent inhibition for SARS-CoV-2 M^{Pro}. The crystal structure of the protein in complex with N3 reveals that the C_β atom of the inhibitor vinyl group is covalently bound to the sulfur of Cys145 of the protein catalytic dyad. As the formation of this covalent bond requires a deprotonated sulfur, the functional PT from Cys145 to His41 is required for the covalent binding of the inhibitor.²²

In order to compute the energy variation upon PT in the presence of the N3 inhibitor inside the substrate binding pocket, we simulate the complex between the non-covalently bound inhibitor and the enzyme. Given the above considerations about the comparison between the His41D and His41E reactant states in the apo protein and given that previous computational results showed that the inhibitor N3 better interacts with His41D,³⁸ the computation of the PT energy in the presence of N3 focuses on the His41D reactant state. Therefore, two MD simulations are performed, one in the reactant ensemble (Cys145H + His41) with His41 protonated at N_δ and one in the product ensemble (Cys145⁻ + His41H⁺).

From the calculation (see Methods section in the SI) we obtain that, in the presence of the inhibitor in the active-site, the PT reaction free energy is significantly lower, changing

from $\Delta G^0 = 31 \pm 6$ kJ/mol to $\Delta G^0 = -2 \pm 6$ 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, with the reactant and product state being almost isoenergetic for the former (see Scheme 1). Therefore, the PT reaction is thermodynamically feasible in the presence of N3. This result is in line with a recent computational work²² showing that in the presence of N3 the reactants and products state are almost isoenergetic, with the former slightly thermodynamically favored (by ≈ 5 kJ/mol). Our estimate of a slightly negative PT free energy in the presence if the inhibitor is in agreement with the experimental observation of the covalent binding of N3 to the catalytic cysteine and to its potent inhibitory activity. 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.⁵ A mechanism similar to the one here proposed, has also been recently suggested for the covalent binding of a different inhibitor, 13b α -ketoamide.³⁹



Scheme 1: Representation of the free energy change upon PT in the apo state and in the presence of inhibitor N3

Contribution of protein residues and a conserved water molecule 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 simulations with both His41E and His41D in the apo state. 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 QCs³⁰ (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 the crystal structure of SARS-CoV-2 M^{Pro} a highly buried water molecule is present, which is packed in a tight hydrogen bond (HB) network involving His41 N_δ, His41 backbone NH group, the side chain oxygens of Asp187 and His164 N_δ. This HB network is conserved along other available crystal structures (e.g., 6y2e, 7bqy, 6y2g, 6yb7) and the buried water molecule, Cwat, has been suggested to play an active role in the PT reaction.⁴⁰ The important role of HB networks in proton transfer and transient proton binding is well recognized^{41,42} and thus the contribution of Cwat was also included as an additional “residue” in the analysis.

In Figure 2, we report qV , i.e., the contribution to the PT energy due to the electrostatic potential, of each residue in the MD simulation with His41E (A) and His41D (B). 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 highlighted with colored dots, being those that most contribute to determine the PT energy.

The residues that most relevantly contribute to the PT energy are the same for both His41E and His41D: Arg40, Glu166, Asp187 and Arg188. These are charged residues in the

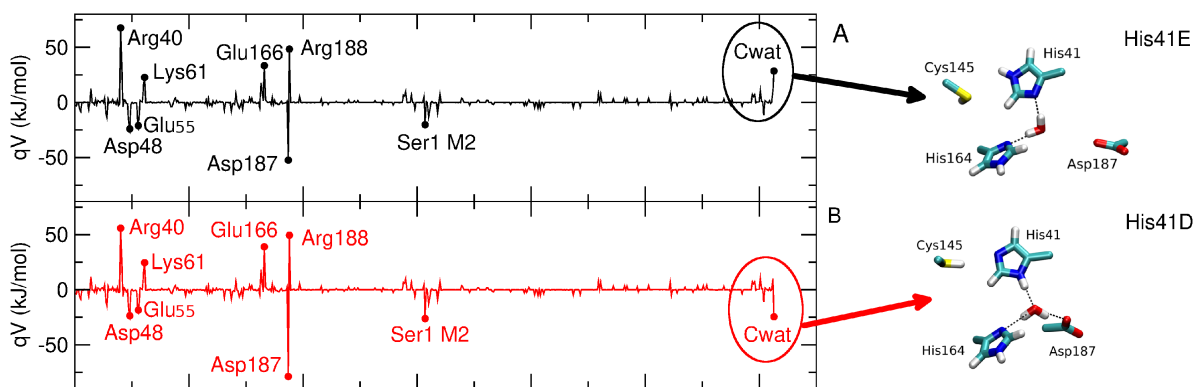


Figure 2: qV is plotted for each protein residue and Cwat for the reactant state with His41E (A) and His41D (B). qV is the mean value along the MD trajectories of the contribution due to the electrostatic potential to the PT energy. The residues featuring an absolute value of qV higher than 20 kJ/mol are highlighted with colored dots. 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 residues of the catalytic dyad (His41 and Cys145) is not included in the plot. Two representative structures of the different hydrogen bonding pattern involving Cwat in the MD simulations in the reactant state with His41E and His41D are reported on the right.

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.⁴⁰ The contribution of Ser1 of the other monomer, related to the one of Glu166 (see above), can be also seen. Very interestingly, the only significant difference between the His41E and His41D reactant state is in the contribution of Cwat: in the His41E state this contribution disfavors the PT reaction, while in the His41D state the contribution of Cwat favors the PT reaction.

The analysis of the HB network involving Cwat along the MD simulations with His41D and His41E reveals relevant differences. In the His41D simulation the HB network observed in the crystal structures is essentially maintained in the MD (see Figure S2A). In contrast, in the MD simulations with His41E the interaction between Cwat and His41 N_δ is maintained,

the one between Cwat and His164 N_δ is only partially maintained and those with Asp187 side chain oxygens and His41 backbone N are lost (see Figure S2B). In fact, as a consequence of the protonation of His41 at N_ε, Cwat changes its orientation, assuming a new configuration that prevents formation of the HBs with Asp187 and the the backbone of His41 (see representative structures in Figure 2). This rotation implies a different electrostatic effect on the side chain of His41, favoring the PT reaction in the His41D reactant state and disfavoring it in the His41E reactant state. This result further supports the above conclusion that the His41D state is the most probable reactant state for the PT reaction, both from an energetic and from a structural point of view.

Three additional minor contributions can also be observed in Figure 2: Asp48, Glu55 and Lys61. These charged residues, that are located on the protein surface not far from the catalytic dyad cleft, are an example of how the present approach can be used to enhance or inhibit the catalytic activity. These residues are in fact good candidates for point mutations. Differently from the other residues emerging from the above analysis, they are in fact located outside the active site pocket. Therefore, mutations at these sites should not relevantly perturb the active site structure. Nevertheless, mutation of one of these residues with a neutral residue would result in a non-negligible change in the PT energy favoring/disfavoring the PT reaction. In addition, the contribution of these residues shows how the catalytic activity can be modulated by allosteric control, acting on regions outside the active site cleft.

The analysis of the residues that contribute the most to the PT energy was also performed for the MD simulation in the reactant ensemble in the presence of N3. The results, reported in Figure 3B, 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 3C 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 3C are those that contribute to lowering the PT energy in the presence of the inhibitor with respect to the apo state.

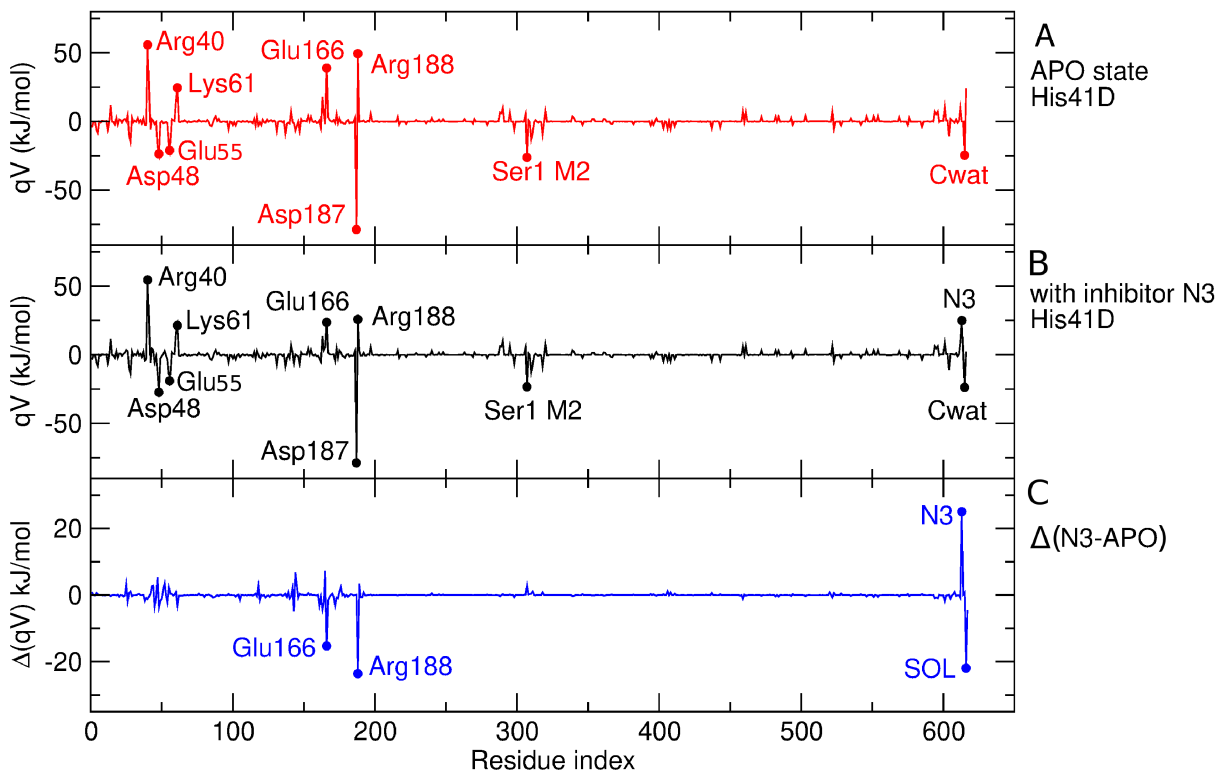


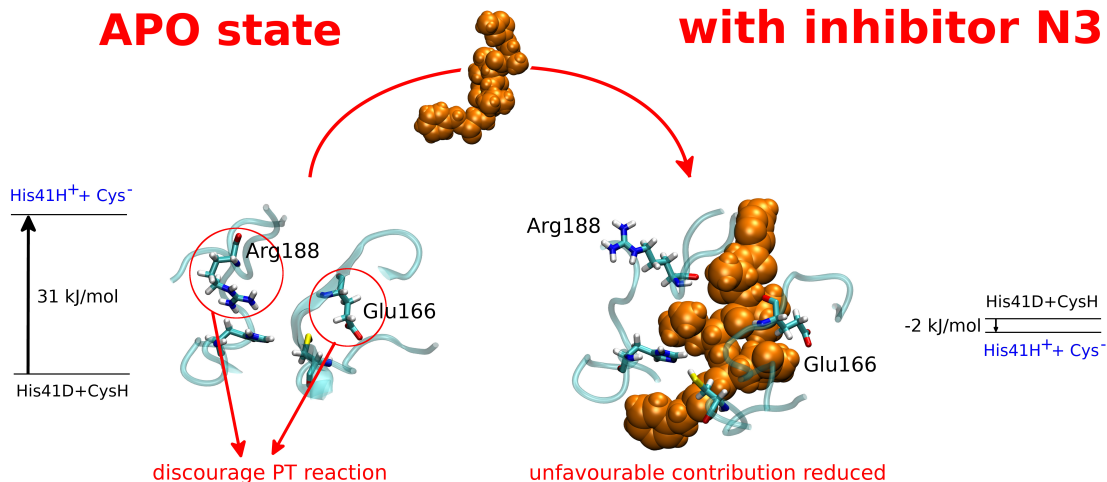
Figure 3: A and B: qV is plotted for each protein residue, Cwat and all the other water molecules as an additional virtual residue SOL for the apo state (A) and in the presence of the inhibitor N3 (B). qV is the mean value along the MD trajectories of the contribution due to the electrostatic potential to the PT energy. The residues featuring an absolute value of qV higher than 20 kJ/mol are highlighted with colored dots. 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. C: $\Delta(qV) = qV(N3) - qV(apo)$ is plotted for each protein residue and Cwat. The residues featuring an absolute value of qV higher than 10 kJ/mol are highlighted with colored dots. 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.

The most relevant (positive and negative) contributions are exerted by Glu166, Arg188, N3 and the solvent. The electrostatic contribution of N3 in Figure 3C is positive and therefore does not favor the reduction 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, likely arising from the fact that when N3 is located inside the active site pocket the catalytic dyad is less hydrated (see Figure S3). It should also be noted that the contribution of the solvent, as highlighted in Figure 3, does not include the contribution of C_{wat} , that is calculated separately (see Figure 3A and B). The contribution of C_{wat} is essentially the same in the apo state and in the presence of N3. This is in agreement with the fact that in the MD simulation in the presence of the inhibitor the same HB pattern observed in the MD simulation in the apo state with His41D is present.

Interestingly, the residues that contribute the most to the reduction of the PT energy are 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 enable 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. This reduced electrostatic effect contributes to the observed lowering of the PT energy (see Scheme 2).

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 Cys145 sulfur and the N_ϵ atom of His41 is lower and less variable in the MD simulation with N3 (see Figure S4).



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

Discussion

The present results support the hypothesis that in the apo state of SARS-CoV-2 M^{Pro} 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 favored with respect to the protonation at N_δ. However, the two tautomers can easily interconvert, given that the tautomerization free energy difference is rather small (5 kJ/mol). We also investigated which protein regions have an important role in the energy change upon PT. We show that a number of key groups (Arg40, Glu166, Asp187, Arg188 and the conserved water molecule) 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. We believe that these findings can be very useful in the screening of potential inhibitors. In fact, for covalent inhibition going through a direct PT mechanism, the deprotonation of the sulfhydryl group of Cys145 is required. The most viable route for that 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 protein regions that, in the apo state, enhance or reduce the thermodynamic feasibility of this PT reaction, 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 formation of a covalent bond.

To show that the residues that we identify here can be reliably used for the screening of new inhibitors, we turn as a test case to a previously designed Michael acceptor inhibitor, N3, that covalently binds to the Cys145 sulfur and exhibits potent inhibition for SARS-CoV-2 M^{pro}. In the presence of N3 the PT reaction free energy decreases significantly and the zwitterionic product state is almost isoenergetic with the Cys145H-His41D reactant state. The free energy decrease depends on local conformational changes induced by the inhibitor: the presence of N3 favors a lower average Cys145S-His41N_ε distance and the electrostatic contribution of two residues (Glu166 and Arg188), which was found to disfavor the PT in the apo state, is reduced in the presence of the inhibitor.

The approach presented in this work could be of significant aid in the design of new inhibitors by means of computer simulations. In particular, our approach could assist in the identification of compounds that can promote the catalytic PT reaction and, therefore, might 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. Design of analogues of known covalent inhibitors (as for example 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 with the available computational procedures for the screening of covalent 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

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) 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.
- (6) 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.
- (7) 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. **2020**,
- (8) Ramos-Guzman, C. A.; Ruiz-Pernia, J. J.; Tunon, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, doi:10.1021/acscatal.0c03420.
- (9) 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**, doi:10.1021/acs.jmedchem.0c01063.
- (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) Yang, Z.; Zhou, Y.; Liu, K.; Cheng, Y.; Liu, R.; Chen, G.; Jia, Z. Computational study on the function of water within a β -helix antifreeze protein dimer and in the process of ice-protein binding. *Biophys. Journ.* **2003**, *85*, 2599–2605.
- (14) Vreven, T.; Morokuma, K. Chapter 3 Hybrid Methods: ONIOM(QM:MM) and QM/MM. *Ann. Rep. Comp. Chem.* **2006**, *2*, 35–51.
- (15) 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.
- (16) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- (17) 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.
- (18) Gao, J.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Ann. Rev. Phys. Chem.* **2002**, *53*, 467–505.
- (19) 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.
- (20) Faraji, S.; Zhong, D.; Dreuw, A. Characterization of the Intermediate in and Identifi-

- cation of the Repair Mechanism of (6-4) Photolesions by Photolyases. *Angew. Chem. Int. Ed.* **2016**, *55*, 5175–5178.
- (21) 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.
- (22) 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.
- (23) 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.
- (24) Lu, Y.; Zhong, D. Nonequilibrium dynamics of photoinduced forward and backward electron transfer reactions. *J. Chem. Phys.* **2020**, *152*, 065102.
- (25) 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.
- (26) 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.
- (27) 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.

- (28) 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.
- (29) 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.
- (30) 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.
- (31) 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.
- (32) 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.
- (33) 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.
- (34) Zanetti-Polzi, L.; Daidone, I.; Amadei, A. Fully Atomistic Multiscale Approach for pKa Prediction. *J. Phys. Chem. B* **2020**, *124*, 4712–4722.
- (35) 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.

- (36) 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.
- (37) 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.
- (38) 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.
- (39) Poater, A. Michael Acceptors Tuned by the Pivotal Aromaticity of Histidine to Block COVID-19 Activity. *J. Phys. Chem. Lett.* **2020**, *11*, 6262–6265.
- (40) 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.
- (41) Lazaratos, M.; Karathanou, K.; Bondar, A.-N. Graphs of dynamic H-bond networks: from model proteins to protein complexes in cell signaling. *Curr. Opin. Struct. Biol.* **2020**, *64*, 79 – 87.
- (42) Karathanou, K.; Lazaratos, M.; Bertalan, E.; Siemers, M.; Buzar, K.; Schertler, G. F.; del Val, C.; Bondar, A.-N. A graph-based approach identifies dynamic H-bond communication networks in spike protein S of SARS-CoV-2. *J. Struct. Biol.* **2020**, *212*, 107617.
- (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.

Graphical TOC Entry

