Nutho, B., Mulholland, A. J., & Rungrotmongkol, T. (2019). Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations Support a Concerted Reaction Mechanism for the Zika Virus NS2B/NS3 Serine Protease with Its Substrate. *Journal of Physical Chemistry B*, *123*(13), 2889-2903. https://doi.org/10.1021/acs.jpcb.9b02157

Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations Support a Concerted Reaction Mechanism for the Zika Virus NS2B/NS3 Serine Protease with Its Substrate

Bodee Nutho,[1] Adrian J. Mulholland,[2*] and Thanyada Rungrotmongkol[3,4*]

[1]Program in Biotechnology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

[2]Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK

[3]Structural and Computational Biology Research Group, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

[4]Ph.D. Program in Bioinformatics and Computational Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

[*]Corresponding authors. TR Fax: +66 2 218-5418; Tel: +66 2 218-5426, AJM Fax: +44(0)117-925-0612; Tel: +44(0)-117-928-9097

E-mail address: thanyada.r@chula.ac.th, adrian.mulholland@bristol.ac.uk

Abstract

Zika virus (ZIKV) is mainly transmitted to humans by *Aedes* species mosquitoes and is associated with serious pathological disorders including microcephaly in newborns and Guillain-Barré syndrome in adults. Currently, there is no vaccine or anti-ZIKV drug available for preventing or controlling ZIKV infection. An attractive drug-target for ZIKV treatment is a two-compartment (NS2B/NS3) serine protease that processes viral polyprotein during infection. Here, conventional molecular dynamics (MD) simulations of the ZIKV protease in complex with peptide substrate (TGKRS) sequence at the C-terminus of NS2B show that the substrate is in the active conformation for cleavage reaction by ZIKV protease. Hybrid quantum mechanics/molecular mechanics (QM/MM) umbrella sampling simulation (PM6/ff14SB) of acylation results reveal that proton transfer from S135 to H51 and nucleophilic attack on the substrate by S135 are concerted. The rate-limiting step involves the formation of a tetrahedral intermediate. In addition, the single-point energy QM/MM calculations, precisely at the coupled cluster theory (LCCSD(T)/(aug)-cc-pVTZ), were calculated to correct the potential energy profiles for the first step of the acylation process. The average computed activation barrier at this level of theory is 16.3 kcal mol^{-1}. Therefore, the computational approaches presented here are helpful for further designing of NS2B/NS3 inhibitors based on transition state analogues.

Introduction

Zika virus (ZIKV) is primarily transmitted to humans through the bite of infected *Aedes* mosquitoes and has become a global public health problem. ZIKV infection appears to produce higher risk of complications including neurologic symptoms in adults and microcephaly in newborns. ZIKV was first isolated from rhesus monkey cells in 1947 in Uganda.[1] Additionally, it can be further classified into two genotypes, African and Asian, based on phylogenetic analysis.[2] ZIKV is a member of the *Flavivirus* genus and Flaviviridae family. Its genome is similar to other flaviviruses, including West Nile virus (WNV), Dengue virus (DENV), Japanese encephalitis virus (JEV) and yellow fever virus (YFV).[3] ZIKV is an RNA virus containing a single-stranded positive-sense RNA (ssRNA) genome of approximately 11 Kb. After infection of ZIKV into the host cells, its ssRNA genome is then translated into a large single polyprotein precursor. This polyprotein precursor is subsequently processed into three structural proteins (capsid (C), premembrane/membrane (prM), and envelope (E)), and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) through the activity of both host and viral proteases. Note that the structural and nonstructural proteins are responsible for forming the virus particle and viral replication process, respectively.[4]

Clinically, the symptoms of the ZIKV infection have been reported to be similar to those observed for other arboviruses, which are commonly known as Zika fever.[5] Due to asymptomatic or mild symptoms of ZIKV infection, the disease was not considered originally as a relevant human pathogen and consequently research into the virus was neglected for more than 50 years by the scientific community.[1, 6] This situation changed dramatically in recent years, as in 2015 the ZIKV infection increased suddenly, becoming epidemic regions in South America especially in Brazil, and spread quickly to other countries around the world.[7, 8] More seriously, there is now growing

scientific evidence indicating a relationship between ZIKV infection and fetal microcephaly in newborns,[9, 10] and Guillain–Barré syndrome in adults.[11, 12] Thus, these phenomena have driven researchers to look at this 'relatively harmless' virus with new perspective.[13-15] Unfortunately, neither vaccine nor antiviral drug is currently available to combat with ZIKV, an intended attempt of researchers is now focused on a rapid understanding of the structural biology of the ZIKV.

Similar to other flaviviral NS3 proteins, the ZIKV NS3 protein (~70 kDa) consists of two functional domains comprising a protease domain and a helicase domain. The C-terminal region of NS3 contains a helicase–nucleoside triphosphatase (NTPase), whereas the N-terminal region of the NS3 protein encodes a chymotrypsin-like serine protease that specifically recognizes substrates containing amino acid residues with basic character such as arginine at the P1 position and lysine at the P2 position.[16, 17] Based on the common structures of flavivirus proteases, the active form of the ZIKV protease comprises the N-terminal protease domain of NS3 protein, which carries the catalytic triad (H51, D75 and S135), and the membrane-bound NS2B protein. Assembly of NS3 protease with the membrane-bound NS2B rearranges the NS3 active site to achieve the optimal geometrical configuration for catalysis, as has been observed for related flavivirus proteases.[18, 19] The NS2B/NS3 serine protease is necessary for viral polyprotein processing to yield mature structural and nonstructural proteins; hence, this enzyme is a particularly promising flavivirus therapeutic target for the design of novel antiviral agents.[20-22]

Recently, several research groups have reported the X-ray crystal structures of the NS3 protease domain of ZIKV in complex with the membrane-bound NS2B (~40 amino acids), covalently bonded via a glycine-rich linker to the N-terminus of NS3, which shows strong peptidolytic activity.[16, 23] Interestingly, a peptidomimetic boronic acid inhibitor (cn-716) covalently linked to the ZIKV protease induces a conformational change in NS2B protein

structure, yielding a more compact 'closed' form, with the NS2B polypeptide wrapped around NS3.[16] In addition, a high-resolution of 1.8 Å crystal structure of ZIKV NS2B/NS3 protease bound with the last four amino acids T^{127}-G^{128}-K^{129}-R^{130} (TGKR) of the NS2B protein has also been solved (**Figure 1A**).[17] This 'TGKR' proteolytic product could bind to S1, S2 and only partially to S3 pockets of the ZIKV protease through mainly charge-charge and hydrogen bonding interactions. More recently, the unlinked ZIKV NS2B/NS3 protease construct in apo form and bound form with a reversely oriented peptide has been produced and revealed by X-ray crystallography.[24] The closed conformation of this construct suggests that NS2B is necessary for NS3 to generate an unoccupied substrate-binding pocket. This is distinguished from the relaxed or open form, as found in the NS2B-(linker)-NS3 construct[23] and as has been reported for other flavivirus NS2B/NS3 proteases.[19] Furthermore, it was observed from the study reported in ref 24 that four amino acids K^{14}-K^{15}-G^{16}-E^{17} (KKGE) from the neighboring NS3 could bind to the ZIKV protease and resist proteolytic activity. Altogether, this information has provided a good starting point for further design of anti-ZIKV agents with high specificity and efficiency.

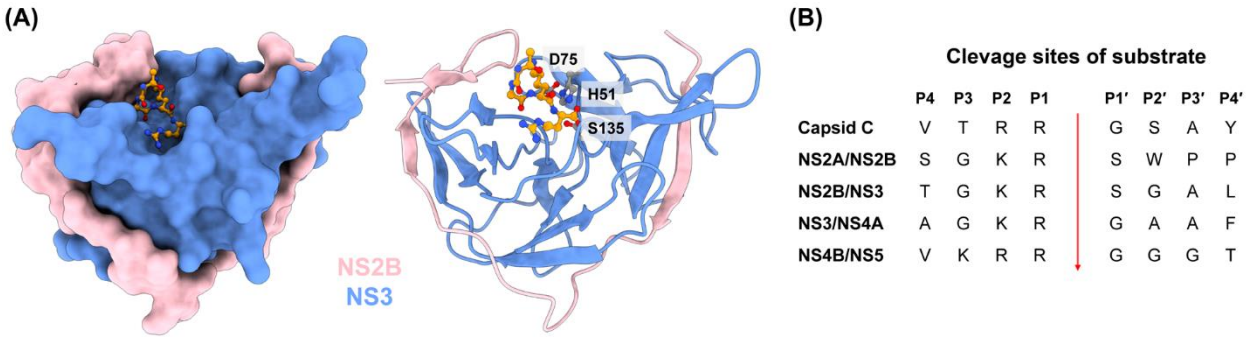
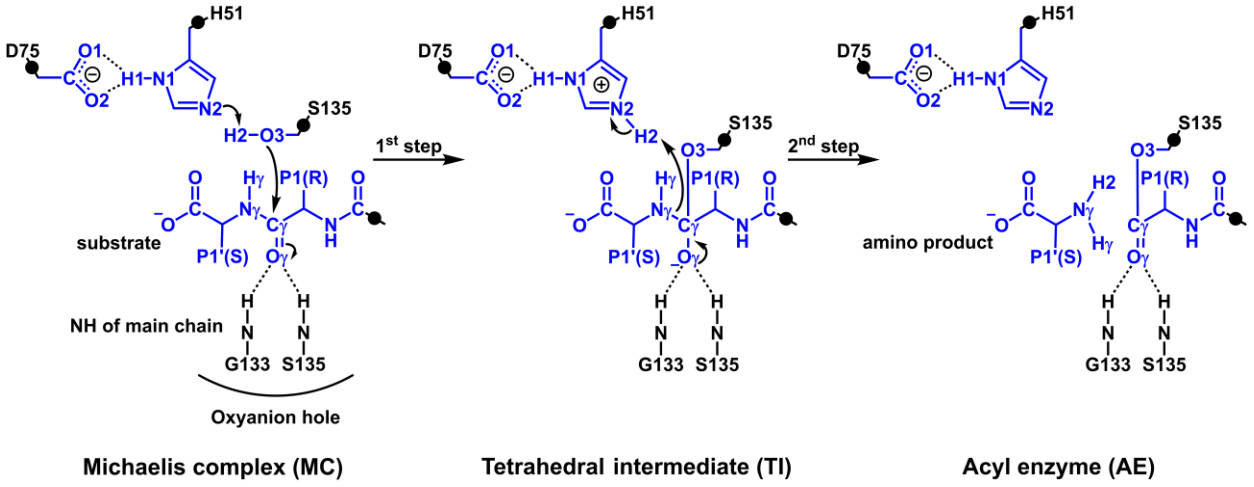


Figure 1. (**A**) 3D-structures of the ZIKV protease in complex with the tetrapeptide 'TGKR' of the NS2B protein (PDB ID: 5GJ4). The catalytic triad residues (D75, H51 and S135) and the co-

crystalized tetrapeptide (orange) are shown. (**B**) The cleavage sites between P1 and P1′ junction for each substrate are represented by red arrow.[25]

The NS2B/NS3 serine protease exhibits a preferential cleavage site of a peptide bond within a peptide substrate, which has an amino acid sequence commonly represented by -P4-P3-P2-P1-P1′-P2′-P3′-P4′-. Note that the peptide bond is cleaved at the P1 and P1′ boundary by the enzyme (**Figure 1B**). With the structural similarities of flaviviral proteases to other proteins in the chymotrypsin family,[26] it has been proposed that these enzymes should present a similar catalytic mechanism. Theoretical and experimental evidence provide a general reaction mechanism for chymotrypsin–like serine proteases composed of two steps; acylation and deacylation processes.[27-29] In the acylation step, the imidazole ring of H51 is in the position to deprotonate the nucleophilic hydroxyl group of S135 in the Michaelis complex (MC), and the D75 side chain is in the orientation to stabilize the positively charged H51 after accepting the proton from S135. Subsequently, the peptide bond is cleaved by a nucleophilic addition reaction on the substrate carbonyl carbon by the oxygen atom on the hydroxyl group of S135, generating a tetrahedral intermediate (TI) formation that is stabilized by hydrogen bonds with some residues of the oxyanion hole (formed by the backbone amides of G133 and S135). The peptide bond is then broken, liberating the N-terminus of the peptide substrate and acyl enzyme (AE) (**Scheme 1**). The deacylation process involves the hydrolysis reaction on the AE, leading to a release of the C-terminus of the substrate and the catalytic residues of the enzyme. On the basis of the reaction mechanism proposed for serine proteases, the acylation process is considered as the rate-limiting step, in which the reaction rate is determined by formation of a relatively stable TI.[28, 30, 31] In addition, for most serine proteases, the acylation process is known to be the rate-limiting step for

hydrolysis of amide bonds, while the deacylation process is rate-limiting step for hydrolysis of ester bonds.[32-34]



Scheme 1. Proposed reaction mechanism for the reaction of the acylation process catalyzed by the ZIKV protease. This process involves proton transfer and nucleophilic attack, leading to the formation of a tetrahedral intermediate (TI), and thereafter the peptide bond breaking to produce acyl enzyme (AE). The QM region (labeled in blue) consists of a fragment of the catalytic residues (H51, D75 and S135) of the NS3 protease and its substrate. 4 hydrogen link atoms are displayed in a closed circle. The atomic labels involved in the reaction mechanism are also given.

Although serine proteases have been reviewed extensively,[35-37] the structural dynamics and the enzymatic reaction catalyzed by ZIKV NS2B/NS3 serine protease have not been described and established yet. Therefore, the purpose of the current study is to gain a detailed picture of the dynamical behavior and the mechanistic reaction of the acylation process catalyzed by NS2B/NS3 serine protease of ZIKV. Classical molecular dynamics (MD) simulations with a molecular mechanics (MM) potential function were used to explore the dynamics of the active enzyme-substrate complex, whereas combined quantum mechanics/molecular mechanics (QM/MM) methods were applied to determine the mechanistic reaction for the enzymatic cleavage of

substrate. Here, we performed an extensive semiempirical QM/MM umbrella sampling simulation to elucidate the reaction energy profiles of the two steps of the acylation process catalyzed by the ZIKV protease. The corresponding rate-limiting step of the acylation process, which involves the TI formation, was further clarified from the potential energy profiles using DFT QM/MM level. The corrections with high level QM/MM calculations (precisely at the coupled cluster method) were also applied to estimate multiple QM/MM potential energy surfaces for the first step of the acylation process. It should be noted that coupled cluster calculations are considered as the "gold standard" of *ab initio* methods for QM/MM calculations of potential energy surfaces for enzyme catalysis.[38] The results shown here give a detailed insight into the structural dynamics and reaction mechanism that could be useful for future inhibitor design based on transition state (TS) analogues.

Computational Methods

System Preparation

The coordinates of the ZIKV protease enzyme in complex with TGKR peptide from the X-ray structure (Protein Data Bank (PDB), code 5GJ4)[17] were employed. Note that a tetrapeptide 'TGKR' corresponds to the -P4-P3-P2-P1- positions of the NS2B/NS3 cleavage site, respectively. The carboxylate group presented on the peptide C-terminus was replaced by the P1′ serine (S) residue, corresponding to the NS2B/NS3-mediated cleavage site.[25] The addition of missing hydrogen atoms and protonation states of ionizable amino acid residues were assigned at pH = 8.5, using the H++ server (http://biophysics.cs.vt.edu/H++),[39] except for H51 which was set as the neutral form with protonated δ-NH (HID type) due to mechanistic consideration. In addition, their environments were considered based on the possibility of hydrogen bond formation with the surrounding residues. The ff14SB[40] AMBER force field was used to assign bonded and non-

bonded parameters to the protein and peptide substrate. The system was solvated in an octahedron box of TIP3P[41] water model and neutralized with sodium ions (Na^+),[42] consisting of ~6000 water molecules and 3 Na^+ ions.

Molecular Dynamics Simulations

MD simulations of the ZIKV NS2B/NS3−TGKRS complex were performed with the AMBER16 software package coupled with the *sander* and *pmemd* modules.[43] The system was simulated under periodic boundary conditions and the electrostatic interactions were treated using the Particle Mesh Ewald method.[44] Temperature was maintained by the Langevin dynamics technique with a collision frequency of 2 ps^{-1},[45] and pressure was controlled by the simple Berendsen barostat,[46] with a pressure-relaxation time of 1 ps. A cut-off distance of 10 Å was set for non-bonded interactions, while all covalent bonds involving hydrogen atoms were constrained by the SHAKE[47] algorithm, allowing a simulation time step of 2 fs. Prior to performing MD simulations, the water molecules and counter ions were energy-minimized by 1000 iterations of steepest descent (SD), followed by 2000 iterations of conjugate gradient (CG) methods, while the rest of the molecules were restrained. Afterward, the protein was energy-minimized by SD (1000 steps) and CG (2000 steps) methods together with restrained solvent. Finally, the entire system was fully energy-minimized using the same minimization procedure. After an initial energy minimization step, the complex was slowly heated up to 300 K in 100 ps using positional restraints of 50.0 kcal mol^{-1} $Å^{-2}$ for the alpha carbon (Cα) atoms of protein. The system was further equilibrated for 1000 ps with positional restraints, which were consequently decreased by 5.0 kcal mol^{-1} $Å^{-2}$ every 100 ps, and another 100 ps without any positional restraints. Subsequently, the whole system was simulated under an isothermal-isobaric (*NPT*) ensemble for 200 ns (300 K, 1

atm) in three independent simulations, leading to 600 ns in total of simulation time. In addition, a single MD simulation was performed for the apo form for 1 μs.

QM/MM MD Free Energy Calculations

QM/MM MD simulations were carried out using the *sander* module implemented in the AMBER16 package. One suitable representative conformation from the previous MM MD simulations that fulfilled the following criteria: (*i*) *d*(N2–H2) < 2.0 Å and *d*(O3–Cγ) < 3.0 Å, and (*ii*) the nucleophilic attack angle, θ_y(O3–Cγ–Oγ), close to 90°, was selected as the starting geometry for the QM/MM simulations. The QM region consisting of 60 atoms included the side-chains of the catalytic triad (H51, D75 and S135) and the substrate amino acids (P1 and P1′) close to the scissile bond (blue label in **Scheme 1**; see **Figure 2**). Moreover, a negative total charge (–1e) was set for the QM region and 4 hydrogen link atoms (LAs)[48, 49] were applied to saturate the QM/MM boundary. The SHAKE algorithm was not calculated in the QM region. The QM region was treated with the PM6 semiempirical method,[50] while the remaining region was described at the MM level using the ff14SB force field.

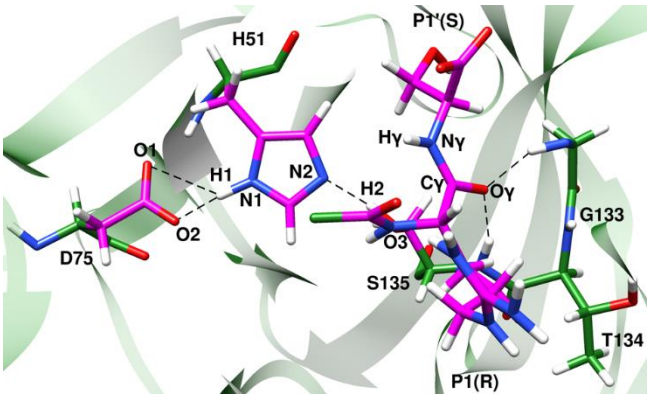


Figure 2. Active site of ZIKV protease with the fragment of substrate. The QM atoms and the MM regions are displayed as magenta and green carbon atoms, respectively. Four hydrogen link atoms were used to saturate the QM/MM boundary, where the color is switched from magenta to green.

The hydrogen bonds are represented by dashed lines. The atomic labels involved in the reaction mechanism are also given for further discussion.

QM/MM umbrella sampling MD simulations were performed for each step of the acylation process, harmonically restraining the reaction coordinate (RC) with a force constant of 100 kcal mol^{-1} $Å^{-2}$. Each window was composed of 50 ps of restrained QM/MM MD. The RC was scanned at 0.1 Å intervals between adjacent windows, using the previous geometry as the starting point for the subsequent window. The free energy profiles for each step were calculated by the weighted histogram analysis method (WHAM).[51-53] The acylation process was categorized into two steps, in which the first and second steps were defined by TI formation and the breaking of peptide bond, respectively. The following reaction coordinates (RCs) were used: (*i*) *the first step of the acylation* – a 2D umbrella sampling simulation involving the proton transfer from S135 oxygen to H51 ε nitrogen, defined as *d*(O3−H2)−*d*(N2−H2), ranging from −1.0 to 1.0 Å, and the nucleophilic attack of S135 oxygen to the carbonyl carbon of substrate, defined as *d*(O3−Cγ), varying from 3.0 to 1.5 Å. In order to calculate a less constrained pathway, a 1D umbrella sampling for the reaction was run using a RC described by the difference of the breaking and forming bonds (*d*(O3−H2)−*d*(N2−H2)−*d*(O3−Cγ), from −3.8 to −0.6 Å). For (*ii*) *the second step of the acylation* – 2D umbrella sampling with two RCs accounting for the breaking of Cγ−Nγ peptide bond, defined as *d*(Cγ−Nγ), with values ranging from 1.5 to 1.9 Å, and the transfer of proton from the ε nitrogen of the H51 to the peptide bond nitrogen of the substrate, defined as *d*(Nγ−H2)−*d*(N2−H2), varying from 1.6 to −1.0 Å, was simulated. Similar to the first step of the acylation process, a single RC described by the antisymmetric combination of *d*(Cγ−Nγ) and *d*(Nγ−H2)−*d*(N2−H2), (*d*(Cγ−Nγ)−*d*(Nγ−H2)+*d*(N2−H2), from −0.2 to 2.8 Å) was used.

High Level QM/MM Calculations for the First Step of the Acylation Process

High-level QM/MM calculations (up to the LCCSD(T) level, which can provide 'chemically accurate' barriers for reactions[54-56]) were performed to obtain accurate potential energy profiles for the reaction. TS-like conformations obtained from PM6/ff14SB 1D umbrella sampling MD simulations were used as the starting geometries for an adiabatic mapping approach along the RC of *d*(O3–H2)–*d*(N2–H2)–*d*(O3–Cγ) for the first step of the acylation process. Five initial geometries of TS structure were selected from 30–50 ps simulation at every 5 ps of the TS-sampling window at the PM6/ff14SB of stationary point (*d*(O3–H2)–*d*(N2–H2)–*d*(O3–Cγ) at –1.8 Å). To reduce the size of QM/MM calculations, all water molecules above the sphere of the protein were removed, and so the resulting protein was in a 10 Å layer of water surrounding protein. The QM region consisted of the atoms that were used in QM/MM umbrella sampling MD simulations as mentioned above. All protein residues plus water molecules within a 5-Å sphere of the substrate were included in the active region of the optimization process and allowed to move freely, while the rest of atoms were kept frozen. All QM/MM calculations were performed using ChemShell,[57, 58] by combining the Orca package[59] for the QM part and DL_POLY[60] for the MM part. A full electrostatic embedding scheme was employed to calculate the polarizing influence of the enzyme environment on the QM region. Hydrogen LAs with the charge-shift model[44, 61] were applied to treat the QM/MM boundary. During the QM/MM geometry optimizations, all atoms in QM region were described at the BH&HLYP-D3/6-31G(d) level of theory. The D3 in the acronym indicates that an empirical dispersion correction was applied to the DFT calculations.[62] The inclusion of dispersion correction plays an important in DFT QM and QM/MM calculations, particularly in those reactions catalyzed by enzymes.[63, 64] The MM region was described by the ff14SB AMBER force field implemented in the DL_POLY code. A RC restraint was used to drive the reaction from the TS toward the reactant and product states at 0.1 Å intervals during the geometry optimizations.

Note that the BH&HLYP[65-67] is known to provide better results than B3LYP[65, 68, 69] for some proton and charge transfers[56, 70] and also a better description of hydrogen bonding.[71-73] In addition, the BH&HLYP method gave the good results in line with *ab initio* methods.[74] Herein, single-point energy QM/MM calculations on the BH&HLYP-D3/6-31G(d)/ff14SB optimized geometries were performed at the LMP2/(aug)-cc-pVTZ/ff14SB, SCS-LMP2/(aug)-cc-pVTZ/ff14SB, and LCCSD(T)/(aug)-cc-pVTZ/ff14SB levels of theory using ChemShell together with MOLPRO for the QM region.[75] SCS denotes the spin component scaled method[76] applied for the MP2[77] calculations. It is worth noting that the combination of SCS and MP2 calculations provided results in good agreement with the coupled cluster methods for the reaction mechanisms catalyzed by other enzymes.[56, 79, 80] The (aug) in the notation of the (aug)-ccpVTZ[81] basis set means that a basis set augmented with diffuse functions were only used for the oxygen atoms. Meanwhile, the L in these acronyms for the correlated ab initio methods indicates that local approximations[38, 82, 83] were adopted in the QM/MM calculations.

Results and Discussion

MM MD Simulations

To monitor the structural stability of the three independent MD simulations, the root-mean-square deviation (RMSD) relative to the initial minimized structure for the backbone atoms of protein was evaluated and plotted along the simulation time (see **Figure S1**). The results show that the protein is stable and shows similar behavior in the three independent replicas (average RMSD of 1.5 Å). Upon substrate binding, the active site (residues within 4 Å of substrate) does not show significant structural changes caused by ligand binding, in comparison with the apo form (average RMSD value of 1.0 Å for all heavy atoms of residues in the active site). In addition, the structure

of the apo state stays in a closed conformation over the course of the simulation (1 μs), and exhibits a stable substrate-binding pocket, which is very similar to the X-ray crystal structure of ZIKV protease as a free enzyme.[24] To evaluate the structural compactness at the protein active site, radius of gyration (R_g) of amino acids within 4 Å of substrate was calculated. The result shows relatively constant R_g values (average value of 8.7 Å) for both forms, reflecting a closed conformation of ZIKV protease (see **SI**, **Figure S2**). Representative structures for each replica in the bound state and apo state (obtained by cluster analysis) of the binding pocket are superimposed to illustrate the closed conformation of enzyme with and without substrate bound are shown in **SI**, **Figure S3**. Our MD simulation results support the closed conformation of ZIKV protease in both apo and bound forms that evidently found in the experiments based on the X-ray structures with and without peptide/inhibitor bound.[24, 84]

Distances at the Catalytic Triad and Oxyanion Hole

The distances of the catalytic residues H51, D75 and S135 (*d1–d3*), nucleophilic attack (*d4*) and oxyanion hole constructed by the backbone amides of G133 and S135 (*d5* and *d6*), are illustrated in **Figure 3A**. Histograms of these important distances (*d1–d6*) are also plotted in **Figure 3B**, where the data presented here is a combination of all three independent MD simulations collected from the last 100 ns of each replica. The histograms for the distances of *d1* and *d2* show a peak in the population at ~3.0 Å for *d1* and ~2.7 Å for *d2* (**Figure 3B**), as monitored from the carboxylate oxygen atoms (O1 and O2) of D75 to the N1-imidazole nitrogen of H51, indicating the presence of two hydrogen bonds. Similar to the X-Ray structures of WNV NS2B/NS3 protease in complex with peptide inhibitors, the O1 atom was located closer to the N1 atom than the O2 atom (*d1* of ~3.3 Å and *d2* of ~2.8 Å).[19, 85] The histogram of the distance between the N2-imidazole nitrogen of H51 and the O3-hydroxyl oxygen of S135 (*d3*) displays a peak at ~2.7 Å, involving the forming of a strong hydrogen bond and feasibly facilitating the proton transfer from S135 to

H51. Besides, the nucleophilic attack involving the first step of the acylation process is defined as the distance between the hydroxyl oxygen atom (O3) of S135 and the carbonyl carbon (Cγ) of the scissile peptide P1(R) position (*d4*). Based on the histogram for this distance, there is a maximum in the population at ~3.0 Å. This finding demonstrates that all three catalytic triad residues and the substrate are in an appropriate configuration for initiating the cleavage reaction.

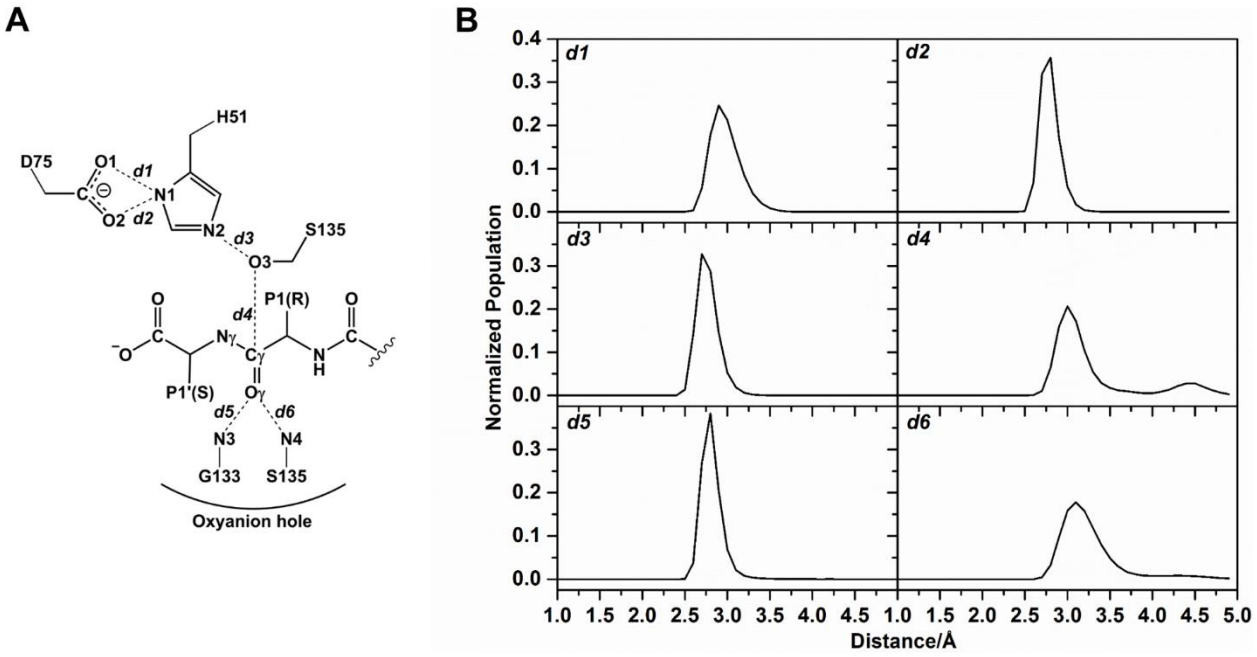


Figure 3. (**A**) Definition of the interatomic distances involved in the cleavage reaction (*d1–d6*). (**B**) Histograms of key interatomic distances, *d1–d6*, from MM MD simulations, sampling from 100 to 200 ns.

The oxyanion hole of the flavivirus NS2B/NS3 protease is formed by the backbone amides of residues G133 and S135, which interact with the Oγ–carbonyl oxygen of the P1(R) reacting residue in the cleavage reaction (*d5* and *d6*). The histograms of these two distances indicate a maximum in the population at ~3.1 Å for *d5* and ~2.8 Å for *d6*, showing stable interactions. Altogether, the overall substrate conformation and interatomic interactions present in the active site of ZIKV NS2B/NS3 protease over the course of MD simulations are relatively similar to those

of the WNV NS2B/NS3 X-ray structures[19, 85] and our previous MD simulations of DENV type 2 NS2B/NS3.[86] This demonstrates that the simulation models are stable.

In addition to these six important distances, the angles related to nucleophilic attack, defined as θ_x and θ_y (see **SI**), are measured. The histograms of both angles show a peak with the highest population at about 98° and 82° for θ_x and θ_y, respectively (**Figure S4**). These two calculated parameters that describe the nucleophilic attack had been introduced from the average value of superimposed 79 serine protease complexes with inhibitors, consistently close to 90°.[87] Additionally, the molecular geometry found in these protease-inhibitor complexes might be considered as a universal hypothesis of good models for the reactive MC, resulting in rapid progression to the AE. Therefore, both parameters, *d4* and θ_y, would be helpful to guide in selecting an initial structure for the QM/MM simulations in the next step (see **Methods**).

PM6/MM Free Energy Profiles

QM/MM approaches can be used with enhanced sampling methods, e.g. the commonly used umbrella sampling technique, to calculate activation free energies in good agreement with experimental data.[88-91] Herein, we applied QM/MM (PM6/ff14SB) MD umbrella sampling simulations to model the reaction involving the acylation step catalyzed by the ZIKV protease and to map the corresponding free energy surfaces (potential of mean force, PMF). The free energy profiles consist of the TI formation and the breakdown of peptide bond, which are calculated by iteratively changing the value of a RC (typically defined as a combination of relevant interatomic distances; see **Computational Methods**). The computational time for extensive sampling is very costly, the use of semiempirical level is more practical. In this study, we selected the PM6 semiempirical method to treat the QM region. PM6 has been successfully applied to study reaction mechanisms involved in enzyme-catalyzed reactions.[92-94] Moreover, our preliminary results show that the energy barrier obtained at the PM6/ff14SB level of theory is in more reasonable agreement

with the experimental activation free energy compared to PM3, AM1/d and SCC-DFTB results. The resulting free energy profiles at PM6/ff14SB level are depicted in **Figures 4** and **5**.

The 2D free energy surfaces for the corresponding reaction are calculated along the RCs. The minimum energy path (MEP)[95-97] on the surface for the first step of acylation (**Figure 4A**) shows an activation free energy ($\Delta G^{\ddagger}$) of 10.1 kcal mol^{-1} relative to the MC (TS1; Table 1). The apparent experimental activation free energy is ~18.2 kcal mol^{-1} (k_{cat} of 0.95 s^{-1} at 37 °C)[98]; hence, the PM6/ff14SB reaction barrier is too low compared to the experimental data. This is due to limitations of the PM6 semiempirical QM method for this reaction, as shown by higher level QM/MM calculations (see below).[99] Afterward, this TS further evolves to the TI formation with a lower barrier of 7.1 kcal mol^{-1} relative to the MC. Only one TS is found in the first step of the acylation, i.e. the reaction occurs in a concerted manner,[100] in which the transfer of proton from S135 to H51 and the nucleophilic addition of S135 oxygen to the carbonyl carbon on P1(R) of the substrate occur together. This is similar to the *ab initio* QM/MM results from MD-FEP calculations obtained by Kato and Ishida on trypsin.[100] Moreover, the concerted reaction mechanism of the TI formation has been studied extensively by theoretical investigations on other serine proteases.[30, 100-104] However, this is in contrast to the QM/MM (PDDG-PM3/ff99SB) study suggesting a stepwise mechanism of the TI formation catalyzed by DENV type 2 NS2B/NS3 serine protease, in which the nucleophilic attack takes place only after the proton transfer.[105] Thus, we additionally performed QM/MM MD umbrella sampling simulations using only one RC involving the proton transfer from S135 oxygen to H51 ε nitrogen (d(O3−H2)−d(N2−H2)) in the MC. During the proton transfer from S135 to H51, the nucleophilic attack on the substrate happens spontaneously. Therefore, this is a further evidence to support a concerted mechanism. The resulting free energy profile at the PM6/ff14SB level has an activation barrier of 9.9 kcal mol^{-1} (See **Figure S5**), which

is slightly lower than that of the energy barrier obtained by using combination of the interatomic distances at the same level of theory (**Figure 5**). Nevertheless, Warshel and colleagues suggested a stepwise reaction with prior proton transfer in the MC.[106] Indeed, the question "How do serine proteases really work?",[107] is still open, and the exact sequence of steps is under debate.

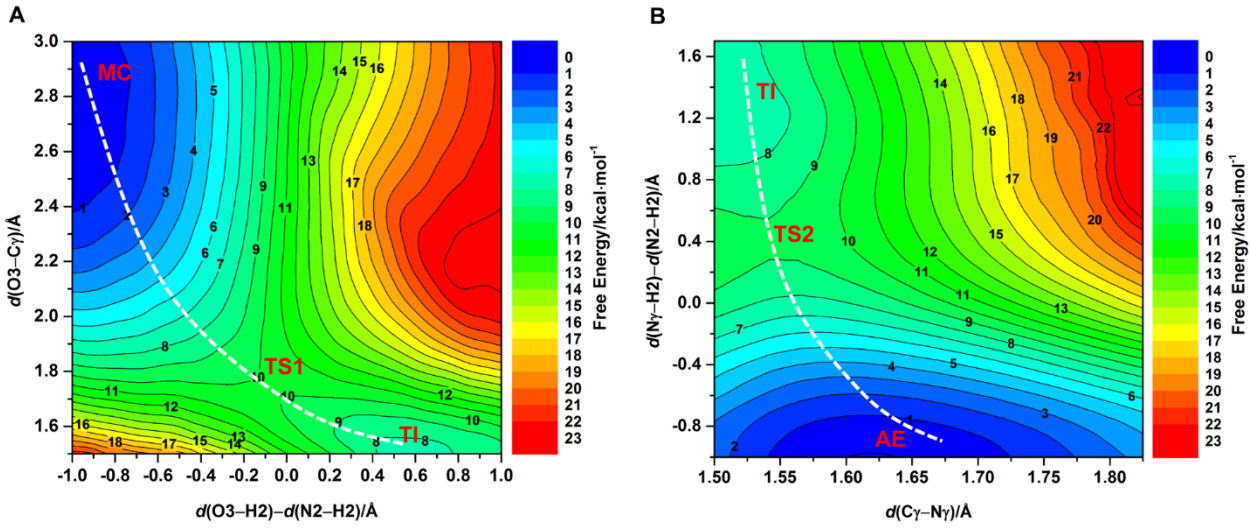


Figure 4. (**A**) 2D free energy surfaces for the TI formation and (**B**) the breakdown of the peptide bond for the reaction of the ZIKV protease with its substrate, calculated from umbrella sampling simulations at the PM6/ff14SB level of theory.

Table 1. Free energies, relative to the MC, obtained at the PM6/ff14SB QM/MM level from two-dimensional umbrella sampling MD simulations (i.e. from the free energy surfaces shown in Figure 4).

	Free energy (kcal mol^{-1})
MC	0.0
TS1	10.1
TI	7.1
TS2	9.1
AE	0.5

Using the combination of the interatomic distances, **Figure 5** shows 1D free energy profiles obtained from QM/MM MD 1D umbrella sampling simulations for both steps of the acylation. All of the intermediates and reaction energy barriers observed in 2D free energy surfaces are apparent. The free energy profile for the first step of the acylation has the same energy barrier of 10.1 kcal mol^{-1} with 2D free energy surfaces at the RC of ~1.8 Å. The distance calculations also support that the first step of the acylation process takes place in a concerted manner, as can be seen from a simultaneous decrease in *d*(N2−H2) and *d*(O3−Cγ) involving the distances of the proton transfer and nucleophilic attack, respectively (**Figure 6A**). Moreover, the alterations of the two hydrogen bond distances between H51 and D75, defined as *d*(O1−H1) and *d*(O2−H1), are monitored along the RC. The calculated distances show that *d*(O1−H1) increases from ~2.0 up to ~2.4 Å at the RC accounting for TS1, while *d*(O2−H1) shortens from ~2.0 to ~1.6 Å with the progress of the reaction (**Figure 6B**). This finding is described by the conformational change of the imidazole ring of H51 that prepares its geometry to accept the proton from S135. The rotation of the H51 imidazole ring also results in the formation of two strong hydrogen bonds with the carboxylate group of D75. Apart from these two hydrogen bonds, other hydrogen bonds are also detected in the oxyanion hole. In the MC complex, two hydrogen bonds are formed between the carbonyl oxygen on the scissile substrate and the NH group of G133 and S135, defined as *d*(NH_{G133}−Oγ) and *d*(NH_{S135}−Oγ) (**Figure 6B**). As the reaction proceeds, both distances continuously decrease along the RC. They remain approximately constant at ~1.9 Å to stabilize negative charge centered on the carbonyl oxygen when the TI is formed (for comparison: the Mulliken charges of the carbonyl oxygen are −0.69 and −0.93 a.u. in the MC and the TI formation, respectively, **Table 2**). In addition, as expected, the Mulliken charge analysis revealed that the electronic properties of the $N2_{H51}$ atom change from basic to acidic after receiving the proton from S135. On the other hand,

as the $N2_{H51}$ becomes more acidic (more positive charge), the Mulliken charge on the substrate's Nγ atom decreases. This event plays an important role for the peptide bond breaking in the next step, where the H2 hydrogen is transferred from $N2_{H51}$ to the Nγ nitrogen on the TI. However, the atomic charge of the $O3_{S135}$ atom remains approximately constant (around –0.6 a.u.) suggesting that the reaction proceeds by a concerted pathway rather than by a stepwise manner as mentioned previously. Also, this information is in good agreement with the single TS found in the 1D and 2D free energy profiles.

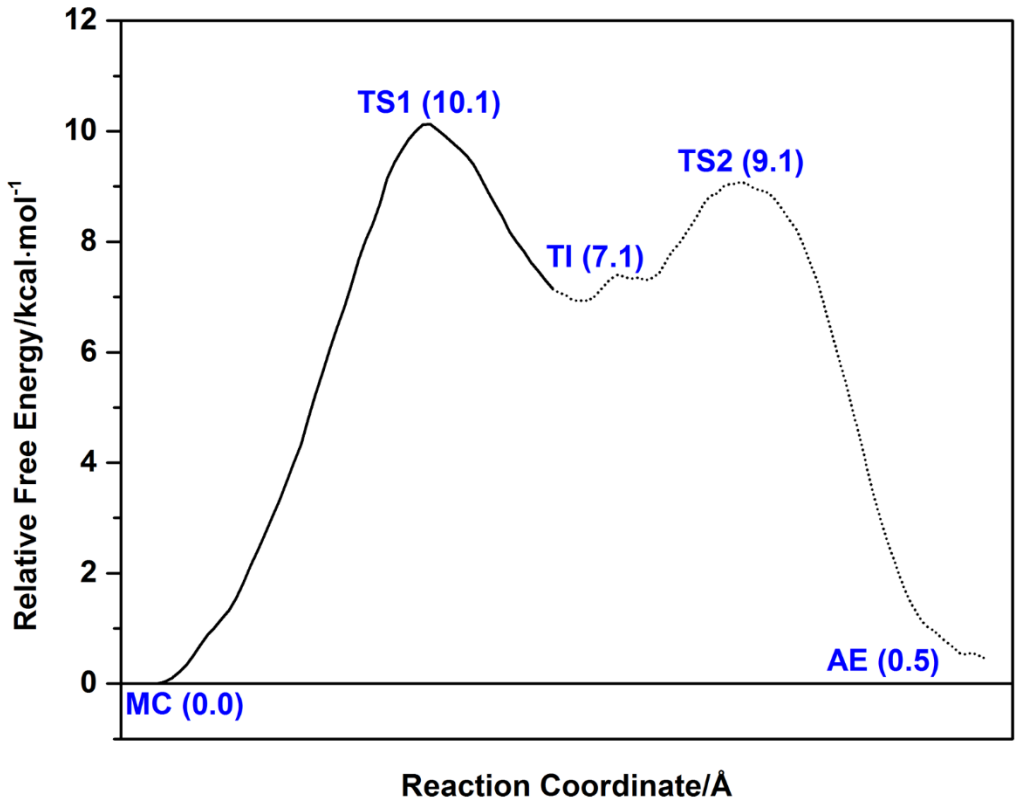


Figure 5. 1D free energy profiles for the TI formation and the breakdown of the peptide bond relative to the MC for the reaction of the ZIKV NS2B/NS3 serine protease with its substrate using the combined RCs at the PM6/ff14SB level of QM/MM theory.

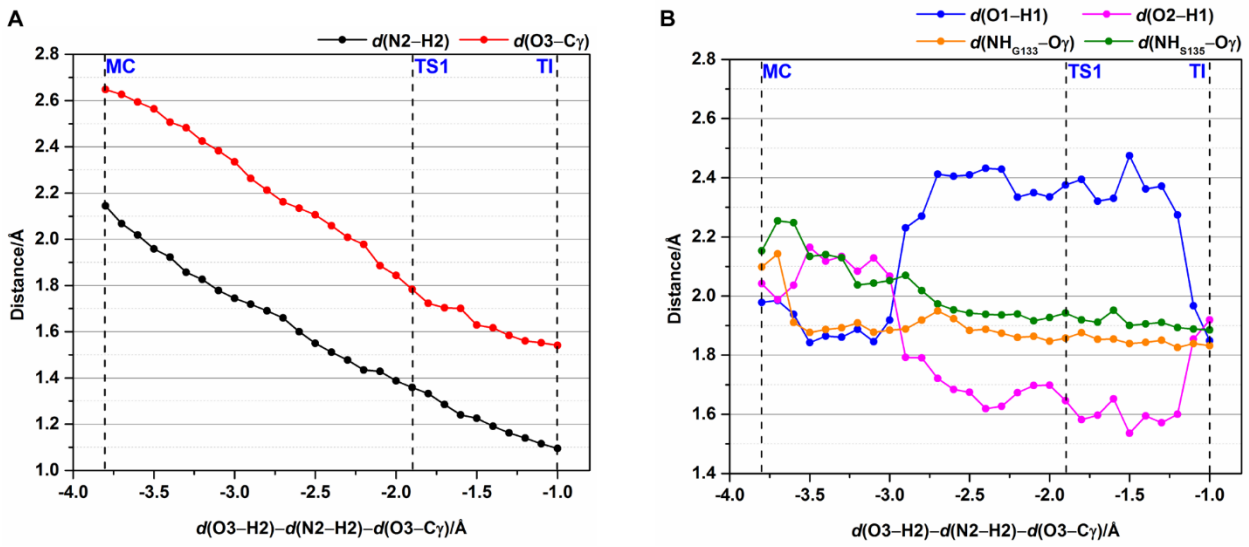


Figure 6. Average values of (**A**) the distances involving proton transfer and nucleophilic attack and (**B**) the hydrogen bond distances between QM and MM atoms along the reaction pathway.

Table 2. Average Mulliken charges (PM6/ff14SB) in atomic units for atoms involving the formation of the TI, averaged from the QM/MM free energy profile.

Relevant atom	Mulliken charge		
	MC	TS1	TI
Nγ	–0.45	–0.56	–0.61
Oγ	–0.69	–0.82	–0.93
Cγ	0.61	0.72	0.74
N2	–0.40	–0.34	–0.18
H2	0.35	0.41	0.38
O3	–0.60	–0.64	–0.63

The free energy profile for the second step of the acylation process connecting the TI to the AE product is shown in **Figures 4B** and **5**. The simulations show that the H2 atom is transferred from $N2_{H51}$ to the substrate's Nγ, and thereafter the peptide bond breaks. The calculated distances along the reaction pathway evidently show that *d*(Nγ–H2) continuously decreases indicating the progression of the proton transfer to Nγ, while *d*(Cγ–Nγ) involving peptide bond breakdown

slightly increases up to ~1.6 Å when the AE product is formed (**Figure 7**). The energy difference between TI and TS2 is considered as the energy barrier for this step and is only 2.0 kcal mol^{-1}. This value is lower than that of the barrier height observed in the TI formation, such that the TI formation is the rate-limiting step for the acylation process.

During the peptide bond breaking process, the bond between $O3_{H51}$ and Cγ of the substrate becomes stronger, which is shown by the acyl–enzyme bond, *d*(O3–Cγ), approaching its lowest value of 1.41 Å upon reaching the AE product (**Figure 7**). Furthermore, as the reaction progresses, hydrogen bonds between H51 and D75 show the similar situation as found in the first step of acylation involving the rotation of the imidazole ring of H51 to facilitate the proton transfer from $N2_{H51}$ to the Nγ nitrogen on the substrate. The calculated distances demonstrate that *d*(O1–H1) increases up to ~2.4 Å, whereas *d*(O2–H1) decreases from ~2.1 to ~1.5 Å. In this event, the position of H2 attached to $N2_{H51}$ atom is in close proximity to the Nγ nitrogen of substrate, allowing easy transfer the proton. The other two hydrogen bonds formed in the oxyanion hole are maintained along the reaction pathway, shown by the *d*(NH_{G133}–Oγ) and *d*(NH_{S135}–Oγ) distances of around 1.9 Å. This indicates that the AE formation is still stabilized by both hydrogen bonds. It is worth noting that structural changes are also found in the carbonyl group on the acyl-portion. As the reaction progresses, the carbonyl bond length becomes shorter in AE with the similar value to that found in MC complex (~1.24 Å), reflecting the conversion of Cγ hybridization from a sp^3 (the carbonyl bond length of ~1.29 Å at the TI stationary point) to a sp^2 hybridization upon AE product.

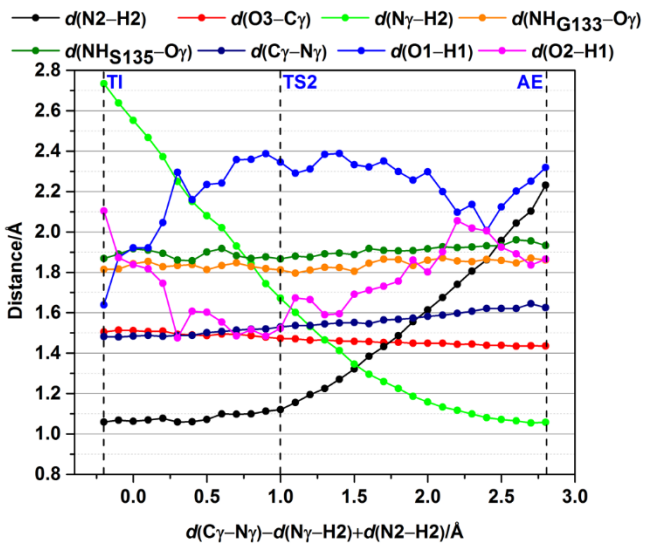


Figure 7. Average values of the distances along the reaction pathway involving the second step of acylation process. Interatomic distances are shown in different colors, as indicated.

Potential Energy Surface for Reaction and Effect of Basis Set Size

To examine the influence of the basis set on the energy profiles for the first step of the acylation process, single point QM/MM energy calculations were performed on the BH&HLYP-D3/6-31G(d)/ff14SB optimized geometries (see below) using a variety of polarization and diffuse functions, as well as a larger basis set for the DFT method (6-311++G(d,p)) (**Figure 8**). The results show that the energy profiles are very similar in terms of shape, reaction energy barriers and energies of the TI formation, except for 6-31G. The BH&HLYP-D3 energy profiles calculated with different basis sets, including the 6-31G(d), 6-31G(d,p), 6-31+G(d), 6-31++G(d,p) and 6-311++G(d,p) are all similar, with energy differences within 1.5 kcal mol^{-1} at transition states (TSs). The addition of polarizable functions (i.e., 6-31G(d) and 6-31G(d,p)) is clearly important for correct description of reaction energetics, changing the energies of the TS and TI by ~9 kcal mol^{-1}, compared to the calculation with 6-31G basis set. However, additional inclusion of diffuse functions (i.e., 6-31+G(d) and 6-31++G(d,p)) does not significantly change the calculated

energies. Thus, the inclusion of only polarization functions is sufficient for giving the reasonable BH&HLYP-D3 energy profiles in this case; hence, the 6-31G(d) basis set provides a good choice for geometry optimization, balancing accuracy and computational cost.

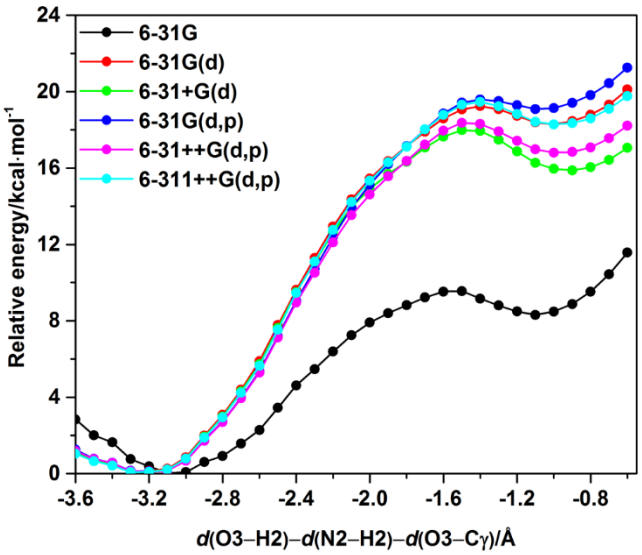


Figure 8. QM/MM energy profiles for the first step of the acylation process along the reaction pathway obtained at BH&HLYP-D3 QM/MM energy calculations using different basis sets for snapshot 1.

Conformational Sampling and *Ab Initio* Energy Calculations

Using TS-like conformations from the 1D PMF profile as starting points, potential energy profiles were calculated with an adiabatic mapping procedure along the RC at the BH&HLYP-D3/6-31G(d)/ff14SB level of theory. Afterward, *ab initio* QM/MM (LMP2/(aug)-cc-pVTZ/ff14SB, SCS-LMP2/(aug)-cc-pVTZ/ff14SB and LCCSD(T)/(aug)-cc-pVTZ/ff14SB) single point energy calculations were used to calculate higher-level potential energy surfaces (PESs).

As can be seen from the PESs, there are two local energy minima corresponding to the MC complex and the TI with a single approximate TS along the MEP at all levels of QM/MM theory (**Figure 9**). This indicates that this step of the acylation occurs in a concerted manner, as also found

in the PM6/MM free energy profiles. To account for conformational sampling, the results were averaged over multiple pathways using five different starting structures. Based on the QM/MM optimization at the BH&HLYP-D3/6-31G(d)/ff14SB level (**Figure 9A**), the average activation energy barrier is 19.3 kcal mol^{-1} (**Table 3**). The location of energy barrier is shifted to the product side at the RC of ~–1.5 Å at higher level QM/MM calculations in comparison with the location of barrier obtained at the PM6/ff14SB level (~–1.8 Å). All structures for each stationary point (i.e., MC, TS and TI structures) were superimposed on top of the QM/MM structures calculated at the BH&HLYP-D3/6-31G(d)/ff14SB level to verify the consistency of the five calculated pathways (**Figure S6**). All five of the optimized MC structures show similar configurations of the substrate and catalytic and important residues in the active site of enzyme. As the reaction proceeds, the transfer of proton from S135 to H51 and the nucleophilic attack of S135 oxygen to the carbonyl carbon on substrate occur simultaneously (see the detail in structural analysis section below). The TS and TI geometries are very similar in structure (**Figure S6**).

Single point energy calculations were performed on the geometries optimized with BH&HLYP-D3/6-31G(d)/ff14SB level of theory along the reaction path using LMP2, SCS-LMP2, and LCCSD(T) methods with the Dunning correlation consistent (aug)-cc-pVTZ basis set for the QM region. The results obtained for the five different initial structures with the correlated *ab initio* methods are shown in **Figure 9B-D** and **Table 3**. From conventional transition state theory (TST),[108] the apparent activation free energy barrier from the experiment is ~18.2 kcal mol^{-1} (converted from the value of k_{cat} = 0.95 s^{-1} at 37 °C).[98] Note that the barriers here are activation potential energy that cannot be compared directly to a free energy barrier; this would require the calculation of the entropic, tunneling and zero-point energy contributions.[109] The most accurate calculations here, from first principles, are at the LCCSD(T)/(aug)-cc-pVTZ level (barrier height

of ~16.3 kcal mol^{-1}), which is also in good agreement with experiment (from a rough estimate of these additional contributions) and we use it as a reference for the comparisons. The average barriers from LMP2/(aug)cc-pVTZ and SCS-LMP2/(aug)cc-pVTZ are close (within 1.5 kcal mol^{-1}) to the LCCSD(T)/(aug)-cc-pVTZ/ff14SB results. The LMP2 result underestimates the average activation energy barrier compared to LCCSD(T)/(aug)-cc-pVTZ level by only 0.7 kcal mol^{-1}, while SCS-LMP2 and BH&HLYP-D3 overestimate the barrier by 1.5 and 3.0 kcal mol^{-1}, respectively. Meanwhile, the relative energy of the TI shows an obviously higher reaction energy than the MC for all levels of QM/MM calculations, indicating that formation of the TI is a relatively unstable species compared to the MC, and the reaction is endothermic. The TI has average energies relative to the MC of 18.2 ± 1.8 kcal mol^{-1}, 14.2 ± 1.7 kcal mol^{-1} and 16.1 ± 1.8 kcal mol^{-1} with BH&HLYP-D3/6-31G(d), LMP2/(aug)cc-pVTZ and SCS-LMP2/(aug)cc-pVTZ, respectively. However, these reaction energies are larger than that of the energy obtained at the LCCSD(T)/(aug)-cc-pVTZ level (reaction energy of 13.0 ± 1.7 kcal mol^{-1}). In addition, the energy barriers found in each snapshot linearly correlate with the reaction energies (i.e. the energy of the TI relative to the MC) at all levels of QM/MM theory presented here (see **Figure S7**). We suggest here that higher-level QM calculations provide significantly less deviation of the observed activation energy barrier and reaction energy from the LCCSD(T)/(aug)-cc-pVTZ energy description of the reaction. Even though the value for the BH&HLYP-D3/6-31G(d) average reaction barrier is higher than those of the LMP2 and SCS-LMP2 methods, it is somewhat close to the SCS-LMP2 value indicating that the BH&HLYP functional in combination with dispersion correction performs well for this particular reaction. It should be noted that dispersion-corrected density functional theory (DFT-D3) is very efficient to treat general intermolecular interactions at a reasonably accurate level and has emerged as a suitable approach for modeling the electronic

structure of large molecular systems.[110] Moreover, BH&HLYP has been shown previously to give better results than the widely used B3LYP for some proton and charge transfers[70] and hydrogen bonding.[71, 73] Nevertheless, the calculation at the B3LYP-D3/6-31G(d)/ff14SB level of theory should be tested, as we do here. We tested by selecting three from five different starting structures (snapshots 1-3) to perform adiabatic mapping at this level of theory. The computed potential energy profiles show that the energy barrier and the reaction energy are not predicable (see **Figure S8**), suggesting that the calculation at the B3LYP-D3/6-31G(d)/ff14SB level fails to describe and locate the TI minimum. Additionally, the QM/MM calculations at the MP2 and SCS-MP2 levels with and without local approximations were tested to examine the effect of local approximations on their accuracy and so clarify their use at the coupled cluster level. Calculations at the (L)MP2 and SCS-(L)MP2 levels of theory with and without local approximations show that the local approximations do not significantly affect the results, with the largest difference around only ~0.5 kcal mol^{-1} (see **Figure S9**). It can be inferred from this that local approximations introduce only very small errors at the LCCSD(T) level, indicating that the use of local approximations is an effective way to reduce computational cost to achieve coupled cluster accuracy.

The negative charge on the TI formation is produced from the substrate carbonyl oxygen atom. This high-energy species is stabilized by proton donating groups of amino acid backbone nearby, termed as the oxyanion hole which, in the case of flavivirus NS2B/NS3 protease, is formed by the backbone amide groups of G133 and S135 of the protease. It has been known that the most important effect of stability on the TI is the electrostatic interactions between the oxyanion hole region and the negatively charged carbonyl oxygen of the scissile substrate. The contributions of the oxyanion hole to electrostatic stabilization were evaluated at the SCS-LMP2/(aug)cc-pVTZ/ff1SB level. The energy contribution of the oxyanion hole region was calculated by

subtracting the energy of the full QM region from the QM region without the residues G133, T134 and main chain NH of S135. The oxyanion hole region stabilizes both the TS1 and TI (on average by ~1.5 kcal mol^{-1} and ~7.0 kcal mol^{-1} for TS1 and TI, respectively), as shown in **SI** (**Figure S10**). The residues of the oxyanion hole stabilize the TI significantly more than TS1, because charge is more localized in the TI than in the TS1, and stabilizes both relative to the neutral components in the MC complex. Besides, preliminarily tests of the influence of enlarging the minimal QM region (see **Computational Methods**) by an additional inclusion of the oxyanion hole region, QM/MM geometry optimizations on the snapshot 1 at the BH&HLYP-D3/6-31G(d)/ff14SB level were carried out by including fragments of G133, T134 and S135 in the QM region, leading to a total of 79 atoms. In the energy profiles calculated with two different sizes of the QM region (with and without residue fragments of oxyanion hole in the QM region), the TS1 and TI formation are found at almost the same values of the RC, as shown in **SI** (**Figure S11**). Furthermore, as expected, the computed potential energy barrier and the reaction energy were lowered by 1.8 kcal mol^{-1} and 2.5 kcal mol^{-1}, respectively for the larger QM subsystem, compared with the minimal QM subsystem. This is possibly due to the fact that the hydrogen bonds between the backbone amides of residues G133 and S135 and the carbonyl oxygen of the scissile substrate are stronger when these residues are treated quantum mechanically. However, it can be seen that addition of the residues located in the oxyanion hole that stabilize the TS1 and TI formation in the QM region does not significantly affect the activation and reaction energies, compared to the minimal QM subsystem. In particular, the difference in activation energy computed with both choices of the size of the QM region is relatively small (1.8 kcal mol^{-1}), indicating that an MM treatment of the oxyanion hole region is still acceptable in terms of giving the similar barrier shapes and reducing computational expenses.

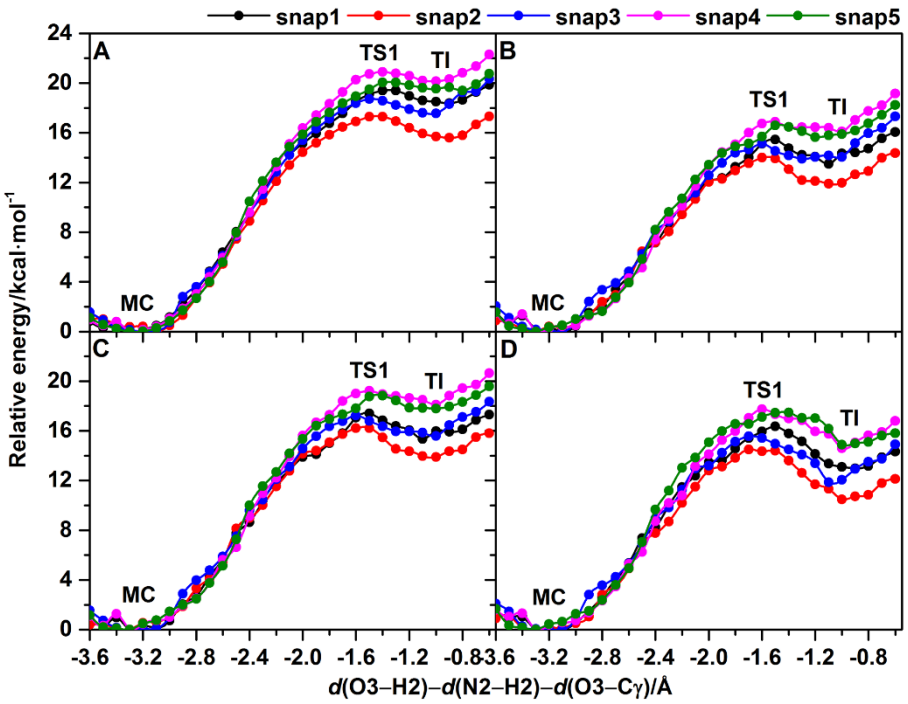


Figure 9. Potential energy profiles for the first step of the acylation process calculated at the (**A**) BH&HLYP-D3/6-31G(d)/ff14SB, (**B**) LMP2/(aug)-cc-pVTZ/ff14SB, (**C**) SCS-LMP2/(aug)-cc-pVTZ/ff14SB, and (**D**) LCCSD(T)/(aug)-cc-pVTZ/ff14SB QM/MM levels. All geometries were optimized at the BH&HLYP-D3/6-31G(d)/ff144SB level.

Table 3. Energy barriers (kcal mol^{-1}) for the TI formation calculated with the BH&HLYP-D3/6-31G(d), LMP2/(aug)-cc-pVTZ, SCS-LMP2/(aug)-cc-pVTZ, and LCCSD(T)/(aug)-cc-pVTZ QM/MM methods[a]

calculation method	**activation energy**					**average**[c]
	snap 1[b]	snap 2	snap 3	snap 4	snap 5	
BH&HLYP-D3	19.4	17.3	18.7	20.9	20.0	19.3 ± 1.4
LMP2	15.5	13.4	15.1	16.9	16.6	15.6 ± 1.2
SCS-LMP2	17.4	16.2	17.2	19.2	18.8	17.8 ± 1.2
LCCSD(T)	16.4	14.5	15.6	17.7	17.5	16.3 ± 1.3

[a] The L in these acronyms indicates that the calculation of local approximations was included in the ab initio methods and (aug) indicates that a basis set augmented with diffuse functions were only applied for the oxygen atoms. [b]The conformational details of MC, TS and TI structures are

presented in Figure 10. [c]The average value is calculated through the simple arithmetic mean of 5 data points.

Structural Analysis and Hints for Designing New Inhibitors

The geometries of the MC, TS and TI in the first step of the acylation process obtained from the representative snapshot 1 of the combined interatomic distances are depicted in **Figure 10**. Again, it is noticeable that the two RC distances involving the formation of TI simultaneously decrease (d(O3–H2)–d(N2–H2): –0.6 to 0.6 Å; and d(O3–Cγ): 2.5 to 1.5 Å) during the reaction, reflecting a concerted reaction mechanism. Moreover, the main chain NH groups of G133 and S135 located in the oxyanion hole retain strong hydrogen bonding interactions with the carbonyl oxygen of the scissile bond and tend to exhibit the stronger electrostatic stabilization when the TI is formed (**Figure S10**). Interestingly, the step of proton transfer from S135 to H51 shows a complete proton transfer between the nitrogen and oxygen atoms at the transition state (TS1), as indicated by the decrease in d(N2–H2) from 1.6 Å (MC) to 1.0 Å (TS1). At the same time, the oxygen nucleophilic attack on the carbon of the substrate's carbonyl group occurs in concert with the proton transfer step, in which the nucleophilic attack distance, d(O3–Cγ), shortens in the order of 2.5 Å (MC), 2.0 Å (TS1) and 1.5 Å (TI). Thus, the reaction mechanism for the TI formation observed in common serine proteases (e.g. trypsin and chymotrypsin) and here in ZIKV NS2B/NS3, is consistent with the generally accepted reaction mechanism. Nevertheless, the acyl enzyme hydrolysis or deacylation process is typically rate-limiting for the hydrolysis of certain ester substrates by several serine proteases such as chymotrypsin,[111] trypsin,[112] subtilisin,[113] and elastase.[114] Likewise, a kinetic study on DENV type 4 protease with its substrates revealed that the deacylation is rate-limiting for ester bond hydrolysis, whereas the acylation is the rate-determining step for amide bond hydrolysis.[115] From this point of view, it can be postulated that the acylation

reaction is the rate-limiting step for substrate amide bond hydrolysis by ZIKV protease and thus the acyl enzyme hydrolysis has not been considered in this study.

In this present study, we have succeeded in detecting the TI, which has a very short lifetime; a very hard task experimentally. The determination of the reaction mechanism of ZIKV protease here could help to design new inhibitors based on TS analogues.[116] Moreover, it seems that the stability of the TI can be a crucial factor in biocatalysis of several enzymes. Therefore, compounds which effectively mimic such an intermediate are considered as good candidates to become potential inhibitors of the targeted enzymes. It is worth noting that tetrahedral intermediates occur in many enzymes, particularly proteases and metallo-enzyme catalyzed reactions involved in hydrolysis of ester, amide or other acyl bonds. For instance, to resemble the tetrahedral TS of serine proteases, reversible covalent inhibitors able to form hemiketal formation have been introduced.[117] This attempt has led to two FDA approved drugs, telaprevir and boceprevir, which are covalent inhibitors against hepatitis C virus (HCV) protease through an α-ketoamide warhead.[118] Another example is proteasome inhibitor bortezomib that imitates the tetrahedral TS via boronic acid warhead and binds covalently to a threonine residue in the proteasome catalytic site. This inhibitor has been approved by FDA and used for treating multiple myeloma.[119] One more example is from the discovery and development of β-lactamase inhibitors. These inhibitors mimic tetrahedral TS catalyzed by β-lactamase and inactivate enzyme activity.[120] Similar to other serine proteases, β-lactamases catalyze an amide bond hydrolysis, and their catalytic mechanisms involve the occurrence of tetrahedral intermediates along the reaction pathway.[121] Currently clinically approved β-lactamase inhibitors are clavulanate, tazobactam, sulbactam, avibactam and vaborbactam.[122] In the case of ZIKV protease, potential inhibitors could form a covalent bond through their warheads with the catalytic residue S135 and mimic the TI structure. Development

of peptidomimetic inhibitors has been successful, and some of them derived from flavivirus protease substrates have been shown to be active toward both WNV and DENV proteases.[123-125] In addition, some of the available inhibitors against WNV and DENV proteases have also been shown to be active toward ZIKV protease.[16, 84] This is due to the high structural similarities among DENV, WNV, and ZIKV proteases.[126,19, 23] Structures of ZIKV NS2B/NS3 serine protease complexed with several inhibitors are available, which assist antiviral inhibitor design.[16, 17, 24] For example, the dipeptidic inhibitor acyl-KR-aldehyde derived from P1 and P2 residues of the substrate is a potent competitive inhibitor, in which the aldehyde moiety forms a covalent bond with the catalytic S135 of NS3.[84] Although it shows high ligand efficiency, it is a big challenge in developing them into drug-like molecules and for use in clinical applications because of their limitations such as cell penetration and stability caused by their positive charge,[84] which has stimulated medicinal chemists to further find new inhibitors.[127, 128] Small-molecule inhibitors such as pyrazole ester derivatives identified from high-throughput screening of a small molecule library from the National Institutes of Health are much more interesting in terms of drug-likeness with small-molecular weight molecules and being potent inhibitors against both WNV and DENV proteases.[129, 130] Recently, one of the pyrazole ester derivatives, namely 5-amino-1-((4-methoxyphenyl)sulfonyl)-1*H*-pyrazol-3-yl benzoate, in complex with ZIKV protease has been revealed by crystallographic study, in which the benzoyl moiety of this compound could form a covalent bond with the catalytic residue S135.[131] These examples provide structural information for covalent inhibitor design. Simulations such those presented here can help guide design[132] of efficient protease inhibitors against ZIKV and other flaviviruses.

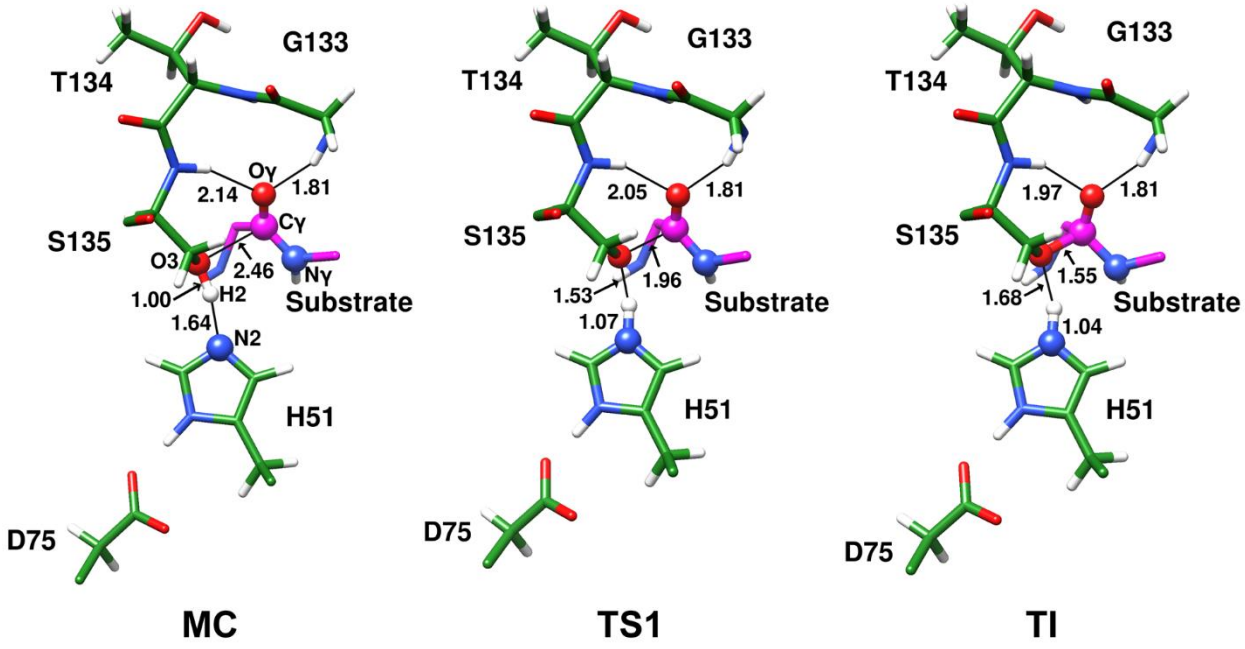


Figure 10. QM/MM (BH&HLYP-D3/6-31G(d)/ff14SB) optimized structures of the Michaelis complex (MC), transition state (TS1), and tetrahedral intermediate (TI) in the active site of ZIKV NS2B/NS3 protease obtained from the combined interatomic distances (in Å) for snapshot 1.

Conclusions

In this present study, QM/MM methods have elucidated the reaction mechanism of the acylation process of ZIKV protease and its substrate cleavage reaction. Although the reaction mechanism of common serine proteases has been widely studied, understanding the mechanistic details of the reaction catalyzed by ZIKV NS2B/NS3 serine protease is still necessary to provide fundamental knowledge of catalysis and may also help in the design and development of potential antiviral agents using TS analogues.

By combining MM MD simulations, QM/MM MD free-energy simulations and QM/MM adiabatic mapping, we investigated ZIKV protease in terms of both dynamical and mechanistic

properties. Our results show that the mechanism of the acylation reaction catalyzed by ZIKV protease behaves according to the proposed reaction mechanism for general serine proteases. Only one TS is determined along the reaction pathway suggesting a concerted reaction mechanism, in which the transfer of proton from S135 to H51 takes place in synchrony with the nucleophilic attack on the substrate. The formation of the TI is the rate-determining step of the acylation process. The calculations reveal that the PM6/ff14SB level underestimates the activation free energy in comparison with experimentally determined apparent catalytic rate, but even so the barrier shapes have been found to be similar to those of other serine proteases.[30, 100-104, 133] Thus, we have performed multiple fully optimized potential energy profiles for the first step of the acylation, using the correlated *ab initio* QM/MM methods to obtain accurate energetics. We have optimized geometries with BH&HLYP-D3/6-31G(d)/ff14SB level of theory. Similar to the free energy profiles, the formation of TI is concluded to occur through a concerted manner that involves a single transition state (TS1), where a stable species of TI has been found at all levels of theory. The potential energy barriers are 16.3 kcal mol^{-1} at the LCCSD(T)/(aug)-cc-pVTZ/ff14SB//BH&HLYP-D3/6-31G(d)/ff14SB level of theory. DFT and *ab initio* results presented here are reasonably close to the coupled cluster energies. The LMP2 result slightly underestimates the barrier compared to the BH&HLYP-D3 and SCS-LMP2 results. However, these results are in good agreement with the apparent experimental free energy of the activation of ~18.2 kcal mol^{-1} (at 310 K). It is indeed of significance that potential energy barriers here cannot directly be compared to activation free energies derived from experimental kinetics data. This is due to the fact that the activation free energies include additional effects such as entropy and quantum tunneling.[55]

By comparisons with the most accurate (LCCSD(T)/(aug)-cc-pVTZ/ff14SB) results, although the BH&HLYP-D3/6-31G(d) result gives a good description, the *ab initio* LMP2/(aug)-cc-pVTZ and SCS-LMP2/(aug)-cc-pVTZ methods give results much more similar to the LCCSD(T) result. We suggest that it is necessary to use higher-level QM methods for obtaining accurate energetics (precisely at the coupled cluster level, or the more practicable, SCS-LMP2) for ZIKV protease and for other related enzymes having the similar reaction mechanism. We hope that this computational information can be a good starting point to design and develop novel inhibitors toward ZIKV protease and other flavivirus proteases based on TS analogues.

Supporting Information Available

Time evolution of the RMSD; time evolution of the radius of gyration; superposition of the representative structures for each replica; geometric parameters describing the nucleophilic attack trajectory; 1D free energy profiles for the tetrahedral intermediate formation using one reaction coordinate; superposition of the QM/MM minimized structures of the five MD snapshots; correlation between energy barriers and reaction energies found in each snapshot; potential energy profiles calculated at the B3LYP-D3/6-31G(d)/ff14SB level of theory; comparison of the average reaction profiles calculated with MP2 and SCS-MP2 with and without local approximations; comparison of the average reaction profiles with full QM region and without the oxyanion hole region; potential energy profiles with two different QM subsystems; Cartesian coordinates of snapshot 1 corresponding to MC, TS1 and TI structures.

Acknowledgements

This research was supported by the Rachadaphiseksomphot Endowment Fund, Chulalongkorn University (CU). B.N. is thankful for the Royal Golden Jubilee PhD Program (PHD/0020/2558) for Ph.D. scholarship, and the 90th Anniversary of CU (Rachadaphiseksomphot Endowment Fund) for research funding. We also thank the Structural and Computational Biology Research Group, Special Task Force for Activating Research (STAR), Faculty of Science, CU, and the Research Chair Grant, the National Science and Technology Development Agency (NSTDA), Thailand. AJM thanks EPSRC for support (grant number EP/M022609/1).

References

1. Dick, G. W. A.; Kitchen, S. F.; Haddow, A. J. Zika Virus (I). Isolations and Serological Specificity. *Trans. R. Soc. Trop. Med. Hyg.* **1952,** *46*, 509-520.
2. Faye, O.; Freire, C. C. M.; Iamarino, A.; Faye, O.; de Oliveira, J. V. C.; Diallo, M.; Zanotto, P. M. A.; Sall, A. A. Molecular Evolution of Zika Virus during Its Emergence in the 20th Century. *PLoS Neglected Trop. Dis.* **2014,** *8*, e2636.
3. Ledermann, J. P.; Guillaumot, L.; Yug, L.; Saweyog, S. C.; Tided, M.; Machieng, P.; Pretrick, M.; Marfel, M.; Griggs, A.; Bel, M.; et al. Aedes hensilli as a Potential Vector of Chikungunya and Zika Viruses. *PLoS Neglected Trop. Dis.* **2014,** *8*, e3188.
4. Fernandez-Garcia, M. D.; Mazzon, M.; Jacobs, M.; Amara, A. Pathogenesis of Flavivirus Infections: Using and Abusing the Host Cell. *Cell Host Microbe* **2009,** *5*, 318-328.
5. Wong, P.-S. J.; Li, M.-z. I.; Chong, C.-S.; Ng, L.-C.; Tan, C.-H. Aedes (Stegomyia) albopictus (Skuse): A Potential Vector of Zika Virus in Singapore. *PLoS Neglected Trop. Dis.* **2013,** *7*, e2348.
6. Bearcroft, W. G. C. Zika Virus Infection Experimentally Induced in a Human Volunteer. *Trans. R. Soc. Trop. Med. Hyg.* **1956,** *50*, 442-448.
7. Petersen, E.; Wilson, M. E.; Touch, S.; McCloskey, B.; Mwaba, P.; Bates, M.; Dar, O.; Mattes, F.; Kidd, M.; Ippolito, G.; et al. Rapid Spread of Zika Virus in The Americas - Implications for Public Health Preparedness for Mass Gatherings at the 2016 Brazil Olympic Games. *Int. J. Infect. Dis.* **2016,** *44*, 11-15.
8. Petersen, L. R.; Jamieson, D. J.; Powers, A. M.; Honein, M. A. Zika Virus. *N. Engl. J. Med.* **2016,** *374*, 1552-1563.

9. Mlakar, J.; Korva, M.; Tul, N.; Popović, M.; Poljšak-Prijatelj, M.; Mraz, J.; Kolenc, M.; Resman Rus, K.; Vesnaver Vipotnik, T.; Fabjan Vodušek, V.; et al. Zika Virus Associated with Microcephaly. *N. Engl. J. Med.* **2016,** *374*, 951-958.
10. Cugola, F. R.; Fernandes, I. R.; Russo, F. B.; Freitas, B. C.; Dias, J. L. M.; Guimarães, K. P.; Benazzato, C.; Almeida, N.; Pignatari, G. C.; Romero, S.; et al. The Brazilian Zika Virus Strain Causes Birth Defects in Experimental Models. *Nature* **2016,** *534*, 267-271.
11. Cao-Lormeau, V. M.; Blake, A.; Mons, S.; Lastère, S.; Roche, C.; Vanhomwegen, J.; Dub, T.; Baudouin, L.; Teissier, A.; Larre, P.; et al. Guillain-Barré Syndrome Outbreak Associated with Zika Virus Infection in French Polynesia: A Case-Control Study. *Lancet* **2016,** *387*, 1531-1539.
12. Weaver, S. C.; Costa, F.; Garcia-Blanco, M. A.; Ko, A. I.; Ribeiro, G. S.; Saade, G.; Shi, P.-Y.; Vasilakis, N. Zika virus: History, Emergence, Biology, and Prospects for Control. *Antiviral Res.* **2016,** *130*, 69-80.
13. Carteaux, G.; Maquart, M.; Bedet, A.; Contou, D.; Brugières, P.; Fourati, S.; Cleret de Langavant, L.; de Broucker, T.; Brun-Buisson, C.; Leparc-Goffart, I.; et al. Zika Virus Associated with Meningoencephalitis. *N. Engl. J. Med.* **2016,** *374*, 1595-1596.
14. Driggers, R. W.; Ho, C.-Y.; Korhonen, E. M.; Kuivanen, S.; Jääskeläinen, A. J.; Smura, T.; Rosenberg, A.; Hill, D. A.; DeBiasi, R. L.; Vezina, G.; et al. Zika Virus Infection with Prolonged Maternal Viremia and Fetal Brain Abnormalities. *N. Engl. J. Med.* **2016,** *374*, 2142-2151.
15. Zhang, Y.; Chen, W.; Wong, G.; Bi, Y.; Yan, J.; Sun, Y.; Chen, E.; Yan, H.; Lou, X.; Mao, H.; et al. Highly Diversified Zika Viruses Imported to China, 2016. *Protein Cell* **2016,** *7*, 461-464.
16. Lei, J.; Hansen, G.; Nitsche, C.; Klein, C. D.; Zhang, L.; Hilgenfeld, R. Crystal Structure of Zika Virus NS2B-NS3 Protease in Complex with a Boronate Inhibitor. *Science* **2016,** *353*, 503-505.
17. Phoo, W. W.; Li, Y.; Zhang, Z.; Lee, M. Y.; Loh, Y. R.; Tan, Y. B.; Ng, E. Y.; Lescar, J.; Kang, C.; Luo, D. Structure of the NS2B-NS3 Protease from Zika Virus After Self-Cleavage. *Nat. Commun.* **2016,** *7*, 13410.
18. Leung, D.; Schroder, K.; White, H.; Fang, N. X.; Stoermer, M. J.; Abbenante, G.; Martin, J. L.; Young, P. R.; Fairlie, D. P. Activity of Recombinant Dengue 2 Virus NS3 Protease in the Presence of a Truncated NS2B Co-Factor, Small Peptide Substrates, and Inhibitors. *J. Biol. Chem.* **2001,** *276*, 45762-45771.
19. Erbel, P.; Schiering, N.; D'Arcy, A.; Renatus, M.; Kroemer, M.; Lim, S. P.; Yin, Z.; Keller, T. H.; Vasudevan, S. G.; Hommel, U. Structural Basis for the Activation of Flaviviral NS3 Proteases from Dengue and West Nile Virus. *Nat. Struct. Mol. Biol.* **2006,** *13*, 372-373.
20. Chappell, K. J.; Stoermer, M. J.; Fairlie, D. P.; Young, P. R. Mutagenesis of the West Nile Virus NS2B Cofactor Domain Reveals Two Regions Essential for Protease Activity. *J. Gen. Virol.* **2008,** *89*, 1010-1014.

21. Skoreński, M.; Grzywa, R.; Sieńczyk, M. Why should We Target Viral Serine Proteases when Developing Antiviral Agents? *Future Virol.* **2016,** *11*, 745-748.
22. Patick, A. K.; Potts, K. E. Protease Inhibitors as Antiviral Agents. *Clin. Microbiol. Rev.* **1998,** *11*, 614-627.
23. Chen, X.; Yang, K.; Wu, C.; Chen, C.; Hu, C.; Buzovetsky, O.; Wang, Z.; Ji, X.; Xiong, Y.; Yang, H. Mechanisms of Activation and Inhibition of Zika Virus NS2B-NS3 Protease. *Cell Res.* **2016,** *26*, 1260-1263.
24. Zhang, Z.; Li, Y.; Loh, Y. R.; Phoo, W. W.; Hung, A. W.; Kang, C.; Luo, D. Crystal Structure of Unlinked NS2B-NS3 Protease from Zika Virus. *Science* **2016,** *354*, 1597-1600.
25. Shiryaev, S. A.; Farhy, C.; Pinto, A.; Huang, C.-T.; Simonetti, N.; Ngono, A. E.; Dewing, A.; Shresta, S.; Pinkerton, A. B.; Cieplak, P.; et al. Characterization of the Zika Virus Two-Component NS2B-NS3 Protease and Structure-Assisted Identification of Allosteric Small-Molecule Antagonists. *Antiviral Res.* **2017,** *143*, 218-229.
26. Aleshin, A. E.; Shiryaev, S. A.; Strongin, A. Y.; Liddington, R. C. Structural Evidence for Regulation and Specificity of Flaviviral Proteases and Evolution of the Flaviviridae Fold. *Protein Sci.* **2007,** *16*, 795-806.
27. Blow, D. M.; Birktoft, J. J.; Hartley, B. S. Role of a Buried Acid Group in the Mechanism of Action of Chymotrypsin. *Nature* **1969,** *221*, 337.
28. Blow, D. M. Structure and Mechanism of Chymotrypsin. *Acc. Chem. Res.* **1976,** *9*, 145-152.
29. Warshel, A.; Russell, S. Theoretical Correlation of Structure and Energetics in the Catalytic Reaction of Trypsin. *J. Am. Chem. Soc.* **1986,** *108*, 6569-6579.
30. Rodríguez, A.; Oliva, C.; González, M.; van der Kamp, M.; Mulholland, A. J. Comparison of Different Quantum Mechanical/Molecular Mechanics Boundary Treatments in the Reaction of the Hepatitis C Virus NS3 Protease with the NS5A/5B Substrate. *J. Phys. Chem. B* **2007,** *111*, 12909-12915.
31. Zhou, Y.; Zhang, Y. Serine Protease Acylation Proceeds with a Subtle Re-Orientation of the Histidine Ring at the Tetrahedral Intermediate. *Chem. Commun.* **2011,** *47*, 1577-1579.
32. Zerner, B.; Bender, M. L. The Kinetic Consequences of the Acyl-Enzyme Mechanism for the Reactions of Specific Substrates with Chymotrypsin. *J. Am. Chem. Soc.* **1964,** *86*, 3669-3674.
33. Hedstrom, L.; Szilagyi, L.; Rutter, W. Converting Trypsin to Chymotrypsin: the Role of Surface Loops. *Science* **1992,** *255*, 1249-1253.
34. Harris, J. L.; Peterson, E. P.; Hudig, D.; Thornberry, N. A.; Craik, C. S. Definition and Redesign of the Extended Substrate Specificity of Granzyme B. *J. Biol. Chem.* **1998,** *273*, 27364-27373.
35. Perona, J. J.; Craik, C. S. Structural Basis of Substrate Specificity in the Serine Proteases. *Protein Sci.* **1995,** *4*, 337-360.
36. Polgár, L. The Catalytic Triad of Serine Peptidases. *Cell. Mol. Life Sci.* **2005,** *62*, 2161-2172.
37. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002,** *102*, 4501-4523.

38. Werner, H. J.; Schütz, M. An Efficient Local Coupled Cluster Method for Accurate Thermochemistry of Large Systems. *J. Chem. Phys.* **2011,** *135*.
39. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating pKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005,** *33*, W368-W371.
40. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
41. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.
42. Åqvist, J. Ion-water Interaction Potentials Derived from Free Energy Perturbation simulations. *J. Phys. Chem.* **1990,** *94*, 8021-8024.
43. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. *AMBER 16*; University of California, San Francisco, 2016.
44. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
45. Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of Trajectories in Canonical Molecular-Dynamics Simulations: Observation, Explanation, and Exploitation. *J. Chem. Phys.* **2004,** *120*, 6363-6374.
46. Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to An External Bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.
47. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.
48. Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem. A* **2000,** *104*, 1720-1735.
49. Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990,** *11*, 700-733.
50. Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007,** *13*, 1173-1213.
51. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The method. *J. Comput. Chem.* **1992,** *13*, 1011-1021.
52. Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* **1995,** *91*, 275-282.

53. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **1995,** *16*, 1339-1350.
54. Helgaker, T.; Ruden, T. A.; Jørgensen, P.; Olsen, J.; Klopper, W. A Priori Calculation of Molecular Properties to Chemical Accuracy. *J. Phys. Org. Chem.* **2004,** *17*, 913-933.
55. Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H. J. High-Accuracy Computation of Reaction Barriers in Enzymes. *Angew. Chem.* **2006,** *45*, 6856-6859.
56. Van Der Kamp, M. W.; Zurek, J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Testing High-Level QM/MM Methods for Modeling Enzyme Reactions: Acetyl-CoA Deprotonation in Citrate Synthase. *J. Phys. Chem. B* **2010,** *114*, 11303-11314.
57. Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; et al. QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis. *J. Mol. Struct.: THEOCHEM* **2003,** *632*, 1-28.
58. Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell—A Modular Software Package for QM/MM Simulations. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014,** *4*, 101-110.
59. Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012,** *2*, 73-78.
60. Smith, W.; Forester, T. R. DL_POLY_2.0: A General-Purpose Parallel Molecular Dynamics Simulation Package. *J. Mol. Graphics Modell.* **1996,** *14*, 136-141.
61. Senn, H. M.; Thiel, W. QM/MM Studies of Enzymes. *Curr. Opin. Chem. Biol.* **2007,** *11*, 182-187.
62. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011,** *32*, 1456-1465.
63. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Inclusion of Dispersion Effects Significantly Improves Accuracy of Calculated Reaction Barriers for Cytochrome P450 Catalyzed Reactions. *J. Phys. Chem. Lett.* **2010,** *1*, 3232-3237.
64. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J. Chem. Theory Comput.* **2012,** *8*, 4637-4645.
65. Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988,** *37*, 785-789.
66. Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988,** *38*, 3098-3100.
67. Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993,** *98*, 1372-1377.
68. Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.

69. Kohn, W.; Becke, A. D.; Parr, R. G. Density Functional Theory of Electronic Structure. *J. Phys. Chem.* **1996,** *100*, 12974-12980.
70. Zhao, Y.; Truhlar, D. G. Benchmark Databases for Nonbonded Interactions and Their Use to Test Density Functional Theory. *J. Chem. Theory Comput.* **2005,** *1*, 415-432.
71. Zhao, Y.; González-Garda, N.; Truhlar, D. G. Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and Its Use to Test Theoretical Methods. *J. Phys. Chem. A* **2005,** *109*, 2012-2018.
72. Dilabio, G. A.; Johnson, E. R.; Otero-De-La-Roza, A. Performance of Conventional and Dispersion-Corrected Density-Functional Theory Methods for Hydrogen Bonding Interaction Energies. *Phys. Chem. Chem. Phys.* **2013,** *15*, 12821-12828.
73. Gkionis, K.; Hill, J. G.; Oldfield, S. P.; Platts, J. A. Performance of Becke's Half-and-Half Functional for Non-Covalent Interactions: Energetics, Geometries and Electron Densities. *J. Mol. Graphics Modell.* **2009,** *15*, 1051-1060.
74. Yu, W.; Liang, L.; Lin, Z.; Ling, S.; Haranczyk, M.; Gutowski, M. Comparison of Some Representative Density Functional Theory and Wave Function Theory Methods for the Studies of Amino Acids. *J. Comput. Chem.* **2009,** *30*, 589-600.
75. Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: A General-Purpose Quantum Chemistry Program Package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012,** *2*, 242-253.
76. Grimme, S. Improved Second-Order Møller-Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003,** *118*, 9095-9102.
77. Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934,** *46*, 618-622.
78. Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989,** *157*, 479-483.
79. Lawan, N.; Ranaghan, K. E.; Manby, F. R.; Mulholland, A. J. Comparison of DFT and Ab Initio QM/MM Methods for Modelling Reaction in Chorismate Synthase. *Chem. Phys. Lett.* **2014,** *608*, 380-385.
80. Antony, J.; Grimme, S. Is Spin-Component Scaled Second-Order Møller-Plesset Perturbation Theory an Appropriate Method for the Study of Noncovalent Interactions in Molecules? *J. Phys. Chem. A* **2007,** *111*, 4862-4868.
81. Dunning Jr, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989,** *90*, 1007-1023.
82. Pulay, P. Localizability of Dynamic Electron Correlation. *Chem. Phys. Lett.* **1983,** *100*, 151-154.
83. Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for Ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions. *J. Chem. Phys.* **1989,** *90*, 4916-4926.

84. Li, Y.; Zhang, Z.; Phoo, W. W.; Loh, Y. R.; Wang, W.; Liu, S.; Chen, M. W.; Hung, A. W.; Keller, T. H.; Luo, D.; et al. Structural Dynamics of Zika Virus NS2B-NS3 Protease Binding to Dipeptide Inhibitors. *Structure* **2017,** *25*, 1242-1250.
85. Robin, G.; Chappell, K.; Stoermer, M. J.; Hu, S. H.; Young, P. R.; Fairlie, D. P.; Martin, J. L. Structure of West Nile Virus NS3 Protease: Ligand Stabilization of the Catalytic Conformation. *J. Mol. Biol.* **2009,** *385*, 1568-1577.
86. Yotmanee, P.; Rungrotmongkol, T.; Wichapong, K.; Choi, S. B.; Wahab, H. A.; Kungwan, N.; Hannongbua, S. Binding Specificity of Polypeptide Substrates in NS2B/NS3pro Serine Protease of Dengue Virus Type 2: A Molecular Dynamics Study. *J. Mol. Graphics Modell.* **2015,** *60*, 24-33.
87. Radisky, E. S.; Koshland Jr, D. E. A Clogged Gutter Mechanism for Protease Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99*, 10316-10321.
88. Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **2004,** *303*, 186-195.
89. Warshel, A., Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425-443.
90. Van Der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013,** *52*, 2708-2728.
91. Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009,** *48*, 1198-1229.
92. Jindal, G.; Warshel, A. Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region. *J. Phys. Chem. B* **2016,** *120*, 9913-9921.
93. Dhoke, G. V.; Davari, M. D.; Schwaneberg, U.; Bocola, M. QM/MM Calculations Revealing the Resting and Catalytic States in Zinc-Dependent Medium-Chain Dehydrogenases/Reductases. *ACS Catal.* **2015,** *5*, 3207-3215.
94. Korth, M.; Thiel, W. Benchmarking Semiempirical Methods for Thermochemistry, Kinetics, and Noncovalent Interactions: OMx Methods are Almost as Accurate and Robust as DFT-GGA Methods for Organic Molecules. *J. Chem. Theory Comput.* **2011,** *7*, 2929-2936.
95. Chudyk, E. I.; Limb, M. A. L.; Jones, C.; Spencer, J.; Van Der Kamp, M. W.; Mulholland, A. J. QM/MM Simulations as an Assay for Carbapenemase Activity in Class A β-lactamases. *Chem. Commun.* **2014,** *50*, 14736-14739.
96. Callegari, D.; Ranaghan, K. E.; Woods, C. J.; Minari, R.; Tiseo, M.; Mor, M.; Mulholland, A. J.; Lodola, A. L718Q Mutant EGFR Escapes Covalent Inhibition by Stabilizing a Non-Reactive Conformation of the Lung Cancer Drug Osimertinib. *Chem. Sci.* **2018,** *9*, 2740-2749.
97. Fritz, R. A.; Alzate-Morales, J. H.; Spencer, J.; Mulholland, A. J.; Van Der Kamp, M. W. Multiscale Simulations of Clavulanate Inhibition Identify the Reactive Complex in Class A β-Lactamases and Predict the Efficiency of Inhibition. *Biochemistry* **2018,** *57*, 3560-3563.

98. Kuiper, B. D.; Slater, K.; Spellmon, N.; Holcomb, J.; Medapureddy, P.; Muzzarelli, K. M.; Yang, Z.; Ovadia, R.; Amblard, F.; Kovari, I. A.; et al. Increased Activity of Unlinked Zika Virus NS2B/NS3 Protease Compared to Linked Zika Virus Protease. *Biochem. Biophys. Res. Commun.* **2017,** *492*, 668-673.
99. Korth, M. Third-generation Hydrogen-bonding Corrections for Semiempirical QM Methods and Force Fields. *J. Chem. Theory Comput.* **2010,** *6*, 3808-3816.
100. Ishida, T.; Kato, S. Theoretical Perspectives on the Reaction Mechanism of Serine Proteases: The Reaction Free energy Profiles of the Acylation Process. *J. Am. Chem. Soc.* **2003,** *125*, 12035-12048.
101. Daggett, V.; Schröder, S.; Kollman, P. Catalytic Pathway of Serine Proteases: Classical and Quantum Mechanical Calculations. *J. Am. Chem. Soc.* **1991,** *113*, 8926-8935.
102. Nemukhin, A. V.; Grigorenko, B. L.; Rogov, A. V.; Topol, I. A.; Burt, S. K. Modeling of Serine Protease Prototype Reactions with the Flexible Effective Fragment Potential Quantum Mechanical/Molecular Mechanical Method. *Theor. Chem. Acc.* **2004,** *111*, 36-48.
103. Guo, H.; Wlodawer, A.; Guo, H. A General Acid-base Mechanism for the Stabilization of a Tetrahedral Adduct in a Serine Carboxyl Peptidase: A Computational Study. *J. Am. Chem. Soc.* **2005,** *127*, 15662-15663.
104. Rodríguez, A.; Oliva, C.; González, M. A Comparative QM/MM Study of the Reaction Mechanism of the Hepatitis C Virus NS3/NS4A Protease with the Three Main Natural Substrates NS5A/5B, NS4B/5A and NS4A/4B. *Phys. Chem. Chem. Phys.* **2010,** *12*, 8001-8015.
105. Lima, M. C.; Seabra, G. M. Reaction Mechanism of the Dengue Virus Serine Protease: A QM/MM Study. *Phys. Chem. Chem. Phys.* **2016,** *18*, 30288-30296.
106. Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. Hybrid Ab Initio Quantum Mechanics/Molecular Mechanics Calculations of Free Energy Surfaces for Enzymatic Reactions: The Nucleophilic Attack in Subtilisin. *J. Phys. Chem. B* **1998,** *102*, 2293-2301.
107. Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J. K. How do Serine Proteases Really Work? *Biochemistry* **1989,** *28*, 3629-3637.
108. Laidler, K. J.; King, M. C. The Development of Transition-State Theory. *J. Phys. Chem.* **1983,** *87*, 2657-2664.
109. Ranaghan, K. E.; Morris, W. G.; Masgrau, L.; Senthilkumar, K.; Johannissen, L. O.; Scrutton, N. S.; Harvey, J. N.; Manby, F. R.; Mulholland, A. J. Ab Initio QM/MM Modeling of the Rate-Limiting Proton Transfer Step in the Deamination of Tryptamine by Aromatic Amine Dehydrogenase. *J. Phys. Chem. B* **2017,** *121*, 9785-9798.
110. Antony, J.; Grimme, S.; Liakos, D. G.; Neese, F. Protein-Ligand Interaction Energies with Dispersion Corrected Density Functional Theory and High-Level Wave Function Based Methods. *J. Phys. Chem. A* **2011,** *115*, 11210-11220.
111. Bender, M. L.; Kezdy, F. J.; Gunter, C. R. The Anatomy of an Enzymatic Catalysis. α-Chymotrypsin. *J. Am. Chem. Soc.* **1964,** *86*, 3714-3721.

112. Christensen, U.; Ipsen, H. H. Steady-State Kinetics of Plasmin- and Trypsin-Catalysed Hydrolysis of a Number of Tripeptide-p-Nitroanilides. *Biochim. Biophys. Acta, Enzymol.* **1979,** *569*, 177-183.
113. Philipp, M.; Bender, M. L. Kinetics of Subtilisin and Thiolsubtilisin. *Mol. Cell. Biochem.* **1983,** *51*, 5-32.
114. Stein, R. L. Catalysis by Human Leukocyte Elastase. Role of Secondary-Subsite Interactions. *J. Am. Chem. Soc.* **1985,** *107*, 5767-5775.
115. Li, J.; Lim, S. P.; Beer, D.; Patel, V.; Wen, D.; Tumanut, C.; Tully, D. C.; Williams, J. A.; Jiricek, J.; Priestle, J. P.; et al. Functional Profiling of Recombinant NS3 Proteases from All Four Serotypes of Dengue Virus Using Tetrapeptide and Octapeptide Substrate Libraries. *J. Biol. Chem.* **2005,** *280*, 28766-28774.
116. Schramm, V. L., Enzymatic Transition States, Transition-State Analogs, Dynamics, Thermodynamics, and Lifetimes. *Annu. Rev. Biochem.* **2011,** *80*, 703-732.
117. Bandyopadhyay, A.; Gao, J. Targeting Biomolecules with Reversible Covalent Chemistry. *Curr. Opin. Chem. Biol.* **2016,** *34*, 110-116.
118. Pan, Q.; Peppelenbosch, M. P.; Janssen, H. L. A.; de Knegt, R. J. Telaprevir/Boceprevir Era: From Bench to Bed and Back. *World J. Gastroenterol.* **2012,** *18*, 6183-6188.
119. Goldberg, A. L. Development of Proteasome Inhibitors as Research Tools and Cancer Drugs. *J. Cell Biol.* **2012,** *199*, 583-588.
120. Kuzin, A. P.; Nukaga, M.; Nukaga, Y.; Hujer, A.; Bonomo, R. A.; Knox, J. R. Inhibition of the SHV-1 β-lactamase by Sulfones: Crystallographic Observation of Two Reaction Intermediates with Tazobactam. *Biochemistry* **2001,** *40*, 1861-1866.
121. Beadle, B. M.; Trehan, I.; Focia, P. J.; Shoichet, B. K. Structural Milestones in the Reaction Pathway of an Amide Hydrolase: Substrate, Acyl, and Product Complexes of Cephalothin with AmpC β-lactamase. *Structure* **2002,** *10*, 413-424.
122. Docquier, J. D.; Mangani, S. An Update on β-lactamase Inhibitor Discovery and Development. *Drug Resist. Updates* **2018,** *36*, 13-29.
123. Knox, J. E.; Ma, N. L.; Yin, Z.; Patel, S. J.; Wang, W. L.; Chan, W. L.; Rao, K. R. R.; Wang, G.; Ngew, X.; Patel, V.; et al. Peptide Inhibitors of West Nile NS3 Protease: SAR Study of Tetrapeptide Aldehyde Inhibitors. *J. Med. Chem.* **2006,** *49*, 6585-6590.
124. Yin, Z.; Patel, S. J.; Wang, W. L.; Wang, G.; Chan, W. L.; Rao, K. R. R.; Alam, J.; Jeyaraj, D. A.; Ngew, X.; Patel, V.; et al. Peptide Inhibitors of Dengue Virus NS3 Protease. Part 1: Warhead. *Bioorg. Med. Chem. Lett.* **2006,** *16*, 36-39.
125. Kang, C.; Gayen, S.; Wang, W.; Severin, R.; Chen, A. S.; Lim, H. A.; Chia, C. S. B.; Schüller, A.; Doan, D. N. P.; Poulsen, A.; et al. Exploring the Binding of Peptidic West Nile Virus NS2B-NS3 Protease Inhibitors by NMR. *Antiviral Res.* **2013,** *97*, 137-144.
126. Kang, C.; Keller, T. H.; Luo, D. Zika Virus Protease: An Antiviral Drug Target. *Trends Microbiol.* **2017,** *25*, 797-808.
127. Nitsche, C.; Holloway, S.; Schirmeister, T.; Klein, C. D. Biochemistry and Medicinal Chemistry of the Dengue Virus Protease. *Chem. Rev.* **2014,** *114*, 11348-11381.

128. Luo, D.; Vasudevan, S. G.; Lescar, J. The Flavivirus NS2B-NS3 Protease-Helicase as a Target for Antiviral Drug Development. *Antiviral Res.* **2015,** *118*, 148-158.
129. Johnston, P. A.; Phillips, J.; Shun, T. Y.; Shinde, S.; Lazo, J. S.; Huryn, D. M.; Myers, M. C.; Ratnikov, B. I.; Smith, J. W.; Su, Y.; et al. HTS Identifies Novel and Specific Uncompetitive Inhibitors of the Two-Component NS2B-NS3 Proteinase of West Nile virus. *Assay Drug Dev. Technol.* **2007,** *5*, 737-750.
130. Koh-Stenta, X.; Joy, J.; Wang, S. F.; Kwek, P. Z.; Wee, J. L. K.; Wan, K. F.; Gayen, S.; Chen, A. S.; Kang, C.; Lee, M. A.; et al. Identification of Covalent Active Site Inhibitors of Dengue Virus Protease. *Drug Des., Dev. Ther.* **2015,** *9*, 6389-6399.
131. Li, Y.; Zhang, Z.; Phoo, W. W.; Loh, Y. R.; Li, R.; Yang, H. Y.; Jansson, A. E.; Hill, J.; Keller, T. H.; Nacro, K.; et al. Structural Insights into the Inhibition of Zika Virus NS2B-NS3 Protease by a Small-Molecule Inhibitor. *Structure* **2018,** *26*, 555-564.
132. Amaro, R. E.; Mulholland, A. J. Multiscale Methods in Drug Design Bridge Chemical and Biological Complexity in the Search for Cures. *Nat. Rev. Chem.* **2018,** *2*, 0148.
133. Rungrotmongkol, T.; Decha, P.; Sompornpisut, P.; Malaisree, M.; Intharathep, P.; Nunthaboot, N.; Udommaneethanakit, T.; Aruksakunwong, O.; Hannongbua, S. Combined QM/MM Mechanistic Study of the Acylation Process in Furin Complexed with the H5N1 Avian Influenza Virus Hemagglutinin's Cleavage Site. *Proteins: Struct., Funct., Bioinf.* **2009,** *76*, 62-71.

TOC Graphic