Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Chemical Physics Letters 750 (2020) 137489

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Research paper

# Identification of potential binders of the main protease $3CL^{pro}$ of the COVID-19 via structure-based ligand design and molecular modeling

Marina Macchiagodena, Marco Pagliai, Piero Procacci*

*Dipartimento di Chimica "Ugo Schiff", Universitá degli Studi di Firenze, Via della Lastruccia 3, Sesto Fiorentino I-50019 Italy*

ARTICLE INFO

*Keywords:*
COVID-19
SARS-CoV2
Coronavirus
3CL-PRO
Coronavirus main protease
Molecular docking
Molecular dynamics
3CL-PRO inhibitor
Binding affinity

ABSTRACT

We have applied a computational strategy, using a combination of virtual screening, docking and molecular dynamics techniques, aimed at identifying possible lead compounds for the non-covalent inhibition of the main protease $3CL^{pro}$ of the SARS-CoV2 Coronavirus. Based on the X-ray structure (PDB code: 6LU7), ligands were generated using a multimodal structure-based design and then docked to the monomer in the active state. Docking calculations show that ligand-binding is strikingly similar in SARS-CoV and SARS-CoV2 main proteases. The most potent docked ligands are found to share a common binding pattern with aromatic moieties connected by rotatable bonds in a pseudo-linear arrangement.

At the beginning of this year, the world was dismayed by the outbreak of a severe viral acute respiratory syndrome (SARS), known as COVID-19, that rapidly spreads from its origin in the Hubei Chinese district to virtually whole China and, as of today, to more than 200 countries and territories around the world [1]. The new coronavirus, named SARS-CoV2 and believed to have a zoonotic origin, has infected thus far about 2450000 people worldwide with thousands in critical conditions, causing the death of 166000 people. The SARS-CoV2's genome [2,3] has a large identity [4] with that of the SARS-CoV whose epidemic started in early in 2003 and ended in the summer of the same year.

Most of the Coronaviridae genome encodes two polyproteins, pp1a and, through ribosomal frameshifting during translation [5], pp1ab. These polyproteins are cleaved and transformed in mature non-structural proteins (NSPs) by the two proteases $3CL^{pro}$ (3C-like protease) and $PL^{pro}$ (Papain Like Protease) encoded by the open reading frame 1 [6]. NSPs, in turn, play a fundamental role in the transcription/replication during the infection [5]. Targeting these proteases may hence constitute a valid approach for antiviral drug design. The catalytically active $3CL^{pro}$ is a dimer. Cleavage by $3CL^{pro}$ occurs at the glutamine residue in the P1 position of the substrate via the protease CYS-HIS dyad in which the cysteine thiol functions as the nucleophile in the proteolytic process [7]. While dimerization is believed to provide a substrate-binding cleft between the two monomers [8], in the dimer the solvent-exposed CYS-HYS dyads are symmetrically located at the opposite edges the cleft, probably acting independently [9]. As no host-cell proteases are currently known with this specificity, early drug discovery was directed towards the so-called covalent Michael inhibitors [10], via electrophilic attack to the cysteinyl residue. On the other hand, the consensus in drug discovery leads to excluding electrophiles from drug candidates for reasons relating to safety and adverse effects such as allergies, tissue destruction, or carcinogenesis [11].

In spite of the initial effort in developing compounds with anti-coronavirus activity following the SARS outbreak [12], no anti-viral drug was ever approved or even reached the clinical stage due to a sharp decline in funding of coronavirus research after 2005–2006, based on the erroneous conviction that chance of a repetition of a new zoonotic transmission was extremely unlikely [6], The most potent non-covalent inhibitor for $3CL^{pro}$, ML188, was reported nearly ten years ago [13] with moderate activity in the low micromolar range [14].

According to the latest report of the structure of $3CL^{pro}$ from SARS-CoV2 [15] (PDB code 6LU7) and the available structure of $3CL^{pro}$ from SARS-CoV [12], (PDB code 1UK4), the two proteases differ by only 12 amino acids, with $\alpha$ carbon atoms all lying at least 1 nm away from the $3CL^{pro}$ active site (see Fig. 1a). The substrate-binding pockets of two coronavirus main proteases are compared in Fig. 1b, exhibiting a strikingly high level of alignment (RMSD = 0.99 Å) of the key residues involved in substrate binding, including the CYS145$\cdots$HIS41 dyad, and THR45, MET49, PHE140, ASN142, ASP187, ARG188, GLN189, MET165, HIS172, GLU166. The latter are believed to provide the

---

* Corresponding author.
*E-mail address:* procacci@unifi.it (P. Procacci).

https://doi.org/10.1016/j.cplett.2020.137489
Received 31 March 2020; Received in revised form 14 April 2020; Accepted 15 April 2020
Available online 18 April 2020
0009-2614/ © 2020 Elsevier B.V. All rights reserved.

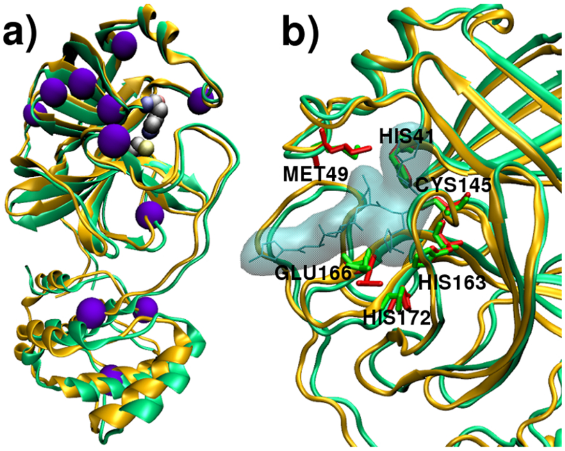

**Fig. 1.** (a) SARS-CoV2 (orange) and SARS-CoV (green) main proteases. Violet spheres correspond to the alpha carbons of the 12 differing residues. Grey spheres indicate the CYS-HIS dyad (b) binding pocket with the main residues in bond representation (green and red for SARS-CoV2 and SARS-CoV, respectively). The shaded mark the binding site for the substrate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

opening gate for the substrate in the active state [12].

Fig. 1(a,b) strongly suggest that effective non-covalent inhibitors for SARS-CoV and SARS-CoV2 main proteases should share the same structural and chemical features. In order to investigate this matter, we have performed a molecular modeling study on both the 6LU7 and 1UK4 structures. 6LU7 is the monomer of the main protease in the active state with the N3 peptidomimetic inhibitor [15] while 1UK4 is the dimer with the protomer chain A in the active state [12]. The main protease monomer contains three domains. Domains I and II (residues 8-101 and 102-184) are made of antiparallel $\beta$-barrel structures in a chymotrypsin-like fold responsible for catalysis [16].

The 6LU7 structure was first fed to the *PlayMolecule* web application [17] using a novel virtual screening technique for the multimodal structure-based ligand design [18], called Ligand Generative Adversarial Network (LIGANN). Ligands in LIGANN are generated so as to match the shape and chemical attributes of the binding pocket and decoded into a sequence of SMILES enabling directly the structure-based *de novo* drug design. SMILES codes for ligands were obtained using the default LIGANN values for shapes and channels with the cubic box center set at the midpoint vector connecting the SH and NE atoms of the CYS-HIS dyad in the 6LU7 structure. The *PlayMolecule* interface delivered 93 optimally fit non-congeneric compounds, spanning a significant portion of the chemical space, whose SMILES and structures are reported in the Supporting Information (SI). Each of these compounds was docked to the 6LU7 and to the 1UK4 structures, using Autodock4 [19] with full ligand flexibility. For both structures, the docking was repeated by setting the dyad with the residue in their neutral (CYS-HIS) and charged state ($CYS^{-}/HIS^{+}$). Details on Docking parameters are given in the SI.

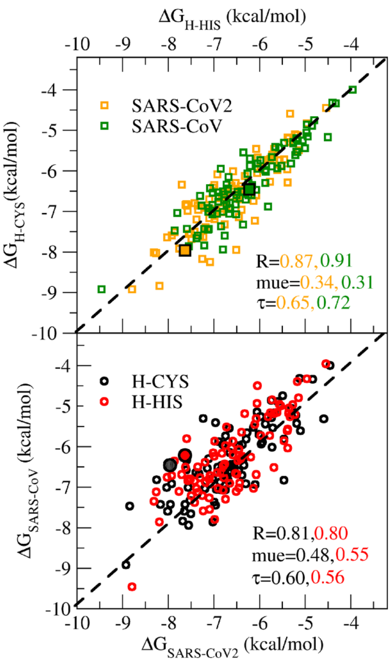

**Fig. 2.** Correlation diagrams of autodock-computed binding free energies for 93 ligands of the SARS-CoV and SARS-CoV2 3CL$^{\text{pro}}$ structures. $R$, mue, $\tau$ indicate the Pearson correlation coefficient, the mean unsigned error, and the Kendall rank coefficient, respectively. Upper panel: correlation diagram between ligand free energies obtained with the charged CYS$^{-1}$-HIS$^{+}$ and with neutral CYS-HIS dyad. Lower panel: correlation diagram between ligand free energies of SARS-CoV2 and SARS-CoV. Larger symbols refer to ML188.

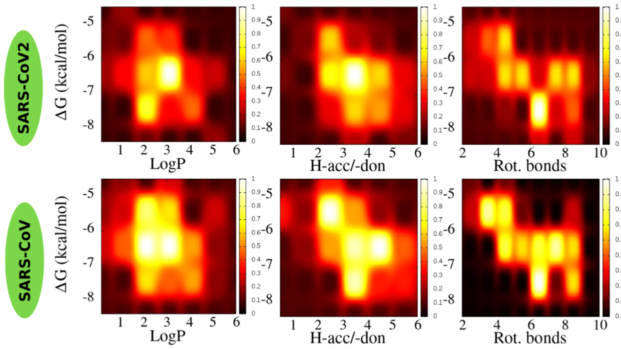

**Fig. 3.** 2D probability histograms $\Delta G$ with LogP (left), H-bond acceptors or donors (center) and rotatable bonds (right) for the 93 compounds of Table S1 of the SI. The common color-coded $z$-scale on the right corresponds to the 2D probability. Results for SARS-CoV2 and SARS-CoV in upper and lower panels, respectively.

Results for the binding free energies of the 93 3CL$^{\text{pro}}$ ligands are reported in Fig. 2. Binding free energies are comprised in the range 4–9 kcal/mol and are found to be strongly correlated for the two protonation states of the CYS-HIS dyad. Correlation is still high when ligand binding free energies for the main proteases are compared, confirming that good binders for SARS-CoV are, in general, also good binders for SARS-CoV2 3CL$^{\text{pro}}$. For each of these compounds, using the XLOGP3 methodology [20], we computed the octanol/water partition coefficient (LogP) to assess the distribution in hydrophobic and cytosolic environments. LogP values range from $-0.5$ to 5 with a number of rotatable bonds from 2 to 12. Most of the LIGANN compounds bear from 2 to 5 H-bond acceptor or donors (Table S1 of the SI). In Fig. 3, we show the probability distributions for $\Delta G$ correlated in turn to the LogP, number of H-bond donor/acceptors and number of rotatable bonds. We note, on the left and central panel, sharp maxima for $\text{LogP} = 3{:}\ 4$, $\Delta G = -7{:}\ -8$ and for $\text{H} - \text{acc/don} = 3$, $\Delta G = -6{:}\ -7$, respectively, suggestive of a ligand-protein association driven mostly by hydrophobic interactions. We must stress that the computed $\Delta G$ pertains to the associations of the ligand with *one* protein whatever the state of association of the protein. At free ligand concentration equal to $K_d \equiv e^{-\Delta G/RT}$, i.e. when half of the protein molecules are inhibited, the probability to have *both* monomers inhibited is equal to 1/4, whatever the dissociation constant of the dimer [21], hence the need for identifying nanomolar or subnanomolar inhibitors of 3CL$^{\text{pro}}$.

Table 1 shows the chemical structures of the five compounds exhibiting the highest binding affinity to the main protease of SARS-CoV2 when the CYS-HIS dyad is in the neutral state. None of these

**Table 1**
Computed binding free energies (kcal/mol), $\Delta G$, of the best five binders for SARS-CoV2 $3CL^{pro}$. $\Delta G$ values are reported for the two protonation states of the dyad and for SARS-CoV and SARS-CoV2 main protease. Below the 2D structures and $k_d$ values.

| | CoV19 | | SARS | | |
|---|---|---|---|---|---|
| Comp. | H-CYS | H-HIS | H-CYS | H-HIS | LogP |
| 27 | −8.92 | −8.79 | −8.92 | −9.46 | 4.90 |
| 30 | −8.84 | −8.19 | −7.47 | −7.86 | 3.74 |
| 39 | −8.25 | −7.08 | −6.82 | −6.72 | 6.06 |
| 77 | -8.17 | −7.25 | −7.43 | −7.21 | 2.03 |
| 19 | −8.03 | −8.26 | −7.01 | −7.12 | 5.58 |
| ML188 | −7.96 | −7.63 | −6.46 | −6.22[a] | 4.97 |

[a] Experimental value for ML188 is[14] $\Delta G = -7.98$ kcal/mol.

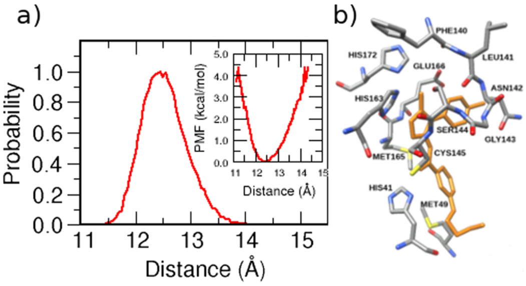

**Fig. 4.** (a) Probability distribution of the distance between the CoM of compound **27** and domain I + II of $3CL^{PRO}$, as obtained from MD simulations (inset: the corresponding PMF). (b) Binding pocket of 6LU7 with ligand **27**. The time record of the minimal distances between ligand and the depicted nearby residues are reported in Figs. S6, S7 of the SI.

compounds is commercially available, although some of them (**27**, **31**, **40**) show a high degree of similarity with known structures according to the Tanimoto metrics.[22] The structures of Table 1, as well as many of those reported in Figs. S1-S5 of the SI, seem to share a common pattern with aromatic moieties connected by rotatable bonds in a pseudo-linear arrangement. Table 1 shows the binding free energy data of these five best ligands for both CoV proteases and both protonation states of the dyad.

Inspection of Table 1 confirms that SARS-CoV2 best binders **27**, **29**, **39**, **77**, **19** are also good binders for SARS-CoV $3CL^{pro}$. Remarkably, compound **27** is consistently the most potent ligand for the two proteases, irrespective of the dyad protonation state. In the Table 1 we also report the Autodock4-computed binding free energy for ML188 ($-6.2$ and $-6.5$ kcal/mol for the H-HIS and H-CYS tautomers), not too distant indeed from the experimentally determined value of $-8$ kcal/mol, lending support for the LIGANN-Autodock4 protocol used in identifying the lead compounds of Table 1.

In order to assess the stability of the $3CL^{pro}$-**27** association, we have performed extensive molecular dynamics simulations[23,24] of the bound state with explicit solvent. The overall structural information was obtained by combining data from three independent simulations (for a total of about 120 ns), all started from the best docking pose of **27** on the 6LU7 monomeric structure. Further methodological aspects [25] are provided in the SI. In Fig. 4, we show the probability distribution, $P(R)$, of the distance $R$ between the center of mass (CoM) of the ligand and that of the domains I + II. The distribution has a Gaussian shape with a half-width of about 1 Å, exhibiting only a minor positive skewness and defining a tight binding site volume $V_{\rm site}$ of few Å$^3$ at most.[26] The MD-determined $P(R)$ shows that the ligand never leaves the binding pocket at any stage during the whole simulation. In the inset of Fig. 4a, we show the potential of mean force (PMF) along the ligand-protein CoM distance $R$, computed as $v(R) = -RT\log(P(R)/\max[P(r)])$. As $1/K_d = \int_{V_{\rm site}} e^{-\beta v(R)} d\mathbf{R}$,[27], the steepness of the curve is suggestive of a profound minimum and hence of a large association constant, confirming the indication obtained from the Docking calculations. Fig. 4b shows polar and hydrophobic residues found in contact with the ligand **27**. All essential residues for binding are included, with the addition of MET165, PHE140 and LEU141 hydrophobic residues, consistently lingering near the pyrazolic or the chlorinated phenyl rings of **27**, in agreement with the hydrophobic character of the interaction.

Figs. 3 and 4 show possible avenues for improvement. For example, forcing the L-shaped binding structure in bulk also, by redesigning the rotatable connectors in the ligand, may reduce the penalty due conformational entropy loss upon binding,[26] hence boosting the ligand affinity for $3CL^{pro}$. Building upon this knowledge, we plan to optimize the lead using MD simulations coupled to relative binding free energy calculation on congeneric variants [28], eventually providing *in silico* determined anti-viral compounds to be synthesized an experimentally tested *in vitro* and *in vivo*.

The infection rate for COVID-19 in China is currently declining for days. As the road for delivering an effective anti-viral drug is still a long one indeed, the SARS-CoV2 harsh lesson, nonetheless, should not be forgotten once the emergency will end, hoping that our contribution can pave the way for the design of effective non-covalent antiviral drugs for the present and future Coronavirus emergencies.

## CRediT authorship contribution statement

**Marina Macchiagodena:** Data curation, Investigation, Writing-review & editing. **Marco Pagliai:** Investigation, Writing-review & editing. **Piero Procacci:** Conceptualization, Investigation, Writing- review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

The authors thank MIUR-Italy ("Progetto Dipartimenti di Eccellenza 2018-2022" allocated to Department of Chemistry "Ugo Schiff").

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cplett.2020.137489.

## References

[1] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, Lancet Infect. Dis. (2020).
[2] 2020; Viralzone News, https://viralzone.expasy.org.
[3] 2020; The National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov.
[4] A. Shanker, D. Bhanu, A. Alluri, Analysis of whole genome sequences and homology modelling of a 3C like peptidase and a non-structural protein of the novel coronavirus COVID-19 shows protein ligand interaction with an Aza-peptide and a noncovalent lead inhibitor with possible antiviral properties, ChemRxiv (2020).
[5] V. Thiel, K.A. Ivanov, A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weißbrich, E.J. Snijder, H. Rabenau, H.W. Doerr, A.E. Gorbalenya, J. Ziebuhr, Mechanisms and enzymes involved in SARS coronavirus genome expression, J. Gen. Virol. 84 (2003) 2305–2315.
[6] R. Hilgenfeld, From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design, FEBS J. 281 (2014) 4085–4096.
[7] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CLpro) structure: basis for design of Anti-SARS drugs, Science 300 (2003) 1763–1767.
[8] C.-P. Chuck, L.-T. Chong, C. Chen, H.-F. Chow, D.C.-C. Wan, K.-B. Wong, Profiling of substrate specificity of SARS-CoV 3CL, PloS One 5 (2010) e13197.
[9] J. Shi, J. Sivaraman, J. Song, Mechanism for controlling the dimer-monomer switch

and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease, J. Virol. 82 (2008) 4620–4629.

[10] M.H. Johansson, Reversible michael additions: covalent inhibitors and prodrugs, Mini-Rev. Med. Chem. 12 (2012) 1330–1344.

[11] A. Vasudevan, M.A. Argiriadi, A. Baranczak, M.M. Friedman, J. Gavrilyuk, A.D. Hobson, J.J. Hulce, S. Osman, N.S. Wilson, in: D.R. Witty, B. Cox (eds.), Chapter One – Covalent Binders in Drug Discovery. Prog. Med. Chem., vol. 58. Elsevier, 2019; pp 1–62.

[12] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G.F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z. Rao, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, Proc. Natl. Acad. Sci. USA 100 (2003) 13190–13195.

[13] J. Jacobs, S. Zhou, E. Dawson, J.S. Daniels, P. Hodder, V. Tokars, A. Mesecar, C.W. Lindsley, S.R. Stauffer, Discovery of Non-Covalent Inhibitors of the SARS Main Proteinase 3CLpro. Probe Reports from the NIH Molecular Libraries Program 2010, https://www.ncbi.nlm.nih.gov/books/NBK133447/.

[14] J. Jacobs, V. Grum-Tokars, Y. Zhou, M. Turlington, S.A. Saldanha, P. Chase, A. Eggler, E.S. Dawson, Y.M. Baez-Santos, S. Tomar, A.M. Mielech, S.C. Baker, C.W. Lindsley, P. Hodder, A. Mesecar, S.R. Stauffer, Discovery, synthesis, and structure-based optimization of a series of N-(tert-Butyl)-2-(N-arylamido)-2-(pyridin-3-yl) Acetamides (ML188) as potent noncovalent small molecule inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL protease, J. Med. Chem. 56 (2013) 534–546.

[15] X. Liu, B. Zhang, Z. Jin, H. Yang, Z. Rao, The Crystal Structure of 2019-nCoV Main Protease in Complex with an Inhibitor N3. RSCB PDB, pdbode: 6LU7.

[16] T. Hu, Y. Zhang, L. Li, K. Wang, S. Chen, J. Chen, J. Ding, H. Jiang, X. Shen, Two adjacent mutations on the dimer interface of SARS Coronavirus 3C-like protease cause different conformational changes in crystal structure, Virology 388 (2009) 324–334.

[17] PlayMolecule™, https://www.acellera.com, accessed 20 February 2020.

[18] M. Skalic, D. Sabbadin, B. Sattarov, S. Sciabola, G. De Fabritiis, From target to drug: generative modeling for the multimodal structure-based ligand design, Mol. Pharm. 16 (2019) 4282–4291.

[19] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791.

[20] T. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y. Li, R. Wang, L. Lai, Computation of octanol-water partition coefficients by guiding an additive model with knowledge, J. Chem. Inf. Model. 47 (2007) 2140–2148.

[21] V. Graziano, W.J. McGrath, L. Yang, W.F. Mangel, SARS CoV main proteinase: the monomer-dimer equilibrium dissociation constant, Biochemistry 45 (2006) 14632–14641.

[22] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, S.H. Bryant, PubChem substance and compound databases, Nucleic Acids Res. 44 (2016) D1202–D1213.

[23] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845.

[24] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.

[25] M. Macchiagodena, M. Pagliai, C. Andreini, A. Rosato, P. Procacci, Upgrading and validation of the AMBER force field for histidine and cysteine zinc(ii)-binding residues in sites with four protein ligands, J. Chem. Inf. Model. 59 (2019) 3803–3816.

[26] P. Procacci, R. Chelli, Statistical mechanics of ligand-receptor noncovalent association, revisited: binding site and standard state volumes in modern alchemical theories, J. Chem. Theory Comput. 13 (2017) 1924–1933.

[27] M.K. Gilson, J.A. Given, B.L. Bush, J.A. McCammon, The statistical-thermodynamic basis for computation of binding affinities: a critical review, Biophys. J. 72 (1997) 1047–1069.

[28] M.R. Shirts, D.L. Mobley, An introduction to best practices in free energy calculations, Methods Mol. Biol. 924 (2013) 271–311.