Full Articles

Supercomputer simulation of the covalent inhibition of the main protease of SARS-CoV-2*

A. V. Nemukhin,[a,b]★ B. L. Grigorenko,[a,b] S. V. Lushchekina,[b] and S. D. Varfolomeev [a,b]

[a] *Department of Chemistry, Lomonosov Moscow State University, Build. 3, 1 Leninskie Gory, 119991 Moscow, Russian Federation. Fax: +7 (495) 932 8846. E-mail: anem@lcc.chem.msu.ru*
[b] *N. M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, 4 ul. Kosygina, 119334 Moscow, Russian Federation. Fax: +7 (499) 137 4101*

Molecular modeling tools were applied to design a potential covalent inhibitor of the main protease (M^{pro}) of the SARS-CoV-2 virus and to investigate its interaction with the enzyme. The compound includes a benzoisothiazolone (BZT) moiety of antimalarial drugs and a 5-fluoro-6-nitropyrimidine-2,4(1*H*,3*H*)-dione (FNP) moiety mimicking motifs of inhibitors of other cysteine proteases. The BZT moiety provides a fair binding of the ligand on the protein surface, whereas the warhead FNP is responsible for efficient nucleophilic aromatic substitution reaction with the catalytic cysteine residue in the M^{pro} active site, leading to a stable covalent adduct. According to supercomputer calculations of the reaction energy profile using the quantum mechanics/molecular mechanics method, the energy of the covalent adduct is 21 kcal mol^{-1} below the energy of the reactants, while the highest barrier along the reaction pathway is 9 kcal mol^{-1}. These estimates indicate that the reaction can proceed efficiently and can block the M^{pro} enzyme. The computed structures along the reaction path illustrate the nucleophilic aromatic substitution (S_NAr) mechanism in enzymes. The results of this study are important for the choice of potential drugs blocking the development of coronavirus infection.

Key words: cysteine proteases, main protease of SARS-CoV-2, covalent inhibition, S_NAr reaction, molecular modeling, quantum mechanics/molecular mechanics method.

The problems associated with the spread of COVID-19 gave a new impetus for studying enzymes and enzymatic reactions. Recently, unprecedented efforts have been made to study the structures of SARS-CoV-2 components and the mechanisms of interaction of these components with potential drugs. The main protease of SARS-CoV-2 M^{pro} (or $3CL^{pro}$), belonging to the class of cysteine proteases, is considered as a possible target for the therapy of COVID-19.[1] The goal of the studies is to select inhibitors

* Dedicated to Academician of the Russian Academy of Sciences O. M. Nefedov on the occasion of his 90th birthday.

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 11, pp. 2084—2089, November, 2021.
1066-5285/21/7011-2084

of this enzyme in order to prevent its functioning necessary for virus replication. The scientific literature of the last months presents the results of experimental and computational studies that propose both noncovalent and covalent M^{pro} inhibitors. In the former case, the inhibitor molecule is attached to the protein surface without the formation of protein—ligand chemical bonds and blocks the enzyme activity by either preventing the entry of natural substrates into the active site or through allosteric effects. In the case of covalent inhibitors of M^{pro}, the molecule of a prospective drug binds to the catalytic cysteine amino acid residue (see, *e.g.*, Refs 2—4), which makes impossible the enzyme functioning.

It is noteworthy that studying the irreversible inhibitors of cysteine proteases is an extensive area of enzymatic catalysis.[5] A new trend in this area is the idea to use reagents that contain aromatic groups capable of adding to the cysteine residue *via* nucleophilic aromatic substitution (S_NAr) reaction. A number of such compounds were synthesized and tested for inhibition of papain-like cysteine proteases.[6] In this study, we use the 5-fluoro-6-nitropyrimidine-2,4(1*H*,3*H*)-dione (FNP), which mimics the motifs of compounds proposed previously,[6] for the computer design of a covalent inhibitor of M^{pro} of the new SARS-CoV-2 virus. The second building block of the potential inhibitor contains a benzoisothiazolone (BZT) moiety, present in modern antimalarial drugs.[7] In the early period of fighting COVID-19, antimalarial drugs were considered as possible therapeutic means. Furthermore, benzoisothiazolones are well attached to the surface of protein macromolecules. In this study, covalently bound BZT and FNP moieties were combined into a single molecule by molecular modeling methods, and the compound thus obtained was considered as a ligand able to bind to the M^{pro} enzyme and then to react with molecular groups of the enzyme active site.

Currently, computer simulation is an important method for the search for most efficient drugs.[8,9] It is widely used to study the structure of biomolecular systems and reactions involving them. Studies on the modeling of cysteine proteases, including the main proteases of SARS and SARS-CoV-2, were reported in the literature.[10—16] According to the established practice, we investigated the interaction of the BZT—FNP compound with M^{pro} using quantum chemical methods, molecular mechanics, docking, molecular dynamics, and quantum mechanics/molecular mechanics (QM/MM) approaches. After the construction of all-atom 3D model systems, we calculated the reactant structures and energy profile of the formation of covalent adduct with the cysteine amino acid residue of the M^{pro} catalytic dyad. The calculation results show that the proposed compound can efficiently block M^{pro}. The calculated structures along the reaction path illustrate the nucleophilic aromatic substitution S_NAr mechanism, which is of interest not only for protein reactions.[17,18]

Calculation Procedure

The structural model of the compound designated in this study as BZT—FNP was initially constructed by molecular mechanic approach using the Discovery Studio Visualizer software (Accelrys Software Inc.) with subsequent quantum chemical optimization of geometric parameters by the DFT(PBE0)/6-31G* method. The all-atom 3D model of the enzyme in the apo form was designed on the grounds of the crystal structure of 6YB7 with 1.25 Å a resolution available from the protein data bank PDB.[19] The hydrogen atoms were added in such a way that the protonated amino acids with ionogenic groups corresponded to neutral pH. The Lys and Arg residues were protonated, while Glu and Asp occurred in the deprotonated state. The His residues were neutral, the hydrogen atom was attached to N_δ or N_ε, depending on the local environment, thus providing the maximum number of hydrogen bonds. The resulting model of M^{pro} in the apo form was fully surrounded by water molecules (the VMD program was used).[20] The molecular dynamic calculations carried out by the NAMD program[21] with the CHARMM36 force field used fixed protein coordinates for relaxation of the solvation shell. The Autodock4 program[22] was applied for primary estimation of BZT—FNP position on the M^{pro} surface using molecular docking.

The major part of calculations was done using the QM/MM approach and the NWChem software program.[23] The quantum subsystem included all atoms of the BZT—FNP compound, atoms of the Gly143—Ser144—Cys145 moiety and His41 side chain of the protein, and water molecules of the active site. The energies and forces in the QM part were calculated by the electron density functional theory using the PBE0 functional;[24] the MM subsystem was described by the AMBER99 force field. The minima in the potential energy surface were determined by the QM/MM unconstrained optimization of the whole system. The reaction coordinate at the nucleophilic attack step corresponded to the distance between the Cys145 sulfur atom and the C(5) carbon atom of the 5-fluoro-6-nitropyrimidine-2,4(1*H*,3*H*)-dione moiety and that at the product formation step corresponded to the C(5)—F distance. The structures of the **TS1** and **TS2** transition states were determined by series of QM/MM constrained minimizations at variable reaction coordinates. The transition state positions were verified by descending from the saddle point in the forward and back directions.

The atomic interactions in the enzyme active site were characterized using results of analysis of the Laplacian of electron density $\nabla^2\rho(\mathbf{r})$ (see Ref. 25) in the plane through the nucleophilic S atom and the C(5)—F group, representing the electrophilic site.

Results and Discussion

The molecular model of the compound consisting of two covalently bound BZT and FNP moieties, the geometry of which was determined by quantum chemistry methods, is shown in Fig. 1.

The use of the BZT moiety was justified by the following preliminary molecular docking calculations. One of the antimalarial compounds that functions as a covalent inhibitor of the malaria parasite enzyme (*Pf*IspD)[7] contains

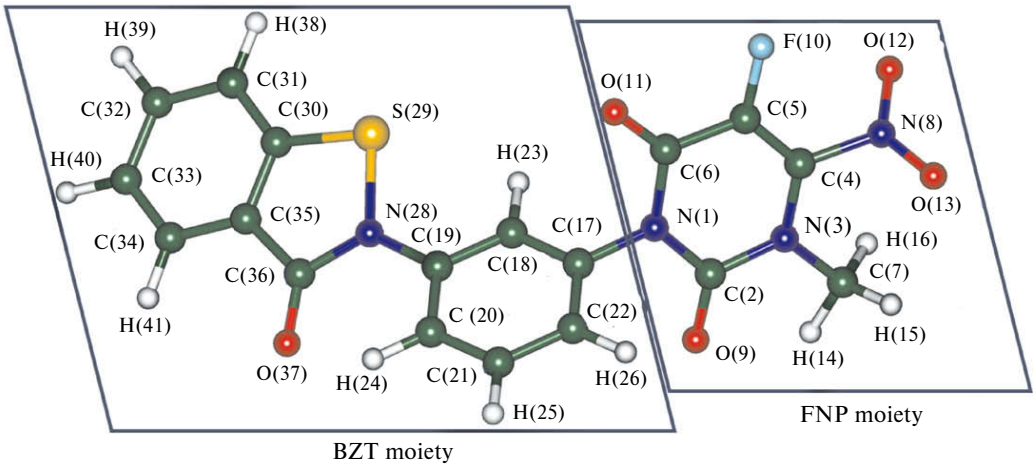

Fig. 1. Molecular model of the BZT—FNP compound.

a BZT group. Docking of this inhibitor into the M^{pro} catalytic site showed that, unlike binding to *Pf*IspD, no covalent interaction with the catalytic cysteine to form a disulfide bond takes place in this case. Instead, the BZT group binds to the M^{pro} pocket adjoining the active site.

In this study, the attention was focused to the FNP part of the molecule, which can actively interact with the Cys145 catalytic residue of the M^{pro} cysteine protease. After all-atom three-dimensional model was constructed for M^{pro} in the apo form by the molecular docking technique, the position of the ligand (*i.e.*, BZT—FNP molecule) on the protein surface near the enzyme active site was tentatively determined. An important factor is that the position of the BZT moiety almost did not change in comparison with the results of docking of the *Pf*IspD inhibitor, while the FNP moiety was located in the close vicinity of the catalytic residues. Then the QM/MM calculations were used to determine more accurately the atomic coordinates of the whole protein—ligand system in solvation shells. This system configuration will be referred to as reactants.

The left-hand part of Fig. 2 shows the general view of the protein, clearly demonstrating the Cys145—His41 catalytic dyad and the ligand molecule; the right-hand part shows the positions of the reactants in the active site of the enzyme. Configuration of the Cys145—His41 catalytic dyad is convenient for bringing the system to the reactive state *via* proton transfer from the cysteine residue to histidine: the distance between the cysteine sulfur atom and the N_ε nitrogen atom in histidine is 3.26 Å. The S(Cys145)—C(5)(FNP) distance is used in this study as the reaction coordinate for nucleophilic attack; at the starting point, this distance is 3.79 Å. An important role in the reaction is performed by the oxyanion hole, including the N—H groups of the Gly143—Cys145 peptide chain.

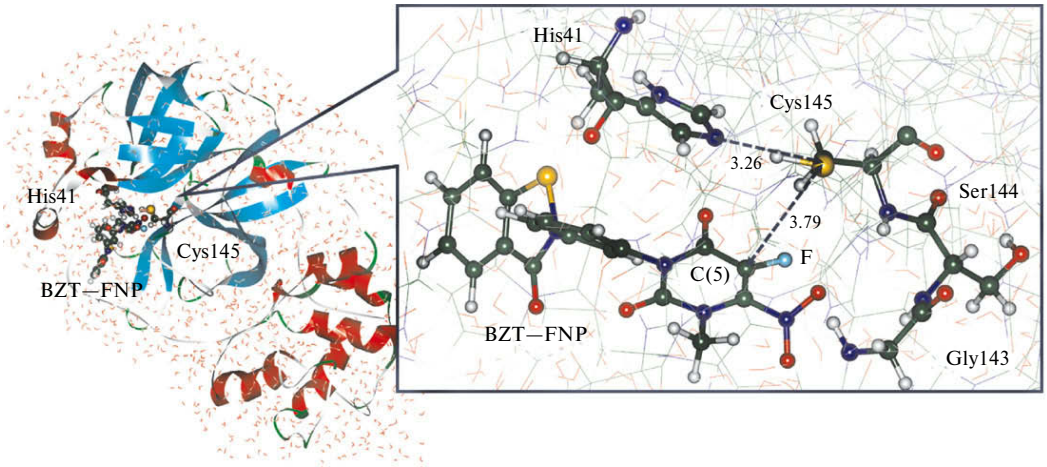

Fig. 2. On the left: general view of the enzyme with the Cys145—His41 catalytic dyad and the BZT—FNP ligand on the protein surface. On the right: structure of the active site before the reaction starts. Here and in Figs 3 and 5, the distances between heavy atoms are given in Angstroms.

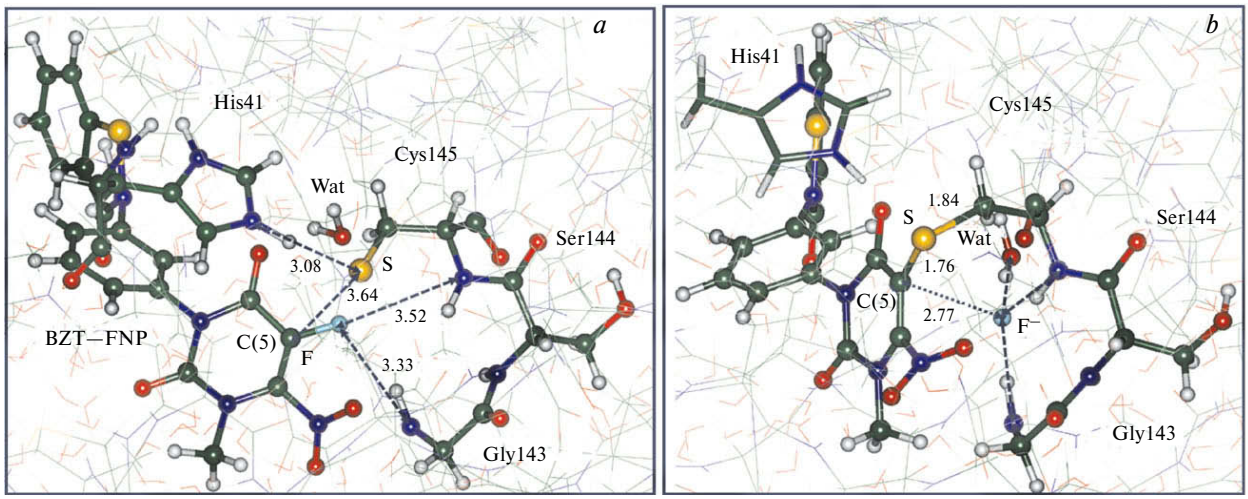

Fig. 3. Reactive state of the system (Cys^-—His^+) (*a*) and reaction products (*b*).

An actively debated issue is the state of the catalytic dyad and the relationship of configurations with a pair of neutral Cys—His amino acid residues and a Cys^-—His^+ ion pair in the cysteine protease catalysis: whether the ion pair is formed in the apo form of the enzyme or only after binding of the substrate.[26] It is also known form published computational data that the barrier for proton transfer between Cys and His does not exceed[13] 2 kcal mol^{-1}.

In our case, QM/MM calculations show that both systems (Cys—His and Cys^-—His^+) correspond to the minima in the potential energy surface, with the energy of the ion pair system being 4 kcal mol^{-1} higher. The Cys^-—His^+ configuration is, in some cases, referred to as the pre-reaction complex.[6] We will use the term "reactive state" of the system.

Figure 3 illustrates the arrangement of molecular groups in the enzyme active site in the reactive state of the reactants and in the products formed upon the covalent addition of the BZT—FNP molecule.

As can be seen from Fig. 3, *a*, the S—C(5) distance for the nucleophilic attack of cysteine has decreased to 3.64 Å in comparison with the configuration of the neutral dyad system (see Fig. 2). A further consecutive decrease in this distance accompanied by optimization of all geometric parameters of the system makes it possible to calculate the minimum energy profile along the reaction path (shown in Fig. 4) and to identify the stationary points in the potential energy surface, which correspond to minima and transition states (**TS1** and **TS2**).

The structure of the enzyme active site at the final point of the reaction (see Fig. 3, *b*) indicates that BZT—FNP is covalently bound to the side chain of Cys145: the S—C(5) distance (1.76 Å) implies a covalent bond. The F^- anion (leaving group) is fully separated from the starting compound and well incorporated into the oxyanion hole, with the distances to the nitrogen atoms of the peptide chain and the oxygen atom of the water molecule (Wat) being 2.5—2.6 Å. The energy of the products is 21 kcal mol^{-1} lower than the reactant energy. The energy barriers of the elementary steps do not exceed 9 kcal mol^{-1}. Thus, the computation results indicate that the enzyme can be irreversibly inhibited by BZT—FNP.

The results of this study also illustrate the mechanism of aromatic nucleophilic substitution S_NAr taking place in a protein environment. The mechanistic details of this reaction in solution, important for organic chemistry, are actively discussed in the literature.[15,16] In particular, researchers are interested in the nature of the Meisenheimer complex: whether it is an intermediate or a transition state. Fragments of intermediate structures (at the center) and the transition states separating this intermediate from the reactant (**TS1**) and product (**TS2**) sides are shown in Fig. 5. According to our results, the Meisenheimer complex corresponds to a fairly stable minimum in the energy surface, separated by a high barrier (more than 22 kcal mol^{-1}) from

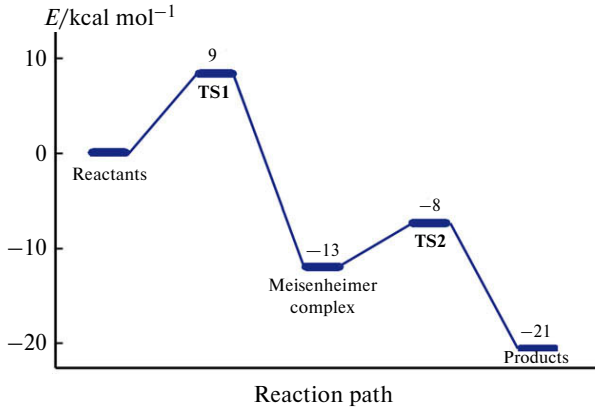

Fig. 4. QM/MM-calculated energy profile of the formation of covalent adduct in the reaction between BZT—FNP and M^{pro}.

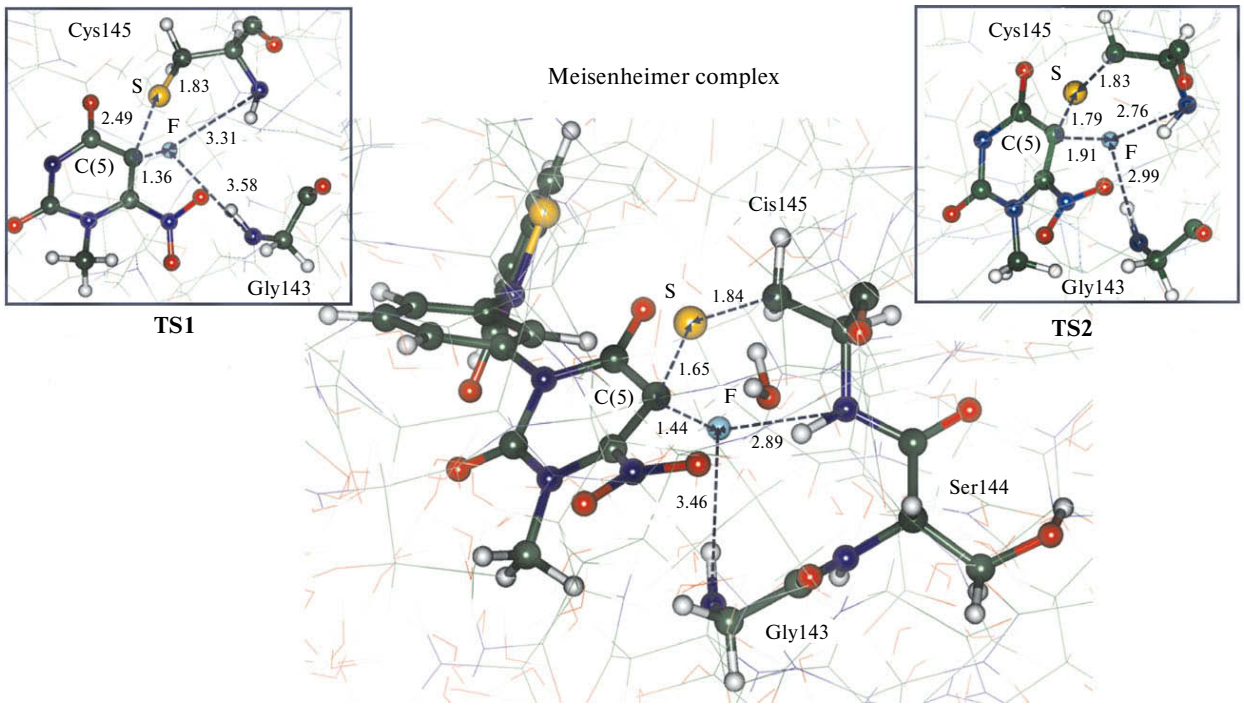

Fig. 5. Molecular structures corresponding to the intermediate of the nucleophilic aromatic substitution in the reaction of BZT—FNP with M^{pro} (in the middle) and to the **TS1** and **TS2** transition states separating the intermediate.

the reactants and by a lower (5 kcal mol^{-1}) from the products.

It is pertinent to analyze the topological properties of electron density in the region directly involved in the S_NAr nucleophilic substitution reaction in the aromatic system. Figure 6 shows the Laplacian of electron density maps in the plane through the C(5)—F—S atoms, calculated for the geometric configurations of reactants, **TS1**, and the

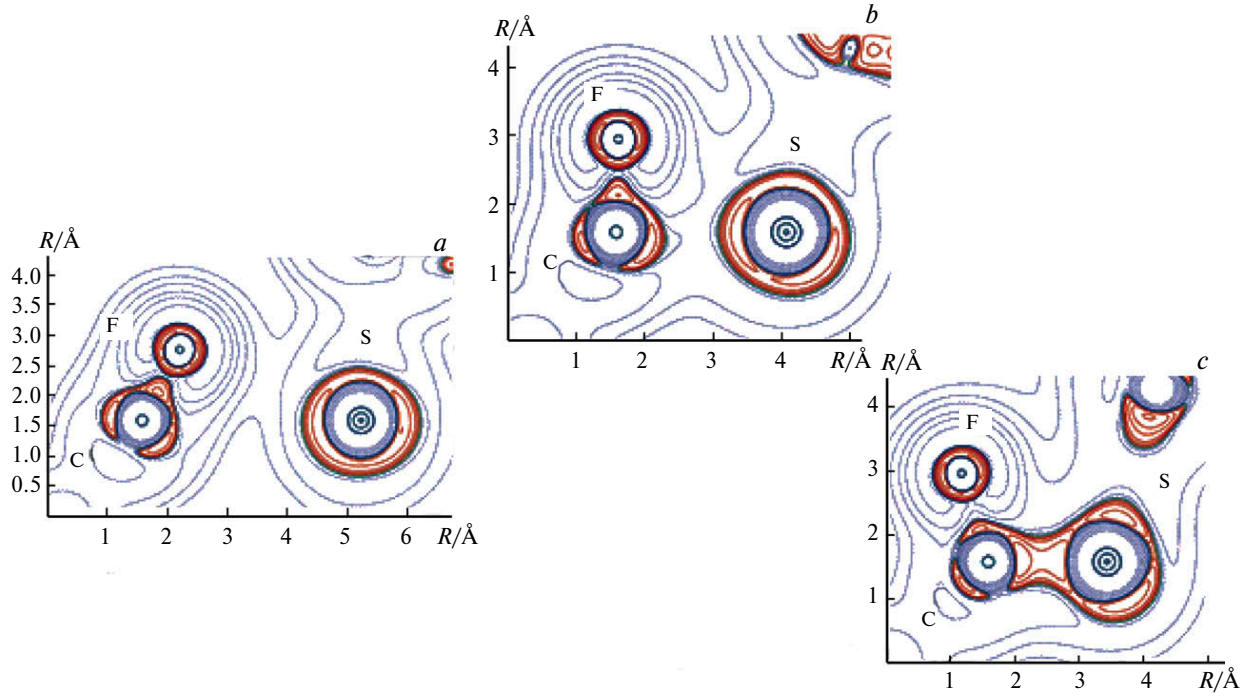

Fig. 6. Maps of the Laplacian of electron density $\nabla^2\rho(\mathbf{r})$ in the plane though the C(5)—F—S atoms for geometric configurations of reactants (*a*), **TS1** (*b*), and the intermediate Meisenheimer complex (*c*). The contour lines correspond to $\pm(2; 4; 8)\cdot 10^n$ au values.

intermediate. The bordered contour lines enclose the electron density concentration areas. It can be seen that the C(5)—F chemical bond with the leaving group (here the fluoride anion) is retained up to the **TS1** transition state without any signs of binding of the nucleophile to the ligand. In the Meisenheimer complex, reorganization has taken place: the C(5)—F bond has cleaved and the C(5)—S bond has formed.

The computer simulation results described in this study and reported in the literature[27,28] could be useful for the development of drugs against the SARS-CoV-2 virus.

Thus, the supercomputer molecular modeling of the reaction of the BZT—FNP compound proposed in this study with the SARS-CoV-2 main protease M^{pro} provides the conclusion that this reaction with the limiting energy barrier of not more than 9 kcal mol^{-1} is possible and that the resulting covalent adduct can irreversibly block functioning of the enzyme and, hence, the virus. The reaction follows the nucleophilic aromatic substitution mechanism S_NAr giving a Meisenheimer complex as a stable intermediate. Structure and electron density analysis for the region of the intermediate complex indicates that the bond to the leaving group in this complex has cleaved and the covalent bond between the reactants has formed. These results, which give the idea of the molecular mechanism of inhibition of the SARS-CoV-2 main protease, can serve as the basis for the design of new cysteine protease inhibitors, particularly M^{pro} inhibitors, an important component suppressing the virus.

The authors are grateful to M. G. Khrenova and I. V. Polyakov for help in the study and some valuable remarks. The calculations were carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University. The authors also acknowledge the use of supercomputer resources of the Joint Supercomputer Center of the Russian Academy of Sciences.

This study was financially supported by the Russian Science Foundation (Project No. 19-73-20032).

This paper does not contain descriptions of studies on animals or humans.

The authors declare no competing interests.

References

1. S. Ullrich, C. Nitsche, *Bioorg. Med. Chem. Lett.*, 2020, **30**, 127377.
2. L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, *Science*, 2020, **368**, 409.
3. D. Mondal, A. Warshel, *Biochemistry*, 2020, **59**, 4601.
4. Z. Jin, Y. Zhao, Y. Sun, B. Zhang, H. Wang, Y. Wu, Y. Zhu, C. Zhu, T. Hu, X. Du, Y. Duan, J. Yu, X. Yang, X. Yang, K. Yang, X. Liu, L. W. Guddat, G. Xiao, L. Zhang, H. Yang, Z. Rao, *Nat. Struct. Mol. Biol.*, 2020, **27**, 529.
5. J. C. Powers, J. L Asgian, Ö. D. Ekici, K. E. James, *Chem. Rev.*, 2002, **102**, 4639.
6. P. Klein, P. Johe, A. Wagner, S. Jung, J. Kühlborn, F. Barthels, S. Tenzer, U. Distler, W. Waigel, B. Engels, U. A. Hellmich, T. Opatz, T. Schirmeister, *Molecules*, 2020, **25**, 1451.
7. K. E. Price, C. M. Armstrong, L. S. Imlay, D. M. Hodge, C. Pidathala, N. J. Roberts, J. Park, M. Mikati, R. Sharma, A. S. Lawrenson, N. H. Tolia, N. G. Berry, P. M. O´Neill, A. R. Odom John, *Sci. Rep.*, 2016, **6**, 36777.
8. S. D. Varfolomeev, S. V. Lushchekina, A. V. Nemukhin, *Herald Russ. Acad. Sci.*, 2016, **86**, 185.
9. S. D. Varfolomeev, S. V. Lushchekina, A. V. Nemukhin, A. M. Kulakova, E. D. Kots, G. F. Makhaeva, E. Delacour, O. Lockridge, P. Masson, *Russ. Chem. Bull.*, 2016, **65**, 1592.
10. A. G. Taranto, P. Carvalho, M. A. Avery, *J. Mol. Graph. Model.*, 2008, **27**, 275.
11. K. Arafet, S. Ferrer, V. Moliner, *Biochemistry*, 2015, **54**, 3381.
12. K. Arafet, S. Ferrer, V. Moliner, *ACS Catal.*, 2017, **7**, 1207.
13. K. Arafet, K. Świderek, V. Moliner, *ACS Omega*, 2018, **3**, 18613.
14. J. R. A. Silva, L. Cianni, D. Araujo, P. H. J. Batista, D. de Vita, F. Rosini, A. Leitão, J. Lameira, C. A. Montanari, *J. Chem. Inf. Model.*, 2020, **60**, 1666.
15. K. Świderek, V. Moliner, *Chem. Sci.*, 2020, **11**, 10626.
16. A. V. Nemukhin, B. L. Grigorenko, I. V. Polyakov, S. V. Lushchekina, *Supercomput. Front. Innov.*, 2020, **7**, 25.
17. N. A. Senger, B. Bo, Q. Cheng, J. R. Keeffe, S. Gronert, W. Wu, *J. Org. Chem.*, 2012, **77**, 9535.
18. A. J. J. Lennox, *Angew. Chem., Int. Ed. Engl.*, 2018, **57**, 14686.
19. H. M. Berman, J. Westbrook, J. Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235.
20. W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph. Model.*, 1996, **14**, 33.
21. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781.
22. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *Comput. Chem.*, 2009, **16**, 2785.
23. M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, W. de Jong, *Comput. Phys. Commun.*, 2010, **181**, 1477.
24. C. Adamo, V. Barone, *J. Chem. Phys.* 1999, **110**, 6158.
25. M. T. Carroll, J. R. Cheeseman, R. Osman, H. Weinstein, *J. Phys. Chem.*, 1989, **93**, 5120.
26. X. Zhai, T. D. Meek, *Biochemistry*, 2018, **57**, 3176.
27. F. N. Novikov, V. S. Stroylov, I. V. Svitanko, V. E. Nebolsin, *Russ. Chem. Rev.*, 2020, **89**, 858.
28. V. S. Stroilov, I. V. Svitanko, *Mendeleev Commun.*, 2020, **30**, 419.

Received August 4, 2021;
in revised form September 19, 2021;
accepted September 20, 2021