Theor Chem Acc (2004) 111:36–48
DOI 10.1007/s00214-003-0556-x

Theoretical
Chemistry Accounts

Regular article

Modeling of serine protease prototype reactions with the flexible effective fragment potential quantum mechanical/molecular mechanical method

Alexander V. Nemukhin[1], Bella L. Grigorenko[1], Alexander V. Rogov[1], Igor A. Topol[2], Stanley K. Burt[2]

[1] Department of Chemistry, M.V. Lomonosov Moscow State University, 119992, Moscow, Russian Federation
[2] Advanced Biomedical Computing Center, SAIC Frederick, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702-1201, USA

Received: 16 September 2003 / Accepted: 18 October 2003 / Published online: 6 February 2004
© Springer-Verlag 2004

Abstract. A complete cycle of chemical transformations for the serine protease prototype reaction is modeled following calculations with the flexible effective fragment quantum mechanical/molecular mechanical (QM/MM) method. The initial molecular model is based on the crystal structure of the trypsin–bovine pancreatic trypsin inhibitor complex including all atoms of the enzyme within approximately 15–18 Å of the oxygen center O_γ of the catalytic serine residue. Several selections of the QM/MM partitioning are considered. Fractions of the side chains of the residues from the catalytic triad (serine, histidine and aspartic acid) and a central part of a model substrate around the C–N bond to be cleaved are included into the QM subsystem. The remaining part, or the MM subsystem, is represented by flexible chains of small effective fragments, whose potentials explicitly contribute to the Hamiltonian of the QM part, but the corresponding fragment–fragment interactions are described by the MM force fields. The QM/MM boundaries are extended over the C_α–C_β bonds of the peptides assigned to the QM subsystem in the enzyme, C–C and C–N bonds in model substrates. Multiple geometry optimizations have been performed by using the RHF/6-31G method in the QM part and OPLSAA or AMBER sets of MM parameters, resulting in a series of stationary points on the complex potential-energy surfaces. All structures generally accepted for the serine protease catalytic cycle have been located. Energies at the stationary points found have been recomputed at the MP2/6-31 + G* level for the QM part in the protein environment. Structural changes along the reaction path are analyzed with special attention to hydrogen-bonding networks. In the case of a model substrate selected as a short peptide $CH_3(NHCO\text{-}CH_2)_2$–HN–CO–$(CH_2$–$NHCO)CH_3$ the computed energy profile for the acylation step shows too high activation energy barriers. The energetics of this rate-limiting step is considerably improved, if more realistic model for the substrate is considered, following the motifs of the ThrI11–GlyI12–ProI13-–CysI14–LysI15–AlaI16–ArgI17–IleI18–IleI19 sequence of the bovine pancreatic trypsin inhibitor.

Keywords: Serine protease – Reaction path – Quantum mechanical/molecular mechanical modeling

Correspondence to: I. A. Topol
e-mail: topol@ncifcrf.gov

Introduction

Serine proteases catalyzing hydrolysis of peptide or ester bonds constitute one of the most experimentally studied families of enzymes. The general features of the reaction mechanism are widely accepted; however, the important details are still controversial and are subjects of extensive research. A good illustration is the content of Sect. III.C from the recent review article by Hedstrom [1] devoted to functioning of the catalytic triad in serine proteases. Molecular modeling and computer simulations are valuable tools to help to discriminate between competitive mechanisms and to provide activation energies of the rate-limiting stages. In this work we consider several models for the trypsin-catalyzed hydrolysis of peptide bonds by using modern versions of the hybrid quantum mechanical/molecular mechanical (QM/MM) approach.

The entire catalytic cycle of serine proteases (Fig. 1) is usually subdivided into the acylation and deacylation steps [1], and they both proceed through the formation of the so-called tetrahedral intermediates. His57 acts as a general base abstracting a proton from Ser195 in the acylation step or from the hydrolytic water molecule in the deacylation step (the chymotrypsinogen numbering system is used throughout the paper). Subsequently, the protonated imidazolium donates a proton to the

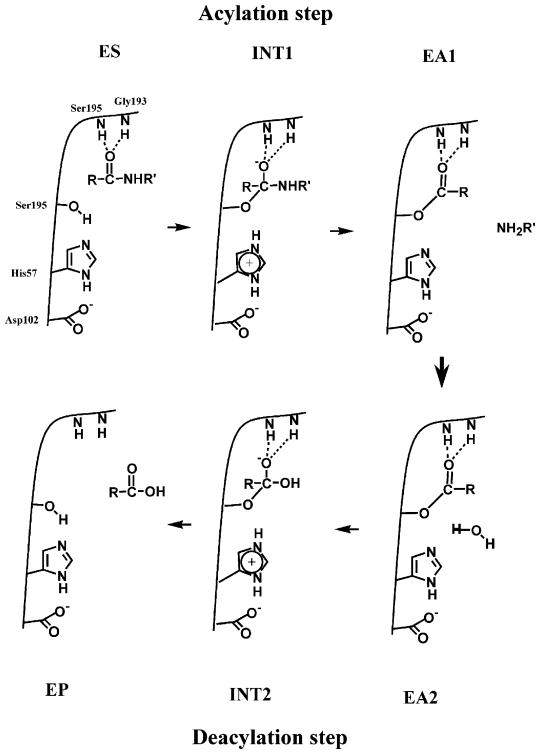

Fig. 1. Steps along the reaction pathway of serine protease catalyzed peptide bond cleavage. *ES* and *EP* refer to the enzyme–substrate and enzyme–product complexes, *INT1* and *INT2* refer to the tetrahedral intermediates presumably formed during the reaction in the acylation and deacylation stages, and *EA1* and *EA2* denote configurations of acylenzyme intermediates, which are distinguished by the positions of the leaving groups and the hydrolytic water molecule

substrate in the first step or gives it back to Ser195 in the second step. In the enzyme–substrate (ES) complex, the substrate *R*CONH*R′* is captured in the position convenient for nucleophilic attack of Ser195 on the carbonyl carbon of the substrate, leading to formation of the intermediate state complex with a tetrahedral coordination of the substrate carbon atom (INT1). The main-chain NH groups of Gly193 and Ser195 refer to the so-called oxyanion hole which assists in stabilizing the negatively charged oxyanion of the tetrahedral intermediate. After donation of the proton from imidazolium to the amide nitrogen of the substrate the peptide bond is broken, and the system proceeds to the fairly stable acylenzyme intermediate. We shall distinguish two possible structures of the acylenzyme intermediate, EA1 and EA2. In EA2, the water molecule replaces in the hydrogen-bonding network the half-product NH_2R' formed after the bond rupture in EA1. The deacylation step of the reaction repeats the sequence in the sense that the lytic water attacks the acylenzyme with the assistance of His57, leading to a second tetrahedral intermediate, INT2. The second tetrahedral intermediate collapses by expelling Ser195 and carboxylic acid product, and the enzyme–product (EP) complex is formed.

In simulations, it should be taken into account that the catalytic effect of the enzyme is to a large extent due to interactions between the active site and the rest of the molecular system, and adequate models should be carefully selected. For these reasons, QM calculations that consider only the catalytic triad (and model substrate) without the surrounding protein cannot be used for quantitative conclusions. The first semiquantitative analyses for catalytic reactions of serine proteases were performed by Warshel and coworkers [2, 3] on the basis of the empirical valence bond method. One of the conclusions was that the most important catalytic factor in serine proteases referred to the electrostatic stabilization of the negatively charged tetrahedral intermediate INT1 by the oxyanion hole. In these publications as well as in the subsequent publication [4], strong evidence was provided against the so-called double-proton mechanism, according to which the proton originally residing on the N_δ atom of His57 should transfer to Asp102 upon formation of the first tetrahedral intermediate, INT1.

The use of combined QM/MM methods [5, 6, 7, 8, 9] is one of the most appealing approaches for an adequate description of chemical transformations occurring in the active part of the system in the protein environment. The paper of Bentzien et al. [10] described the first application of the QM(ab initio)/MM method for modeling the nucleophilic attack in subtilisin. The authors separated this elementary reaction form the proton transfer from serine to histidine on the segment ES $\rightarrow$ INT1 (Fig. 1). In simulations, the quantum subsystem was selected as a reactive part of serine and a fragment from the tyrosine–glycine dipeptide, which was used to model substrate. The QM part considered at the HF/4-31G and MP2/6-31+G* levels contained 14 atoms, including four link atoms. The remaining atoms within 18 Å of the O_γ atom of the catalytic serine were treated with the ENZYMIX force field [11]. The free-energy profiles for the reaction in the protein and for the reference reaction in water were compared, demonstrating a lowering of the activation barrier for this half of the segment ES $\rightarrow$ INT1. For our purposes, it is important to note that these QM/MM calculations resulted in an exothermic energy balance of about 10 kcal/mol.

The QM/MM calculations of Topf et al. [12, 13] were directed towards analysis of the deacylation stage EA2 $\rightarrow$ EP of the serine protease catalytic cycle. In Ref. [12], the authors estimated the energy difference between INT2 and EA2 as 28 kcal/mol. In the recent study of Ref. [13], ab initio QM/MM dynamics simulations were exploited to examine the stability of INT2 along the 26-ps trajectory.

Another approach in modeling serine protease prototype reactions, based on the combined use of ab initio quantum chemistry and free-energy perturbation theory, was explored in the work by Stanton et al. [14]. In order to estimate the energetics of the acylation step ES $\rightarrow$ INT1, the following model system was considered. The QM part consisted of methanol and imidazole representing serine and histidine, respectively, and *N*-methylacetamide for the substrate. The computed QM energy difference between the analogues of INT1 and ES

configurations (considered as the activation energy) ranging from 71 to 54 kcal/mol, at different theoretical levels, was corrected by estimating empirical solvation interactions of both structures with either the protein environment or bulk water. In the calculation considered as the most successful one, the free energy of activation, ΔG^*, in the protein was estimated to be 16–18 kcal/mol, which was "in good agreement with the experimental value of around 15.1 kcal/mol for the acylation by trypsin (using the single Eyring equation relating k_{cat} to ΔG^*)" [14]. The same approximation was used by Peräkylä and Kollman [15] in order to explain the slow cleavage of the bovine pancreatic trypsin inhibitor (BPTI) by trypsin. One of the reasons for such slowness was attributed to a higher energy difference between INT1 and ES in the case of BPTI compared with regular substrates.

An important contribution to this field is the works of Štrajbl et al. [16, 17] in which the energy diagrams for the solution reactions of base-catalyzed methanolysis of formamide were constructed by combined ab initio/Langevin dipoles calculations. The basic motivation was to provide data for a reference prototype reaction in aqueous solution. The calculated activation barrier for the histidine-catalyzed reaction was found around 26 kcal/mol.

With few exceptions we do not include in our analysis papers in which simplified molecular models for the serine protease type systems (without adequate treatment of the protein environment) are considered at various quantum chemistry levels. In 1991 Daggett et al. [18] described a complete reaction pathway in serine proteases by using semiempirical PM3 calculations for a simplified model system. In the acylation step the authors considered the $CH_3CO_2^-$ species for aspartic acid, CH_3OH for serine, methylimidazol for histidine, and *N*-methylacetamide CH_3–CO–NH–CH_3 for the substrate. Two water molecules were added to the system in order to mimic an important effect of the oxianion hole. In the deacylation step the product CH_3NH_2 was removed, and a hydrolytic water molecule was introduced to the system. In a series of constrained-geometry optimizations for the reacting complex, the authors estimated the energy diagram for the reaction and analyzed the structural changes during the transformations. As expected, the energy barriers found in these calculations were too high, for example, 31.2 kcal/mol for the rate-limiting stage ES $\rightarrow$ INT1.

Also, the simplified model system similar to that described by Daggett et al. [18] was used in our previous ab initio quantum chemical calculations [19] only for the stage ES $\rightarrow$ INT1 of the serine protease prototype reaction. These calculations for the minimum energy profile were carried out at the Hartree–Fock level with the Stevens–Bash–Krauss effective core potentials and the corresponding basis sets. The sets of geometry parameters obtained along the reaction path were used in subsequent calculations at the B3LYP/6-31+G** level both in a vacuum and in the continuum model employing the dielectric constant of pure water. According to these calculations, the energy barrier height for this stage was 27–28 kcal/mol depending on the theoretical level, while the tetrahedral intermediate INT1 was placed 21–23 kcal/mol above the global minimum (ES).

Finally, we mention several theoretical papers addressing the problem of the low-barrier hydrogen bond often introduced as a hypothesis (although controversial) of the catalytic power of serine proteases [20, 21, 22, 23, 24, 25]. Energy profiles for proton transfer between N_δ of His57 and oxygen of Asp102 were analyzed with the goal to estimate the effect of the strength of the corresponding hydrogen bond on the stability of INT1. In a recent paper, Westler et al. [23] used density functional theory approaches to characterize properties of the fragments of the active site of chymotrypsin and of the chymotrypsin–trifluromethyl ketone complex, with the latter case mimicking the INT1 structure [23]. The properties of the hydrogen bond between His57 and Asp102 in the active sites of the low-pH form of α-chymotrypsin were carefully considered by Molina and coworkers [24, 25] by using, among other approaches, the QM/MM model based on the effective fragment potential (EFP) theory [26].

As shown by this brief analysis of the literature, only separate stages of the entire catalytic cycle have been the subject of a particular modeling. To the best of our knowledge, the only exception is the paper of Daggett et al. [18], which describes the results of low-level PM3 quantum chemical calculations for a severely simplified gas-phase system. In this work, we apply a "brute-force" QM/MM modeling of the entire cycle of serine protease catalyzed hydrolysis of a peptide bond (Fig. 1) and especially of its rate-limiting acylation step in a realistic protein environment. Our implementation of the QM/MM theory [27, 28, 29] is essentially based on the EFP method [26] and its implementation in the GAMESS(US) [30] and the Intel-specific version, PC GAMESS [31], quantum chemistry programs.

Originally, the EFP method was formulated as a tool to simulate solvation phenomena at the cluster level of modeling [32]. In this approach the solvent molecules are considered explicitly, and their positions in the system can be optimized. The influence of solvent molecules on the quantum solute species and the backward response are described through the potentials, the parameters of which should be determined at preliminary stages by the results of ab initio calculations. The EFP parameters for water were carefully selected and included into the database of the GAMESS(US) program. It was demonstrated in a series of papers [26, 27, 32, 33, 34, 35, 36, 37, 38, 39, 40] that applications of these EFPs led to very successful results.

The EFP technique was also used for simulations of chemical transformations in peptide environments [24, 25, 41, 42, 43, 44, 45, 46, 47, 48]; however, in these applications geometry configurations of effective fragments were supposed to be frozen in the course of the

optimization of other geometry parameters in the system. Therefore, we suggested an improvement of the method by introducing flexible chains of small rigid effective fragments, the relative positions of which were predicted by conventional MM force fields [27, 28, 29]. The concept of effective fragments was also utilized to solve the QM/MM boundary problem across covalent bonds [28, 29].

There is an important issue about the QM/MM method based on EFPs. Unlike many other QM/MM versions, its application does not assume explicit formulae for the QM/MM cross-terms for the energy, which are strongly dependent on the MM parameters. Instead, one-electron EFPs, the parameters of which are determined in independent preliminary ab initio calculations, directly contribute to the quantum Hamiltonian. If effective fragments describe a spectator part of an enzyme then, in this scheme, polarization of the QM active part by the protein environment is fully taken into account. The only use of empirical parameters is assumed in that part of the calculation scheme which accounts for interactions solely in the MM subsystem. In this sense, calculations with the flexible EFP method can be characterized as nonempirical.

Having in mind such benefits of the flexible EFP QM/MM method, we initially intended to apply it in order to construct a nonempirical energy diagram for the entire catalytic cycle of serine proteases, providing a uniform theoretical level for all stages. Following the work of Peräkylä and Kollman [15] we considered a model system taking the coordinates of the trypsin–BPTI complex as an initial guess for the ES complex. In a series of minimizations for different parts of the potential-energy surface we located all the structures generally accepted for the serine protease catalytic cycle (Fig. 1). These calculations are referred to in the forthcoming sections as our MODEL1 system. The energies at the stationary points found were computed at the MP2/6-31+G*//RHF/6-31G level for the QM part in the protein environment; however, for all levels of the theory and several modifications of MODEL1, the computed energy profile for the acylation step showed activation barriers that were too high. After that, we concluded that the main reason for this problem was an improper choice of the model substrate, selected in MODEL1 as a short peptide $CH_3(NHCO–CH_2)_2$–HN–CO–$(CH_2–NHCO)CH_3$. We improved the energetics of the acylation step with a MODEL2 system, in which a larger substrate was considered, keeping motifs of the ThrI11–GlyI12–ProI13–CysI14–LysI15–AlaI16–ArgI17–IleI18–IleI19 sequence of BPTI.

Both models took into account a considerable fraction of trypsin atoms within approximately 15–18 Å of the oxygen O_γ of Ser195. Fractions of the side chains of the residues from the catalytic triad (serine, histidine, and aspartic acid) and a central part of a model substrate around the scissile C–N bond were assigned to the QM part. The remaining parts, the MM subsystems, were represented by flexible chains of small effective fragments. The QM/MM boundaries were extended over the C_α–C_β bonds of the peptides assigned to the QM subsystem in the enzyme, as well as across the C–C and C–N bonds in the model substrates.

Details of calculations

The details of the flexible effective fragment QM/MM method are presented in Refs. [27, 28, 29], and we summarize here only its basic features. As in the original EFP formulation [26, 32], the Hamiltonian of a molecular system composed of the ab initio region, H_{AR}, and the environmental, V, parts is written as

$$H = H_{AR} + V, \tag{1}$$

where V represents the potentials due to the effective fragments. In our realization we consider the MM subsystem as a connected collection of relatively small effective fragments interacting with the ab initio region by quantum rules, but interactions between effective fragments is described by MM force fields.

We illustrate in Fig. 2 the QM/MM partitioning scheme using a substrate molecule for MODEL1 of this work as an example. This species consists of the QM part (CH_3–NH–CO–CH_3) at the center (distinguished by balls and sticks) where the broken valences are saturated by hydrogen link atoms. To the left and to the right of this ab initio region sequences of small effective fragments (CH_3, NHCO, CH_2) are introduced. One-electron potentials V from each effective fragment are subdivided into the electrostatic, V^{ELEC}, and the remaining, V^R, parts

$$V_\mu(r) = \sum_{k=1}^{K} V_{\mu,k}^{ELEC}(r) + V_\mu^R(r). \tag{2}$$

According to the implementation of the EFP method in the GAMESS(US) quantum chemistry package, the electrostatic potential acting on the quantum subsystem is represented by distributed multipoles centered at each atom and each bond midpoint. The multipole expansions are extended from charges up to octupoles, and the corresponding parameters can be created in preliminary ab initio calculations using GAMESS. The potentials V^R are approximated by the Gaussian-type functions located at atomic centers

$$V^R(r) = \sum_{m=1}^{M} \sum_{k=1}^{k_{max}} c_{mk} \exp\left(-\alpha_{mk} r_m^2\right). \tag{3}$$

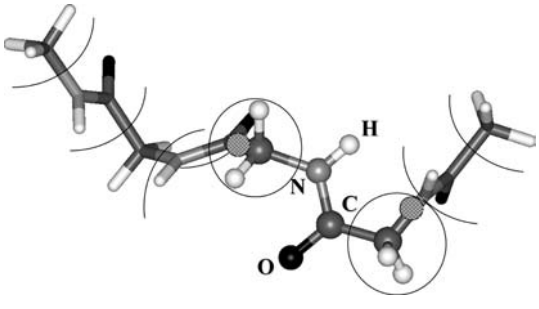

Fig. 2. Quantum mechanical(*QM*)/molecular mechanical (*MM*) (effective fragment potential, *EFP*) partitioning exemplified for a substrate of MODEL1. A *ball-and-stick* representation is used for the atoms assigned to the QM part. *Sticks* denote the MM subsystem subdivided into small rigid effective fragments: CH_3, NHCO, CH_2, NHCO on the left from the quantum region, and CH_3, COHN on the right from the quantum region. *Circles* distinguish buffer fragments including hydrogen link atoms

In Eqs. (2) and (3), r denotes electronic coordinates originating from the corresponding expansion points, K is the number of such expansion points for a distributed multipolar analysis, and the parameters c_{mk} and α_{mk} should be optimized by using reference ab initio data. The terms of Eqs. (2) and (3) are added to the one-electron operators in the Hamiltonian H_{AR} of the ab initio subsystem. We selected parameters c_{mk} and α_{mk} of the potentials of Eq. (3) for all the needed effective fragments by the following procedure. For biomolecular systems, description of hydrogen bonding seems to be of primary importance and, therefore, a water molecule can serve as a probing vehicle in the procedure of adjustment of parameters of EFPs. We considered a variety of directions along which the water molecule could reach each effective fragment and carried out ab initio calculations in order to provide reference data. Then we utilized the computer code REPGEN [49] to perform least-squares optimization of parameters of the potentials $V^R(r)$ (Eq. 3) by using two-term expansions ($k_{max} = 2$ in Eq. 3). The sum over k in Eq. (2) included all atomic centers and midbond points, and the sum over m in Eq. (3) extended over atomic centers. The Hartree–Fock approximation with the conventional 6-31G** basis sets was used for creation of multipole expansion parameters in $V^{ELEC}(r)$ as well as for creating reference ab initio interaction energies. For exactly the same coordinates the sets of QM/EFP interaction energies were produced and the best fitted coefficients c_{mk} and α_{mk} were selected.

Our treatment of the QM/MM boundary across the covalent bonds is also essentially based on the concept of effective fragments. In Fig. 3 two such cuts are shown. The key point is the introduction of a buffer fragment as a group of atoms belonging to both QM and MM subsystems. In this case the $-CH_2-$ group is assigned to the buffer, and a usual maneuver is used to saturate the broken valence by the closing (link) hydrogen atoms. Therefore, in the QM part we distinguish the buffer (in this case, CH_3) as a special group of the quantum subsystem. The same geometry configuration of the buffer fragment is assumed in the MM part. In the MM subsystem, which is a collection of effective fragments, the buffer is a special fragment as well. The position of the link atom is formally considered as an additional expansion point (as midbond points in "normal" effective fragments) which actually holds no multipoles. This device, essentially based on the GAMESS(US) implementation of the EFP method, helps us to keep the link atom precisely along the broken C–C bond during geometry optimizations of the entire system. In our scheme this empty expansion point and the neighboring CH group of the MM peptide chain form an effective fragment, which interacts with the buffer fragment according to the MM force fields and, as a consequence, the link atom cannot leave the C–C axis. The contributions to the one-electron Hamiltonian matrix of the QM subsystem from the MM(EFP) species are added only to the matrix elements of the quantum piece without buffer centers, and the electronic density in the buffer region was described by the minimal basis set of atomic orbitals.

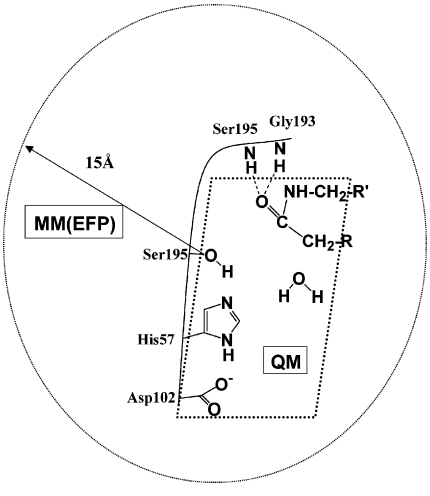

Fig. 3. Partitioning of the model reacting system into QM and MM parts for MODEL1

The forces on atoms in the QM part beyond the buffer are computed as the sums of quantum forces, F^{QM}, and contributions from effective fragments F^{QM-EFP}. Both types of forces are routinely computed in GAMESS(US). The forces acting on the centers of an effective fragment in the MM subsystem arise partly from other MM fragments, F^{MM}. These quantities are computed by using the molecular modeling program TINKER [50]. The forces from the QM subsystem, F^{EFP-QM}, act on the atomic centers and on the midbond expansion centers of this effective fragment as coded in GAMESS(US). All the forces acting on each center are summed up and finally applied to the center of mass of the fragment, resulting in the total force, F, and the total momentum, M. Finally, all energy contributions and energy gradients are collected in GAMESS(US), and the algorithms of this molecular modeling program are employed to find equilibrium geometry parameters of the entire QM/MM system. The current implementation of the flexible effective fragment QM/MM method allows the user to select which of the fragments can move, and which can be frozen during optimization.

Results for MODEL1

The initial atom positions for both model reacting system considered in this work were taken from the coordinates of the X-ray structure (Brookhaven Protein Data Bank code 2PTC) of trypsin inhibited by the BPTI determined to 1.9-Å resolution [51]. The substrate molecule for MODEL1 was taken as a peptide chain $CH_3(NHCO\text{-}CH_2)_2$**–HN–CO–**$(CH_2–NHCO)CH_3$ following the motifs of the LysI15–AlaI16–ArgI17 main-chain sequence of BPTI. In this chemical formula for a model substrate we distinguish in bold the fragment containing the N–C bond to be cleaved by the enzyme. The model for the latter was simplified as follows. We selected all residues of the protein contained within approximately 15 Å of the O_γ atom of the catalytic Ser195 and collected them into six pieces of peptide chains each terminated by the CH_3 groups.

As shown in Fig. 3, we included into the QM subsystem the main fragments of the residues from the catalytic triad Asp102, His57, Ser195, the central part of the model substrate, and the lytic water molecule. In MODEL1, this water molecule was introduced to the reactive region from the very beginning. It stayed inactive, i.e. not involved in the hydrogen-bonding network, in the acylation step and became active in the deacylation step, which was modeled at the EA1 → EA2 segment of the reaction path. In total, 40 atoms were treated quantum mechanically, including five hydrogen link atoms. The MM subsystem comprised 621 atoms combined into 191 small rigid effective fragments. The QM/MM boundaries were extended over the C_α–C_β covalent bonds of the aspartic acid, histidine, and serine residues and the C–C and C–N bonds from the central part of the model substrate.

Multiple geometry optimizations were performed by using the RHF/6-31G approximation in the QM part, and

OPLSAA force field parameters [52] for interactions in the MM part. As a result, a set of minimum-energy points on the complex potential-energy surface was obtained.

In these simulations we imposed some restrictions on the chains of effective fragments in the MM part, which were kept the same in the first and all subsequent minimizations. Namely, we set free all fragments of the model substrate, several effective fragments immediately following the buffer fragments at the QM/MM boundaries, the chain around Ser195 and Gly193, constituting an oxyanion hole, and also fragments representing Ser214. By doing this, we intended to provide enough flexibility to the fragments forming the most important region of the hydrogen-bonding network around the active site. The usual threshold for the gradient minimization (10^{-4} au) was assumed in the location of stationary points. During minimization the positions of all 40 atoms in the quantum part were completely optimized, as well as the positions of 33 effective fragments out of a total of 191.

The first found configuration corresponds to the ES complex for this model system. Superposition of the starting experimental (2PTC) and calculated ES structures is shown in Fig. 4. As expected, the largest differences are obtained for a relative position of the potential substrate. The QM/MM procedure introduced no dramatic changes for an overall representation of the active site.

Figures 5, 6, 7, 8, 9, 10, and 11 illustrate how chemical transformations during the reaction flow are reflected by changes in geometry configurations of the model system migrating over the QM/MM potential-energy surface from one stationary point to another.

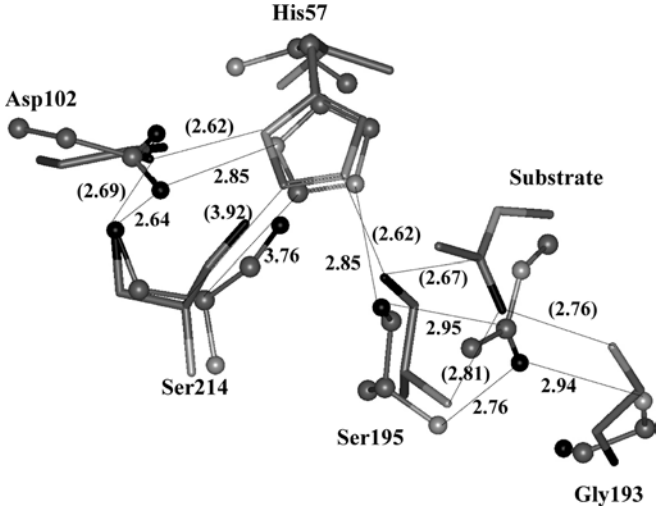

Fig. 4. Superposition of the experimental (2PTC, [51]) and calculated (ES configuration) structures of the central part of the system showing only heavy atoms. The representation with *sticks* refers to the X-ray structure of the trypsin–bovine pancreatic trypsin inhibitor complex, the representation with *balls and sticks* describes the configuration of the global minimum on the QM/MM potential-energy surface. Selected distances are given in angstroms. The values in *parentheses* refer to the crystal structure

The configuration of the ES complex is shown in Fig. 5. It is worth noting that a water molecule (depicted in sticks in this figure) stays above the plane of the

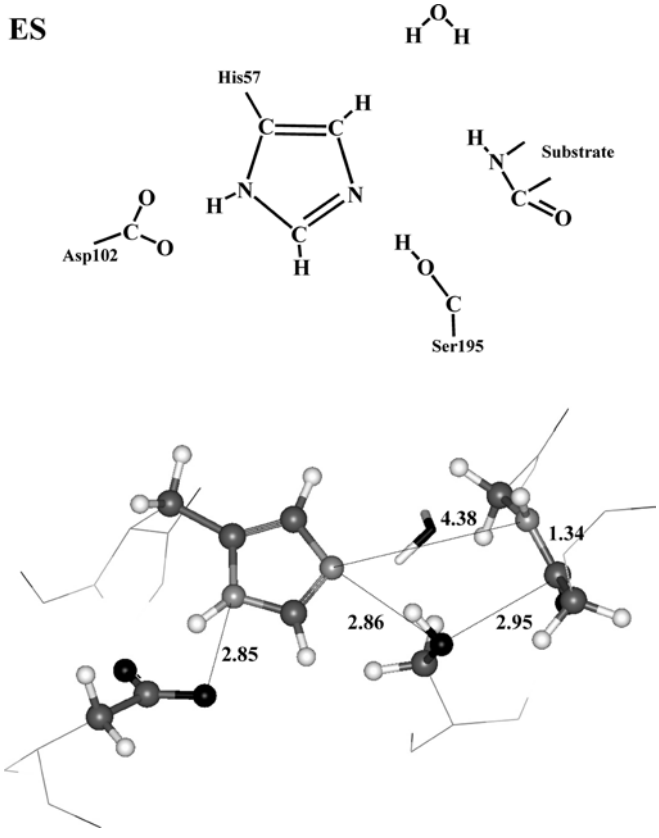

Fig. 5. ES complex

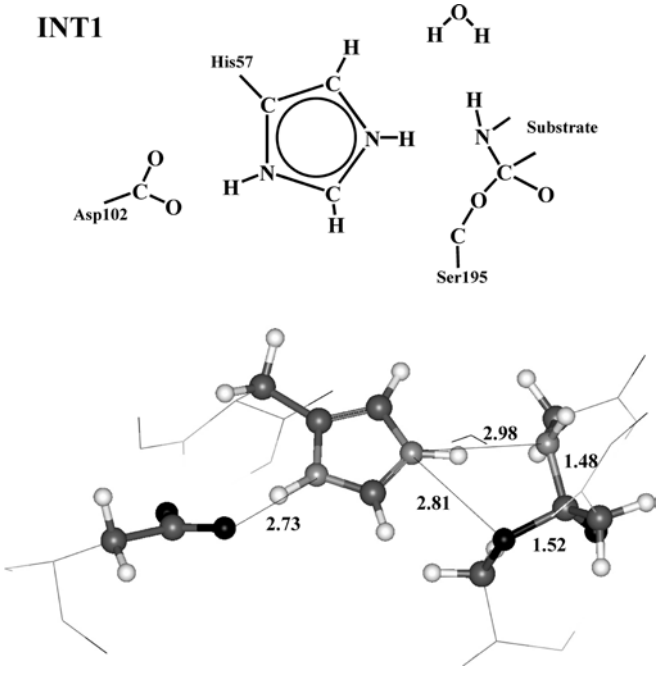

Fig. 6. Structure with the first tetrahedral intermediate, INT1

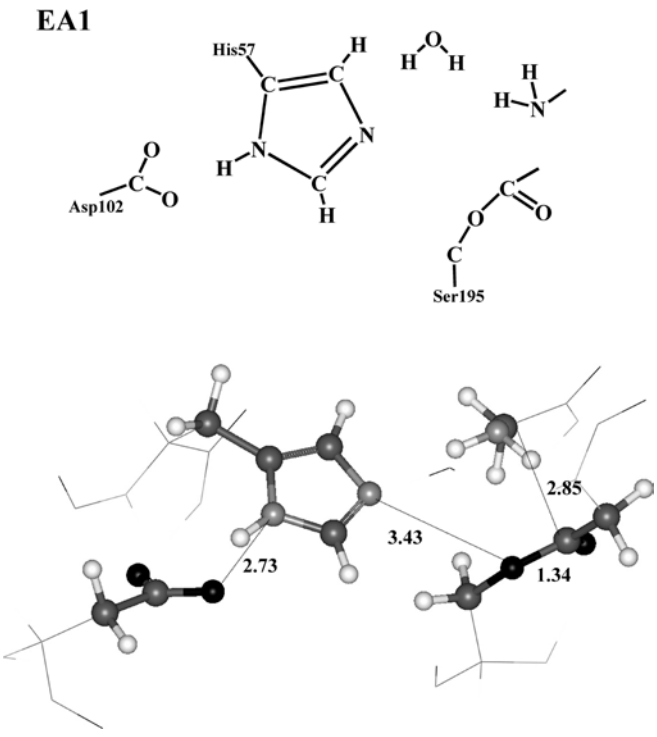

Fig. 7. Acylenzyme complex EA1 with a water molecule not participating in hydrogen bonding with the active site

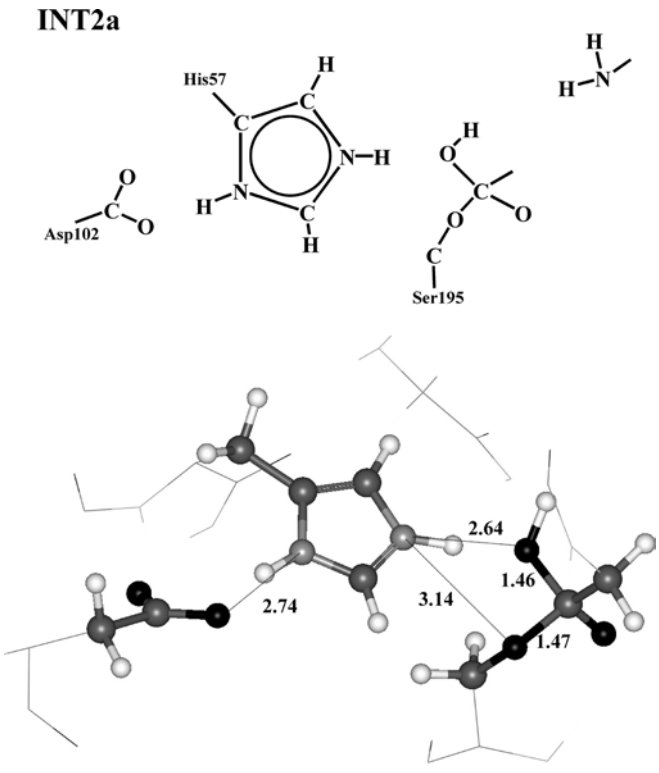

Fig. 9. Structure with the tetrahedral intermediate for the deacylation step (first minimum *INT2a*)

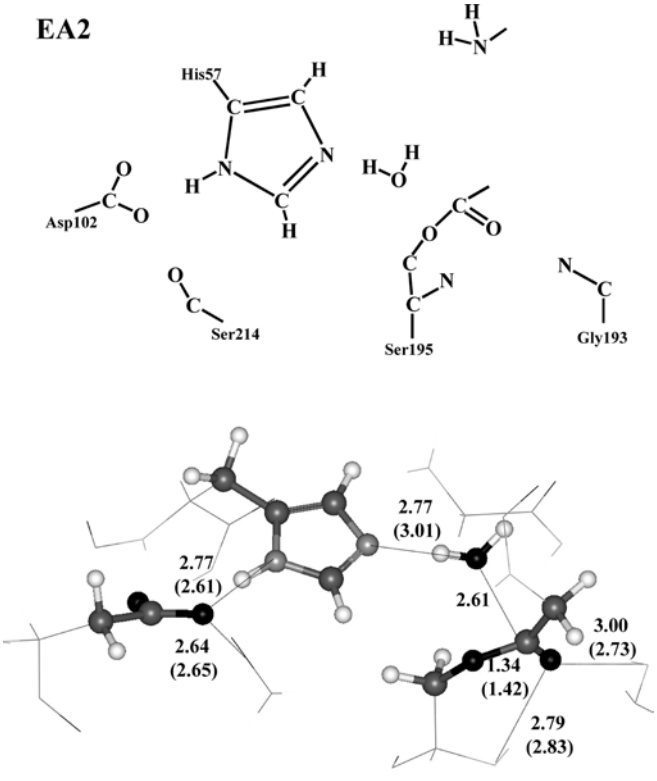

Fig. 8. Acylenzyme complex EA2 with a lytic water molecule involved in the hydrogen-bonding network of the active site. The distances in *parentheses* refer to the X-ray structure of the related complex 1QIX [53]

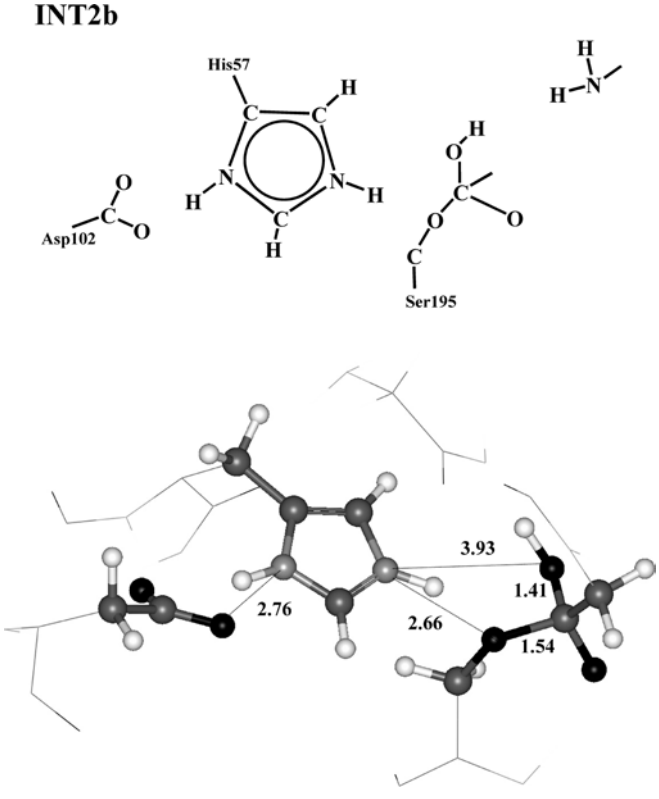

Fig. 10. Structure with the second tetrahedral intermediate for the deacylation step (second minimum *INT2b*)

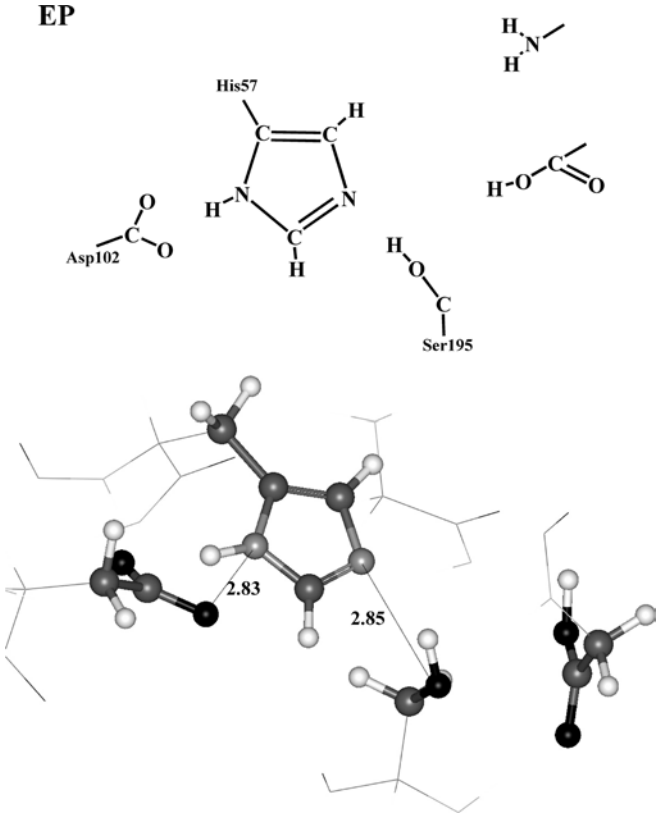

Fig. 11. EP complex

imidazole ring of His57 and does not participate in the hydrogen-bonding network with the active participants of the reaction. It is interesting to note that the Asp102–His57 dyad is not arranged in the perfect position for productive hydrogen bonding unlike the Ser195–His57 pair.

We obtained the configuration of the first tetrahedral intermediate, INT1, when moving from the ES in a straightforward manner [17, 18, 29]. The distance, R(CO), between the O_γ atom of Ser195 and the carbonyl carbon of the substrate was selected as a reaction coordinate, and the slow decrease of its value from initial 2.95 Å accompanying optimizations of other internal coordinates finally allowed us to reach the point where the unconstrained minimization specified the structure shown in Fig. 6. Without imposing additional constraints on the coordinates the proton from Ser195 moved to the N_ϵ atom of His57.

The transformation ES $\rightarrow$ INT1 is the most energy consuming step of the entire reaction path. It requires a severe distortion of the originally planar structure of the HN–CO group of the substrate and elongation of its N–C bond from 1.34 to 1.48 Å. Instead of the N_ϵ(His57)...HO_γ(Ser195) hydrogen bond, a longer N_ϵ(His57)...HN(substrate) hydrogen bond is formed. Remarkable changes are noticed in the Asp102–His57 dyad, and formation of a perfect hydrogen-bonded O(Asp102)...HN_δ(His57) moiety is seen. We shall return to the discussion of the ES $\rightarrow$ INT1 step later.

A proton transfer from N_ϵ(His57) to the nitrogen atom of the substrate resulted in the acylenzyme complex EA1, as illustrated in Fig. 7. The target peptide bond was broken and the leaving group moved away from the parent species [R(N–C) = 2.85 Å]. The water molecule, which was present in the quantum region during all the transformations described, continued to not participate actively in the reaction.

At this stage, the position of this water molecule was corrected in order to facilitate its binding by a hydrogen bond to N_ϵ(His57). The unconstrained minimization which followed such a rearrangement allowed us to obtain the new structure of the acylenzyme complex EA2 shown in Fig. 8.

A close configuration for this acylenzyme complex was described by Topf et al. [12, 13]. These authors used it as a starting point for QM/MM modeling of the deacylation step. The crystal structure of a specific acylenzyme complex between β-casomorphin 7 and porcine pancreatic elastase (Protein Data Bank entry 1QIX [53]) is believed to be a proper experimental reference for acylenzyme. Therefore, we draw in Fig. 8 additional fragments of the system (Ser214 and Gly193) and we added in parentheses the distances referring to the 1QIX crystal. A resemblance of the calculated EA2 configuration to the crystal structure 1QIX is practically the same as that of the ES to 2PTC (Fig. 4).

Preparation of the configuration for the second tetrahedral intermediate, INT2, was performed in the same manner as for the acylation step. We decreased the distance between the oxygen of the water molecule and the carbonyl carbon of the acylenzyme intermediate from the initial value 2.61 Å and adjusted all other internal coordinates. The structure shown in Fig. 9 is a result of an unconstrained minimization finally performed at this step.

The structure, called INT2a in this work, was discussed by Topfet al. [13] as a minimum which was closer to the reactant in their QM/MM dynamics simulations. Another minimum for the second tetrahedral intermediate, INT2b, can be obtained in accord with the findings of Ref. [13] by twisting the protonated imidazolium in such a manner that the N_ϵ(His57)H...O_γ(Ser195) hydrogen bond is formed, as illustrated in Fig. 10.

The proton transfer from N_ϵ(His57) to O_γ(Ser195) is followed by immediate breaking of the O_γ(Ser195)–carbonyl carbon bond and formation of the carboxylic acid product. This EP structure is illustrated in Fig. 11. Comparison of the initial (Fig. 5) and final (Fig. 11) configurations shows that the initial state of the catalytic triad is restored.

Thus, the simulations performed with the MODEL1 system completely described the entire catalytic cycle of serine proteases. We successfully located all minimum-energy points on the complex potential-energy surface which were compared with the intermediate structures accepted for this enzymatic reaction: however, the energy diagram, which was one of the objectives of these calculations, could not be considered as a success of

modeling. We show the results of calculations for MODEL1 at various segments of the reaction route along with the results of PM3 calculations of Daggett et al. [18] in Table 1. We point out that by using ab initio quantum chemical "gas-phase" calculations for the strongly simplified system (methanol and imidazole for serine and histidine and *N*-methylacetamide for the substrate), Stanton et al. [14] obtained $\Delta E(\mathrm{ES} \rightarrow \mathrm{INT1}) = 59$ kcal/mol at the MP2/6-31 + G*//HF/6-31 + G* level. This energy was converted to the free energy of activation, $\Delta G^{*} \approx 16$–18 kcal/mol, by applying the free-energy perturbation approach and estimating differences in interaction of ES and INT1 with the rest of the enzyme [14]. By using combined ab initio/Langevin dipoles calculations, Štrajbl et al. [17] estimated the free-energy activation barrier, 26 kcal/mol, for the histidine-catalyzed methanolysis of formamide in solution, obviously assuming that the activation barrier in the enzyme should be substantially less.

In our QM/MM(EFP) approach we assumed that a considerable fraction of the solvation energy from the spectator part of the protein was included in the calculation, and therefore, it was discouraging to obtain the energy increase of 38 kcal/mol for the segment ES → INT1, especially taking into account that a transition state connecting ES and INT1 must be even higher in energy.

We considered several reasons why MODEL1 failed to reproduce correctly the energetics of this stage. First of all we examined the case when a water molecule is present in the active site in the acylation step at all. Therefore, we carried out a new series of calculations for another model system, called here MODEL1-NW. In the latter no water molecule was introduced in the reaction region, 37 atoms were treated quantum mechanically, and the MM subsystem included 621 atoms collected in 191 effective fragments as before. For this system we reoptimized geometry configurations corresponding to the ES, INT1, and EA1 structures. The configurations obtained did not differ noticeably from those located before with MODEL1. For example, the distances specified in Figs. 5, 6, and 7 changed by 0.02 Å at most. However, the energy differences changed considerably as shown in the corresponding row of Table 1. About 10-kcal/mol reduction of the initial value of 38 kcal/mol for $\Delta E(\mathrm{ES} \rightarrow \mathrm{INT1})$ may be interpreted as follows. In the acylation step, the presence of a water molecule in the immediate vicinity of the reactive region does not benefit stabilization of the tetrahedral intermediate INT1, even if it is not involved in the hydrogen-bonding network with active participants of the reaction. This is actually assumed when the serine protease reaction mechanism is discussed at the qualitative level [1]; however, our calculations provide some quantitative estimates. An additional check was performed for the ES and INT1 structures, when we again reoptimized geometry configurations using the larger basis set 6–31 + G* in the QM/MM minimization, unlike in all previous calculations with the 6-31G basis. These QM/MM calculations resulted in practically the same sets of coordinates and a further slight reduction in $\Delta E(\mathrm{ES} \rightarrow \mathrm{INT1})$ as shown in the last row of Table 1.

In spite of such substantial improvement of the initial ES → INT1 energy difference, the final result, 26 kcal/mol, cannot be considered as a satisfactory one, since it means that the computed activation barrier is still too high.

Results for MODEL2

The motivation to consider MODEL2 for simulations of the ES → INT1 segment in the acylation step was as follows. It is known that trypsin cleaves a peptide bond near a charged residue. The model substrate chosen for MODEL1 does not satisfy such a requirement. Therefore, we returned to the structure of the trypsin–BPTI complex (Brookhaven Protein Data Bank code 2PTC) [51] and extracted a larger peptide chain ThrI11–GlyI12–ProI13–CysI14–LysI15–AlaI16–ArgI17–IleI18–IleI19 from the sequence of the BPTI. The C–N scissile bond was supposed to be between LysI15 and AlaI16. This follows the strategy used previously by Peräkylä and Kollman [15]. These authors considered the chain acetyl–Pro–Cys–Lys–Ala–Arg–NH_2 from the BPTI sequence as a model substrate for trypsin hydrolysis.

As in the case of MODEL1 (Fig. 3) we divided the model substrate into QM and MM parta, as shown in Fig. 12. We included in the QM subsystem the residues around the scissile C–N bond, namely, alanine and a fraction of lysine, and also fractions of cysteine and arginine, terminated by the link hydrogen atoms. The

Table 1. Relative energies (kcal/mol) on the reaction pathway estimated in the QM(MP2/6-31 + G*//RHF/6-31G)/MM(EFP; OPLSAA) approximation at the stationary points located. Shown are the results of Daggett et al. [18], the results of our MODEL1 with the water molecule in the active part, the results of simulations without the water molecule in the active part (MODEL1-NW), and the result of QM(MP2/6-31 + G*//RHF/6-31 + G*)/MM(EFP; OPLSAA) calculations for MODEL1-NW

Model	Acylation step		Deacylation step		
	ES → INT1	INT1 → EA1	EA2 → INT2a	EA2 → INT2b	INT2b → EP
PM3 results of Ref. [18]	31.2	−24.8	–	10.3	−16.1
MODEL1	38.0	−11.8	6.7	2.0	−20.1
MODEL1-NW	28.5	−8.1			
MODEL1-NW (6-31 + G* geometry optimization)	25.9				

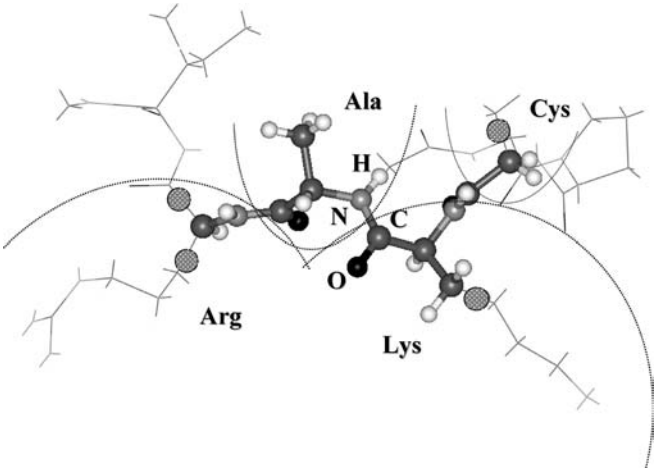

Fig. 12. Substrate molecule for MODEL2. A *ball-and-stick* representation is used for the QM part. *Lines* designate the MM part of the rest of the molecule described by a collection of 34 effective fragments. *Dashed lines* distinguish the scissile C–N bond. Link atoms at the borders of lysine (*Lys*), arginine (*Arg*), and cysteine (*Cys*) are specified by *circles with adornment*

remaining part of the model substrate was represented by a collection of 34 flexible effective fragments. The protonated forms of lysine and arginine were assumed.

We also enlarged a spectator part from the protein and included in the MM part all residues at a distance of 18 Å from the O_γ atom of the catalytic Ser195 as well as the ten water molecules closest to the reaction center from the 2PTC crystal structure. Water–water interactions were described by the TIP3P potential [54], while interactions of water molecules with other QM or MM atoms were computed with the EFP theory [26]. In total, 221 flexible effective fragments were included in the MM subsystem, comprising 776 atoms. Since the quantum part was considerably enlarged owing to some expansion of the model substrate, we decided to assign the entire Asp102 residue from the catalytic triad to the MM part during optimizations of stationary points on the potential-energy surface. In total, 56 atoms were assigned to the QM part, including 20 link atoms. In this series of calculations we considered two sets of force field parameters in the MM part: OPLSAA [52] and AMBER [55].

Using MODEL2 we located three points on the potential-energy surface for the initial stage of the acylation step. First, the ES configuration was optimized at the QM(6-31G)/MM(AMBER) level, taking the starting atomic coordinates from the crystal structure. Then we gradually decreased the distance R(CO) between the O_γ atom of Ser195 and the carbonyl carbon of the substrate from the initial value of 2.52 Å and carried out geometry optimizations with a single constraint R(CO) = constant. The energy curve proceeded through a maximum at R(CO) = 1.81 Å, and after this point, called TOP in the rest of the paper, unconstrained minimization allowed us to arrive at the first tetrahedral intermediate, INT1, with R(CO) = 1.47 Å. From previous ab initio calculations for smaller model systems [19] we know that a saddle point located in such a manner is fairly close to the true transition-state configuration. As with MODEL1 the proton transfer from O_γ of Ser195 to N_ϵ of His57 was accomplished without imposing additional constraints.

The geometry configuration in Fig. 13 (to be compared with Fig. 4) shows positions of some heavy atoms in the vicinity of the reaction center for the computed ES complex. Selected distances in Fig. 13 show the results of MODEL1 (top), MODEL2 (middle), and the experimental crystal values (bottom, in parentheses). First of all we note that the description of Asp102 by effective fragments results in practically the same arrangement of the Asp102–His57 dyad, as in the case of MODEL1, when Asp102 was included in the QM subsystem. Another important observation is that MODEL2 accounts for a considerably better description of the reactive part of the system, His57–Ser195–Gly193–substrate, than MODEL1. The anchoring parts of the lysine and arginine residues from the model substrate help to orient the substrate in a position better suited for nucleophilic attack of Ser195.

In Table 2 we present the results of energy calculations for the differences ΔE(ES → TOP) and ΔE(ES → INT1). Geometry optimizations were carried out for the total energy, composed of the energy of the quantum subsystem in the field of effective fragments and of the MM energy, computed with specified force field parameters. We show total energies as well as the energies of the QM part in the field of effective fragments in Table 2.

We notice that the energy barrier heights, ΔE(ES → TOP), computed with MODEL2 seem very reasonable—around 10 kcal/mol. We already referred to the paper of Stanton et al. [14] in which the experimental value of the free-energy activation barrier of around 15.1 kcal/mol for the acylation by trypsin was given, apparently by relating k_{cat} to ΔG^*. We also make refer-

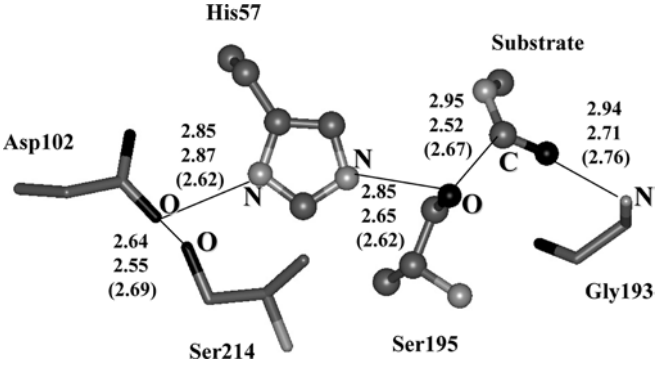

Fig. 13. ES configuration for MODEL2 showing the positions of some heavy atoms in the vicinity of the reaction center. The representation with *balls and sticks* refers to the atoms included in the QM subsystem; the representation with *sticks* refers to the atoms of the MM part. Selected distances in angstroms are arranged as follows: results of MODEL1 (*top*), results of MODEL2 (*middle*), experimental values from the crystal 2PTC structure [51] (*bottom*, in *parentheses*)

Table 2. Relative energies (kcal/mol) ΔE(ES → TOP) and ΔE (ES → INT1) at the points located, ES, TOP, INT1, in the QM (RHF/6-31)/MM(EFP; AMBER) approximation with MODEL2. The calculation with the OPLSAA set of force field parameters is specified explicitly in the third row; otherwise, AMBER force field parameters were used

Mode of calculation	Energies of QM part in the field of effective fragments		Total energies	
	ES → TOP	ES → INT1	ES → TOP	ES → INT1
RHF/6-31G	13.5	0.5	10.0	−5.1
RHF/6-31G (MM = OPLSAA)	13.5	0.5	7.4	−6.5
MP2/6-31 + G*//RHF/6-31G	12.8	0	9.3	−5.6

ence to the work of Park and Chi [56] in which the trypsin-catalyzed hydrolysis of N^{α}–Cbz–L-lysine *p*-nitrophenyl ester in a medium of variable dielectric constant was studied. It was found that the activation energy, E_a, determined with the Arrhenius plot for k_{cat} showed linear decreases with decreasing dielectric constant of the solvent from around 14.2 kcal/mol at $\epsilon = 70$ to 13.7 kcal/mol at $\epsilon = 65$. We stress that the flexible EFP QM/MM method contains no empirical adjustable parameters when computing QM energies in the field of effective fragments. We can see the effect of different force fields by comparing the second and third rows of Table 2—relative energies change by 1.4–2.6 kcal/mol when the OPLSAA set is used instead of the AMBER set of parameters.

In calculations with MODEL2 we observed that the energy of the first tetrahedral intermediate, INT1, was almost the same as that of the ES complex if the energies of the QM part in the field of effective fragments were compared. When MM contributions were added the energy of INT1 was even lower than that of ES. A similar result was obtained in the first paper on the QM/MM modeling of the serine protease prototype reaction by Bentzien et al. [10]. The energy diagram for a part of the ES → INT1 stage showed that this process was exothermic with an energy gain of approximately 10 kcal/mol. In our simulations INT1 is efficiently stabilized by the protein, although it is interesting to note that a substantial contribution to such stabilization arises from some lower value of pure MM energy.

Discussion and conclusions

In this work we utilized a new realization of the QM/MM method [27, 28, 29], from which we expect results fairly close to those of ab initio modeling. This flexible EFP QM/MM method contains no empirical adjustable parameters when computing properties of the QM subsystem in the field of effective fragments. All parameters of EFPs were obtained in preliminary ab initio calculations and were not subject to any changes from one application to another. In all previous calculations carried out with this method [28, 29, 57] we kept the same one-electron potentials, specified by Eqs. (2) and (3). The same holds for the calculations described in this paper. In this method there are no explicit formulae for the QM/MM cross-terms in the Hamiltonian, and therefore there is no temptation, for example, to correct or remove charges on MM atoms in the vicinity of the QM/MM interface or to modify other parameters in these terms to gain better coincidence with available experimental data. Polarization of the electron density of the quantum subsystem due to effects of the MM subsystem is taken into account through the one-electron potentials from effective fragments contributing to the quantum Hamiltonian. Forces from the QM part are added to the forces acting on the effective fragments and finally on the MM atoms. In many respects, the method belongs to the highest level of the hybrid QM/MM approaches. Of course, the computed energies of the pure MM part are dependent on the force field parameters, although not substantially, as demonstrated by the data in Table 2.

The success of the modeling with such a QM/MM method depends on the reliable choice of the molecular model, including partitioning of the entire system into the QM and MM parts. This is clearly demonstrated in this work. With MODEL1 we succeeded in a uniform description of the complete reaction path for the serine protease prototype transformations from ES to EP complexes. Geometry parameters and energies were calculated with the same set of atoms in the QM part and the same spatial restrictions on the displacements of effective fragments constituting the MM part. Qualitatively, the close correspondence of the stationary points found on the complex potential-energy surface to the structures consistent with the generally accepted mechanism of the serine protease catalytic cycle is a clear success of the approach, since no constraints were imposed on the positions of the atoms in the active region. However, MODEL1 failed to describe the activation barrier for the rate-limiting acylation step of the catalytic cycle.

MODEL2 was selected to satisfy a requirement that a realistic species for the substrate should be considered. Obviously at such a level of modeling a balanced treatment of both participants, enzyme and substrate, must be provided. When a model substrate with a charged residue (lysine) near a scissile C–N bond was considered, the energetics of the acylation step was computed with a reasonable accuracy.

Finally, we comment on several controversial details of the mechanism of catalysis by serine proteases [1] on the basis of these simulation results. First, our calculations do not support the double-proton mechanism, in

agreement with the reasons of Refs. [2, 3, 4]. In the QM/MM modeling we could not locate another local minimum of the first tetrahedral intermediate, namely, a configuration corresponding to the double proton transfer (from Ser195 to N_ϵ of His57 and from N_δ of His57 to aspartic acid) upon formation of INT1. However, such a local minimum ($AspH^+$...His) could be easily located, along with that described earlier (Asp...$HisH^+$), if the quantum part was separated from the rest of protein, i.e. if all EFPs were eliminated from the system. The importance of Ser214, mentioned in the work of Warshel et al. [4], seems justified (Figs. 4, 13). If this residue is present in the system, even in the form of a set of effective fragments, Asp102 cannot be protonated.

The role of Asp102 in stabilizing the tetrahedral intermediate INT1 and the transition state on the way ES → TOP → INT1 seems to be only electrostatic, again in agreement with the works of Warshel and coworkers [2, 3, 4], as illustrated by the data of Table 3.

We compared the energies ΔE(ES → TOP) and ΔE(ES → INT1) obtained in the original MODEL2, and upon "mutations" of the charged residues: lysine and arginine from the substrate, and Asp102 from the protein. Mutation means that the charge on the corresponding effective fragment is formally set to zero. We see that while mutations of lysine and arginine lead to almost negligible effects, setting formal zero charge on Asp102 results in a dramatic increase of both energy gaps, ΔE(ES → TOP) and ΔE(ES → INT1).

Another intriguing hypothesis for the serine protease catalytic cycle is an assumption of the so-called "histidine flip" mechanism according to which after formation of the tetrahedral intermediate, His57–H^+ flips to place the N_δ (but not the N_ϵ!) proton near the leaving group. This would require of around 180° rotation of the imidazole ring about the histidine C_β–C_γ bond [58, 59]. Our simulations with MODEL1 show that the imidazole of His57 is really mobile during the reaction cycle, but not to the extent needed for the histidine flip hypothesis. We plot changes in the $C_\alpha C_\beta C_\gamma N_\delta$ torsional angle of His57 when the reaction proceeds from ES to EP in Fig. 14. These data are basically consistent with the findings of Ref. [13] relative to the deacylation step.

Table 3. Relative energies (kcal/mol) ΔE (ES → TOP) and ΔE (ES → INT1) obtained in the original MODEL2, and upon "mutations" of the charged residues. By mutations we mean zeroing the corresponding charge Q: +1 for lysine (*Lys*) and arginine (*Arg*) from the substrate, and −1 for Asp102. The energies of the quantum subsystem in the field of effective fragments are shown

Mode of calculation	ES → TOP	ES → INT1
MP2/6-31+G*//RHF/6-31G	12.8	0
MP2/6-31+G*//RHF/6-31G [Q(Lys) = 0]	14.2	0.3
MP2/6-31+G*//RHF/6-31G [Q(Arg) = 0, Q(Lys) = 0]	17.5	3.1
MP2/6-31+G*//RHF/6-31G [Q(Asp102) = 0]	25.1	17.6

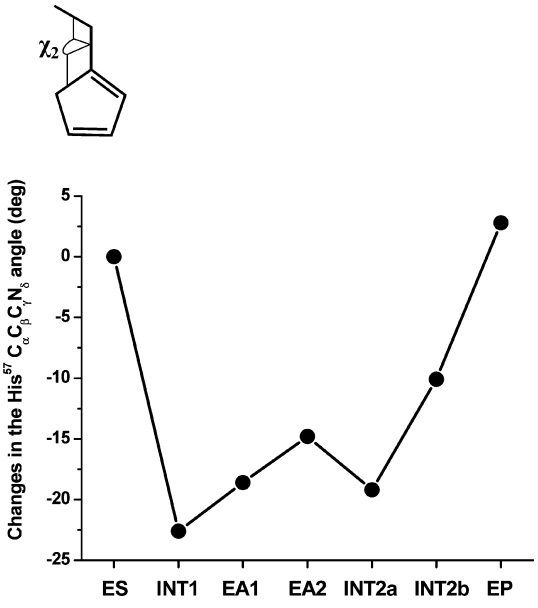

Fig. 14. Changes in the $C_\alpha C_\beta C_\gamma N_\delta$ torsional angle of His57 in the reaction

As shown in Fig. 14, the most intense movements of the imidazole ring of His57 happen at the segments ES → INT1 and INT2 → EP; however, they are not as dramatic as required for the histidine flip mechanism [58, 59].

Acknowledgements. The research described in this publication was made possible in part by award no. RC1-2350-MO-02 of the U.S. Civilian Research and Development Foundation for the Independent States of Former Soviet Union. This work was also supported in part by a grant from the Russian Foundation for Basic Researches (project 01-03-32071). We thank the staff and administration of the Advanced Biomedical Computing Center for their support of this project. This project was funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract no. NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

1. Hedstrom L (2002) Chem Rev 102:4501
2. Warshel A, Sussman F (1986) Proc Natl Acad Sci USA 83:3806–3810
3. Warshel A, Russel S (1986) J Am Chem Soc 108:6569–6579
4. Warshel A, Naray-Szabo G, Sussman F, Hwang J-K (1989) Biochemistry 28:3629–3637
5. Warshel A, Levitt M (1976) J Mol Biol 103:227
6. Field M J, Bash P A, Karplus M (1990) J Comput Chem 11:700
7. Gao JL, Xia XF (1992) Science 258:631
8. Zheng Y-J, Merz KM (1992) J Am Chem Soc 114:10498
9. Monard G, Loss M, Thery V, Baka K, Rivail J-L (1996) Int J Quantum Chem 58:153
10. Bentzien J, Muller RP, Florián J, Warshel A (1998) J Phys Chem B 102:2293
11. Lee FS, Chu ZT, Warshel A (1993) J Comput Chem 14:161
12. Topf M, Várnai P, Richards W G (2001) Theor Chem Acc 106:146

13. Topf M, Várnai P, Richards WG (2002) J Am Chem Soc 124:14780
14. Stanton RV, Peräkylä M, Bakowies D, Kollman PA (1998) J Am Chem Soc 120:3448
15. Peräkylä M, Kollman PA (2000) J Am Chem Soc 122:3436
16. Štrajbl M, Florián J, Warshel A (2000) Int J Quantum Chem 77:44
17. Štrajbl M, Florián J, Warshel A (2000) J Am Chem Soc 122:5354
18. Daggett V, Schröder S, Kollman P A (1991) J Am Chem Soc 113:8926
19. Nemukhin AV, Topol IA, Burt SK (2002) Int J Quantum Chem 88:34
20. Frey PA, Whitt SA, Tobin JB (1994) Science 264:1927
21. Warshel A, Papazyan A, Kollman P A (1995) Science 269:102
22. De Santis L, Carloni P (1999) Proteins Struct Funct Genet 37:611
23. Westler WM, Weinhold F, Markley JL (2002) J Am Chem Soc 124:14373
24. Molina PA, Sikorski RS, Jensen JH (2003) Theor Chem Acc 109:100–107
25. Molina PA, Jensen JH (2003) J Phys Chem B 107:6226–6233
26. Gordon MS, Freitag MA, Bandyopadhyay P, Jensen JH, Kairys V, Stevens WJ (2001) J Phys Chem A 105:293
27. Nemukhin AV, Grigorenko BL, Bochenkova AV, Topol IA, Burt SK (2002)J Mol Struct (THEOCHEM) 581:167
28. Grigorenko BL, Nemukhin AV, Topol IA, Burt SK (2002) J Phys Chem A 106:10663
29. Nemukhin AV, Grigorenko BL, Topol IA, Burt SK (2003) J Comput Chem 24:1410
30. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL, Dupuis M, Montgomery JA (1993) J Comput Chem 14:1347
31. Granovsky A http://lcc.chem.msu.ru/gran/gamess/index.html
32. Day PN, Jensen JH, Gordon MS, Webb SP, Stevens WJ, Krauss M, Garmer D, Basch H, Cohen D (1996) J Chem Phys 105:1968
33. Chen W, Gordon MS (1996) J Chem Phys 105:11081
34. Day PN, Pachter R (1997) J Chem Phys 107:2990
35. Krauss M, Webb SP (1997) J Chem Phys 107:5771
36. Merrill GN, Gordon MS (1998) J Phys Chem A 102:2650
37. Webb SP, Gordon MS (1999) J Phys Chem A 103:1265
38. Petersen CP, Gordon MS (1999) J Phys Chem A 103:4162
39. Day PN, Pachter R, Gordon MS, Merrill GN (2000) J Chem Phys 112:2063
40. Nemukhin AV, Topol IA, Grigorenko BL, Burt SK (2002) J Phys Chem B 106:1734
41. Wladkowski BD, Krauss M, Stevens WJ (1995) J Am Chem Soc 117:10537
42. Krauss M (1995) Comput Chem 19:199
43. Worthington SE, Krauss M (2000) Comput Chem 24:275
44. Roitberg AE, Worthington SE, Holden MJ, Mayhew MP, Krauss M (2000) J Am Chem Soc 122:7312
45. Worthington SE, Roitberg AE, Krauss M (2001) J Phys Chem B 105:7087
46. Worthington SE, Krauss M (2001) J Phys Chem B 105:7096
47. Kairys V, Jensen JH (2000) J Phys Chem A 104:6656
48. Minikis RM, Kairys V, Jensen JH (2001) J Phys Chem A 105:3829
49. Stevens W (1991) REPGEN for optimizing effective fragment repulsive potentials. CARB
50. Ponder JW, Richards FM (1987) J Comput Chem 8:1016
51. Marquart M, Walter J, Deisenhofer J, Bode W, Huber R (1983) Acta Crystallogr Sect B 39:480
52. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 117:11225
53. Wilmouth RC, Clifton IJ, Robinson CV, Roach PL, Aplin RT, Westwood NJ, Najdu J, Schofield CJ (1997) Nature Struct Biol 4:456
54. Jorgensen WL, Chandrasekhar J, Madura J, Impley RW, Klein ML (1983) J Chem Phys 79:926
55. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179
56. Park H, Chi YM (2001) Biochim Biophys Acta 1568:53
57. Nemukhin AV, Grigorenko BL, Topol IA, Burt SK (2003) J Phys Chem B 107:2958
58. Bachovchin WW (1986) Biochemistry 25:7751
59. Ash EL, Sudmeier JL, Day RD, Vincent M, Torchilin EV, Haddad RC, Bradshow EM, Sanford DG, Bachovchin WW (2000) Proc Natl Acad Sci USA 97:10371