Applications of density functional theory in COVID-19 drug modeling

Naike Ye [a],[*], **Zekai Yang** [b], **Yuchen Liu** [c]

[a] Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[b] Department of Physics, Imperial College London, London SW7 2AZ, UK
[c] College of Engineering, University of Michigan, Ann Arbor, MI 48109, USA

The rapidly evolving Coronavirus 2019 (COVID-19) pandemic has led to millions of deaths around the world, highlighting the pressing need to develop effective antiviral pharmaceuticals. Recent efforts with computer-aided rational drug discovery have allowed detailed examination of drug–macromolecule interactions primarily by molecular mechanics (MM) techniques. Less widely applied in COVID-19 drug modeling is density functional theory (DFT), a quantum mechanics (QM) method that enables electronic structure calculations and elucidations of reaction mechanisms. Here, we review recent advances in applying DFT in molecular modeling studies of COVID-19 pharmaceuticals. We start by providing an overview of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) drugs and targets, followed by a brief introduction to DFT. We then provide a discussion of different approaches by which DFT has been applied. Finally, we discuss essential factors to consider when incorporating DFT in future drug modeling research.

Keywords: COVID-19; SARS-CoV-2; Quantum mechanics; Density functional theory; Rational drug design; Molecular modeling; QM/MM; Molecular mechanics

Introduction

The ongoing COVID-19 global pandemic has caused millions of deaths around the globe. Although effective COVID-19 vaccines have been developed within unprecedented timeframes, it is only recently that two small-molecule antiviral drug candidates were shown clinically to be capable of inhibiting the disease-causing coronavirus, SARS-CoV-2.[1–3] The lack of effective antiviral pharmaceuticals highlights the pressing need to develop novel antiviral drugs to assist in combating the COVID-19 pandemic, to prevent more severe COVID-19 resulting in hospitalization and death.[3]

However, developing a lead molecule and an effective drug is challenging, even with known biological targets.[4] During the early stage of drug development, rational drug design is an essential step for preclinical evaluations of drug properties.[4]

Fortunately, because of advances in both structural biology and computer-aided methodologies, drug discovery has evolved from the empirical-design era to the current rational-design era, making the process more efficient.[5] Once the structures of biological targets are determined, modern computing technology enables the applications of chemo/bioinformatics and molecular modeling techniques to the study of drugs and medicines, which all significantly accelerated drug discovery efforts toward combating COVID-19.[5]

Molecular modeling of pharmaceuticals typically requires computations of large, complex, and often solvated biological systems, such that techniques with high computational costs (time and computing resources) are rendered impractical. Therefore, force field-based MM is an attractive approach given its low costs and relatively high calculation speed.[6] MM is based on

* Corresponding author.Ye, N. (nye@mit.edu)

Newtonian-level calculations, such that its algorithm disregards changes at the electronic level and treats nuclei and electrons surrounding nuclei as perfect spheres. Atoms are treated as balls, and bonds as springs. Nonbonding atoms interact through van der Waals and electrostatic interactions.[7] Therefore, biological systems comprising hundreds or even thousands of atoms are computationally suitable for MM treatments to describe receptor–ligand interactions. A major drawback of MM is its lack of chemical accuracy and, thus, it is not amenable for calculations of electronic properties. When chemical bonds break and new bonds form, MM can no longer afford energetic or geometric predictions because these changes occur at an electronic (quantum) level. Thus, QM-based calculations are favored for such detailed studies. In the context of the molecular modeling of drugs, QM studies allow close examination of the enzyme catalytic processes at the active site, and the mechanisms by which the inhibitor inhibits the catalytic activity of an enzyme. QM methods are also capable of providing energetic landscapes when such reactions happen.

DFT is the most popular and widely applied quantum theory to calculate the electronic structures of atoms and molecules. In drug modeling research, DFT is applied to study detailed electronic properties of isolated drug molecules and drug delivery systems, and is also applied in synergy with MM-based methodologies to study drug–receptor interactions. DFT provides the chemical accuracy otherwise unattainable by MM, making it particularly useful for describing reaction mechanisms when drug molecules act on enzyme active sites. Therefore, DFT serves well to both describe drug properties and their inhibitory actions on drug targets. To the best of our knowledge, no articles have been published that systematically review applications of DFT in COVID-19 drug modeling research. Therefore, here we introduce SARS-CoV-2 drug targets and drug compounds, fundamentals of DFT, followed by a review of literature grouped by applications of DFT in these studies. Finally, we discuss insights from COVID-19 drug modeling research so far and suggest essential factors to consider when setting up and incorporating DFT in future computational studies.

COVID-19 drug discovery: Biological targets and their inhibitors

The key initial step in drug discovery is identifying target biomolecules that drug molecules bind to, and inhibit. Examination of the SARS-CoV-2 virology and its lifecycle reveals a range of proteins promising as drug targets, which are widely used for such purposes.[1] However, not every protein encoded by the SARS-CoV-2 genome is suitable for drug design; for instance, accessory proteins, which are involved in viral replication, assembly, and virus–host interactions, are poor drug targets because of their high mutation rates.[8] Among all SARS-CoV-2 drug targets, six targets were found to be most popular among COVID-19 drug discovery research (Fig. 1).

Among the six targets, the main protease (M^{pro}) of SARS-CoV-2, also known as the chymotrypsin-like cysteine protease ($3CL^{pro}$), is the most promising target, and is targeted the most in drug modeling studies. M^{pro} [Protein Data Bank (PDB) ID: 6LU7] is a crucial protease of coronaviruses and has the essential role of mediating viral replication and transcription, making it one of the most attractive drug targets for SARS-CoV-2.[9–10] Coronaviruses are subject to extensive mutagenesis; however, key proteins, such as M^{pro}, are highly conserved, because mutations in key proteins are often lethal to the virus. Thus, drugs that target the highly conserved M^{pro} are often effective in preventing viral proliferation processes and display broad-spectrum antiviral activity.[11] The M^{pro} active site contains a Cys − His catalytic dyad, and drugs that inhibit M^{pro} activity either form an irreversible covalent linkage with this dyad, or reversibly block substrate entry to the active site.

The RNA-dependent RNA polymerase (RdRp) is another promising macromolecular target: remdesivir, as the first US Food and Drug Administration (FDA)-approved antiviral drug, targets this enzyme. RdRp (PDB ID: 6XEZ) catalyzes the synthesis of viral RNA and, thus, also has a central role in the viral transcription cycle.[12] Drugs that target RdRp generally achieve inhibition via different mechanisms; remdesivir is a nucleotide analog and, when RdRp effectively incorporates its active remdesivir-triphosphate form, the RNA synthesis process is stalled.[13] DFT has been applied the most in COVID-19 drug modeling studies that target these two enzymes: M^{pro} (e.g.,[14–17]) and RdRp (e.g.,[18–19]).

Small molecules have also shown promise as drug lead compounds for COVID-19 by acting on these targets (Fig. 1). The ligands (drugs) in these studies are diverse: natural products (e.g., embelin, hypericin, and naringin), repurposed antiviral pharmaceuticals (e.g., remdesivir, lopinavir, galidesivir, and amodiaquine), metal complexes of known ligands (e.g., ATA metal complexes and NHE–metal complexes), and newly synthesized compounds (e.g., M1BZP). Fig. 1 provides the molecular structures of these compounds and references to corresponding literature. In addition to small-molecule antiviral drugs, properties of drug delivery systems, such as C60 fullerene[20] or metallo-fullerenes,[21] are also evaluated by DFT to assess their abilities as carriers of pharmaceutical compounds.

DFT overview

Here, we provide an overview of DFT to introduce fundamental concepts to nonexpert readers, especially those that are essential to understand in the context of computational drug modeling. We recommend that readers refer to tutorials or books, such as[38], for a more comprehensive guide to DFT.

Since Hohenberg and Kohn published their theorems of DFT during the 1960s, DFT has become a common computational method for electronic structure calculations. Instead of solving the many-electron Schrodinger Equation (wave function theory), which is a 3N-spatial-coordinate problem for N electrons, DFT shifts the focus to solving electron density $\rho(r)$ with only three spatial coordinates.[34–35] Under the Born–Oppenheimer approximation, electronic energy and electronic properties of a system could be uniquely defined by the ground-state electron density $\rho(r)$.[35–37] Therefore, DFT has become one of the most popular methods over traditional wave function theory in computational chemistry for its low complexity and high accuracy. Here, we discuss Kohn–Sham (KS) DFT, the most common DFT formalism that approximates the kinetic energy in terms of fictitious
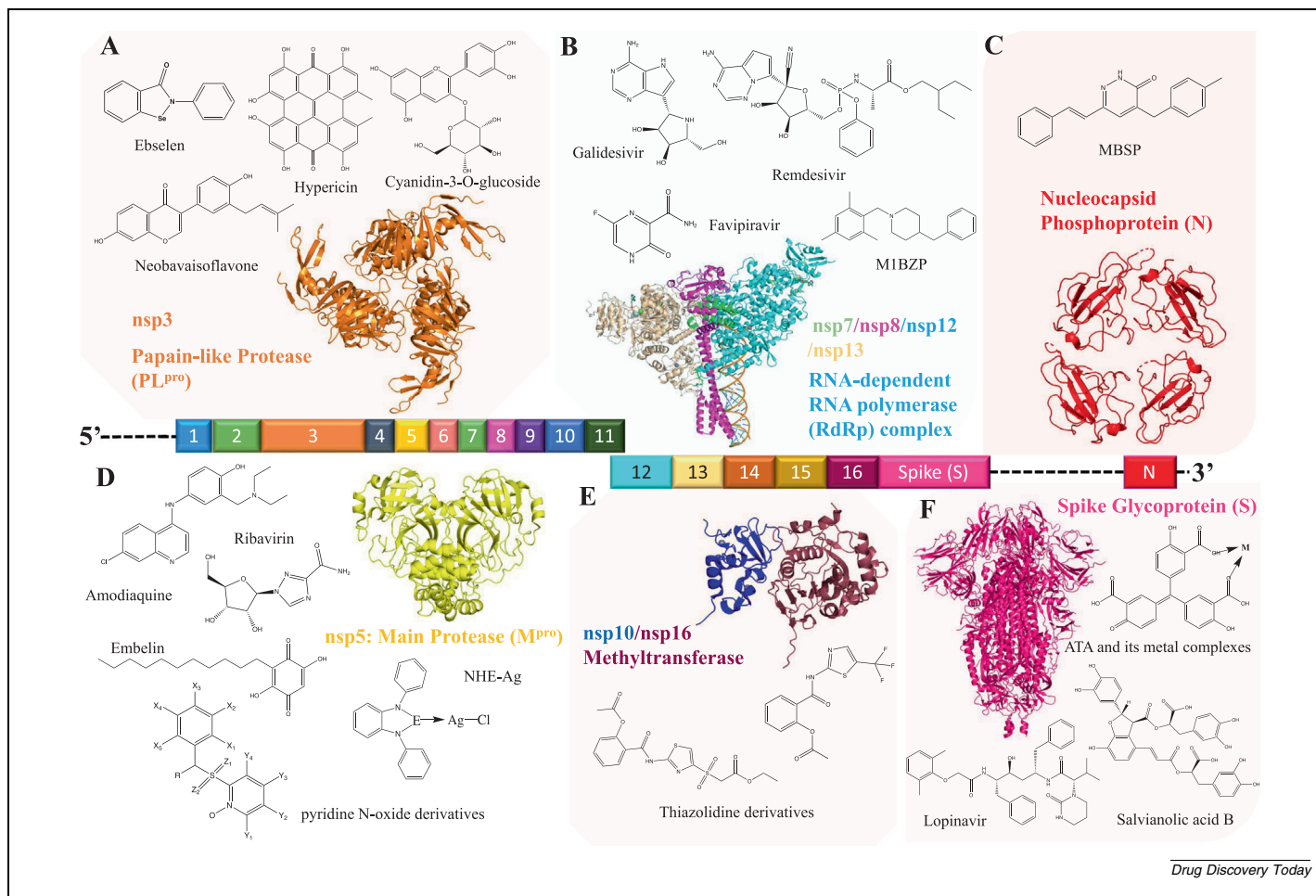
FIGURE 1

Map of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) protein targets, their positions on the viral RNA genome, and molecular structures of drug molecules targeting these proteins. The string represents the open reading frame (ORF) genome, with numbers corresponding to nonstructural protein (nsp) gene pieces. **(a)** The papain-like protease (PL[pro]) [Protein Data Bank (PDB) ID: 6W9C] (J. Osipiuk et al., unpublished data, 2021) ebselen,[17] hypericin and cyanidin-3-O-glucoside,[22] neobavaisoflavone,[23] **(b)** The RNA-dependent RNA polymerase complex, with its co-factor proteins (PDB ID: 6XEZ[12]. Galidesevir,[24] remdesivir and its nucleotide analogs,[18] 4-benzyl-1-(2,4,6-trimethyl-benzyl)-piperidine (M1BZP),[19] favipiravir.[25] **(c)** Nucleocapsid phosphoprotein (N) (PDB ID: 6WKP) (C. Chang et al., unpublished data, 2021). (E)-4-(4-methylbenzyl)-6-styrylpyridazin-3(2H)-one (MBSP).[26] **(d)** Main protease (M[pro]) (PDB ID: 6LU7[27]. Amodiaquine and ribavirin,[14] embelin,[16] NHE–Ag (E = C, Si, and Ge) complex,[28] pyridine N-oxide derivatives.[29] **(e)** Methyltransferase (PDB ID: 6W4H[30]. Thiazolidine derivatives.[15] **(f)** Spike glycoprotein (S) (PDB ID: 6VXX[31]. Lopinavir and aurintricarboxylic acid (ATA) with its metal complexes,[32] salvianolic acid B.[33]

single-particles orbitals within a non-interacting reference system.[34,36,38] For a detailed description of KS-DFT, we refer readers to the book by Koch and Haolthausen.[38]

Exchange-correlation functionals

Here, we provide a brief overview of DFT functionals; for further discussion, we refer readers to[36]. For a given system, the electronic energy could be regarded as the sum of four parts: non-interacting electronic kinetic energy; nuclear–electron attraction; classical electron–electron repulsion; and the exchange–correlation (XC) energy, E_{XC}.[34,36] KS-DFT is formally exact, but the form of XC functional $E_{XC}[\rho(\mathbf{r})]$ is not known exactly, and the search for accurate XC functionals remains one of the greatest challenges in DFT. Generally, the XC functional can be treated as a summation of the exchange term $E_X[\rho(\mathbf{r})]$ and the correlation term $E_C[\rho(\mathbf{r})]$.[35–36]

Types of functionals

The early generation of XC functionals uses the local density approximation (LDA), which assumes the XC energy for any given point in space depends only on the electron density at that point.[34–35] Generalized gradient approximation (GGA) reduces limitations of LDA functionals by considering that the real particle system is spatially inhomogeneous by adding the gradient of the density $\nabla\rho$ into the XC functional,[35] such as the Perdew–Burke–Ernzerhof (PBE) exchange-correlational functional (Table 1). Meta-GGA functionals take two more additional variables from GGA into account: the kinetic energy density τ_σ, and the Laplacian of the density $\nabla^2\rho_\sigma$.[35–37] Typically, Meta-GGA functionals perform better on predicting electronic properties, such as atomization energies.[35]

Functionals that consider Hartree–Fock (HF) exchange are hybrid functionals. Introducing HF exchange to GGA function-

INFORMATICS (ORANGE)

TABLE 1

Summary of the DFT XC functionals in recent (2020–2021) COVID-19 molecular modeling studies.[a,b]

Functional	Type	χ[c]	Exchange functional	Correlation functional	DC class	Description	Selected basis sets in study	Refs
PBE	LDA					Not used in studies reviewed		
	GGA	0	PBE	PBE	None	Used to study inactivation mechanism Good at predicting solid-state properties Inaccurate in determining atomization energies and reaction barriers for molecular reactions.	Not mentioned	35,36,41
	Meta-GGA					Not used in studies reviewed		
B3LYP	Hybrid GGA	20	B88	Lee-Yang-Parr (LYP)	None	Widely used for many different types of ligand because of its relatively high accuracy and low cost; B3LYP is more suitable for main-group elements than for transition metals	6-31G, 6-31G* 6-311G(d,p) 6-311G* LANL2DZ def2-SVP	14,15,18,20,25,32,33,37,42
B3PW91		20	B88	Perdew-Wang91	None	B3PW91 is similar to B3LYP; performs better for systems with uniform density	6-311++G**	36,43
PBE0		25	PBE	PBE	None	PBE0 is PBE functional with a 25% HF exchange functional; used to study inactivation mechanisms through covalent binding	6-31G* 6-311++G**	43,44
PBE0-D3		25	PBE	PBE	III	PBE0-D3 is PBE0 with semiclassical correction on dispersion; describes thermochemistry better with low numerical complexity; dispersion correction does not affect density expression of system; used to study enzyme–substrate complex dynamical properties	6-31G**	39,45
M06-2X	Hybrid meta-GGA	54	M06-2X	M06-2X	II	Performs well in determining main-group thermochemical properties, noncovalent interactions, and barrier heights; however, not suitable for transition metal chemistry	6-311(d,p)	29,37
ωB97X-D	RSH GGA	22, 100[d]	Becke97	Becke97	III	Performs well for bonded interactions and barrier heights	def2-TZVP 6-31+G(d,p)	21,36,43

[a] The types of XC functional are sorted by their levels of accuracy and computational complexity from low (top rows) to high (bottom rows). Some information is from[35–37,39,46].

[b] Abbreviations: DC Class, Dispersion-Correction Class (defined in[39]GGA, generalized gradient approximation; GH, global hybrid; RSH, range-separation hybrid.

[c] χ: magnitude of the percentage of HF exchange functional in the total exchange functional part.

[d] 20 for short range exchange functional and 100 for long range exchange functional.

als significantly enhances their ability to describe self-interactions.[36] One of the most common functionals, B3LYP, is a hybrid GGA, and is used extensively in COVID-19 molecular modeling studies because of its proper balance between accuracy and cost. Global hybrid (GH) functionals treat the HF exchange as a global fraction, replacing a fraction of the exchange functional.[36] To accurately describe self-interaction further, more advanced functionals could be constructed through range separation.[36] Given its high accuracy, many COVID-19 modeling studies use ωB97X-D, an advanced range-separation hybrid (RSH) GGA functional created over the past decade (Table 1).

Dispersion correction

Dispersion interaction being described incorrectly is a major problem for many functionals. For example, B3LYP describes the long-range interaction as purely repulsive, which ignores the dispersion interaction.[39] To reduce such error, dispersion-correction terms are added in density functionals. Four classes of dispersion correction are described by Grimme.[39]

When selecting appropriate functionals for DFT calculations, several factors need to be considered: desired chemical accuracy, computing resources available (computer clusters and time), and the given system (ligand and target). Table 1 details the accuracy and complexity of functionals.

Basis sets

In DFT, basis sets are sets of basis functions that are used to represent electronic wavefunctions. Basis functions are used to construct molecular orbitals (MOs), in which a large number of atomic orbitals (AOs) are expanded in linear combination to achieve a more flexible representation of MOs.[40] In practical density functionals, basis sets often make use of Gaussian-type

functions, basis functions that resemble the true hydrogen AO functions and allow for easier computations. For a detailed description of different types of basis sets, refer to.[40] Table 1 lists basis sets that are selected in COVID-19 drug modeling studies for each corresponding functionals. The selection of basis sets must consider the balance between accuracy and costs. To more adequately describe wavefunctions, two types of addition could be incorporated in the basis sets: polarization functions that describe atomic charge polarizations in AOs, and diffuse functions that more accurately describe the portion of AOs distant from atomic nuclei.[40]

DFT as a stand-alone technique for electronic property calculations

In the COVID-19 molecular modeling research reviewed here, DFT is extensively applied on isolated drug molecules. Molecular properties, such as reactivities, could be determined from DFT calculations, because they have the distinct advantage of providing electronic insight. Without studying protein–ligand interactions, electronic structures provided by DFT calculations could inform how drug compounds preferentially bind to their targets. The relationship between drug properties and resulting biological processes is dependent on the electronic properties of the target compounds.[25] Electronic properties of drug molecules determined by DFT in COVID-19 research are discussed in Table 2; studies that applied DFT for examining these properties of drug compounds are also cited.

Geometry optimization is at the core of small-molecule DFT calculations. Optimized molecular geometry informs the lowest-energy conformation of the drug species, which is especially useful for complex ligand systems. Bui *et al.* examined tetrylene metal complexes as potential COVID-19 drug molecules. Their optimized molecular geometry allowed detailed structural characterizations of ligands binding to the metal centers and the resulting complexes.[28] Similarly, DFT allowed analyses of geometric properties of first-row transition metal-doped metallofullerenes, which are potential COVID-19 drug carriers, and their complexes when drug molecules bind.[21]

Another important property to study by DFT is frontier molecular orbital (FMO) calculations. FMO are the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), which govern the electron donating (HOMO) and accepting (LUMO) capabilities of a molecule. The HOMO–LUMO energy gap (E_{HOMO}-E_{LUMO}) is inversely correlated with the stabilization interactions of the compounds in protein–ligand interactions: a large HOMO–LUMO gap suggests high stability, large stabilizing interactions with the receptor, and poor reactivity of a molecule.[20] By this principle, FMOs are extensively calculated for drug stability and reactivity evaluations (see references in Table 2).

Molecular electrostatic potentials (MEPs) are another useful electronic property in COVID-19 drug modeling.[47] MEP allows for visualizations of molecular charge distributions and charge-related properties, such as reactive sites of the molecule prone to nucleophilic (most positive region) or electrophilic (most negative region) attacks.[14] This informs molecular potentials for electrostatic interactions, such as van der Waals or hydrogen bonding interactions, also suggesting sites of potential covalent interactions upon binding to macromolecules.[20] In an evaluation of *N*-heterocycles as candidates of COVID-19 antiviral drugs,[14] MEP maps identified electrophilic centers of the molecules that are electron-rich and prone to electron density donations. This informs binding modes and reaction potentials when these drugs bind to their targets.

DFT used in synergy with other modeling methodologies

MM techniques: molecular docking and molecular dynamics
Given high computational costs, direct application of DFT in complex biomolecular systems is limited. However, effective drug modeling relies on accurate depiction of drug–receptor interactions; this is when MM-based techniques have a role in COVID-19 drug modeling studies along with DFT.

Molecular docking treats the macromolecule (protein target) as rigid and allows the ligand (drug molecule) to be flexible within the defined grid, which is usually the active site of an

TABLE 2

Molecular electronic properties calculated by DFT.

DFT functions and calculated properties	Descriptions	Refs
Geometry optimization	Generates the lowest energy structures of molecules in a given system from an arbitrary starting state. Together with energy calculations, optimized geometry provides crucial structural insights into the relaxed structure of drug molecules and when they bind to their receptors	14,20–21,28–29,32,42,48
FMO	HOMO and LUMO together form FMOs. FMO governs reactivity of molecules and can be used to calculated multiple chemical reactivity descriptors, such as electronegativity, electrophilicity, and global hardness and softness	14–15,18,20–21,25,28, 32–33,42,48–49
MEPs	EP correlates with dipole moment, electronegativity, and partial charges distributed in a molecule. Therefore, MEP provides a visual method to understand the relative polarity of a molecule	14,20,32,42
Mulliken atomic charges/natural population analysis	Both Mulliken atomic charge and natural population analysis are used to determine the partial charges of a molecule. Natural population analysis is an improved method with increased numerical stability and precision in calculation	14,20,26,28,48
Spectral properties	DFT is applied to assist in interpreting experimental spectral data and provides competent calculations of excited state properties based on a time-dependent DFT (TD-DFT) method	21,25,32

enzyme. To compensate for the unrealistic nature of this rigid system, motions of this protein–ligand complex are then simulated by molecular dynamics (MD) calculations, described using classical Newtonian mechanics. Results of docking and MD studies provide pictures of binding modes between small–molecule drugs and the targets, together with potential noncovalent interactions. These resulting binding poses can be further analyzed in detail by DFT calculations: a common approach found in COVID-19 drug modeling research so far is to analyze using DFT the drug molecules shown to have the highest binding affinities with the target revealed by molecular docking or MD simulations. For instance, Rasool et al. studied the inhibitory potentials of 19 thiazolide derivatives on the Mpro and methyltransferase.[15] DFT was executed for reactivity evaluations of compounds with the highest binding affinities. Evaluations were based on HOMO–LUMO gaps, which identified promising drugs with high reactivities as well as favored binding modes to the targets.[15]

More importantly, MM calculation results can be interpreted by electronic structures and properties of the drug compounds calculated by DFT. For example, MEP provides insights relating to charge distributions along the molecular surface, which informs regions that are prone to electrophilic or nucleophilic attacks during enzymatic reactions. Noncovalent bonding, such as van der Waals interactions and hydrogen bonds, could also be predicted by docking results combined with DFT-calculated electronic properties of drug molecules. Similarly, when comparing charge transfer potentials of drugs, FMO informs likely electron donors and acceptors. In fact, this DFT docking–MD approach is common in COVID-19 drug modeling studies.[14–15]

Such a combination of DFT and MM-based techniques allows for a detailed analysis of both molecular properties of the drug molecule and its interaction with target biomolecules. However, the separate application of DFT and MM-based techniques complicates the computational modeling procedure and reduces chemical accuracy. The more QM is applied, the more accurate depiction of chemical processes it provides at a molecular level, concerning changes of electronic structures. Here, we introduce the combined QM/MM method and its applications in COVID-19 drug modeling research.

QM/MM hybrid approach

QM/MM is an integrated, hybrid mean of modeling chemical reactions in large biomolecular systems.[50] The QM/MM approach uses a QM method, usually DFT, for the chemically active enzymatic catalysis region (where bonds break and new bonds form), and combines it with an MM treatment for the rest of the macromolecule (Fig. 2a).[51] As a result, the complex macromolecular portion (comprising hundreds of atoms) can be effectively simulated by force field-based MM methods; the detailed description of chemical reactions or other electronic-level chemical processes, such as charge transfer at enzymatic active sites, would be calculated by QM methods.[50,51] This hybrid approach ensures the balance between affordable computational costs and necessary levels of chemical accuracy. A distinct advantage of QM/MM is that it directly provides information on both noncovalent (e.g., electrostatic interactions) and covalent interactions that occur between the ligand and the macromolecule.[50]

In COVID-19 drug modeling research, QM/MM is less common than the DFT-MM approach, but is largely applied for the elucidation of reaction mechanisms of two types: catalysis mechanisms of SARS-CoV-2 enzymes[52–53] and inhibitory mechanisms of potential antiviral drugs on these enzymes.[54–57]

Arafet et al. used a hybrid M06-2X/6-31+G(d,p):AM1/MM method to investigate inhibition mechanisms of a known Mpro inhibitor, N3, and created energy profiles for covalent complex formation with the Mpro.[55] Two other designed covalent inhibitors were also studied by applying DFT/MM approach to evaluate their inhibitory potentials and mechanisms. QM/MM approach also enables researchers to gain novel insights into the catalytic mechanism of SARS-CoV-2 enzymes, which are crucial for mechanism-based rational drug design: QM/MM at a DLPNO-C
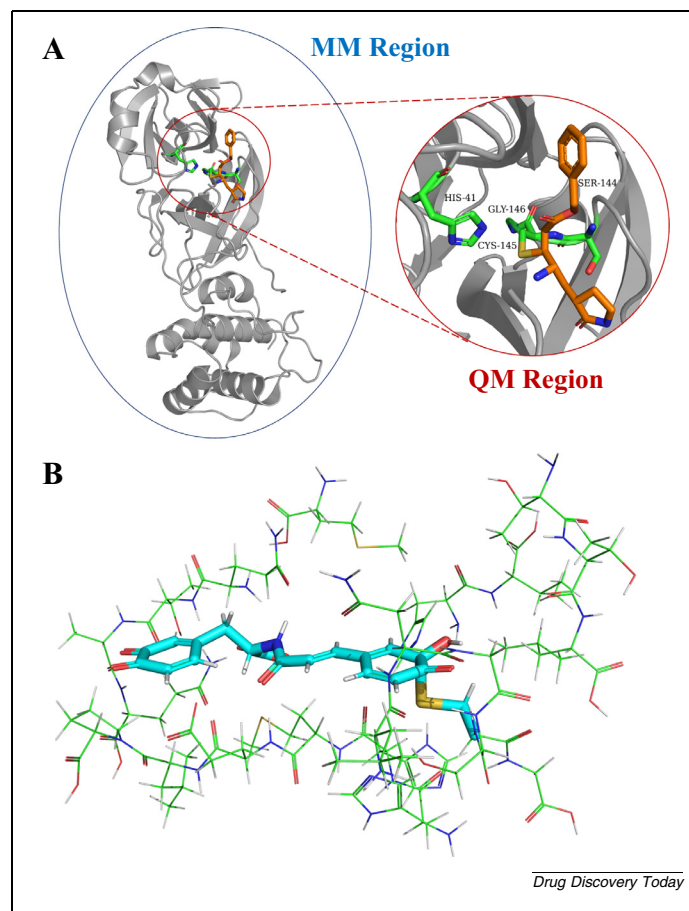

FIGURE 2

Hybrid quantum mechanics/molecular mechanics (QM/MM) and QM-cluster approaches. **(a)** The hybrid QM/MM approach. Blue oval circle indicates the QM region, which includes the whole protein. Shown is the monomer of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Mpro in complex with inhibitor N3 [Protein Data Bank (PDB) ID: 6LU7]. Red circle indicates the QM region, which includes the inhibitor (ligand, shown in orange sticks) and selected amino acid residues (shown in green sticks) in the active site that often involve in the catalytic mechanism. **(b)** The QM-cluster approach. This figure provides a conformation of the active site of the protease–inhibitor complex after density functional theory (DFT) geometry optimization. Shown in green-line representations are the active site residues and in cyan-stick representation is the inhibitor, the clovamide molecule. It clearly suggests that a covalent bond is formed between one carbon on the quinone ring and the sulfur atom in the C145 catalytic residue. For details, refer to Ye et al.[59] Adapted with permission from[59].

CSD(T)/CBS//B3LYP/6-31G(d,p):AMBER level was applied to re-examine the mechanism of M^{pro} catalysis with detailed parameters: transition states, rate-limiting step, energy barriers, and free energy profiles of each step in the catalysis.[52] In other studies, MD simulations are also combined with QM/MM methods to study the noncovalent enzyme–inhibitor/substrate complexes, before close examinations of reaction mechanisms by DFT/MM approaches.[53,57] Khrenova et al. performed a detailed dynamic study of the enzyme–substrate complex of M^{pro}.[45] They evaluated the electronic density features of the complex and showed that QM/MM-MD trajectories disclose substrate reactivity in M^{pro}, and were in good agreement with relevant experimental data.

Less applied in COVID-19 drug modeling studies is the QM-cluster approach, which treats the small-molecule drug and a truncated model of an enzyme receptor (usually dozens of surrounding amino acid residues in the active site) explicitly at an QM level.[58] Compared with QM/MM, this method applies only QM, providing high-level chemical accuracy particularly at enzyme catalytic regions to closely examine protein–ligand interactions and reaction mechanisms. However, one downside of this method is its high computational cost, because as it typically involves hundreds of atoms to be treated by DFT. Ye et al. applied the QM-cluster approach in their modeling of the inhibitory effects of a small molecule, clovamide, on M^{pro}.[59] Fig. 2b shows the 22 amino acids and the inhibitor ligand included in the study, in which the QM-cluster method provided crucial insights into covalent modifications by the inhibitor of the catalytic residues of M^{pro}.[59]

Integrating DFT calculations in drug modeling research

Here, we propose a roadmap that highlights important factors to consider when developing a DFT-based drug modeling methodology. First, when developing an methodology that is based directly on modeling purposes, the general approach of how DFT is applied in a study is strongly dependent on the objectives. These studies can be categorized into three types: (i) studies intended to study solely molecular properties of isolated drug molecules at an electronic level. In this case, DFT is used as a stand-alone methodology[20–21,49]; (ii) studies intended to study drug–receptor interactions would usually incorporate MM methodologies; and (iii) for those that examine mechanisms of inhibition reactions by a potential COVID-19 drug, QM/MM is often used and should be considered for future studies. A review[50] contains detailed information pertaining to the QM/MM method and its set-up, and is a great source for biochemists considering applying QM/MM in their studies.

Next, successful DFT calculations rely on appropriate set-ups. The two most important parameters to consider, functionals and basis sets, were introduced above. A list of computer codes and programs used to run DFT calculations can be found in Table S1 in the supplemental information online. Table S1 in the supplemental information online also supplies information regarding open-access options.

As discussed previously, the highly accurate nature of DFT has determined its limited applicability to macromolecular systems, which can be computationally expensive, time-consuming, and limiting regarding treatable size of given systems. A potential solution to circumvent these limitations is by parallel implementations of DFT programs, which makes DFT applicable to larger systems that encapsulate a larger portion of protein molecules.[60] This could be applied to QM/MM studies and effectively enlarges the QM region treatable by DFT. Linear-scaling DFT approaches (loaded in codes such as ONETEP[61], allow such expansion of scopes and scales. Parallel implementation of DFT on graphical processing units (GPUs) are also shown to be capable of providing rapid and accurate electronic structure calculations.[62] In the list provided in Table S1 in the supplemental information online, TeraChem is a GPU-based program that provides calculations with higher speed and, hence, improved efficiency.

Recently, semiempirical quantum mechanical methods GFNn-xTB (where n = 1, 2) were developed on the basis of density functional tight binding (DFTB) theory for equilibrium structure determination of both organic and metalloproteins.[63–64] These methods, particularly GFD2-xTB, provide an impressive level of accuracy in protein structure optimization with minimal deviation from experimental X-ray crystal structures compared with other force-field based MM methods.[64] The GFNn-xTB methods incorporate highly accurate, well-established D3 or D4 dispersion correction, two advanced successors that belong to the Class III dispersion correction mentioned in Table 1, allowing improved accuracy in macromolecular modeling.[64–65] In addition, these novel methods provide alternative approaches with affordable computational cost (even the largest proteins with up to 5000 atoms could be optimized in a few days on a small workstation), by which proteins and drug modeling research could be performed.[64] Given its low computational cost for a QM method, GFNn-xTB opens up new avenues for computational protein research and in silico rational, structure-based drug discovery.[64]

Concluding remarks and future perspectives

The molecular modeling studies of COVID-19 pharmaceuticals reviewed in this article highlight the extensive application of DFT in the evaluation of molecular electronic properties of drug compounds. Most frequently, DFT is used to calculate electronic structures and properties of a given small molecule (Table 2). These electronic details correlate with properties of drug molecules and directly govern their binding modes with receptors. The most common purpose of application of DFT in studying small molecules is reactivity evaluations revealed by energy gaps of the FMO. The combined application of MM-based techniques along with DFT is common among COVID-19 drug modeling research. Moreover, when it comes to studying reaction mechanisms (inhibitory mechanisms by drug inhibitors or catalytic mechanisms by substrates), hybrid QM/MM is the prevalent approach because of its high efficiency in both tasks: examining drug–receptor bindings (MM) and predicting chemical reactions (QM). Integrating QM and MM calculations into one package, QM/MM offers researchers great convenience and efficiency. However, only those studies that investigate mechanisms are likely to apply this approach because of the complex nature involved in setting up QM/MM calculations in large biological

systems. QM-cluster is a less applied, but relatively accurate approach to examine protein–ligand interactions at enzyme active sites and to predict reaction mechanisms.

Future drug modeling efforts should pay special attention to developing an appropriate methodology with proper set-ups (including codes, programs, levels of theory, etc.). MM techniques are fast, efficient, and easy approaches to describe how drug molecules bind macromolecular targets, but they lack chemical accuracy and are unable to provide electronic insights. QM techniques, represented by DFT, are capable of offering electronic details and predict reaction mechanisms (flow of electrons) between drug molecules and receptors, but are computationally expensive and time-consuming. QM/MM requires expertise in both QM and MM and sufficient computing resources. Developing a method that is suitable for research objectives while balancing computational costs and chemical accuracy is at the core of drug modeling studies.

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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.drudis.2021.12.017.

References

1. J.M. Sanders, M.L. Monogue, T.Z. Jodlowski, J.B. Cutrell, Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review, JAMA 323 (2020) 1824–1836.
2. F. Kabinger, C. Stiller, J. Schmitzová, C. Dienemann, G. Kokic, H.S. Hillen, et al., Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis, Nat Struct Mol Biol 28 (2021) 740–746.
3. D.R. Owen, C.M.N. Allerton, A.S. Anderson, L. Aschenbrenner, M. Avery, S. Berritt, et al., An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19, Science 374 (2021) 1586–1593.
4. S. Mandal, M. Moudgil, S.K. Mandal, Rational drug design, Eur J Pharmacol 625 (2009) 90–100.
5. M.L. Lynch, E.H. Snell, S.E.J. Bowman, Structural biology in the time of COVID-19: perspectives on methods and milestones, IUCrJ 8 (2021) 335–341.
6. S. LaPointe, D. Weaver, A review of density functional theory quantum mechanics as applied to pharmaceutically relevant systems, Curr Comput Aided-Drug Des 3 (2007) 290–296.
7. K. Vanommeslaeghe, O. Guvench, A.D. MacKerell, Molecular mechanics, Curr Pharm Des 20 (2014) 3281–3292.
8. F.K. Yoshimoto, The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19, Protein J 39 (2020) 198–216.
9. E. Hartenian, D. Nandakumar, A. Lari, M. Ly, J.M. Tucker, B.A. Glaunsinger, The molecular virology of coronaviruses, J Biol Chem 295 (2020) 12910–12934.
10. S. Ullrich, C. Nitsche, The SARS-CoV-2 main protease as drug target, Bioorg Med Chem Lett 30 (2020) 127377.
11. B. Goyal, D. Goyal, Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy, ACS Comb Sci 22 (2020) 297–305.
12. J. Chen, B. Malone, E. Llewellyn, M. Grasso, P.M.M. Shelton, P.D.B. Olinares, et al., Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex, Cell 182 (2020) 1560–1573.
13. C.J. Gordon, E.P. Tchesnokov, E. Woolner, J.K. Perry, J.Y. Feng, D.P. Porter, et al., Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency, J Biol Chem 295 (2020) 6785–6797.
14. M. Hagar, H.A. Ahmed, G. Aljohani, O.A. Alhaddad, Investigation of some antiviral N-heterocycles as COVID 19 drug: molecular docking and DFT calculations, Int J Mol Sci 21 (2020) 3922.
15. N. Rasool, F. Yasmin, S. Sahai, W. Hussain, H. Inam, A. Arshad, Biological perspective of thiazolide derivatives against Mpro and MTase of SARS-CoV-2: molecular docking, DFT and MD simulation investigations, Chem Phys Lett 771 (2021) 138463.
16. F. Caruso, M. Rossi, J.Z. Pedersen, S. Incerpi, Computational studies reveal mechanism by which quinone derivatives can inhibit SARS-CoV-2. Study of embelin and two therapeutic compounds of interest, methyl prednisolone and dexamethasone, J Infect Public Health 13 (2020) 1868–1877.
17. P.A. Nogara, F.B. Omage, G.R. Bolzan, C.P. Delgado, M. Aschner, L. Orian, et al., In silico studies on the interaction between Mpro and PLpro From SARS-CoV-2 and Ebselen, its metabolites and derivatives, Mol Inform 40 (2021) 2100028.
18. P.D. Wakchaure, S. Ghosh, B. Ganguly, Revealing the inhibition mechanism of RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2 by remdesivir and nucleotide analogues: a molecular dynamics simulation study, J Phys Chem B 124 (2020) 10641–10652.
19. R. Nandini Asha, B. Ravindran Durai Nayagam, N. Bhuvanesh, Synthesis, molecular docking, and in silico ADMET studies of 4-benzyl-1-(2,4,6-trimethyl-benzyl)-piperidine: Potential Inhibitor of SARS-CoV2, Bioorganic Chem 112 (2021) 104967.
20. S. Bagheri Novir, M.R. Aram, Quantum mechanical simulation of chloroquine drug interaction with C60 fullerene for treatment of COVID-19, Chem Phys Lett 757 (2020) 137869.
21. A.S. Rad, M. Ardjmand, M.R. Esfahani, B. Khodashenas, DFT calculations towards the geometry optimization, electronic structure, infrared spectroscopy and UV-vis analyses of favipiravir adsorption on the first-row transition metals doped fullerenes; a new strategy for COVID-19 therapy, Spectrochim Acta A Mol Biomol Spectrosc 247 (2021) 119082.
22. E. Pitsillou, J. Liang, K. Ververis, K.W. Lim, A. Hung, T.C. Karagiannis, Identification of small molecule inhibitors of the deubiquitinating activity of the SARS-CoV-2 papain-like protease: in silico molecular docking studies and in vitro enzymatic activity assay, Front Chem 8 (2020) 623971.
23. D. Li, J. Luan, L. Zhang, Molecular docking of potential SARS-CoV-2 papain-like protease inhibitors, Biochem Biophys Res Commun 538 (2021) 72–79.
24. S.O. Aftab, M.Z. Ghouri, M.U. Masood, Z. Haider, Z. Khan, A. Ahmad, et al., Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach, J Transl Med 18 (2020) 275.
25. H.M. Yasir, F.H. Hanoon, DFT and TD-DFT study of favipiravir tautomerism as RNA polymerase inhibitors: COVID-19, IOP Conf Ser Mater Sci Eng 928 (2020) 072066.
26. F.E. Kalai, E.B. Ç nar, C.-H. Lai, S. Daoui, T. Chelfi, M. Allali, et al., Synthesis, spectroscopy, crystal structure, TGA/DTA study, DFT and molecular docking investigations of (E)-4-(4-methylbenzyl)-6-styrylpyridazin-3(2H)-one, J Mol Struct 1228 (2021) 129435.
27. Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, et al., Structure of M pro from SARS-CoV-2 and discovery of its inhibitors, Nature 582 (2020) 289–293.
28. T.Q. Bui, H.T.P. Loan, T.T.A. My, D.T. Quang, B.T.P. Thuy, V.D. Nhan, et al., A density functional theory study on silver and bis-silver complexes with lighter tetrylene: are silver and bis-silver carbenes candidates for SARS-CoV-2 inhibition? Insight from molecular docking simulation, RSC Adv 10 (2020) 30961–30974.
29. A. Ghaleb, A. Aouidate, H.B.E. Ayouchia, M. Aarjane, H. Anane, S.-E. Stiriba, In silico molecular investigations of pyridine N-Oxide compounds as potential inhibitors of SARS-CoV-2: 3D QSAR, molecular docking modeling, and ADMET screening, J Biomol Struct Dyn 40 (2020) 143–153.
30. M. Rosas-Lemus, G. Minasov, L. Shuvalova, N.L. Inniss, O. Kiryukhina, J. Brunzelle, et al., High-resolution structures of the SARS-CoV-2 2′-O-methyltransferase reveal strategies for structure-based inhibitor design, Sci Signal 13 (2020) eabe1202.

31. A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2020) 281–292.

32. M.S. Refat, A.A. Sedayo, A. Sayqal, A. Alharbi, H.A. Katouah, H.M. Abumelha, et al., Aurintricarboxylic acid and its metal ion complexes in comparative virtual screening versus lopinavir and hydroxychloroquine in fighting COVID-19 pandemic: synthesis and characterization, Inorg Chem Commun 126 (2021) 108472.

33. A.A. Elmaaty, K.M. Darwish, M. Khattab, S.S. Elhady, M. Salah, M.I.A. Hamed, et al., In a search for potential drug candidates for combating COVID-19: computational study revealed salvianolic acid B as a potential therapeutic targeting 3CLpro and spike proteins, J Biomol Struct Dyn (2021), https://doi.org/10.1080/07391102.2021.1918256. Published online April 30, 2021.

34. W. Kohn, A.D. Becke, R.G. Parr, Density functional theory of electronic structure, J Phys Chem 100 (1996) 12974–12980.

35. S.F. Sousa, P.A. Fernandes, M.J. Ramos, General performance of density functionals, J Phys Chem A 111 (2007) 10439–10452.

36. N. Mardirossian, M. Head-Gordon, Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals, Mol Phys 115 (2017) 2315–2372.

37. Y. Zhao, D.G. Truhlar, Density functionals with broad applicability in chemistry, Acc Chem Res 41 (2008) 157–167.

38. W. Koch, M.C. Holthausen, A chemist's guide to density functional theory (2nd ed.), Wiley-VCH, Weinheim, 2015.

39. S. Grimme, Density functional theory with London dispersion corrections, WIREs Comput Mol Sci 1 (2011) 211–228.

40. K.I. Ramachandran, Basis sets, in: K.I. Ramachandran, D. Gopakumar, K. Namboori (Eds.), Computational chemistry and molecular modeling: principles and applications, Springer, Berlin, 2008, pp. 115–138.

41. J. Tan, K.H.G. Verschueren, K. Anand, J. Shen, M. Yang, Y. Xu, et al., pH-dependent conformational flexibility of the SARS-CoV main proteinase (Mpro) dimer: molecular dynamics simulations and multiple X-ray structure analyses, J Mol Biol 354 (2005) 25–40.

42. H. Noor, A. Ikram, T. Rathinavel, S. Kumarasamy, M. Nasir Iqbal, Z. Bashir, Immunomodulatory and anti-cytokine therapeutic potential of curcumin and its derivatives for treating COVID-19 – a computational modeling, J Biomol Struct Dyn (2021), https://doi.org/10.1080/07391102.2021.1873190. Published online January 25, 2021.

43. A. Abo Elmaaty, M.I.A. Hamed, M.I. Ismail, E. B. Elkaeed, H. S. Abulkhair, M. Khattab, et al., Computational insights on the potential of some NSAIDs for treating COVID-19: priority set and lead optimization, Molecules 26 (2021) 3772.

44 A.V. Nemukhin, B.L. Grigorenko, I.V. Polyakov, S.V. Lushchekina, Computational modeling of the SARS-CoV-2 main protease inhibition by the covalent binding of prospective drug molecules, Supercomput Front Innov 7 (2020) 25–32.

45. M.G. Khrenova, V.G. Tsirelson, A.V. Nemukhin, Dynamical properties of enzyme–substrate complexes disclose substrate specificity of the SARS-CoV-2 main protease as characterized by the electron density descriptors, Phys Chem Chem Phys 22 (2020) 19069–71909.

46. Y. Zhao, D.G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, Theor Chem Acc 120 (2008) 215–241.

47. J.S. Murray, Molecular electrostatic potentials: concepts and applications, Elsevier, Amsterdam, 2011.

48. S. Shahab, M. Sheikhi, R. Alnajjar, S.A. Saud, M. Khancheuski, A. Strogova, DFT investigation of atazanavir as potential inhibitor for 2019-nCoV coronavirus M protease, J Mol Struct 1228 (2021) 129461.

49. S.A. Almalki, T.M. Bawazeer, B. Asghar, A. Alharbi, M.M. Aljohani, M.E. Khalifa, et al., Synthesis and characterization of new thiazole-based Co(II) and Cu(II) complexes; therapeutic function of thiazole towards COVID-19 in comparing to current antivirals in treatment protocol, J Mol Struct 1244 (2021) 130961.

50. H.M. Senn, W. Thiel, QM/MM methods for biological systems, Topics Curr Chem 2687 (2007) 173–290.

51. H. Lin, D.G. Truhlar, QM/MM: what have we learned, where are we, and where do we go from here?, Theor Chem Acc 117 (2006) 185

52. H.S. Fernandes, S.F. Sousa, N.M.F.S.A. Cerqueira, New insights into the catalytic mechanism of the SARS-CoV-2 main protease: an ONIOM QM/MM approach, Mol Divers (2021), https://doi.org/10.1007/s11030-021-10259-7. Published online June 21, 2021.

53. K. Świderek, V. Moliner, Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods, Chem Sci 11 (2020) 10626–10630.

54. V. Frecer, S. Miertus, Antiviral agents against COVID-19: structure-based design of specific peptidomimetic inhibitors of SARS-CoV-2 main protease, RSC Adv 10 (2020) 40244–40263.

55. K. Arafet, N. Serrano-Aparicio, A. Lodola, A.J. Mulholland, F.V. González, K. Świderek, et al., 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, Chem Sci 12 (2021) 1433–1444.

56. E. Awoonor-Williams, A.-A.-A.-A. Abu-Saleh, Covalent and non-covalent binding free energy calculations for peptidomimetic inhibitors of SARS-CoV-2 main protease, Phys Chem Chem Phys 23 (2021) 6746–6757.

57. C.A. Ramos-Guzmán, J.J. Ruiz-Pernía, I. Tuñón, A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design, Chem Sci 12 (2021) 3489–3496.

58. S. Ahmadi, L.B. Herrera, M. Chehelamirani, J. Hostaš, S. Jalife, D.R. Salahub, Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: a tutorial review, Int J Quantum Chem 118 (2018) e25558.

59. N. Ye, F. Caruso, M. Rossi, Mechanistic insights into the inhibition of SARS-CoV-2 main protease by clovamide and its derivatives: in silico studies, Biophysica 1 (2021) 377–404.

60. L. Gundelach, T. Fox, C.S. Tautermann, C.-K. Skylaris, Protein-ligand free energies of binding from full-protein DFT calculations: convergence and choice of exchange-correlation functional, Phys Chem Chem Phys 23 (2021) 9381–9393.

61. J.C.A. Prentice, J. Aarons, J.C. Womack, A.E.A. Allen, L. Andrinopoulous, L. Anton, et al., The ONETEP linear-scaling density functional theory program, J Chem Phys 152 (2020) 174111.

62. K. Yasuda, Accelerating density functional calculations with graphic processing unit, J Chem Theory Comput 4 (2008) 1230–1236.

63. S. Grimme, C. Bannwarth, P. Shushkov, A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z = 1–86), J Chem Theory Comput 13 (2017) 1989–2009.

64. S. Schmitz, J. Seibert, K. Ostermeir, A. Hansen, A.H. Göller, S. Grimme, Quantum chemical calculation of molecular and periodic peptide and protein structures, J Phys Chem B 124 (2020) 3636–3646.

65. E. Caldeweyer, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, et al., A generally applicable atomic-charge dependent London dispersion correction, J Chem Phys 150 (2019) 154122.