

Computational Characterization of the Substrate Activation in the Active Site of SARS-CoV-2 Main Protease

Maria G. Khrenova^{1,2}, *Vladimir G. Tsirelson*^{3,4}, *Alexander V. Nemukhin*^{1,5}

© The Authors 2020. This paper is published with open access at SuperFri.org

Molecular dynamics simulations with the QM(DFT)/MM potentials are utilized to discriminate between reactive and nonreactive complexes of the SARS-CoV-2 main protease and its substrates. Classification of frames along the molecular dynamic trajectories is utilized by analysis of the 2D maps of the Laplacian of electron density. Those are calculated in the plane formed by the carbonyl group of the substrate and a nucleophilic sulfur atom of the cysteine residue that initiates enzymatic reaction. Utilization of the GPU-based DFT code allows fast and accurate simulations with the hybrid functional PBE0 and double-zeta basis set. Exclusion of the polarization functions accelerates the calculations 2-fold, however this does not describe the substrate activation. Larger basis set with d-functions on heavy atoms and p-functions on hydrogen atoms enables to disclose equilibrium between the reactive and nonreactive species along the MD trajectory. The suggested approach can be utilized to choose covalent inhibitors that will readily interact with the catalytic residue of the selected enzyme.

Keywords: SARS-CoV-2 main protease, QM/MM MD, GPU-accelerated algorithms, substrate activation.

Introduction

The combined quantum mechanics/molecular mechanics (QM/MM) approach is a proper tool to study chemical reactions in the active sites of enzymes. It allows considering chemical reaction at the QM level taking into account the electrostatic field and steric constraints coming from the rest of the protein and polar water solvent. Reliable results can be obtained only if a relatively high level of theory for the QM subsystem is applied. It was demonstrated that utilization of the density functional theory (DFT) with the hybrid functionals [23, 24] is suitable for such purpose, whereas simplified semiempirical methods like, for example, DFTB fails [25]. DFT-based QM/MM calculations require utilization of supercomputer facilities and many CPUs. Recently, considerable efforts have been performed to develop a GPU-based DFT code [8, 9, 11, 12, 18, 19]. It was implemented in the TeraChem program package [17]. Later, the interface for the QM/MM calculations was proposed [10]. It is based on the powerful NAMD program [16] and the corresponding supporting features can be utilized in QM/MM simulations. This implementation demonstrates dramatic speedup: the same calculations of energy gradient for the system with about 1000 basis functions can be performed on a GPU in less than 2 minutes or on 100 CPUs in about 5 minutes. This allows not only geometry optimization requiring hundreds of gradient calculations for each stationary point on the potential energy surface, but thousands of calculations to obtain a molecular dynamics trajectory.

Computational studies of molecular mechanisms of enzymatic reactions are important for biomedical applications. Analysis of the substrate specificity of proteases can be utilized to construct covalent inhibitors. Those form covalent bonds with the catalytic residues and com-

¹Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia

²Bach Institute of Biochemistry, Federal Research Center "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia

³Mendeleev University of Chemical Technology, Moscow, Russia

⁴South Ural State University (National Research University), Chelyabinsk, Russia

⁵Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, Russia

pletely abolish enzymatic activity. This is of great importance for the main protease from the SARS-CoV-2 as it is responsible for partitioning of a synthesized viral polypeptides to the specific fragments that forms a set of proteins required for the virus replication. Elimination of enzymatic activity of the main protease leads to the virus death.

It was recently demonstrated for the main protease from the SARS-CoV-2 that the analysis of QM/MM MD trajectories of the enzyme-substrate complexes can explain the substrate specificity of this enzyme [6]. Also, it was shown that the proper theory level is crucial to obtain valuable results. Utilization of the hybrid DFT functional PBE0 [1] allows one to discriminate reactive and nonreactive enzyme-substrate complexes along MD trajectory, whereas GGA-type functional PBE [14, 15] that lacks exact HF exchange fails.

In this communication, we address an important question of the theory level required for the proper description of the dynamic behaviour of the enzyme-substrate complex taking a topical example of the SARS-CoV-2 main protease and its substrate. We test the influence of the basis set, namely presence of polarization functions (d-functions on heavy atoms and p-functions on hydrogen atoms) in QM calculations, as their exclusion speeds up calculations about 2-fold.

1. Methodology

We borrowed the model system of the SARS-CoV-2 main protease with its oligopeptide substrate from the recent study [6]. The main protease exists in the dimeric form and the active site of one of monomers is considered (Fig. 1). The substrate is involved in three key interatomic interactions with the active site: two hydrogen bonds with the oxyanion hole of the enzyme formed by the NH groups of the backbones of glycine G143 and cysteine C145; specific interaction formed prior the nucleophilic attack between the carbonyl carbon atom of the substrate and a sulfur atom of C145 (Fig. 1). Another important quantity considered in this study is the interatomic distance between the hydrogen atom of the SH group of C145 and nitrogen atom of histidine H41 (Fig. 1). This distance characterizes whether the proton forms a covalent bond with the nitrogen or sulfur atom along a MD trajectory. Details of the molecular dynamic protocol can be found in ref [6]. Shortly, the QM/MM MD simulations were performed in NPT ensemble at $T = 300$ K and $p = 1$ atm. The integration time step was set to 1 ps and the trajectory length was 15 ps for each system. The CHARMM36 [2, 3] and CGenFF [20–22] force fields parameters were utilized for the enzyme and substrate molecules and TIP3P [5] parameters for water molecules. The QM subsystem was treated with the hybrid PBE0 [1] DFT functional with empirical dispersion correction D3 [4] with the 6-31G(d,p) or 6-31G basis sets. The QM/MM MD simulations were performed with TeraChem [17] and NAMD [16] programs and their interface proposed in ref [10]. The electron density analysis, namely the Laplacian of electron density, $\nabla^2\rho(r)$, calculations were performed in the Multiwfn program [7].

2. Results and Discussion

The first step of chemical reaction in the active site of cysteine protease is the nucleophilic attack accompanied with the proton transfer from the SH group of the catalytic cysteine residue (C145) to a histidine residue (H41) [13] (Fig. 1). The oxyanion hole formed by the NH fragments of the backbones of C145 and G143 in case of SARS-CoV-2 is supposed to be responsible for the substrate activation [13] (Fig. 1). Therefore, we start with the analysis of these four interatomic distances distributions to compare MD trajectories simulated at the QM(PBE0-

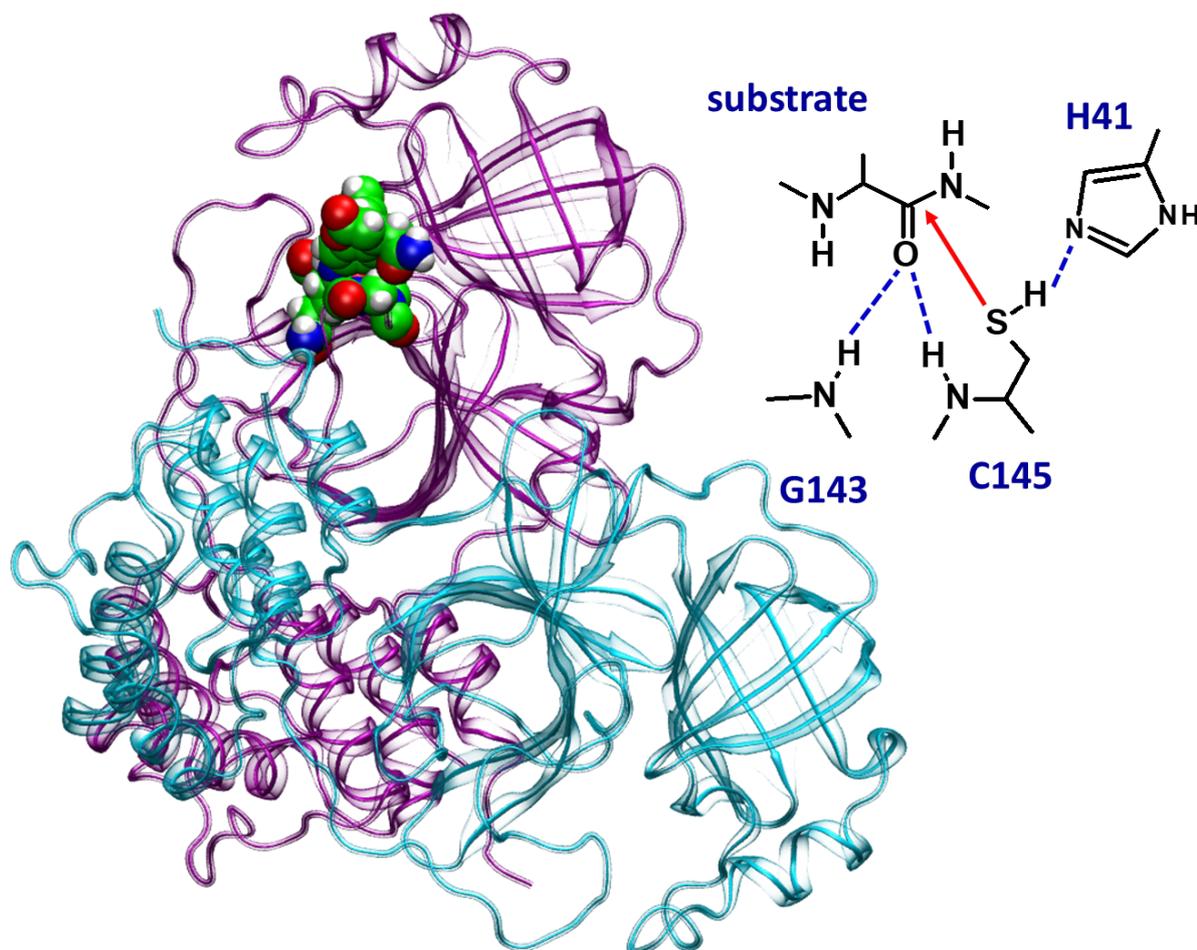


Figure 1. Dimeric main protease from SARS-CoV-2. Monomers are colored cyan and magenta. The active site is shown in colored van der Waals spheres. The 2D structure of the part of substrate and catalytically important parts of the active site are shown. Red arrow depicts the direction of the nucleophilic attack and hydrogen bonds are shown by blue dashed lines

D3/6-31G(d,p))/MM and QM(PBE0-D3/6-31G)/MM levels (Fig. 2, Tab. 1). The hydrogen bonds between the oxygen atom of the substrate and oxyanion hole are distributed similarly in both MD trajectories. The distribution is slightly more narrow and the mean values are slightly shifted to the smaller values in case of QM(PBE0-D3/6-31G)/MM. Similar features are observed for the distribution of the distances of the nucleophilic attack, $d(C\dots S(C145))$. Contrary, the $d(H(C145)\dots N(H41))$ distributions are notably different. In the QM(PBE0-D3/6-31G(d,p))/MM MD trajectory the covalent bond is always formed between a proton and a sulfur atom of C145. In the QM(PBE0-D3/6-31G)/MM MD trajectory, the proton is usually located closely to the nitrogen atom of H41 and the covalent bond is predominantly formed between them.

We analyzed the origin of considerable differences in the dynamic behaviour of model systems treated at different levels of theory. First, we selected the representative MD frame from the QM(PBE0-D3/6-31G(d,p))/MM corresponding to the reactive complex (Fig. 3). The Laplacian of electron density, $\nabla^2\rho(r)$, depicts areas of electron density concentration ($\nabla^2\rho(r) < 0$) and depletion ($\nabla^2\rho(r) > 0$), i.e. electrophilic and nucleophilic sites, respectively. If the electron

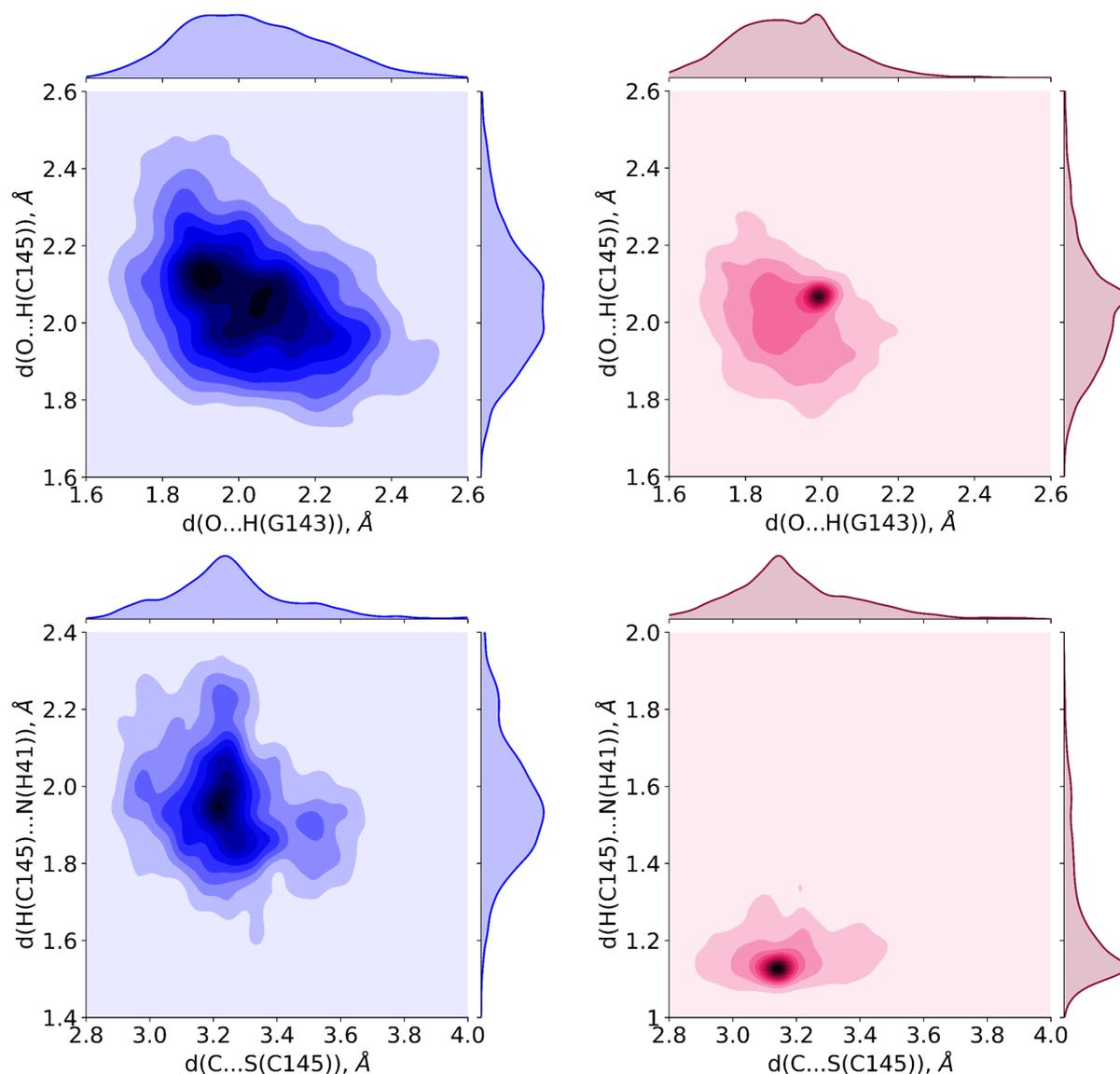


Figure 2. Distributions of key interatomic distances obtained along MD trajectories with the QM(PBE0-D3/6-31G(d,p))/MM (blue) and QM(PBE0-D3/6-31G)/MM (red) potentials

Table 1. The mean values and standard deviations for the key interatomic distances obtained at different levels of theory

Interatomic distance	QM(PBE0-D3/6-31G(d,p))/MM	QM(PBE0-D3/6-31G)/MM
$d(C...S(C145)), \text{ \AA}$	3.26 ± 0.20	3.19 ± 0.19
$d(O...H(G143)), \text{ \AA}$	2.04 ± 0.18	1.91 ± 0.14
$d(O...H(C145)), \text{ \AA}$	2.06 ± 0.16	2.03 ± 0.16
$d(H(C145)...N(H41)), \text{ \AA}$	1.96 ± 0.16	1.25 ± 0.17

density is calculated at the PBE0-D3/6-31G(d,p) level with polarization functions on all atoms, the electrophilic site on the carbonyl carbon atom of the substrate is observed (Fig. 3a). It is the region of positive values of the $\nabla^2\rho(r)$ in the direction of nucleophilic attack between the C and S atoms. If the Laplacian of electron density is recalculated at the same geometry

configuration without polarization functions, no substrate activation is observed (Fig. 3b). The nucleophilic attack is facilitated by the electron lone pair of the sulfur atom that is observed on both $\nabla^2\rho(r)$ maps. It is due to the fact that the substrate activation is a more refined electron density feature than the lone pair electron concentration, and polarization functions are required for its description. To further support this idea, we calculated electron density difference, $\Delta\rho$, maps in the same plane (Fig. 4). The difference is mostly pronounced around the carbonyl carbon atom. Electron density is considerably higher in the direction of the nucleophilic attack if calculated without polarization functions. This explains the absence of the electron density depletion region around C atom on the $\nabla^2\rho(r)$ maps calculated at the PBE0-D3/6-31G level.

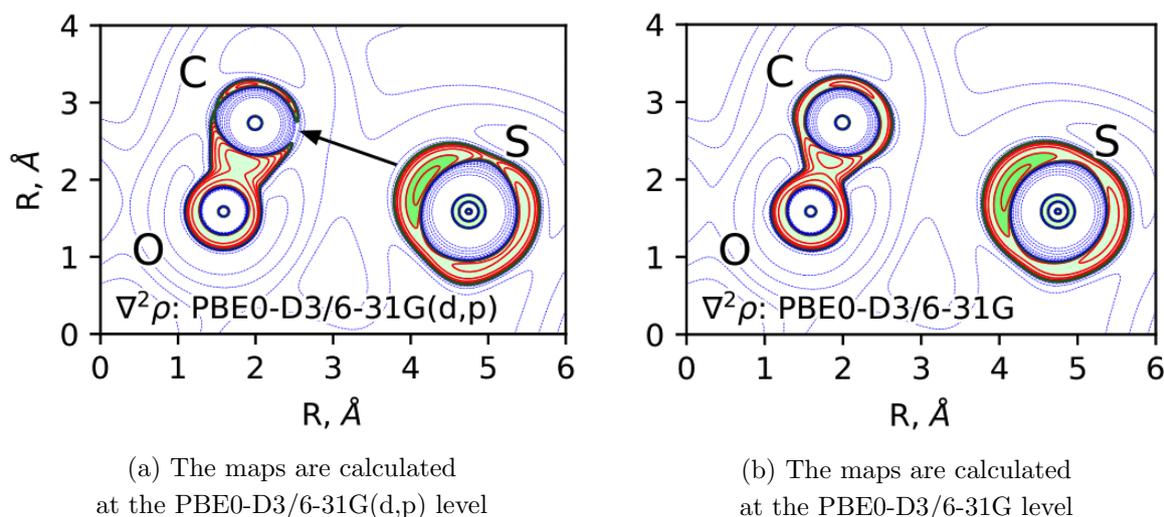


Figure 3. The Laplacian of electron density, $\nabla^2\rho(r)$, maps in the plane S(Cys145)...CO (the carbonyl group of substrate) of the representative reactive enzyme-substrate complex obtained at the QM(PBE0-D3/6-31G(d,p))/MM level. Contour lines are $\pm(2; 4; 8) \times 10^n$ au, $-2 \leq n \leq 1$, blue dashed contour lines indicate the electron density depletion areas ($\nabla^2\rho(r) > 0$) and red solid lines identify the electron density concentration ($\nabla^2\rho(r) < 0$), green solid line is $\nabla^2\rho(r) = 0$. The area with $\nabla^2\rho(r) < 0$ is colored in light green and the lone pair on the sulfur atom ($\nabla^2\rho(r) < -0.2$) is highlighted green. Black arrow indicates the direction of the nucleophilic attack

We also analysed $\nabla^2\rho(r)$ maps at several MD frames from the trajectory calculated at the QM(PBE0-D3/6-31G)/MM level. We selected a frame with the short distances of the nucleophilic attack and hydrogen bonds with the oxyanion hole and a frame where the proton from the SH group is transferred to the histidine residue (in this case the nucleophile is a negatively charged sulfur). In both cases Laplacian of electron density maps demonstrates no substrate activation.

Combining these computational experiments together, we can conclude that the applied computational strategy is useful in predicting substrate specificity of enzymes, if a proper level of theory is applied.

Conclusion

Novel DFT code implemented in the TeraChem [17] and optimized for the GPU discloses new opportunities for the QM(DFT)/MM molecular dynamics simulations. It allows fast and accurate simulations of interactions occurring between the substrate and enzyme in its active site.

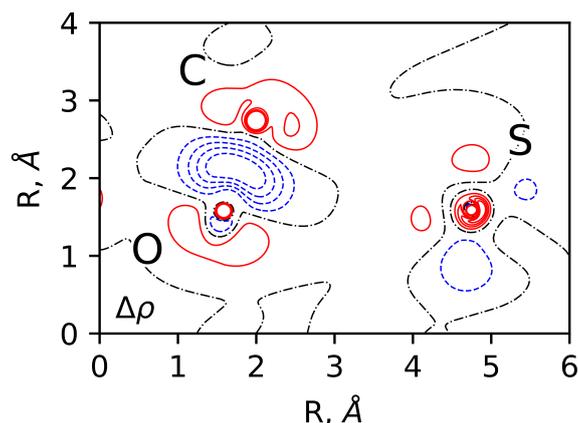


Figure 4. The electron density difference, $\Delta\rho$, maps between ρ calculated at the PBE0-D3/6-31G and PBE0-D3/6-31G(d,p) levels calculated in the same plane and at the same MD frame as on Fig. 3. Blue dashed isolines are -0.028 au, -0.021 au, -0.014 au and -0.007 au, red solid isolines are 0.007 au, 0.014 au, 0.021 au and 0.028 au, black dash-dotted isoline corresponds to $\Delta\rho = 0$

We take the main protease from the SARS-CoV-2 and its substrate to analyze the influence of the basis set on the quality of interaction description. Despite the 2-fold speedup of the calculations, elimination of polarization functions results in the wrong description of the substrate activation and cannot be recommended for such analysis. Utilization of the 6-31G(d,p) basis set enables to disclose equilibrium between the reactive and nonreactive species along the MD trajectory. We proposed the methodology based on the QM/MM MD simulations followed by the electron density analysis that allow one to determine substrate activation in the active site of a protease and we demonstrated the importance of utilization of a proper QM theory level for reliable evaluation of this quantity. This approach can be utilized to choose covalent inhibitors that will readily interact with the catalytic residue of the selected enzyme.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (project № 18-29-13006). We acknowledge the use of supercomputer resources of the Joint Supercomputer Center of the Russian Academy of Sciences and the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University [26].

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Adamo, C., Barone, V.: Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* 110(13), 6158 (1999), DOI: 10.1063/1.478522
2. Anisimov, V.M., Lamoureux, G., Vorobyov, I.V., et al.: Determination of electrostatic

- parameters for a polarizable force field based on the classical drude oscillator. *Journal of Chemical Theory and Computation* 1(1), 153–168 (2005), DOI: 10.1021/ct049930p
3. Best, R.B., Zhu, X., Shim, J., et al.: Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of Chemical Theory and Computation* 8(9), 3257–3273 (2012), DOI: 10.1021/ct300400x
 4. Grimme, S., Antony, J., Ehrlich, S., Krieg, H.: A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* 132(15), 154104 (2010), DOI: 10.1063/1.3382344
 5. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., et al.: Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79(2), 926–935 (1983), DOI: 10.1063/1.445869
 6. Khrenova, M.G., Tsirelson, V.G., Nemukhin, A.V.: Dynamical properties of enzymesubstrate complexes disclose substrate specificity of the SARS-CoV-2 main protease as characterized by the electron density descriptors. *Physical Chemistry Chemical Physics* 22(34), 19069–19079 (2020), DOI: 10.1039/D0CP03560B
 7. Lu, T., Chen, F.: Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry* 33(5), 580–592 (2012), DOI: 10.1002/jcc.22885
 8. Luehr, N., Ufimtsev, I.S., Martínez, T.J.: Dynamic precision for electron repulsion integral evaluation on graphical processing units (GPUs). *Journal of Chemical Theory and Computation* 7(4), 949–954 (2011), DOI: 10.1021/ct100701w
 9. Manathunga, M., Miao, Y., Mu, D., et al.: Parallel implementation of density functional theory methods in the quantum interaction computational kernel program. *Journal of Chemical Theory and Computation* 16(6), 4315–4326 (2020), DOI: 10.1021/acs.jctc.0c00290
 10. Melo, M.C.R., Bernardi, R.C., Rudack, T., et al.: NAMD goes quantum: An integrative suite for QM/MM simulations. *Nature methods* 15(5), 351 (2018), DOI: 10.1038/nmeth.4638
 11. Miao, Y., Merz, K.M.: Acceleration of electron repulsion integral evaluation on graphics processing units via use of recurrence relations. *Journal of Chemical Theory and Computation* 9(2), 965–976 (2013), DOI: 10.1021/ct300754n
 12. Miao, Y., Merz, K.M.: Acceleration of high angular momentum electron repulsion integrals and integral derivatives on graphics processing units. *Journal of Chemical Theory and Computation* 11(4), 1449–1462 (2015), DOI: 10.1021/ct500984t
 13. Otto, H.H.: Cysteine proteases and their inhibitors. *Chemical Reviews* 97(1), 133–172 (1997), DOI: 10.1021/cr950025u
 14. Perdew, J.P., Burke, K., Ernzerhof, M.: Generalized gradient approximation made simple. *Physical Review Letters* 77(18), 3865–3868 (1996), DOI: 10.1103/PhysRevLett.77.3865
 15. Perdew, J.P., Burke, K., Ernzerhof, M.: Generalized gradient approximation made simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Physical Review Letters* 78(7), 1396–1396 (1997), DOI: 10.1103/PhysRevLett.78.1396

16. Phillips, J.C., Braun, R., Wang, W., et al.: Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26(16), 1781–1802 (2005), DOI: 10.1002/jcc.20289
17. TeraChem v 1.9, PetaChem, LLC, www.petachem.com
18. Ufimtsev, I.S., Martínez, T.J.: Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. *Journal of Chemical Theory and Computation* 4(2), 222–231 (2008), DOI: 10.1021/ct700268q
19. Ufimtsev, I.S., Martinez, T.J.: Quantum chemistry on graphical processing units. 2. Direct self-consistent-field implementation. *Journal of Chemical Theory and Computation* 5(4), 1004–1015 (2009), DOI: 10.1021/ct800526s
20. Vanommeslaeghe, K., Hatcher, E., Acharya, C., et al.: CHARMM general force field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* 31(4), 671–690 (2010), DOI: 10.1002/jcc.21367
21. Vanommeslaeghe, K., MacKerell, A.D.: Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *Journal of Chemical Information and Modeling* 52(12), 3144–3154 (2012), DOI: 10.1021/ci300363c
22. Vanommeslaeghe, K., Raman, E.P., MacKerell, A.D.: Automation of the CHARMM general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling* 52(12), 3155–3168 (2012), DOI: 10.1021/ci3003649
23. Vasilevskaya, T., Khrenova, M.G., Nemukhin, A.V., Thiel, W.: Mechanism of proteolysis in matrix metalloproteinase-2 revealed by QM/MM modeling. *Journal of Computational Chemistry* 36(21), 1621–1630 (2015), DOI: 10.1002/jcc.23977
24. Vasilevskaya, T., Khrenova, M.G., Nemukhin, A.V., Thiel, W.: Methodological aspects of QM/MM calculations: A case study on matrix metalloproteinase-2. *Journal of Computational Chemistry* 37(19), 1801–1809 (2016), DOI: 10.1002/jcc.24395
25. Vasilevskaya, T., Khrenova, M.G., Nemukhin, A.V., Thiel, W.: Reaction mechanism of matrix metalloproteinases with a catalytically active zinc ion studied by the QM(DFTB)/MM simulations. *Mendeleev Communications* 26(3), 209–211 (2016), DOI: 10.1016/j.mencom.2016.05.010
26. Voevodin, V., Antonov, A.S., Nikitenko, D.A., et al.: Supercomputer Lomonosov-2: Large scale, deep monitoring and fine analytics for the user community. *Supercomputing Frontiers and Innovations* 6(2), 4–11 (2019), DOI: 10.14529/jsfi190201