Influence of Charge Configuration on Substrate Binding to SARS-CoV-2 Main Protease

Natalia Díaz[*a] and Dimas Suárez[a]

[a] Departamento de Química Física y Analítica. Universidad de Oviedo.

Avda. Julián Clavería 8. 33006. Oviedo. Spain.

While state-of-the-art computational simulations support the neutral state for the catalytic dyad of the SARS-CoV-2 main protease, the recently-reported neutron structure exhibits a zwitterionic form. To better compare the structural and dynamical features of the two charge configurations, we perform a Molecular Dynamics study of the dimeric enzyme in complex with a peptide substrate. The simulations show that the enzyme charge configuration from the neutron structure is not compatible with a catalytically-competent binding mode for peptide substrates.

The 3-chymotrypsin-like or main protease (3CL^{pro} or M^{pro}) of the SARS-CoV-2 pathogen carries out the selective proteolysis of polyproteins expressed after cell infection. It is an attractive target to treat coronavirus associated diseases[1,2] that is being scrutinized by numerous structural and computational studies in response to the current COVID-19 pandemic.[3] The enzyme is active in a homodimeric form and each protomer is constituted by three domains of similar size, the active site being located in a cleft between domains I and II. The selection of the protonation states of the active site residues, particularly of the catalytic dyad formed by Cys$_{145}$ and His$_{41}$, is a critical aspect in molecular modelling given that the details of the catalytic mechanism and/or the stability/selectivity of inhibitor complexes can be largely influenced by the amino acid charges.

The X-ray structures of 3CL^{pro} show relatively-long Cys$_{145}$@Sγ···Nε2@His$_{41}$ distances (3.5-3.9 Å) that have been considered indicative of a neutral Cys$_{145}$/His$_{41}$ pair in the ligand-free enzyme.[4] Molecular dynamics (MD) simulations in explicit solvent performed for the apo and the inhibitor-bound enzyme have also investigated various configurations for His$_{41}$, Cys$_{145}$, His$_{163}$, His$_{164}$, and His$_{172}$, concluding that the zwitterionic state of the catalytic dyad is structurally unstable.[5] Additional MD simulations of the 3CL^{pro} enzyme in complex with oligopeptides have found that the neutral catalytic dyad allows a stable mode of binding of the peptide substrate, which remains favourably oriented for catalysis all along the simulations.[6,7] In consonance with these results, hybrid quantum-mechanical molecular-mechanics (QM/MM) simulations of the catalytic or inhibition mechanisms coincide in assigning the neutral Cys$_{145}$/His$_{41}$ dyad as being energetically more stable than the zwitterionic configuration.[7-9] The same preference has been determined for the structurally-close SARS-CoV enzyme by means of QM and QM/MM calculations on the Cys→His proton transfer.[10]

The SARS-CoV-2 3CL^{pro} activity is thought to exhibit a bell-shaped pH-rate profile as that determined for the SARS-CoV protease.[11] For the latter enzyme, fitting of kinetic data to model equations provides rate constants and two pK_a values (6.2-6.4, 7.7-8.3) that presumably correspond to the catalytically competent His$_{41}$ and Cys$_{145}$ side chains. Further understanding of pH effects can be gained from theoretical pK_a estimations for all the 3CL^{pro} titratable amino acids, which have been also employed to determine their most likely protonation states. Thus, Poisson-Boltzmann calculations on different crystallographic structures yield pK_a values (e.g., <6.2 for His$_{41}$, >9.9 for Cys$_{145}$, < 5.2 for His$_{163}$, < 5.4 for His$_{164}$, and < 6.3 for His$_{172}$) that, within the pH range 6-8, support the neutral state for the catalytic dyad and the nearby His residues. These assignments have been corroborated by more reliable pK_a estimations obtained by continuous constant pH MD simulations (6.6 for His$_{41}$, >9 for Cys$_{145}$, < 5 for His$_{163}$ and His$_{164}$, and 6.6 for His$_{172}$).[12] The latter constant-pH simulations also suggest that proton binding by His$_{41}$ and, probably, His$_{172}$ could be involved in the pH rate effects.

Despite of the aforementioned studies, there is still controversy about the charge configuration of the SARS-CoV-2 3CLpro catalytic dyad. Thus, X-ray structures obtained from crystals grown in the absence of reducing agents have revealed that, at pH 7.0, none of the 3CL^{pro} Cys residues are oxidized excepting Cys$_{145}$,[13] which agrees well with its expected reactivity. However, the oxidation of Cys$_{145}$ is impeded at pH 6.0. The authors have proposed that the oxidation process at pH 7.0 would imply the presence of a Cys$_{145}^{-}$/His$_{41}^{+}$ ion pair in the apo-enzyme, the Cys$_{145}$ side chain becoming protonated at pH 6.0. More significantly, the neutron structure of the native 3CL^{pro} has been recently reported[14] which, in principle, allows the direct determination of H atom positions. The resulting 2.5 Å structure exhibits the catalytic dyad in the zwitterionic state and identifies the protonation states of other important residues (e.g., His$_{64}$, His$_{80}$, and His$_{164}$ are positively charged; His$_{163}$, His$_{172}$ and His$_{246}$ remain neutral; Cys$_{22}$, Cys$_{38}$, Cys$_{44}$, and Cys$_{128}$ also display a deprotonated thiolate).

Clearly, in view of the SARS-CoV-2 3CL^{pro} neutron structure, the results of former computational investigations favouring the neutral catalytic dyad seem questionable. As noticed by Kneller et. al.,[14] the detailed knowledge of the electrostatic environment of the catalytic site as that described in their neutron structure would have implications for computational drug design. Therefore, we investigate herein whether or not the unusual charge configuration of the neutron structure (**ZW**) is more favourable for substrate binding than the standard charge configuration (**STD**) characterized by the neutral catalytic dyad. To this end, we carried out new MD simulations in explicit solvent of the SARS-CoV-2 main protease in complex with an oligopeptide (7-mer) mimicking one of the polyprotein amino acid sequences (Ala$_{3260}$-Val$_{3261}$-Leu$_{3262}$-Gln$_{3263}$-Ser$_{3264}$-Gly$_{3265}$-Phe$_{3266}$) recognized by the enzyme.[15]

Figure 1. a, e) Views of the active site region showing the peptide substrate and the catalytic residues in the most populated cluster representatives (protomer A) of the **7JUN_STD** and **7JUN_ZW** trajectories. The backbone atoms of the scissile $Gln(P_1)\sim Ser(P_1')$ peptide linkage are shown in ball-and-stick. b-c, f-g) Schematic representation of enzyme-substrate and catalytic-dyad interactions. Average values of heavy-atom separation (Å) and % of abundances are indicated for selected contacts. Some abundances are segregated into protomer A/protomer B. d, h) Histogram of $Cys_{145}@S\gamma\cdots Gln(P1)@C$ distances. Fig. S6 in the ESI displays the corresponding models and schemes for the B protomers and the **6LU7** trajectories.
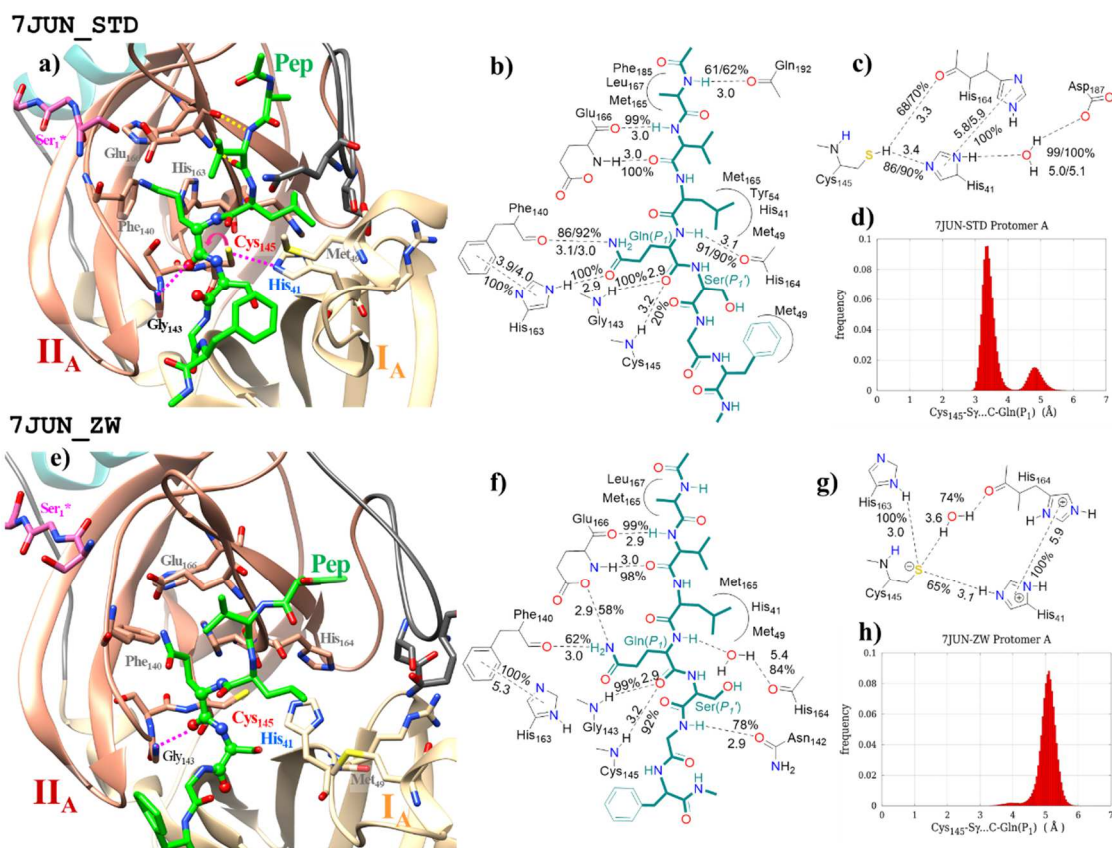
Both the **ZW** and **STD** charge configurations were modelled using an X-ray structure $(6LU7)^{16}$ and the neutron structure (7JUN), resulting in four models (**6LU7_STD**, **6LU7_ZW**, **7JUN_STD**, and **7JUN_ZW**), which were represented by the ff14SB version of the AMBER force field.[17] The selected 7-mer peptide was placed in the active site of both protomers (A and B). The MD production phase comprised 2.0 μs using the PMEMD program included in AMBER18.[18, 19] Full details of the computational settings and broad simulation results are presented in the ESI.

In terms of the global root-mean-squared-deviations (RMSD), the crystallographic structure is more widely relaxed in the aqueous environment (2.5±0.5 Å and 3.6±0.1 Å for **6LU7_STD** and **6LU7_ZW**, respectively) than the neutron structure (1.5±0.2Å and 2.1±0.3Å for **7JUN_STD** and **7JUN_ZW**; see ESI Fig S1). Nonetheless, the small fluctuations in the relative orientation of the principal axis of inertia of the two protomers

(ESI Fig. S2) and the superposition between MD-averaged and the initial structures (ESI Fig. S3) show that the overall protomer⋯protomer orientation remains quite stable in all the simulations. In fact the inter-protomer arrangement is determined by the presence of persistent polar (e.g. $Ala_7@NH/CO⋯NH/CO@Val_{125}$) and hydrophobic (e.g. $Met_6⋯Tyr_{126}$ and $Val_{125}⋯Val_{125}$) contacts (ESI Tables S1-S2) that involve residues in the three domains which, in general, are relatively distant from the charged His and Cys residues.

More pronounced differences among the SARS-CoV-2 3CL^{pro} models arise in the secondary and tertiary structures of the domains I and III. The time plots and the average RMSD values (ESI Fig. S1-S4 and Table S3) reveal that **7JUN_ZW/6LU7_ZW** depart more widely from their initial structures than their **STD** counterparts. The highest RMSD values (2.6±0.8/2.0±0.2 Å for **7JUN_ZW/6LU7_ZW**), which correspond to the domain IB, stem from the short Arg_{40}-Cys_{44} helix and the adjacent Thr_{45}-Pro_{52} loop. Although the Thr_{45}-Pro_{52} loop has been considered inherently flexible,[20] this region comprises the Cys_{38}^-, Cys_{44}^- and His_{41}^+ charged residues in the **ZW** models, which result in the ample and diverse displacements of this loop concomitant with the loss of important and well-conserved interactions involving the catalytic His_{41} (see below). In addition, the *N*-finger element constituted by residues Ser_1-Ala_7 in domain I, which is essential for catalysis, becomes disordered in the **7JUN_ZW/6LU7_ZW** trajectories due to the rupture of important H-bond interactions involving Ser_1 and Arg_4, presumably by the presence of the anionic Cys_{128} that tends to interact closely with Arg_4 (ESI Fig. S4 and Table S1). On the other hand, the **7JUN_STD** and **6LU7_STD** models maintained the relative position of the *N*-finger and the important catalytic residues. Furthermore, clustering calculations and structure superposition analysis further confirm that the four active sites explored in the **7JUN_STD** and **6LU7_STD** simulations have similar shape, flexibility and residue interactions in contrast with the **ZW** trajectories (ESI Fig. S5 and Table S4).

As expected, the impact of the **ZW** charge configuration on the domain I structure modifies the architecture of the active site and influences the peptide binding determinants. Concerning the catalytic dyad, the thiol group of the nucleophile Cys_{145} presents a stable H-bond with $His_{41}@Nε2$ (84-90%) in all the STD models (Fig. 1c; ESI Table S5), what is favourable for catalysis. In contrast, the Cys_{145} thiolate group in the **ZW** state presents more variable H-bond contacts that include the His_{163} and His_{164} side chains (water-mediated in some cases). For His_{41}, the most remarkable change on going from the **STD** to the **ZW** states is the loss of the conserved water-mediated contact with Asp_{187}. Nearby the catalytic dyad, His_{163} presents a different tautomer in the STD (singly protonated at $Nε2$) and the **ZW** (singly protonated at $Nδ1$) configurations, the $Nδ1$-protonated tautomer being incompatible with the observed role of $His_{163}@Nε2$ as a H-bond acceptor for covalent and non-covalent ligands.[21] In addition, some polar contacts like $His_{163}@Nδ1⋯His_{164}@NH$ and $His_{163}@Nε2H⋯Ser_{144}@Oγ$, which are stable in all the **STD** protomers, are lost in the **ZW** state (ESI Table S5). There are also differences in the non-polar interactions that shape the back wall of the S_1 subsite. Compared to the **STD** configuration, we found that the His_{163}/Phe_{140} π-π stacking is partially replaced by the His_{163}/His_{172} contact in some of the **ZW** protomers. For instance, comparing protomer A in the **7JUN_STD** and **7JUN_ZW** simulations, the His_{163}/Phe_{140} and His_{163}/His_{172}

dispersion energies change from -3.2 and -1.3 kcal/mol to -1.3 and -2.4 kcal/mol, respectively (ESI Table S6).

Regardless of the initial structure adopted for the simulations, the 7-mer molecules present in the **STD** Michaelis complexes exhibit a highly-stable and nearly-identical mode of binding (Fig. 1 and ESI Fig. S6). The structural and dynamical features of these complexes, which coincide with those observed in previous simulations,[6] are clearly suitable for catalysis. Both polar contacts like the H-bond interactions between the backbone groups of Glu_{166} and $Val(P_3)$ (99-100% of abundance in all the **STD** protomers), and van der Waals interactions that bind $Ala(P_4)$ and $Leu(P_2)$ within the S_4 (Met_{165}, Leu_{167}, and Phe_{185}) and the S_2 subsites (His_{41} and Met_{165}), respectively, contribute to the binding of the 7-mer substrate (ESI Tables S7-S8). Met_{49} exhibits a remarkable flexibility given that it interacts with either the $Leu(P_2)$ or the $Phe(P_3')$ side chain, depending on the MD trajectory/protomer. The conserved $Gln(P_1)$ is also well anchored within the active site: its backbone amido and carbonyl groups give H-bond contacts with the $His_{164}@O$ (90-96%) and the $Gly_{143}@NH$ (99-100%) sites while the side chain remains H-bonded to the His_{163} imidazole (97-100%) and the $Phe_{140}@O$ (81-92%) groups. Concerning the interatomic $Cys_{145}@S\gamma\cdots Gln(P_1)@C$ distance, the **STD** simulations show bimodal distributions (Fig. 1 and ESI Fig. S6), the largest peak being centred at 3.5 Å and the minor peak around 5.0 Å. As a result, the Cys_{145} thiol group is well-oriented for nucleophilic attack towards the $Gln(P_1)$ carbonyl group and for proton transfer towards the His_{41} imidazol.

The conformational sampling in aqueous solution achieved by the MD simulations reveals that, in the **ZW** state, the average location of the 7-mer substrate is not favourable for catalysis. In the less reactive case, the promoter A in **6LU7_ZW**, water molecules diffuse into the oxyanion hole site, rupturing the bifurcated interaction of the $Gln(P_1)$ backbone with the NH groups of Gly_{143} and Cys_{145} and rearranging the position of the $Gln(P_1)$ side chain and that of the $Ser(P_1')$-$Gly(P_2')$-$Phe(P_3')$-Nme moiety (ESI Fig. S6). As a consequence, the mean $Cys_{145}@S\gamma\cdots Gln(P_1)@C$ distance in the **6LU7_ZW** simulation (protomer A) amounts to 7.9±1.1 Å. In the rest of the **ZW** complexes, 7-mer remains anchored at the oxyanion hole (95-99% of abundance for the $Gln(P_1)@O$ contact with $Gly_{143}@NH$), but the protonation states selected for His_{163} and His_{164} impede their interactions with $Gln(P_1)$ (Fig. 1; ESI Table S7). Moreover, these complexes would be most-likely catalytically unproductive due to their long $Cys_{145}@S\gamma\cdots Gln(P_1)@C$ distances (>5 Å; ESI Fig. S6) which, in turn, reflect the inadequate orientation of the Cys_{145} thiolate group.

In summary, although our simulations do not explain the discrepancy between the neutron structure of the SARS-CoV-2 $3CL^{pro}$ and computational modelling about the protonation state of the catalytic residues, they provide compelling evidence that the **ZW** charge configuration most likely perturbs the structures of the domain I and the active site, and largely impairs the pre-reactive mode of binding of peptide substrates. In contrast, the compatibility of the **STD** protonation state with the X-ray structures and with the stable and productive binding of peptide molecules gains further support. Our results agree with much of the current theoretical and structural data, which favour the neutral state for the Cys_{145}/His_{41} catalytic dyad and support the standard protonation states for other important residues. This charge configuration seems therefore the most appropriate reference for

computer-aided drug design and/or for theoretical studies of the catalytic mechanism. Finally, we note that, although further computational studies may add more conclusive evidence about the SARS-CoV-2 3CL^{pro} charge distribution, an integrative approach combining both theoretical and direct experimental observations may be required to settle this question.

References

1. A. Artese, V. Svicher, G. Costa, R. Salpini, V. C. Di Maio, M. Alkhatib, F. A. Ambrosio, M. M. Santoro, Y. G. Assaraf, S. Alcaro and F. Ceccherini-Silberstein, *Drug Resist. Updat.*, 2020, **53**, 100721.

2. S. Ullrich and C. Nitsche, *Bioorg. Med. Chem. Lett.*, 2020, **30**, 127377.

3. A. Francés-Monerris, C. Hognon, T. Miclot, C. García-Iriepa, I. Iriepa, A. Terenzi, S. Grandemange, G. Barone, M. Marazzi and A. Monari, *J. Proteome Res.*, 2020, **19**, 4291-4315.

4. D. W. Kneller, G. Phillips, H. M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates and A. Kovalevsky, *Nat. Commun.*, 2020, **11**, 3202.

5. A. Pavlova, D. L. Lynch, I. Daidone, L. Zanetti-Polzi, M. D. Smith, C. Chipot, D. W. Kneller, A. Kovalevsky, L. Coates, A. A. Golosov, C. J. Dickson, C. Velez-Vega, J. S. Duca, J. V. Vermaas, Y. T. Pang, A. Acharya, J. M. Parks, J. C. Smith and J. C. Gumbart, *Chem. Sci.*, 2021, **12**, 1513-1527.

6. D. Suárez and N. Díaz, *J. Chem. Inf. Model.*, 2020, **60**, 5815-5831.

7. C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ACS Catal.*, 2020, **10**, 12544-12554.

8. K. Swiderek and V. Moliner, *Chem. Sci.*, 2020, **11**, 10626-10630.

9. K. Arafet, N. Serrano-Aparicio, A. Lodola, A. J. Mulholland, F. V. González, K. Świderek and V. Moliner, *Chem. Sci.*, 2021, **12**, 1433-1444.

10. A. Paasche, T. Schirmeister and B. Engels, *J. Chem. Theory Comput.*, 2013, **9**, 1765-1777.

11. C. Huang, P. Wei, K. Fan, Y. Liu and L. Lai, *Biochemistry*, 2004, **43**, 4568-4574.

12. N. Verma, J. A. Henderson and J. Shen, *J. Am. Chem. Soc.*, 2020, **142**, 21883–21890.

13. D. W. Kneller, G. Phillips, H. M. O'Neill, K. Tan, A. Joachimiak, L. Coates and A. Kovalevsky, *IUCrJ*, 2020, **7**, 1028-1035.

14. D. W. Kneller, G. Phillips, K. L. Weiss, S. Pant, Q. Zhang, H. M. O'Neill, L. Coates and A. Kovalevsky, *J. Biol. Chem.* , 2020, **295**, 17365-17373.

15. T. U. Consortium, *Nucleic Acids Res.*, 2021, **49**, D480-D489.

16. Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao and H. Yang, *Nature*, 2020, **582**, 289-293.

17. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696-3713.

18. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *WIREs Comput. Mol. Sci.*, 2013, **3**, 198-210.

19. S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374-380.

20. T. Muramatsu, C. Takemoto, Y.-T. Kim, H. Wang, W. Nishii, T. Terada, M. Shirouzu and S. Yokoyama, *Proc. Nat. Acad. Sci.*, 2016, **113**, 12997-13002.

21. A. Shitrit, D. Zaidman, O. Kalid, I. Bloch, D. Doron, T. Yarnizky, I. Buch, I. Segev, E. Ben-Zeev, E. Segev and O. Kobiler, *Sci. Rep.*, 2020, **10**, 20808.