A Self-Stabilized Model of the Chymotrypsin Catalytic Pocket. The Energy Profile of the Overall Catalytic Cycle

Péter Hudáky[1,2] **and András Perczel**[2]*
[1]*Department of Theoretical Chemistry, Eötvös Loránd University, Budapest 112, Hungary*
[2]*Department of Organic Chemistry, Eötvös Loránd University, Budapest 112, Hungary*

ABSTRACT **A model of the catalytic triad of chymotrypsin is built assuring the arrangement and properties as they are within the complete enzyme. The model contains 18 amino acid residues of chymotrypsin and its substrate. A total of 135 atoms (including 70 heavy atoms) were subjected to full ab initio geometry optimizations through 127 individual steps along the reaction coordinate of the complete catalytic mechanism. It was shown that the described model of the catalytic apparatus forms a self-stabilized molecule ensemble without the rest of the enzyme and substrate. According to the calculations, the formations of the first and second tetrahedral intermediates in the model have 20.3 and 15.7 kcal/mol activation energy barriers, respectively. Removing elements of the catalytic apparatus such as the (1) catalytic aspartate or (2) the anion hole, as well as (3) inserting a water molecule "early" in the catalytic process, or (4) introducing conformational rigidity of the substrate, results in an increase of the above energy barrier of the first catalytic step in the model by 6.4, 13.7, 3.7, and 4.1 kcal/mol, respectively. Based on the calculated process one can conclude that the catalytic reaction in this model is much more similar to the reaction in the enzyme than to the reference reaction. To our knowledge, this is the first model system that mimics the complete catalytic mechanism. Proteins 2006;62:749–759.**

Key words: ab initio; quantum chemistry; inhibitor; substrate; catalysis

INTRODUCTION

Proteases perform the hydrolysis of ester and peptide bonds. These processes have crucial importance in the digestion of proteins that are part of the diet of metazoan organisms. Chymotrypsin is a member of the protease family involved in the cleavage of specific amide bonds in proteins.[1] The precise details of the catalytic mechanism of chymotrypsin are still under discussion, resulting in theories that are often mutually exclusive. Early models of catalysis at the late 1960s were based on a charge relay model[2] followed by different modified charge relay systems.[3] However, the charge transfer model was later on contravened based on experimental evidences obtained form NMR measurements.[4] A hypothesis of the significance of a low-barrier hydrogen bond[5] was also reported; however, its formation and possible role was than also contravened.[6,7] The possibility of a special ring-flip mechanism was also envisaged involving the 180° rotation of the imidazole ring of the catalytic histidine.[8] Beyond all these theories, the importance of electrostatic mechanisms,[9,10] is unquestionable. Aspects of computational investigation of enzyme reactions is given in recent reviews of Warshel and coworkers.[11,12] It is assumed that for the correct evaluation of the details of the catalytic mechanism multiple methods and processes may be necessary to avoid misleading consequences arising from any known or unknown shortcoming of a particular method.

The inhibition mechanism of the catalytic activity of serine proteases shows a similar mechanism observed for substrates, but the process is stopped/slowed down at a particular step. Canonical inhibitors are thought to ambush the catalytic process at the formation of the first tetrahedral intermediate.[13,14] In addition, serpins form the first tetrahedral intermediate similar to substrates, and are subsequently cleaved.[15] However, the product undergoes to a significant conformational rearrangement that prevents the formation of the second tetrahedral intermediate. Thus, the enzyme–serpin complex does not undergo hydrolysis. There is a third type inhibitory mechanism used by hirudin, where the catalytic serine does not have substantial interaction with the inhibitor.[16]

Experimental and theoretical investigations of Gibbs Free Energy difference between the enzyme substrate complex (ES) and first tetrahedral intermediate (TC1) in chymotrypsin is in the focus of investigation.[17–21]

The kinetics of the catalytic activity is primarily influenced by the activation Gibbs Free Energy (ΔG) as described by the Eyring equation. Quantum chemical optimization methods provide, however, the activation energy (ΔE). Ab initio theoretical determination of ΔG requires

---

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: the Hungarian Scientific Research Foundation; Grant numbers: OTKA T046994 and TS044730; Grant sponsor: Medichem 2; Grant sponsor: ICGEB Hun04-03.

*Correspondence to: András Perczel, Department of Organic Chemistry, Eötvös Loránd University, H-1518 Budapest 112, Hungary. E-mail: perczel@para.chem.elte.hu

Received 13 April 2005; Revised: 21 September 2005; Accepted 23 September 2005

Published online 15 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20872

additional extensive computations. The relation of ΔG and ΔE was investigated, and the ΔG proved to be correlated with ΔE,[22] in cases where the entropy does not undergo major changes. This condition holds, if reacting moieties are and remain in close contact during the reaction. As our study does not cover the processes of substrate binding and release, but the steps of the catalysis, where the substrate is bound, the determination of the value of ΔE provides a good estimate for ΔG.

With regard to serine proteases, the catalytic reaction is initiated by the nucleophilic attack of the alcoholic oxygen of the catalytic Ser on the carbonyl carbon atom of a peptide bond of residues P1 and P1′, followed by breaking the N—C bond of the substrate followed by hydrolysis of the resulting ester. The reference reaction was subjected to theoretical investigations on small model systems in solvent such as hydrolysis[23] and methanolysis of formamide,[24] and hydrolysis of methyl formate,[25] and of N-(o-carboxybenzoyl)-L-amino acid.[26] For better understanding the process of hydrolysis in the enzyme the models must be placed in the catalytic pocket of the enzyme. One way to do it is the formation of a model of the catalytic pocket of the enzyme from carefully selected moieties of the enzyme substrate complex. Such a model typically contains 20–80 atoms, which can be treated quantum chemically.[21,27] Another way is to apply QM/MM approaches,[28–31] where the core of the catalytic pocket and the substrate is investigated quantum chemically, and the rest of the protein with molecular mechanics. The empirical valence bond (EVB) method[32] is an improved QM/MM technique that performs reliably. DFT molecular dynamics methods are also reported[33] handling proteins.

Determination of the energy profile of the formation of the first tetrahedral intermediate of the reaction in the enzyme and in water was covered by numerous articles using a series of techniques. The EVB study of Warshel and Russel[18] (1986) determines the Gibbs free energy of the rate limiting step of the catalysis as 18 kcal/mol, while that of the reference reaction as 25 kcal/mol. Stanton et al.[20] used the QM-FE approach (an alternative to QM/MM) and determined activation free energies of 16 and 18 kcal/mol with two different procedures in protein and 32 and 33 kcal/mol in water. They used cratic free energy correction, which was later criticized because it overestimates the entropic effect.[11] They defined a quantum chemically treated region containing 12 heavy atoms (the imidazole ring for His57, a methanol representing Ser195 and an N-methyl-acetamide as the substrate analog). Bentzien et al. used an ab initio "region" of 14 atoms in a methodological QM(ai)/MM article and determined 19 and 12 kcal/mol for the nucleophilic attack in water and protein, respectively.[34] They emphasize that this result cannot be converted to activation energy (catalytic effect), as it does not contain the proton transfer step. Ishida et al.[35] used a QM/MM approach with a quantum chemical description of 18 heavy atoms in the catalytic process and provided a value of 17.8 kcal/mol as the activation free energy. Daggett et al.[36] reported PM3 calculations and reported a value of 31 kcal/mol in vacuo for the reference reaction on a system containing 19 heavy atoms. They point out that their value is high, which can be due to several reasons as determining potential energy instead of free energy, truncation of the catalytic system, and neglecting of environment. Ab initio methods were combined with Langevin dipoles by Štrajbl et al.[37] on a six heavy atom reference system (methanol, ammonia, formamide) in water, which resulted in an activation barrier of 27 kcal/mol. The Effective Fragment Potential (EFP) theory was recently applied by Nemukhin et al.,[38] where the energy increase of the first tetrahedral complex (TC1) related to enzyme substrate complex (ES) is 38 kcal/mol (40 atoms treated quantum chemically, including all hydrogens). Another calculation was then performed without a water molecule in the model (37 atoms treated quantum chemically, including all hydrogens) and an activation energy of 25.9 kcal/mol was obtained for the same process. Furthermore, a charged residue was also introduced into the substrate to mimic the conditions of trypsin substrates in a new model. In this model a 0 kcal/mol energy difference was found for ES and TC1, and the activation barrier was 13 kcal/mol (56 atoms treated quantum chemically, including all hydrogen atoms). Nemukhin et al.[38] have reported ~28 kcal/mol in the gas phase and 24–28 in solution calculated at midsize quantum chemical levels [B3LYP/6-31+G(d,p)] on a model system containing a total of 18 heavy atoms, where ethanol, methylimidazole, acetic acid molecules represented the catalytic apparatus and formamide was the substrate with two molecules of water.[27]

The reduction of energy barrier of about 8 kcal/mol of the catalytic reaction over the reference one is assured by a multitude of circumstances provided by the enzyme, the origin of all of which are related to electrostatic effects. Such are the preorganized arrangement, stabilization of transition states, H-bonding network, preactivated components. These circumstances of the catalytic pocket of chymotrypsin are assured by the enzyme, the folding of which is very sensitive against any change. Modification of one single amino acid in the vicinity of the catalytic pocket can destroy the suitable arrangement of the catalytic triad.[39]

At present, for the investigation of the catalytic mechanism in the complete enzyme the EVB approach seems to be the most reliable and accurate method.[12] In this approach the catalytic reaction is described in terms of mixtures of diabatic states that correspond to classival valence bond structures.[11] Our aim in the present contribution is to investigate the catalytic apparatus of chymotrypsin in a model system that assures an environment for the catalytic reaction as it is in the enzyme. Such a model will miss several interactions the catalytic pocket has with its environment; however, we attempted to involve all important interactions that have a direct effect on the moieties involved in the catalysis. The model built can be treated uniformly by pure quantum chemical levels. It is self-stabilized, as it maintains its natural-like inner orientations without the presence of any geometrical constraints. The result of such a Hartree-Fock or post-Hartree-Fock

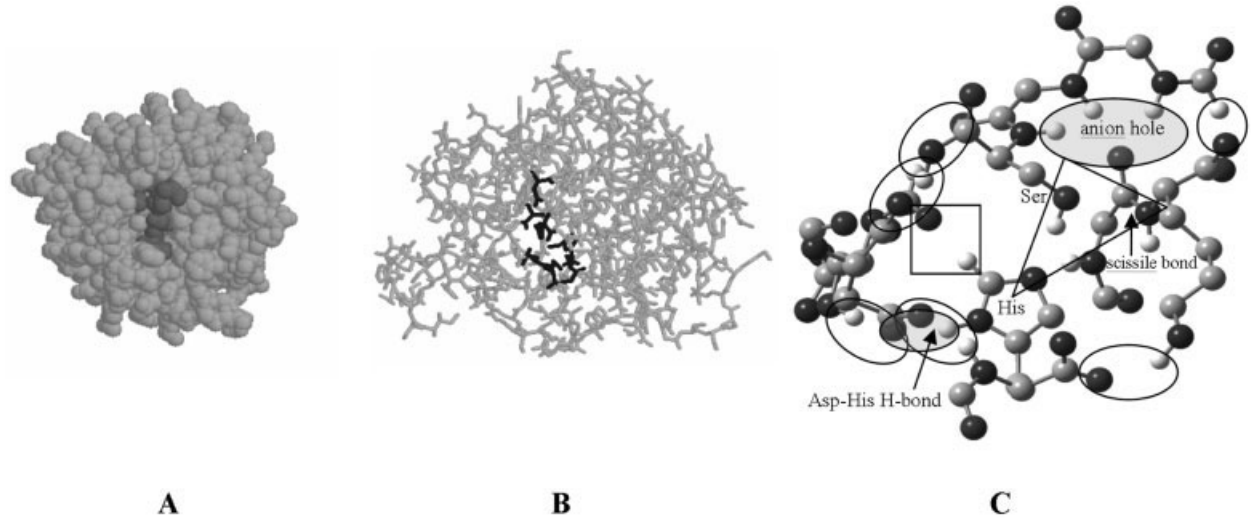


Fig. 1. Catalytic apparatus of chymotrypsin. Location of the catalytic pocket within the enzyme (**A**, **B**). Structure of the model of the catalytic pocket (**C**). To simplify matters, hydrogen atoms that do not have a role in catalysis or form hydrogen bond within the model, in addition to the side chain carbon atoms of Leu in the substrate's P1′ position, are removed. Empty ovals indicate hydrogen bonds present in the structure, gray ovals indicate the hydrogen bond between the catalytic Asp (smaller), and those in the oxyanion hole (larger). The square designates the interaction between C^{α}—H of catalytic His and the carbonyl oxygen of the imidazole ring, while the triangle represents the interaction of the Ser oxygen, His nitrogen, and the atoms of the bond to be cleaved.

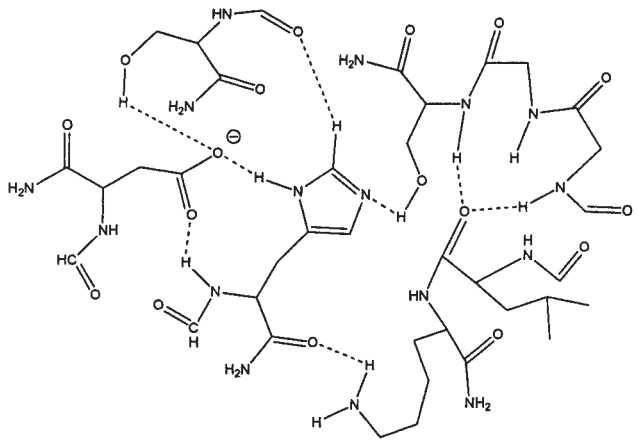
calculation is not dependent on any reparametrization or on alternative handling of boundary regions. The validity of our model was verified through the entire catalytic path of the enzyme and subsequently applied for various modified systems to test the effect of mutations, the role of water molecules and substrate conformational rigidity.

METHODS

All calculations were carried out using the Gaussian 98 software package.[40] Constrained and unconstrained geometry optimizations were completed at the RHF/3-21G level of theory.

The starting structure of the model was taken from the PDB[41] X-ray structure of 1GL1,[42] which was complemented with hydrogen atoms. This model contains the atoms of 18 amino acid residues involved in catalysis (Fig. 1). In our self-stabilized model, a total of eight amino acid residues are present in their entirety (Asp102, His57, Ser214, Gly193, Asp194, Ser195, and Leu30, Lys31, where the latter two are from the substrate at the P1 and P1′ positions, respectively) and the carbonyl or the amide groups of additional 10 residues are incorporated. Asp102, His57, and Ser214 are present as amino acid diamides, Gly193–Asp194–Ser195 is incorporated in form of an N- and C-protected tetra-amide and the Leu30–Lys31 subunit represents the substrate as a triamide unit. These fragments form a system of 135 atoms, among which 70 are heavy atoms. All atoms were treated at ab initio levels of theory. This model was submitted to full geometry optimization. The resultant structure was designated as the initial enzyme–substrate complex (ES). The catalytic mechanism was investigated through a reaction pathway computed in the form of a consecutive series of constrained geometry optimizations, which resulted in a total of 127 distinct structures. In other words, the reaction path is reconstructed from 126 individual steps. During all these optimizations, only one single atom–atom distance was fixed to a characteristic value; this was chosen as the reaction coordinate and all others are fully relaxed (step size was considered to be equal to 0.1 Å.). The minimum energy structures along the reaction pathway were subsequently optimized without any constraints. Although the components of the catalytic region are not connected by true chemical bonds or by dummy atoms, our molecular assembly did not loose its natural spatial orientation.

The energy profile of the catalytic steps along the reaction coordinates was obtained, and the intermediate conformers were analyzed to confirm the self-stabilized nature of our model. As the native conformation was retained in all steps of the catalytic process, we concluded that the model has met the criteria of self-stabilization. With this result at hand, slightly modified and deficient models were created from the original molecular assembly to gain information about the contribution of individual molecules in the assembly to catalysis. To achieve this, elimination of Asp102 (involved in stabilization of the charged His57) or Gly193 and Asp194 (stabilizing the tetrahedral carbon of the substrate) residues were studied. Next, the substrate was conformationally restricted, as well as the effect of the presence or absence of a water molecule in the early stage of the catalysis was evaluated. In the studies where the Asp or the anion hole was eliminated, no further geometry optimizations were performed. Geometries were generated from the optimized structures of the full system and submitted to single point calculations. In the model without Asp, all atoms of the aspartate diamide were deleted. In the model without the anion hole, residues 193, 194, and the backbone nitrogen



Scheme 1. Schematic molecular diagram of the self-stabilized model (model A) of the catalytic apparatus of chymotrypsin with its Lys–Leu substrate. Dashed lines indicate H-bonds, which help to maintain native supermolecular conformation. The side chain of Asp194 was removed to form a Gly.

atom of Ser195 were removed. A proton was then placed on the C^{α} of Ser195 1.07 Å from the carbon and the same bond angles of the original nitrogen atom.

The energy of the first catalytic step, that is, the formation of the first tetrahedral intermediate, was studied at high levels of theory on a model much smaller than the one outlined above, with partially fixed geometry to mimic the structural arrangement of the original assembly. In this model, a total of 27 constraints (bond, angle, and dihedral) were applied to maintain the geometric arrangement of the original model while optimizing the smaller model. Hence, these constraint optimizations mainly affected only the protons introduced at positions where a bond was broken. This smaller system contained only 19 heavy atoms, and could thus be treated at various quantum chemical levels [up to MP2/6-311++G(d,p)].

RESULTS AND DISCUSSION

The Self-Stabilizing Model

The primary goal of the present study was to engineer a model of the active site of chymotrypsin with its molecular environment composed of multiple noncovalently attached subunits, which can maintain the native molecular structure of the enzyme all along the complete catalytic cycle (Fig. 1, Scheme 1). On one hand, all amino acid residues primarily involved in the catalysis or in stabilization of the catalytic residues are present. On the other, the rest of the enzyme that shelters the central residues, and solvent molecules surrounding the system were not considered. Thus, the size of an otherwise large system is reduced to such an extent that the mechanism of chymotrypsin can be studied at the quantum chemical level of theory without making any compromises.

During the design of the model (called Model A), amino acid residues were selected from the enzyme. An amide moiety is present at the termination of each backbone chain included in the model. This form proved to be adequate in the study of conformational units of proteins and peptides.[43] The catalytic triad of Asp102, His57 and Ser195 constitutes the inner part or the "core" of the model. The substrate is included as a triamide with residues at the P1 and P1′ positions of a Leu–Lys unit, and was also taken from the structure of 1GL1. In addition, Ser214 is included in the model since it interacts with the H^{ϵ} of His57.[44] Last but not least, residues forming the anion hole of the tetrahedral intermediate were also incorporated within the model (residues 193 and 194, with the side chain of Asp194 removed in the model so as to form Gly). This model was found to be especially suitable for ab initio investigation as numerous hydrogen bonds are formed among the diamide bonds that maintain the overall structural arrangement of the catalytic apparatus of chymotrypsin in its natural state (Fig. 1). The RMSD value for all heavy atoms between the native protein structure and the first optimized ES model system is 2.15 Å, which is mainly due to the rearrangement of selected atoms to form H-bonds in the outer area of the model. The RMSD of the atoms forming the inner part of the model for the same two structures (His57 side chain, Asp102 side chain, Ser195, Gly194, Gly193, Ser214 backbone Leu30 backbone and Lys31 backbone) is 0.95 Å. For the heavy atoms of optimized structures throughout the catalytic reaction, the RMSD remains below 0.7 Å.

Selection of Theoretical Method

As the size of the designed self-stabilized model (model A) restricts the quantum chemical level that can be applied, we designed a smaller model (model B) based on the geometric arrangement of model A. Model B contains 19 heavy atoms from the core of model A. We have taken both ES and TC1 geometries of model A, applied them for model B and optimized with 27 constraints taken into account to maintain the relative orientation of the heavy atoms in the structure. The energy differences between these two structures were determined (Table I). The values show that calculations at the applied levels of theory provide various energy differences between ES and TC1. However, the values at the highest levels of theory applied [e.g., $\Delta E_{ES\leftrightarrow TC1}$(MP2/6-311++G(d,p)] = 22.9 kcal · mol^{-1}) are similar to that at the RHF/3-21G level of theory (e.g., $\Delta E_{ES\leftrightarrow TC1}$(RHF/3-21G) = 20.3 kcal · mol^{-1}). This phenomenon was observed in earlier investigations; namely, the relative energies found at the RHF/3-21G level of theory are, in several particular cases, as reliable as those from much higher levels due to a fortuitous cancellation of errors.[45,46] We therefore decided to use the RHF/3-21 level of theory, which allows for a complete evaluation of the catalytic cycle.

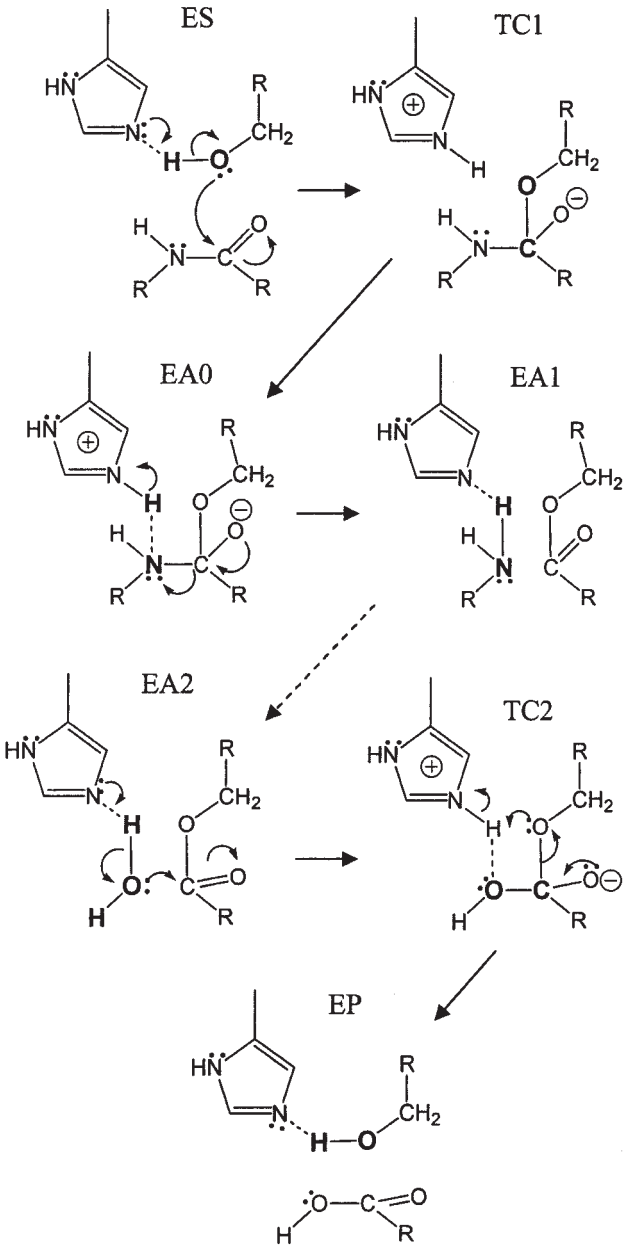
The Basic Catalytic Reaction

Six fully optimized minima were identified along the full reaction path of the catalysis of chymotrypsin at the RHF/3-21G level of theory. These stable supramolecular complexes are labeled from ES to EP (Scheme 2), where ES is the initial Enzyme Substrate complex, TC1 stands for the first tetrahedral complex intermediate, EA1 and EA2 are the Enzyme Adducts (acylenzyme intermediates), TC2

TABLE I. Energy Differences between ES and TC1 for Model B the Chymotrypsin Catalytic Pocket with Geometry Optimization at Various Levels of Theory

	Basis set			
Method	3-21G	6-31+G(d)	6-311++G(d,p)	6-311++G(3df,3pd)
RHF	20.3	36.1	37.0	—
B3LYP	—	29.2	30.4	30.5
PBEPBE	—	24.8	25.8	—
MP2	—	23.6	22.9	—

A number of constraints were set to maintain the geometry arrangement derived from the fully optimised large *ab initio* model. (All values are in kcal/mol)



Scheme 2. The seven structures of catalysis associated with local minima in the supramolecular complex. Arrows show the process of the catalytic reaction, dashed arrow indicate the step, where the Lys31 at the P1′ position is replaced by a water molecule.

is the second Tetrahedral Complex intermediate and EP is the enzyme product complex. The six minima are intercepted by five transition state structures labelled as follows: $TS_{ES\text{-}TC1}$, $TS_{TC1\text{-}EA1}$, etc. Therefore, the potential energy curve associated with the complete catalytic process is subdivided into fractions, or steps, and labeled according to the appropriate minima and/or TS.

The initial enzyme substrate complex (ES) was obtained by fully optimizing the molecular complex as retrieved and truncated from the X-ray structure of 1GL1 deposited in PDB (see Methods). The first part of the potential energy surface along the reaction coordinate (E + S to ES on Fig. 2) was calculated in a "reverse mode." Starting from the fully optimized enzyme–substrate complex ES, disintegration of the substrate from the binding pocket of the enzyme was achieved by separating them from each other with a step size of 1 Å. Geometry optimization was not performed at this step.

The second part of the potential energy curve from ES to TC1 (Fig. 2) was obtained by performing a constrained scan where the distance of O^{γ} (Ser195) and C′ (Leu30 of the substrate) were decreased by 0.1 Å from the initial optimized value of ES: d(O^{γ}—C′) = 3.27 Å (Scheme 2). In each step of the ES to TC1 scan (a total of 16 increments), geometry optimization of the entire molecular complex was completed. By decreasing the distance of the Ser side chain oxygen atom (O^{γ}) from the carbon atom of the carbonyl group (C′) of the substrate, d(O^{γ}—C′), from 3.27 to 1.67 Å, the proton transfer from O^{γ} (Ser195) to N^{ϵ} (His57) was spontaneously occurring. This proton transfer observed is an expected step according to the accepted mechanisms of catalysis. As a result, the imidazole ring of the catalytic histidine residue acquires a net positive charge, which is partly compensated by the catalytic aspartate (Asp102) anion. The tetrahedral intermediate structure, TC1, where d(O^{γ}—C′) = 1.67 Å, was fully optimized without any additional distance restraints. Along the reaction path, ES and TC1 minima are separated by a transition state, $TS_{ES\leftrightarrow TC1}$, the relative energy of which is the rate-limiting step of the overall catalytic reaction. Indeed, $TS_{ES\leftrightarrow TC1}$ was noticed along the potential energy curve and calculated to have an activation energy barrier of 20.3 kcal/mol and the minimum of TC1 is 15.0 kcal/mol above ES at the RHF/3-21G level of theory (Table II).

Although in our original theory we wanted to determine the catalytic efficiency of the catalytic pocket on our model

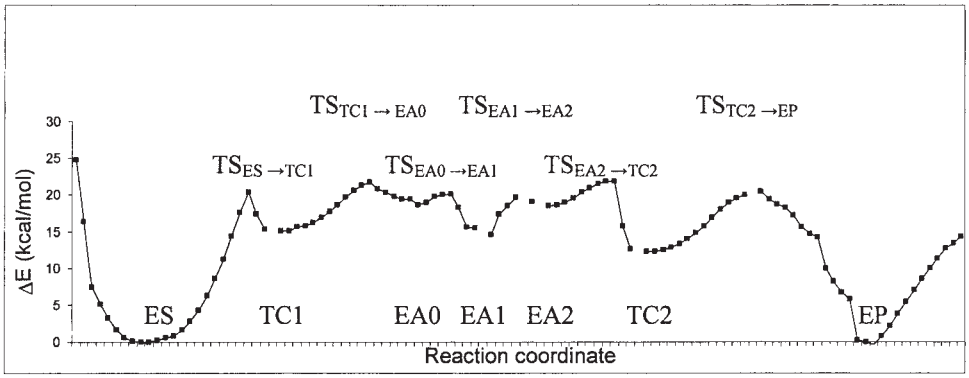


Fig. 2. Energy diagram of the catalytic mechanism of chymotrypsin. A movie showing the structures at all stages of the catalysis is deposited as Supplementary Material.

TABLE II. Activation Energies (ΔE) for the Formation of Tetrahedral Intermediates within the Catalytic Mechanism of the Chymotrypsin Model.

	RHF/3-21G
Tetrahedral intermediate #1 $\Delta E(TS_{ES \rightarrow TC1})$	20.3
Tetrahedral intermediate #2 $\Delta E(TS_{EA1 \rightarrow [EA2] \rightarrow TC2})$	15.7

All values are in kcal/mol.

system without environment, it is to be noted that the vacuum is also a type of "environment." So we carried out PCM calculations using the dielectric constant of water and got 13.2 kcal/mol for TC1, which is only 1.8 kcal/mol different from the calculation in vacuum. This result certifies our assumption, that the environment has no significant effect on the processes in the core of such a large model.

During the second step of the catalysis (TC1 to EA1), the distance between the H^{ϵ} proton of the imidazole ring of His57 and the nitrogen atom of the amide of Lys31 residue in the substrate, d[H^{ϵ}(His57)—N^{amide}(Lys31)] gradually decreases from the initial value of 3.29 to 0.99 Å. At the same time, when the amide N atom acquires the proton in a synchronized manner, the C′—N distance of the scissile amide bond, d[C′(Leu30)—N^{amide}(Lys31)]), increases "automatically" from the equilibrium distance (1.45 Å) to a clearly augmented d[C′—N] value of 1.64 Å. (A bond distance of 1.4 Å corresponds to the equilibrium C—N distance of a primary amide bond.) Furthermore, the true double bond nature of the carbonyl group of the scissile amide is reestablished again in a concerted reaction, as the distance of the carbon and the oxygen atoms, d(O—C′), decreases from 1.34 to 1.31 Å. The energy barrier to break the amide bond in the tetrahedral state (TC1 → EA1 and TC2 → EP) is below 10 kcal/mol.

Following the break of the amide bond, the resultant fragment with a primary amine is replaced by a water molecule, which forms the EA2 supramolecular complex.

In the third step of catalysis (from EA2 to TC2), the distance between the oxygen atom of the water molecule (O^{water}) and the carbon atom of the carbonyl group, C′, of Leu30 is decreased gradually from an initial distance of 2.53 to 1.48 Å. This results in a bond formation between the two atoms, and the formation of the second tetrahedral intermediate (TC2 in Scheme 2). During the CO single-bond formation, one of the protons from the water molecule is spontaneously transferred to the imidazole ring of the central His57. Once again, bond formation and proton transfer are organized in a synchronous manner, which is similar to the observation from the second elementary step of the catalytic reaction (TC1 → EA1). However, as long as a single C—N bond is broken in TC1 → EA1, a C—O single bond is formed in EA2 → TC2. Nevertheless, proton transfer is an induced and concerted rearrangement in both of the elementary catalytic steps, and is the response of a self-stabilized model to the formation or elimination of chemical bonds.

Formation of the second tetrahedral intermediate has an activation energy of 15.7 kcal/mol at theRHF/3-21G. This value is significantly lower then the one calculated for the formation of the first tetrahedral intermediate (Table II).

The fourth step of the catalytic process contains the hydrolysis of the O-acyl serine complex. During this step [TC2 → EP (Scheme 2 and Fig. 2)], the covalent bond connecting the O^{γ} atom of the catalytic serine, Ser195, and the carbon atom of the carbonyl groups of the substrate (Leu30) breaks. As a result, a cleaved substrate with a free carboxylic acid end is generated, in addition to the regeneration of the unsubstituted Ser195 and the neutral His57 amino acid residues. The chosen reaction coordinate along this step was the distance between O^{γ} (Ser195) and C′ (Leu30) atoms, which was increased gradually from 1.45 to 6.48 Å. Again, the increase of the d[O^{γ}—C′] bond distance initiates the proton transfer from the imidazole ring of His57 to the side-chain oxygen atom of Ser195 in a concerted manner. The last minimum of the catalytic cycle is the enzyme product complex, EP, where the d(O^{γ}—C′) distance is 3.48 Å. In comparison to the value of 3.27 Å measured for the initial enzyme substrate complex ES, the difference is only about 0.2 Å, suggesting that the product can easily withdraw from the enzyme. Indeed, the enzyme product complex EP was disintegrated by simply "pulling out" the product from the binding pocket of the enzyme (Fig. 2).

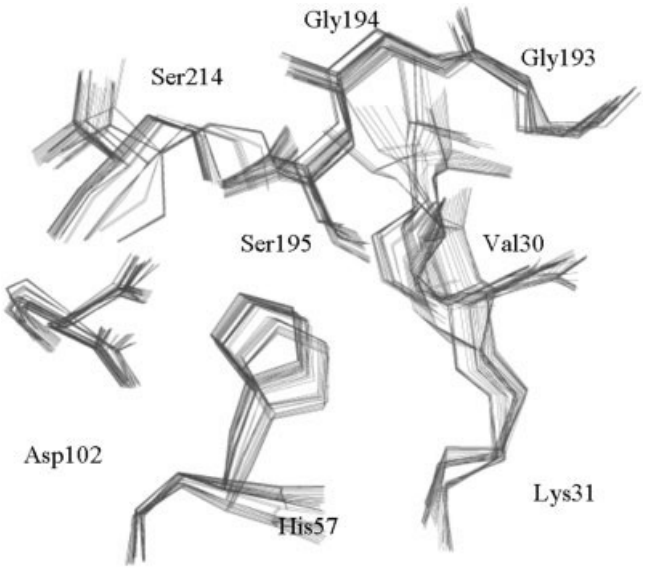
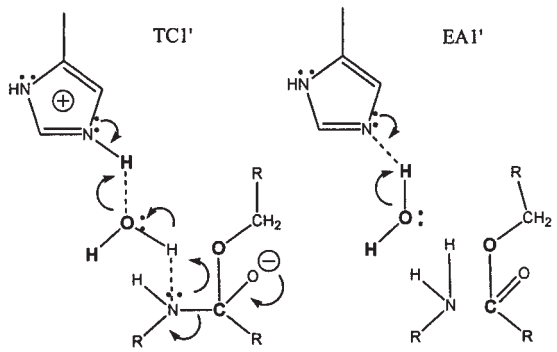


Fig. 3. Variation of the structure of the catalytic apparatus of the chymotrypsin model throughout the catalytic process until the breaking of the amide bond. (The side chain of Asp194 was removed; in effect, Gly was used instead of Asp.)



Scheme 3. Schematic representation of the alternative mechanism where TC1- and EA1-type minima containing the so-called "early water" molecules, called as TC1′ and EA1′, are used to describe the details of the full catalytic cycle. (Remaining structures of this mechanism are the same as presented in Scheme 2.)

It is essential to emphasize that the "truncated protein" model system (model A) was able to simulate each and every step of the catalytic cycle and relax at the end as an integrated entity so that it could be reused at a next cycle. Note that along the reaction coordinates, full optimizations were completed. Nevertheless, the supramolecular complex remained as a complete unit due to the carefully engineered N- and C-protecting groups, hydrogen bonds, and noncovalent interactions. As the atoms of each amino acid residues are allowed to move during optimization, the overall structure of the catalytic apparatus changes slightly throughout the course of the catalytic cycle. However, these variations do not lead to the fundamental restructuring of the model, but rather imply that the enzyme pocket has some plasticity or flexibility according to changing conditions (Fig. 3).

The structures of the catalytic process were collected to a movie. The movie contains the optimized structures at the various stages of the catalysis and shows the process of catalysis as it would look like if the dynamical movements of the protein would be detached (Supplementary Material).

The total heavy atom RMSD value of the overall complex along the reaction coordinates is only 0.68 Å for the 41 geometries between ES and EA1. The same value is 0.48 Å if only the heavy atoms of the enzyme are considered. This value is even smaller (0.34 Å) when only the heavy atoms of the amino acid residues that form the catalytic triad and anion hole are considered. The RMSD values for the backbones of the latter two scenarios are as small as 0.43 and 0.29 Å, respectively. These features of our model system suggest that it is a self-stabilized unit that is a suitable model of the catalytic pocket of chymotrypsin, and could also be used to estimate the energy cost of intriguing modifications such as point mutation(s), introduction of early water molecule mediated mechanism, rigid substrates, etc.

Application of the Self-Stabilized Model

The original, self-stabilized model of chymotrypsin is used after some alterations to determine the role of particular elements and properties within the catalytic pocket. Water is involved in the catalytic processes of chymotrypsin, and may be regarded as a part of the catalytic pocket; however, it can enter the catalytic process in several different stages. We investigated alternatives for the role of water during catalysis to reveal which of the routes requires lower activation energy barriers. The role of two additional elements of the catalytic pocket were studied, aspartate and the anion hole.

These computations include only the first of the two large energy barriers along the reaction coordinates, $TS_{ES \rightarrow TC1}$ and $TS_{EA2 \rightarrow TC2}$ (Fig. 2). Both of these barriers precede the formation of a tetrahedral intermediate (TC1 or TC2). Formation of both of the tetrahedral intermediates is believed to be critical in the overall enzymatic procedure, because natural inhibitors can be found that prevent the formation of these states.

Models for the Catalytic Mechanism with Early Water Insertion

As summarized in Scheme 2, the commonly described fundamental catalytic cycle of chymotrypsin does not involve a water molecule in the formation of the first tetrahedral intermediate, TC1 (Scheme 2). However, an explicit water molecule may be positioned between the imidazole ring of the catalytic histidine and the amide nitrogen atom of the scissile bond (Scheme 3).

According to the widely accepted mechanism of the catalytic process the imidazole ring of the catalytic histidine directly donates its proton, which was acquired from Ser195, to the nitrogen of the scissile amide bond (from TC1 to EA1, Scheme 2). Another way may also be hypothesized, when the proton of the imidazole ring is captured not by the amide, but by a water molecule, which itself donates a proton to the nitrogen atom of the amide group of

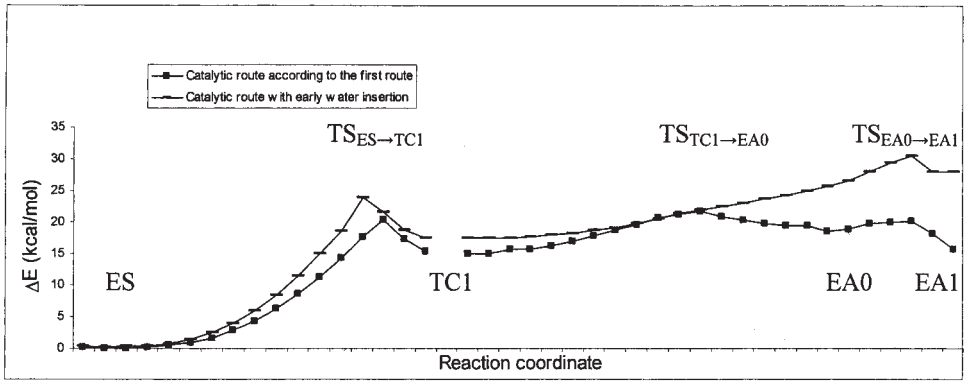


Fig. 4. Catalytic routes of chymotrypsin. Early water insertion to the catalytic process increases $TS_{ES \rightarrow TC1}$ and $TS_{EA1 \rightarrow EA2}$.

the scissile bond (from TC1′ to EA1′, Scheme 3). The latter mechanism results in the same cleaved product. Using the self-stabilized model, one can scan and establish the potential energy curves associated with each route, and decide which of the two is most likely to occur (Fig. 4). By comparing the two energy barriers associated with $TS_{ES \rightarrow TC1}$,and $TS_{ES \rightarrow TC1'}$, it is clear that without the water molecule, the complex is more stable by almost 4 kcal/mol. In addition, the potential energy curve is continuously lower than the one where the explicit water molecule was inserted (Fig. 4). Thus, the model reproduced the energetic advantage of the accepted mechanism of the catalysis reported in Scheme 2.

Effect of Removal of Aspartate and Anion Hole from the Catalytic System

The importance of the elements of the catalytic apparatus was investigated in several experimental and theoretical studies.[9,47,48] In the experimental work of Corey and Craik the D102N mutant enzyme had a decreased k_{cat} value to 0.003% of the native enzyme.[49] The replacement of Asp102 by neutral residues in the work of Warshel and coworkers[9] destabilized the transition state by more than 4 kcal/mol in calculations that consider only electrostatic effects. Using a more sophisticated EVB/FEP calculations in the same article, they reported 6.2 ± 1.6 kcal/mol for the same effect, which including 1.8 kcal/mol for the histidine residue to turn into the adequate orientation. Ishida and Kato[50] reported an increase of the activation energy of 7.3 and 5.1 kcal/mol for an Asn mutant chymotrypsin considering single proton and double proton transfer mechanism, respectively. They used QM/MM calculations combined with MD-FEP simulations defining the QM region mainly for the side chains involved, which involves the known issues arising from difficulties in describing the boundary regions. Shokhen and Albeck[21] determined the contribution of Asp as 3.79 kcal/mol. A study on the on the stability of the tetrahedral complex by Westler et al.[51] determined a change of strength of H-bond between His and Asp was determined to be 11 kcal/mol. (Note that this value is not in direct relationship with the catalytic effect of asp.) These latter two articles[21,51] both used gas phase calculations on a model system, which shall be taken into account in conclusions draw for proteins. In summary, the contribution of the catalytic aspartate was measured and some calculations were able to reproduce it. The stabilizing effect of the oxyanion hole was also investigated and Warshel et al.[52] found that a hydrogen bond between a protein residue and a charged group has a catalytic effect of 4 kcal/mol, which agrees with the genetic modification experiment.

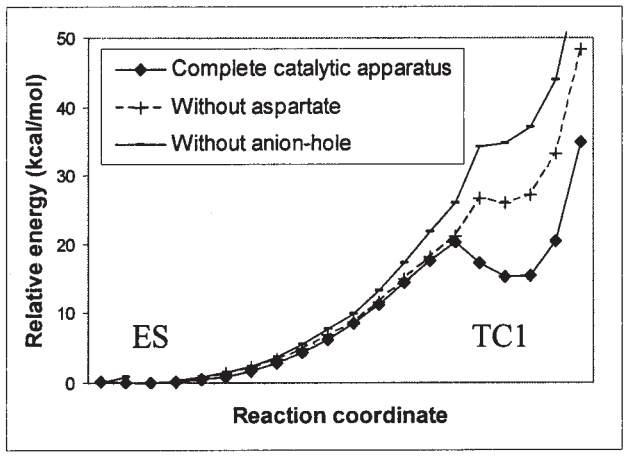


Fig. 5. Different energy profiles for the formation of the first tetrahedral intermediate with complete and incomplete catalytic apparatus. ES represents the initial enzyme substrate complex, while TC1-type molecular conformers are the different variants of the first tetrahedral intermediate structures.

These two effects were investigated in our model of the catalytic pocket and the destabilizing effect of both elements of the catalytic apparatus could be reproduced. The potential energy curve associated with the complete catalytic apparatus in form of the self-stabilized model, as outlined in Scheme 2 and Figure 2, is used as the reference state or reference potential energy curve (Fig. 5).

Both the absence of the aspartate residue, Asp102, and the removal of the anion hole (technical description of removal is given in the Methods) increases the energy barrier, $\Delta E(TS_{ES \rightarrow TC1})$ of the formation of the first tetrahedral intermediate in our model. Relative to the reference potential energy curve, the activation barrier at

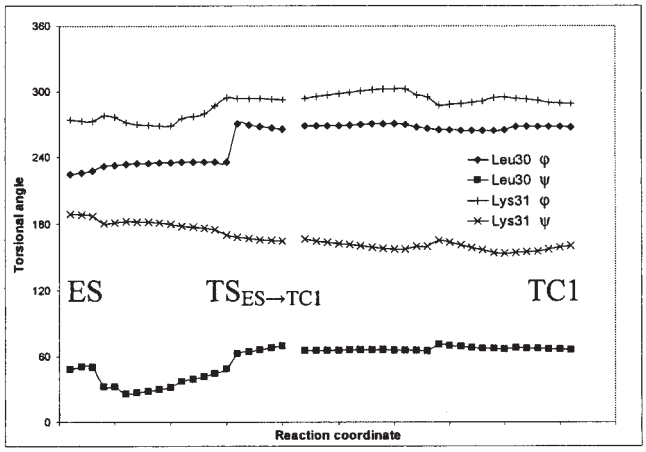


Fig. 6. Variation of the four backbone torsional angles of the substrate residues at the P1 and P1′ positions. Potential energy curves are plotted along the reaction coordinate of the enzyme substrate complex through the tetrahedral intermediate until the amide bond is broken.

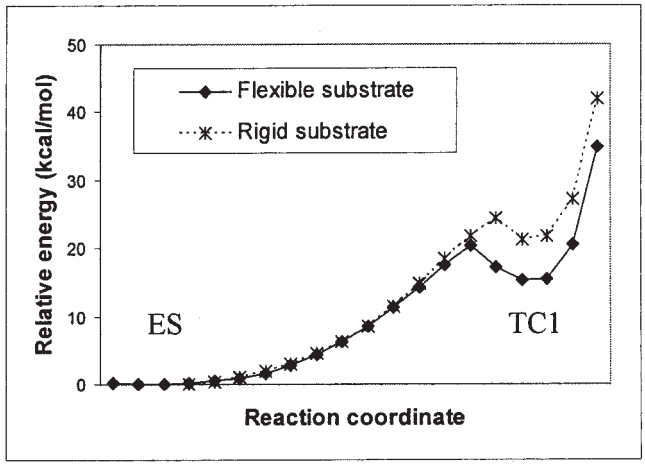


Fig. 7. Energy profiles of the formation of the first tetrahedral intermediate TC1 from ES, with flexible and rigid substrate. Calculations were performed at the RHF/3-21G level of theory.

the RHF/3-21G level of theory increases by 6.4 kcal/mol when Asp102 is removed, and by 13.9 kcal/mol if the anion hole is removed. This reflects a 32 and 68% increase in the activation energy barrier. In addition, the resultant transition state structure has ~0.1 Å shorter serine oxygen and the carbonyl carbon atom distance when one of the two stabilizing factors is eliminated. In the view of theoretical and experimental investigations in the literature, both of these values are high in some degree arising from overestimation of H-bond strength of the applied quantum mechanical level and/or from the applied preorganization coming from the complete catalytic apparatus.

Effect of Rigidity of the Substrate

Several studies indicate that canonical serine protease inhibitors hinder substrate cleavage by their structural rigidity.[53,54] The variation of the φ and ψ torsional angles at the P1 and P1′ positions of the amino acids composing the substrate were therefore monitored (Fig. 6). All four backbone torsional angles of the P1 and P1′ amino acid residues change within a limited range, ζ ±30°, throughout the first half of the catalytic cycle, with the exception of ϕ of Leu30. This dihedral angle suffers major change at the step where the proton "jumps" from the O^{γ} of Ser195 to the N^{ϵ} of His57 on the way of forming TC1. A canonical inhibitor of serine protease is believed to have a reduced conformational mobility.[13,14] Thus, a potent serine protease inhibitor was simply modeled by performing the calculations along the previously described reaction coordinates with fixed torsional angle values. The backbone dihedrals of the substrate model were fixed at that angle values found in the optimized structure of the ES complex. As a result of local backbone conformational restrictions, a considerable amount of conformational rigidity is imposed on the model system (Fig. 7). In these circumstances, one would expect there to be an increase of the relative energy, $\Delta E(TS_{ES \rightarrow TC1})$. Indeed, the energy barrier of the formation of the first tetrahedral intermediate increased by 4.1 kcal/mol. According to the Arrhenius equation this energy value corresponds to three orders of magnitude decrease of reaction rate caused by the rigidity of the inhibitor.

CONCLUSION

A self-stabilized model of the catalytic pocket of chymotrypsin was shown to provide an environment for the catalytic reaction similar to that present in the enzyme. The model maintains its natural-like structural arrangement throughout the geometry optimizations during the complete catalytic cycle without introducing any external constraints. Energy diagrams indicate that the activation energy of the reaction in the model is close to that in the enzyme. Activation energy barriers for the formation of the two tetrahedral intermediates are 20.3 and 15.7 kcal/mol, respectively. The absence of the catalytic aspartate and the anion hole were shown to increase the energy barrier of the formation of the first tetrahedral intermediate. In addition, the conformational rigidity of the substrate also increased the activation energy of the formation of the first tetrahedral intermediate, a phenomenon that reveals the inhibition mechanism of the canonic serine protease inhibitors.

ACKNOWLEDGMENTS

The authors thank Imre G. Csizmadia, Ilona Hudáky, Zoltán Mucsi, and Suzanne K. Lau for helpful discussions. The ELTE computer facility was used for several computations.

REFERENCES

1. Antal J, Pál G, Asbóth B, Buzás Zs, Patthy A, GráfL. Specificity assay of serine proteinases by reverse-phase high-performance liquid chromatography analysis of competing oligopeptide substrate library. Anal Biochem 2001;288:156–167.
2. Birktoft JJ, Kraut J, Freer ST. Detailed structural comparison between charge relay system in chymotrypsinogen and in alpha-chymotrypsin. Biochemistry 1976;15:4481–4485.
3. Hunkapiller MW, Smallcombe SH, Whitaker DR, Richards JH. Carbon nuclear magnetic resonance studies of the histidine resi-

due in α-lytic protease. Implications for the catalytic mechanism of serine proteases. Biochemistry 1973;12:4732–4743.
4. Bachovchin WW, Roberts JD. Nitrogen-15 Nuclear magnetic resonance spectroscopy. The state of histidine in the catalytic triad of α-lytic protease. Implications for the charge-relay mechanism of peptide-bond cleavage by serine proteases. J Am Chem Soc 1978;100:8041–8047l.
5. Frey PA, Whitt SA, Tobin JB. A low-barrier hydrogen bond in the catalytic triad of serine proteases. Science 1994;264:1927–1930.
6. Shutz CN, Warshel A. The low barrier hydrogen bond (LBHB) proposal revisited: the case of the Asp . . . His pair in serine proteases. Proteins Struct Funct Bioinf 2004;55:711–723.
7. Shokhen M, Albeck A. Is there a weak H-bond3LBHB transition on tetrahedral complex formation in serine proteases? Proteins Struct Funct Bioinf 2004;54:468–477.
8. Ash EL, Sudmeier JL, Day RM, Vincent M, Torchilin EV, Haddad KC, Bradshaw EM, Sanford DG, Bachovchin WW Unusual H-1 NMR chemical shifts support (His) C-epsilon 1-H center dot center dot center dot O = C H-bond: proposal for reaction-driven ring flip mechanism in serine protease catalysis. Proc Natl Acad Sci USA 2000;97:10371–10376.
9. Warshel A, Náray-Szabó G, Susman F, Hwang JK. How do serine proteases really work? Biochemistry 1989;28:3629–3637.
10. Náray-Szabó G, Enzyme mechanisms: interplay of theory and experiment. J Mol Struct (Theochem) 2000;500:157–167.
11. Warshel A, Parson WW. Dynamics of biochemical and biophysical reactions: insight from computer simulations. Q Rev Biophys 2001;34:563–679.
12. Villa J, Warshel A. Energetics and dynamics of enzymatic reactions. J Phys Chem B 2001;105:7887–7907.
13. Brauer ABE, Kelly G, Matthews SJ, Leatherbarrow RJ. The H-1-NMR solution structure of the antitryptic core peptide of Bowman-Birk inhibitor proteins: a minimal "canonical loop." J Biomol Struct Dyn 2002;20:59–70.
14. Szenthe B, Gaspari Z, Nagy A, Perczel A, Graf L. Same fold with different mobility: backbone dynamics of small protease inhibitors from the desert locust, *Schistocerca gregaria*. Biochemistry 2004; 43:3376–3384.
15. Kaslik G, Kardos J, Szabó E, Szilágyi L, Závodszky P, Westler WM, Markley JL, Gráf L. Effects of serpin binding on the target proteinase: global stabilization, localized increased structural flexibility, and conserved hydrogen bonding at the active site. Biochemistry 1997;36:5455–5464.
16. Stone SR, le Bonniec BF. Inhibitory mechanism of serpins. Identification of steps involving the active-site serine residue of the protease. J Mol Biol 1997;264:344–362.
17. Bender ML, Kézdy FJ, Gunter CR. The anatomy of an enzymatic catalysis. α-Chymotrypsin. J Am Chem Soc 1964;86:3714–3721.
18. Warshel A, Russel S. Theoretical correlation of structure and energetics in the catalytic reaction of trypsin. J Am Chem Soc 1986;108:6569–6579.
19. Warshel A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. J. Biol Chem 1998;273: 27035–27038.
20. Stanton RV, Peräkylä1 M, Bakowies D, Kollman PA. Combined ab initio and free energy calculations to study reactions in enzymes and solution: amide hydrolysis in trypsin and aqueous solution. J Am Chem Soc 1998;120:3448–3457.
21. Shokhen M, Albeck A. Factors determining the relative stability of anionic tetrahedral complexes in serine protease catalysis and inhibition. Proteins Struct Funct Genet 2000;40:154–167.
22. Perczel A, Hudáky P, Füzéry AK, Csizmadia IG. Stability issues of covalently and noncovalently bonded peptide subunits. J Comput Chem 2004;25:1084–1100.
23. Bakowies D, Kollman PA. Theoretical study of base-catalyzed amide hydrolysis: gas- and aqueous-phase hydrolysis of formamide. J Am Chem Soc 1999;121:5712–5726.
24. Štrajbl M, Florian J, Warshel A. Ab initio evaluation of the potential surface for general base-catalyzed methanolysis of formamide: a reference solution reaction for studies of serine proteases. J Am Chem Soc 2000;122:5354–5366.
25. Pliego JR Jr, Riveros JM. A theoretical analysis of the free-energy profile of the different pathways in the alkaline hydrolysis of methyl formate in aqueous solution. Chem Eur J 2002;8:1945–1953.
26. Wu Z, Ban F, Boyd RJ. Modeling the reaction mechanisms of the amide hydrolysis in an N-(o-carboxybenzoyl)-L-amino acid. J Am Chem Soc 2003;125:6994–7000.
27. Nemukhin AV, Topol IA, Burt SK. Energy profiles for the rate-limiting stage of the serine protease prototype reaction. Int J Quantum Chem 2002;88:34–40.
28. Warshel A, Levitt M. Theoretical studies of enzymic reactions—dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. J Mol Biol 1976;103:227–249.
29. Lee FS, Chu ZT, Warshel A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the polaris and enzymix programs. J Comp Chem 1993;14:161–185.
30. Topf M, Várnai P, Richards WG. Quantum mechanical/molecular mechanical study of three stationary points along the deacylation step of the catalytic mechanism of elastase. Theor Chem Acc 2001;106:146–151.
31. Topf M, Várnai P, Richards WG. Ab initio QM/MM dynamics simulation of the tetrahedral intermediate of serine proteases: insights into the active site hydrogen-bonding network. J Am Chem Soc 2002;124:14780–14788.
32. Warshel A, Weiss RM. An empirical valence bond approach for comparing reactions in solutions and in enzymes. J Am Chem Soc 1980;102:6218–6226.
33. Carloni P. Density functional theory-based molecular dynamics of biological systems. Quantum Struct Act Relat 2002;21:166–172.
34. Bentzien J, Muller RP, Florian 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.
35. 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.
36. Daggett V, Schröder S, Kollman PA. Catalytic pathway of serine proteases: classical and quantum mechanical calculations. J Am Chem Soc 1991;113:8926–8935.
37. Štrajbl M, Florian J, Warshel A. Ab initio/LD studies of chemical reactions in solution: reference free-energy surfaces for acylation reactions occurring in serine and cysteine proteases. Int J Quantum Chem 2000;77:44–53.
38. Nemukhin AV, Grigorenko BL, Rogov AV, Topol IA, Burt SK. Modeling of serine protease prototype reactions with the flexible effective fragment potential quantum mechanical/molecular mechanical method. Theor Chem Acc 2004;111:36–48.
39. Szabó E, Böcskei Zs, Náray-Szabó G, Gráf L. The three-dimensional structure of Asp189Ser trypsin provides evidence for an inherent structural plasticity of the protease. Eur J Biochem 1999;263:20–26.
40. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Salvador P, Dannenberg JJ, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox J, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA. Gaussian 98, Revision A.11.1; Gaussian, Inc.: Pittsburgh, PA, 2001.
41. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235.
42. Roussel A, Mathieu M, Dobbs A, Luu A, Cambillau C, Kellenberger C. Complexation of two proteic insect inhibitors to the active site of chymotrypsin suggests decoupled roles for binding and selectivity. J Biol Chem 2001;276:38893–38898.
43. Perczel A, Ángyán JG, Kajtár M, Viviani W, Rivail JL, Marcoccia JF, Csizmadia IG. Peptide models .1. Topology of selected peptide conformational potential-energy surfaces (glycine and alanine derivatives). J Am Chem Soc 1991;113:6256–6265.
44. Derewenda ZS, Derewenda U, Kobos PM. (His)c-epsilon-h. . .o═c hydrogen-bond in the active-sites of serine hydrolases. J Mol Biol 1994;241:83–93.
45. Perczel A, Farkas Ö, Jákli I, Topol IA, Csizmadia IG. Peptide models. XXXIII. Extrapolation of low-level Hartree-Fock data of peptide conformation to large basis set SCF, MP2, DFT, and

CCSD(T) results. The Ramachandran surface of alanine dipeptide computed at various levels of theory. J Comput Chem 2003;24: 1026–1042.
46. Perczel A, Jákli I, Mcallister MA, Csizmadia IG. Relative stability of major types of beta-turns as a function of amino acid composition: a study based on ab initio energetic and natural abundance data. Chem Eur J 2003;9:2551–2566.
47. Baeten A, Maes D, Geerlings P. Quantum chemical study of the catalytic triad in subtilisin: the influence of amino acid substitutions on enzymatic activity. J Theor Biol 1998;195:27–40.
48. Russell AJ, Fersht AR. Rational modification of enzyme catalysis by engineering surface-charge. Nature 1987;328:496–500.
49. Corey DR, Craik CS. An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. J Am Chem Soc 1992;114:1784–1790.
50. Ishida T, Kato S. Role of Asp102 in the catalytic relay system of serine proteases: a theoretical study. J Am Chem Soc 2004;126: 7111–7118.
51. Westler WM, Weinhold F, Markley JL. Quantum chemical calculations on structural models of the catalytic site of chymotrypsin: comparison of calculated results with experimental data from NMR spectroscopy. J Am Chem Soc 2002;124:14373–14381.
52. Warshel A, Sussman F, Hwang JK. Evaluation of catalytic free energies in genetically modified proteins. J Mol Biol 1988;201:139–159.
53. Chen CP, Hsu CH, Su NY, Lin YC, Chiou SH, Wu SH. Solution structure of a Kunitz-type chymotrypsin inhibitor isolated from the elapid snake *Bungarus fasciatus*. J Biol Chem 2001;276:45079–45087.
54. Kojima S, Hisano Y. Requirement for hydrophobic Phe residues in *Pleurotus ostreatus* proteinase A inhibitor 1 for stable inhibition. Protein Eng 2002;15:325–329.