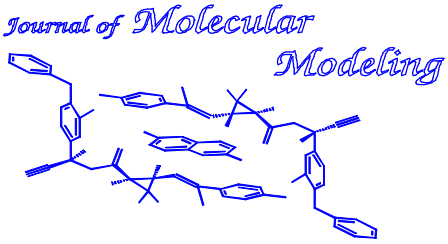
J. Mol. Model. **1996**, *2*, 149 – 159



Substrate Specifity of Chymotrypsin. Study of Induced Strain by Molecular Mechanics

Bernd Kallies*, Rolf Mitzner

Institut für Physikalische und Theoretische Chemie, Universität Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany (kallies@serv.chem.uni-potsdam.de)

Received: 5 March 1996 /Accepted: 4 June 1996 /Published: 13 June 1996

Abstract

Acylenzyme intermediates, produced by transfer of the acyl portions of selected natural substrates onto the catalytic serine hydroxyl of the serine protease chymotrypsin, were modeled with the AMBER force field. The obtained structures were used to calculate interaction and deformation energies. A set of 32 geometry variables were extracted out of each structure. They describe deformation effects specific for each substrate. It is shown by statistical analyses, that the interaction and deformation energies correspond to measured substrate reactivities. The extracted geometry variables are able to reproduce this dependency through multivariante statistical methods. These analyses suggest that there exist specific deformations of both the substrate and the enzyme portion, which are related to substrate reactivity. The geometry changes observed for high specific substrates are interpreted in terms of mechanistical requirements of the enzymatic reaction. The obtained model validates the hypothesis of induced strain as possible source of substrate specifity of chymotrypsin.

Keywords: serine protease mechanism, force field calculation, substrate specifity

Introduction

The specifity of biochemical receptors and its sources has long been subject of investigation. It is given that the fit of the three-dimensional structure and the complementarity of the surface properties of a ligand to a receptor site are necessary conditions for the biological activity of the ligand. These conditions may be associated with Fischer's „lock and key" picture [1]. It holds for most receptors, which only immobilize a ligand. Our current understanding of this principle is the source of the great success of drug design by methods using quantitative structure-activity relationships [2].

If a receptor has to perform a chemical reaction with an immobilized ligand like enzymes do, there exist additional aspects of specifity. The specifity of an enzyme can be expressed by reached reaction rates, not by binding coefficients only. At least two hypotheses were developed accessing sources of enzyme specifity. The first uses Pauling's theory of transition state stabilization [3]. This hypothesis is used mainly to explain the catalytic power of enzymes in comparison to the uncatalyzed reaction. It also explains different reaction rates by different interactions between transition states of several substrates and the enzymatic active site. The second „anti-Pauling" hypothesis states induction of conformational strain into the substrate and the enzyme prior to the reaction by use of binding energy, which is produced by substrate immobilization [4]. This strain may be productive in terms of a catalytic effect, if the conformational changes lie on the reaction coordinate. These two additional sources of receptor specifity are not subject of common QSAR studies, since their implementation requires other structure descriptors than those of isolated substrates.

* *To whom correspondence should be addressed*

The presented work represents an attempt to develop QSAR for an enzymatic reaction, which base on variables describing the substrates in the active site and structure changes induced by substrate binding. These variables were taken from enzyme-substrate adducts, which were obtained through molecular mechanics. We chose the deacylation reaction of the serine protease chymotrypsin acylated with several natural amino acid substrates as object for our studies.

Background [5]

Serine proteases like trypsin, chymotrypsin or subtilisin, as well as serine esterases, cysteine proteases and some mechanistically related lipases are known to function through a two-step mechanism. After immobilization of an ester or amide substrate the acyl portion of the substrate is transferred onto the hydroxymethyl sidechain of an active site serine residue (acylation reaction). This step forms the first hydrolysis product and an ester intermediate (acylenzyme). The latter species is cleaved in a subsequent step by solvent water, releasing carboxylate and the regenerated enzyme (deacylation reaction). Both steps are assisted by a histidine residue hydrogen-bonded to the serine sidechain and an aspartate residue H-bonded to the histidine sidechain (general base catalysis). Each reaction step should proceed via the usual addition-elimination mode of most acyl-transfer reactions, which involves formation of a short-living tetrahedral intermediate by attack of a nucleophile onto a carbonyl carbon. The formal negative charge resting at the carbonyl oxygen (oxyanion) of these intermediates is stabilized by interactions between the oxyanion and enzyme residues, which form the „oxyanion hole“. It is built by NH-portions of enzymatic carboxamides (protein backbone or asparagine sidechain). The stereospecifity of chymotrypsin in discriminating between L- and D-amino acid substrates was explained through various force field calculations [6-8] by interaction of the NH-group of the bound substrate and a C=O group of the protein backbone of the binding pocket.

The rate limiting step of the overall hydrolysis reaction depends on the general carbonyl activity of the substrate. Amides often show rate limiting acylation. In the case of labile ester substrates, deacylation is the rate limiting step. This behaviour enables one to measure individual rate constants for acylation and deacylation. Since large sets of rate constants based on the esterase function of chymotrypsin are published, we chose this system. Among the reactivity data the deacylation and turnover rates of ester substrates are the most reliable ones, so we decided to study this individual reaction.

Premises

Our work is based on several assumptions, which have to be explained first in order to guide through the results. The central idea is that an acylenzyme is formed with every substrate studied, and that this acylenzyme can be handled with common empirical force fields. We assume that the substrate immobilization and the acylation reaction were successful. This assumption holds not, if the free substrate does not fit the enzymatic active site. We performed no docking studies, but utilized the covalent bond between the substrate and enzyme portion of an acylenzyme and the behaviour of the system induced by it, instead. This bond behaves like a true ester bond through force field calculations. The substrate is already placed in the active site, and it can not leave out. In this case, force field calculations will converge to a structural compromise. It includes deformation of both the substrate and the enzyme portion. The careful study of these conformational changes enables us to detect outliers, for which the assumption of successful binding does not hold. It is the basis for the development of variables for QSAR in our study, too.

Another assumption is related to the role of water molecules. The re-solvation of both the substrate and the binding site (disruption of the hydrate shells and solvation by a new micro-environment) is an important contribution to the free association energy produced during the step of immobilization of a ligand at a receptor site [2]. We did not study the initial association process, but calculated an association energy equivalent for the special case of covalent bound substrates. This quantity we define as the sum of the interaction energy between the enzyme and the substrate part of an acylenzyme and deformation energies of these parts, using their free states as reference (see below). Water molecules were not explicitly incorporated in all these calculations. Since the neglect of hydrate shells might be subject of criticism, we consider the errors produced by this approach. At first we look at geometry changes and deformation energy contributions to calculated association energies. Since the structure of the free enzymatic active site without water molecules should differ from that with water molecules included, the calculated deformation energy of the enzymatic part of an acylenzyme bears an error. But its amount should be equal in all acylenzyme structures, so it does not contribute to the relative relationships we derive in this paper. The assumption of equal geometries of isolated substrates in the gas phase and a solvated state is common, it introduces a neglectable error in calculated substrate deformation energies. A similar effect of the used technique is produced by the neglect of dehydration energies, contributing to the association energy. This assumption bases on always equal dehydration energies of the enzymatic active site and equal dehydration energies of different substrates. Now we consider the interaction energy contribution to the association energy. It should be influenced by water molecules remaining in the active site after immobilization of substrates, which fit not the whole binding pocket. It influences the geometries of acylenzymes, too, which are used for the calculation of deformation energies. We assume, that this contribution to the interaction energy can be modelled by damping of electrostatic forces by a distance dependent dielectric constant. In summary, the neglect of explicit water molecules yields errors, which are either

Table 1.
Description of acylated serine monomers SerO-C(O)-CHR$_1$R$_2$

Name	R_1	R_2	Conformations
Ace-OSer	H	H	1
Ac-Gly-OSer	L-NHCOMe	H	1
Ac-Ala-OSer	L-NHCOMe	Me	1
Ac-Val-OSer	L-NHCOMe	i-C_3H_7	3
Ac-Ile-OSer	L-NHCOMe	$CH(CH_3)$-C_2H_5	9
Ac-Leu-OSer	L-NHCOMe	CH_2-i-C_3H_7	9
Ac-Asn-OSer	L-NHCOMe	CH_2-C(O)NH_2	9
Ac-Phe-OSer	L-NHCOMe	CH_2-Ph	1
Ac-Tyr-OSer	L-NHCOMe	CH_2-Ph-4-OH	1
Ac-Trp-OSer	L-NHCOMe	CH_2-indolyl	2

near equal in all structures studied or neglectable in relation to other errors produced by application of molecular mechanics.

Computational Details

Building of acylenzymes and substrates

All studies base on the X-ray structure PDB 1GCT [9] (γ-chymotrypsin acylated with the tetrapeptide Tyr-Ala-Gly-Pro), obtained from the Protein Data Bank [10,11] at Brookhaven National Laboratory. It was modified with SYBYL-6.3 [12] and the AMBER-forcefield [13]. All water molecules included in the original structure were deleted. The rudiment of Ser11 was completed. Other amino acids not visible in the crystal structure were omitted. Hydrogens were added to model protonation states at pH 7. Ser195 acylated with the original tetrapeptide ligand was used as template to define new monomers of a serine residue acylated with various N-acetylated L-amino acids. The monomers used are summarized in table 1. The atom types and charges were chosen consistently with the all-atom model of the AMBER force field. The net atomic charges of the ester group were developed from Mulliken charges taken from AM1 [14] semiempirical quantum mechanical calculations on a methyl acetate patch. They were scaled to give neutral monomers along with charges of other atoms from the AMBER forcefield. This procedure yields the charges -0.484, 0.795 and -0.439 for the carbonyl oxygen (atom type O), carbonyl carbon (atom type C) and ether oxygen (atom type OS), respectively. These charges are consistent with charges of amide groups used in the AMBER force field and with net atomic charge differences between related esters and amides, obtained by ab inito molecular orbital calculations [15]. After deletion of the original ligand and replacement of Ser195 with the new monomers, geometry optimizations followed. We used a distance dependent dielectric constant of 4.0. 1-4-interactions were scaled by a factor of 0.5. A cutoff of nonbonding interactions at 12 Å was applied. Geometry optimizations were done with a conjugate gradient routine in subsequent steps, terminating after reaching an RMS gradient threshold, like outlined in table 2. Through the first step, a distance constraint of 1.8 Å between Hγ(Ser214) and Oδ1(Asp102) was applied in order to produce a hydrogen bond between these residues. After completion of this step, this constraint was obmitted. The scheme was used to obtain geometries of rotamers of the acyl portions of several flexible substrates in the active site. These conformations are produced by rotation around the Cα-Cβ- and Cβ-Cγ-bonds of the amino acid sidechain of the substrates Val, Leu, Ile and Asn in steps of 120°. The number of conformations studied for each substrate is given in table 1.

The free substrates which would form the acylated enzyme after splitting of a leaving group were modelled in a similar manner. The monomer definitions of acylated Ser195 were used as templates for methyl ester substrates by replacing the serine portion by a methyl group. The free substrates were modelled by conformational analyses using a grid search technique and following full geometry optimizations within the AMBER forcefield.

Calculation of energies

From acylenzyme structures obtained after geometry optimizations we calculated interaction energies, deformation and association energies. All energy calculations were done on reduced models. They contain 66 monomers in each

Table 2. *Optimization scheme for acylenzymes.*

Step	Optimized atoms	$\nabla_{RMS}E$ [a]
1	hydrogen atoms	5
2	amino acid sidechains	1
3	all	0.5
4	119 monomers around Ac-Ser195	0.1
5	66 monomers around Ac-Ser195	0.01

[a] in kcal/(mol·Å).

case (residues 16, 17, 30-33, 40-45, 53-60, 94, 99, 102, 138-143, 146, 151, 160, 172, 182-185, 188-198, 212-222, 224-229 + end groups of 16, 146, 151).

The acylenzymes were splitted into an enzyme and a substrate portion by a symbolic cut of the Ca(Ser195)-Cb(Ser195) bond. The resulting model of the enzymatic active site without substrate contains 66 monomers and 880 atoms. Calculation of the force field energy terms with disregard of one portion yields the force field energy of the remaining part in the acylenzyme.

Interaction energies **EInt** between the enzyme and the acyl portion of the substrate including the full ester group are calculated as difference between the non-bonded energy terms of the reduced acylenzyme model and the sum of the non-bonded energies of the two parts. The difference between total energies would include the bond stretching, angle and torsion bendings located in the cutted Ca(Ser195)-Cb(Ser195) bond and neighbouring atoms. This „internal contribution" to the interaction energy is always near equal in magnitude in our models.

Deformation energies of the enzyme portion of the acylenzymes **EDefE** were calculated as difference between the total energies of the corresponding 880-atom models in the acylenzyme and the structure minimized without substrate. The deformation energy of a substrate **EDefS** we define as the difference between total energies of the substrate portion in the acylenzyme and in the free substrate, using its weighted averaged energy as reference. The sum of the obtained interaction and deformation energies yields the association energy EAss.

The weighted averaging of energies bases on simple Boltzmann statistics [16]. The energy of a system with N possible non-degenerate states can be calculated from

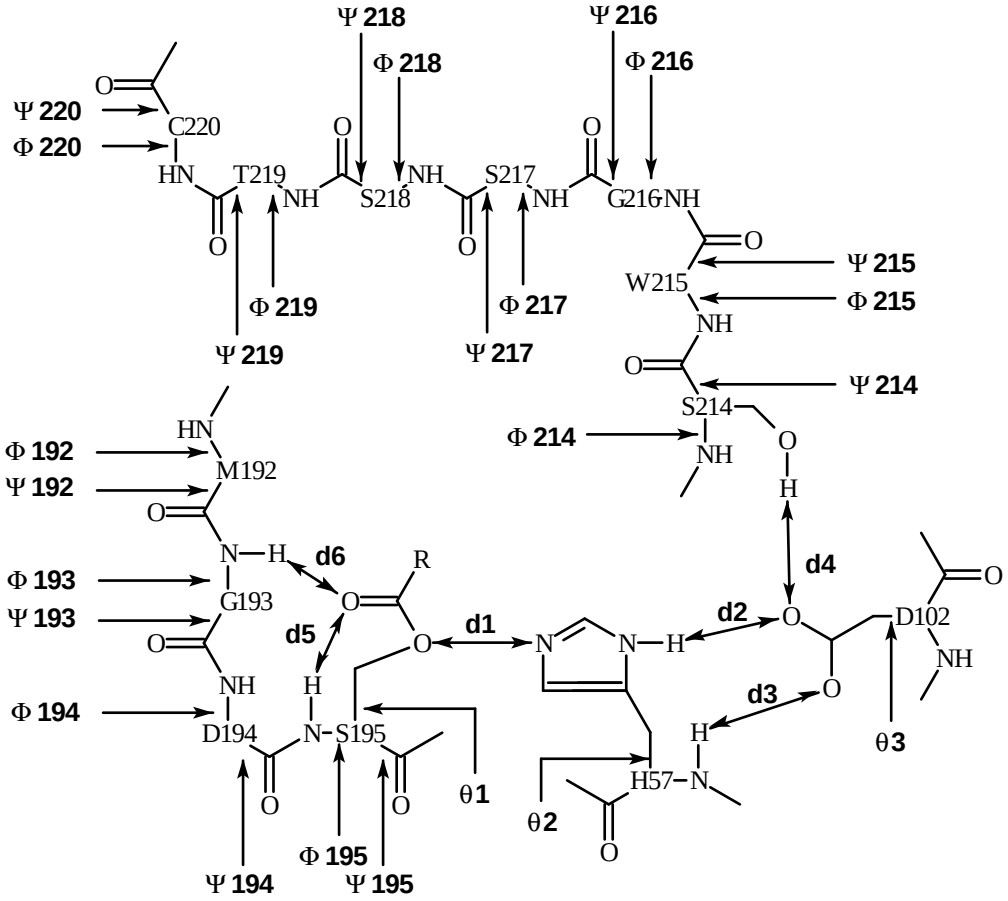


Figure 1. *Definition of geometry variables. Amino acid residues are labeled at Cα with the one-letter code and sequence number. Arrows point to the rotable bond of a dihedral angle. Backbone torsions are defined by* $\Phi_n = C_{n-1}\text{-}N_n\text{-}C\alpha_n\text{-}C_n$, $\Psi_n = N_n\text{-}C\alpha_n\text{-}C_n\text{-}N_{n+1}$.

$$E = \sum_{i=1}^{N} f_i E_i$$

where f_i is a normalized weighting factor for state i, given by

$$f_i = \frac{q_i}{\sum_{i=1}^{N} q_i}$$

The quantity q_i is the individual partition function of state i, which is defined as

$$q_i = \exp(-\Delta E_i / RT)$$

ΔE_i is the energy difference between state i and the ground state measured in J/mol, R is the gas constant and T the temperature. We used T = 298 K. Setting E_i equal to the energy of a particular conformation of a flexible molecule and N equal to the number of possible conformations, E yields the conformational energy of this molecule with respect to statistical occupancies of other than the ground state conformation by neglecting activation barriers. This procedure can also be used to average quantities other than energies such as geometry variables of conformations of a molecule, if the weighting factors were calculated from representative energies.

Statistical analyses

Four different sets of acylenzyme structures were developed. The first contains all 37 structures (set 1). Set 1a is the same like set 1, except outliers. The other sets contain only one representative structure for each substrate. For set 2 structures of the substrates Ac-Ile, Ac-Leu, Ac-Val and Ac-Asn were averaged by their arithmetic mean after exclusion of outliers. Set 3 includes the minimum energy conformations using the total energy of the 66-monomer structure as criterion. Set 4 contains structures which were averaged by Boltzmann statistics as described above. The weighting factors were developed from total energies of the 66-monomer structures. Here only outliers defined as „non-reactive“ are omitted. Outliers which are defined as „inactive“ are automatically eliminated by very small weighting factors (for the definition of outliers see next section).

All sets contain 39 variables for each structure. In addition to the four energy differences a set of 32 geometry variables was taken out of each acylenzyme structure. These variables are summarized in figures 1 and 2. Three sets of experimental data were added to each set, including measured values of log(1/Km) and log(k3) of p-Nitrophenyl esters of the studied acid portions, and log(kcat) of methyl or ethyl esters. The selected p-Nitrophenylesters of amino acid substrates bear the Benzoyloxycarbonyl (Z) N-protecting group, N-acetylated substrates were chosen in the case of

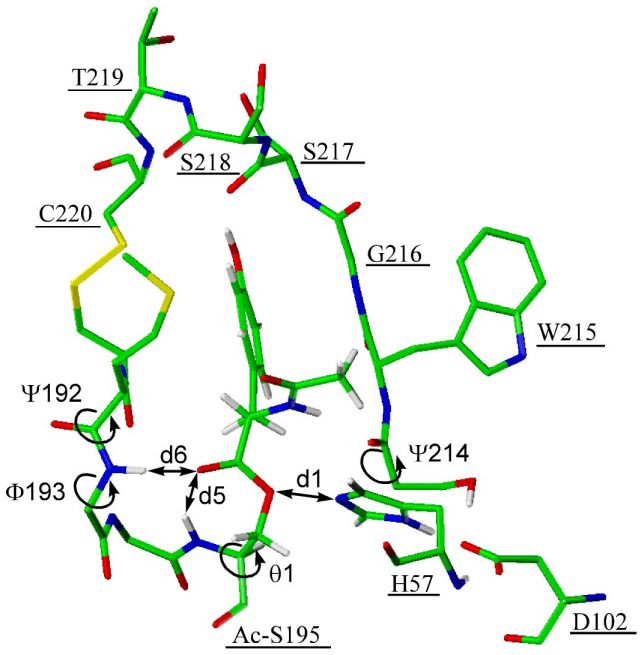


Figure 2. *Active site of chymotrypsin acylated with Ac-Tyr.*

methyl and ethyl esters. The values were taken from [17] and references cited therein.

Statistics performed on these sets included simple linear regression analyses for energies and experimental data and partial least square statistics (PLS) for geometry variables and energies. PLS analyses were done with the QSAR module of SYBYL 6.3. The optimal number of components was determined by 10 subsequent cross-validation runs using the largest possible amount of components as basis and r^2_{PRESS} (q^2) as criterion. Then the analyses were repeated with the number of components which would yield the highest q^2 without further validation of the model. The variables were used unscaled and unweighted. In addition to PLS analyses hierarchical cluster analyses were done, using the complete linkage model. Tables of sets 1 and 4 are given as supplementary material.

Results and Discussion

Rough data

From force field calculations the region which is necessary for enzymatic activity can be determined. Figure 3 shows correlation coefficients for assumed linear dependencies between total energies of several reduced models of acylenzyme structures and the energy of the complete acylenzymes. For our object we find a minimal region of about 30 residues (about 5 Å around acylated Ser195), which contains near the same information like the overall structure. Any further re-

	Set 2	Set 3	Set 4
log (1/Km) = a · EAss + b, (f=6) [b]			
a	-0.09 ± 0.03	-0.09 ± 0.03	-0.09 ± 0.02
b	3.3 ± 0.5	3.2 ± 0.6	3.2 ± 0.5
r^2 / F	0.673 / 12.34	0.692 / 13.47	0.699 / 13.96
log (kcat) = a · EAss + b, (f=5)			
a	-0.14 ± 0.04	-0.15 ± 0.04	-0.15 ± 0.04
b	-2.6 ± 1	-3.0 ± 1	-3.0 ± 1
r^2 / F	0.693 / 11.29	0.735 / 13.89	0.739 / 14.16
log (k3) = a · EAss + b, (f=7)			
a	-0.13 ± 0.04	-0.14 ± 0.04	-0.14 ± 0.04
b	-2.2 ± 0.8	-2.4 ± 0.8	-2.4 ± 0.8
r^2 / F	0.634 / 12.11	0.667 / 14.02	0.660 / 13.56

Table 3. *Parameters of an assumed linear dependency of reactivities on EAss. [a]*

[a] EAss in kcal/mol.
[b] f: degrees of freedom.

duction yields loss of information. So our base model with 66 monomers represents a good choice for calculating interaction and deformation energies.

We now focus our intention on the detection and explanation of outliers. A close inspection of obtained acylenzyme structures led us to the definition of two different outlier cases. The first case we access through anormal geometries. Among the modelled structures there are 7 cases, where the NH-hydrogen of the substrate acyl portions forms a hydrogen bond to His57. This behaviour is coupled with the complete loss of contact between the carbonyl oxygen of the ester bond to cleave and the oxyanion hole (NH-groups of Gly193/Ser195). The substrate conformations exhibiting these features are characterized by a bad steric contact between their acyl portions and the protein backbone around Met192. Since at least the accessibility of Nε2(His57) for water molecules is necessary for deacylation, we defined these structures as „non-reactive“ outliers and did not include them in our data sets.

The second outlier set was detected after looking on deformation energies of the enzyme portions (see figure 4). There are four structures with unusual high deformations of the enzyme. Since this behaviour should be unfavourable, we define them as „inactive“. The high deforming conformations can be used to calculate an allowed and a forbidden substrate volume for chymotrypsin (see figure 5). The obtained forbidden area can be explained by repulsion between the acyl portion and the backbone of Trp215/Gly216. These outliers were produced by simplexing, which we used to reduce highly repulsive contacts prior to geometry optimization.

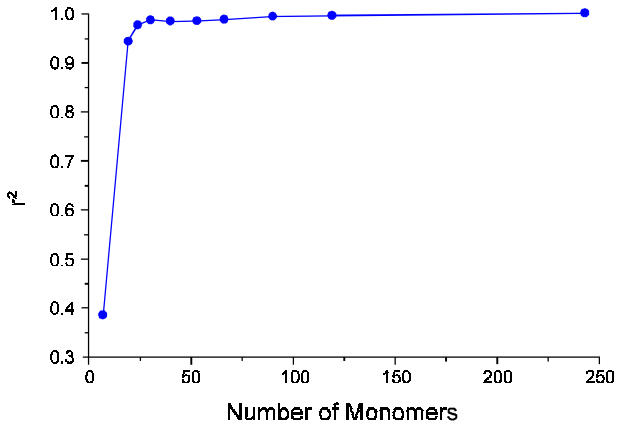


Figure 3. *Correlation coefficients for assumed linear dependencies between total energies of reduced acylenzyme models and the complete structures.*

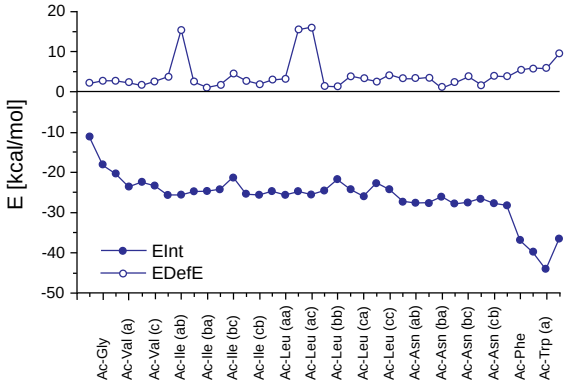


Figure 4. *Interaction energies and deformation energies of the enzyme portions. The order of structures follows that given in table 1.*

Table 4. *Dependency of EAss on geometry variables. [a]*

	Set 2			Set 3			Set 4	
s	2.949			3.845			2.634	
r^2	0.887			0.802			0.907	
F	27.51			14.17			34.22	
x [b]	**coeff.**	**cum.**	**x**	**coeff.**	**cum.**	**x**	**coeff.**	**cum.**
Ψ192	-0.378	0.163	Ψ214	0.428	0.143	Ψ192	-0.372	0.153
Ψ214	0.378	0.325	Ψ192	-0.317	0.282	Ψ214	0.418	0.304
Φ193	0.324	0.468	Φ193	0.268	0.399	Φ193	0.319	0.434
Ψ216	-0.273	0.567	Ψ216	-0.273	0.500	Ψ216	-0.277	0.532
θ1	-0.294	0.646	Ψ215	-0.236	0.593	θ1	-0.312	0.616
Φ215	-0.145	0.715	θ1	-0.266	0.670	Φ215	-0.182	0.691
Φ214	0.188	0.763	Φ216	0.297	0.736	Φ214	0.198	0.740
Φ195	0.167	0.801	Φ214	0.236	0.792	Φ216	0.209	0.785
Φ192	0.191	0.832	Φ217	-0.160	0.837	Φ217	-0.151	0.829
Φ217	-0.113	0.863	Ψ217	0.188	0.876	Φ195	0.178	0.864

[a] PLS with 32 independent geometry variables, 2 components.

[b] Only the 10 regression coefficients with the highest contribution are listed in the order of their normalized values. Normalization was done with respect to the variance of x_i and y. The last column cumulates normalized coefficients wich where scaled to sum to 1.0.

After exclusion of these outliers the number of acylenzymes reduces to 26 structures for 10 substrates. They were used to study relationships between energies and geometries.

Energy relationships

From empirical considerations it is believed that a substrate will be much more strained after immobilization than the enzyme itself [18]. On the other hand, deformation of both the substrate and the enzyme becomes possible, when the interaction is strong enough. This possibility was demonstrated in [19] for the association complex of chymotrypsin with N-Ac-Trp-amide by ab initio and semiempirical molecular orbital calculations. Our models can describe such a relationship for the acylenzyme states of various substrates. We obtain a probably linear dependency between the overall deformation energy and the interaction energy (see figure 6).

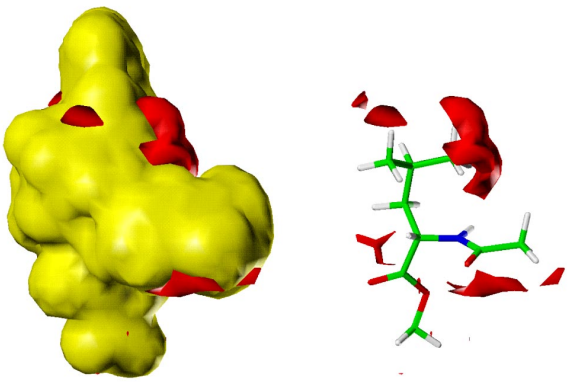


Figure 5. *Left: allowed substrate volume (from 22 „active“ conformations, yellow); right: forbidden substrate volume (from 4 „inactive“ conformations, red) and included inactive Ac-Leu conformer.*

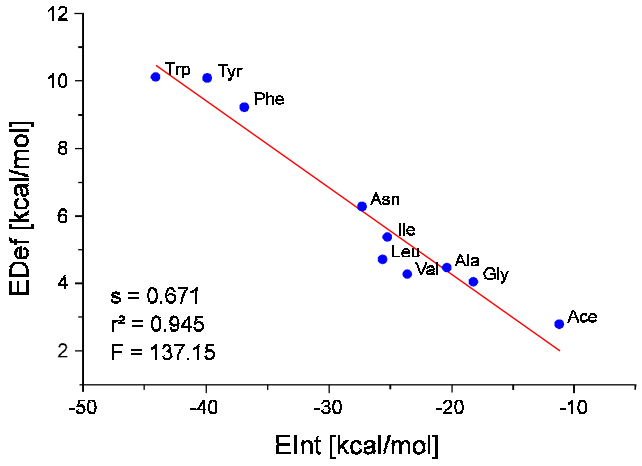


Figure 6. *Parameters for an assumed linear dependency of overall deformation energies on interaction energies (set 4).*

Table 5. *Selected geometries for high specific substrates.*

Distance	Phe, Tyr, Trp	other substrates
d1, Å	2.96 ± 0.01	3.04 ± 0.04
d2, Å	1.901 ± 0.005	1.853 ± 0.006
d5, Å	1.973 ± 0.001	1.94 ± 0.02
d6, Å	1.950 ± 0.001	2.03 ± 0.03

Set 4 is the most predictive in this case. Sets 2 and 3 yield near the same dependency. Sets 1 and 1a bear a high noise level. We do not discuss the latter sets any further in this work.

No linear relationship can be obtained between the energy needed for substrate deformation and that needed for enzyme deformation (see figure 7). We find three groups of different structures here. The first contains substrates with small acyl groups (Ace, Ac-Gly, Ac-Ala) which can not be deformed. The unusual finding that they induce enzyme deformations is an effect of our choice of the reference structure used for calculation of EDefE. The second group contains substrates with aromatic acyl portions (Ac-Phe, Ac-Tyr, Ac-Trp) which exhibit large deformation of both the substrate and the enzyme structure. The third group includes structures with intermediate strength of interaction. They show deformation of the substrate portion rather than of the enzyme part.

These three groups correspond to substrate structures known to have different reactivities in chymotrypsin-catalysed reactions [20]. In order to obtain a quantitative relationship between measured reactivities and calculated energies we assumed a linear dependency. The most predictive model in the majority of cases is the association energy of set 4. The correlation coefficients for all sets and reactivity data are given in table 3. Figure 8 shows the obtained relationship with the deacylation rate constant, for which the most data were available. The correlations are very poor, since the data set is very small and the source and interpretation of experimental data is not out of question [21]. So the derived regression model has to be interpreted in terms of a rough dependency between substrate reactivity and calculated association energies. But it enables one to derive the assumption, that the association energy can be used as a reactivity substitute with some success. We will make use of this working hypothesis, although from outliers of an assumed linear relationship it can be estimated, that a high association energy may not necessarily be coupled with high reaction rates. This behaviour seems to hold for the Ac-Trp substrate, if the measured reactivity is correct. Other substrates may set specific interactions in motion like Ac-Asn, thus reaching a high reaction rate without large deformations of the enzyme.

Geometry analyses

The calculated deformation energies correspond to geometry changes of the enzyme and substrate part of the active site. Performing hierarchical cluster analyses with all 32 geometry variables of set 4, we obtain figure 9. The same clusters of substrate structures is obtained as discussed above. Since geometry variables are dependent from each other, the data set has to be reduced by extraction of principal components. We performed PLS analyses using the association energy as dependent variable. Defining the whole set of 32 geometry variables as X-block, cross-validation runs yielded two principal components. As usual, the scores of cases in

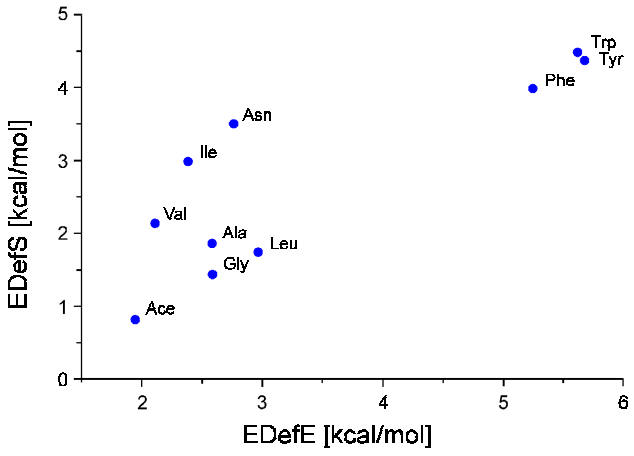


Figure 7. *Relationship between energies of deformation of the substrate and enzyme portions (set 4).*

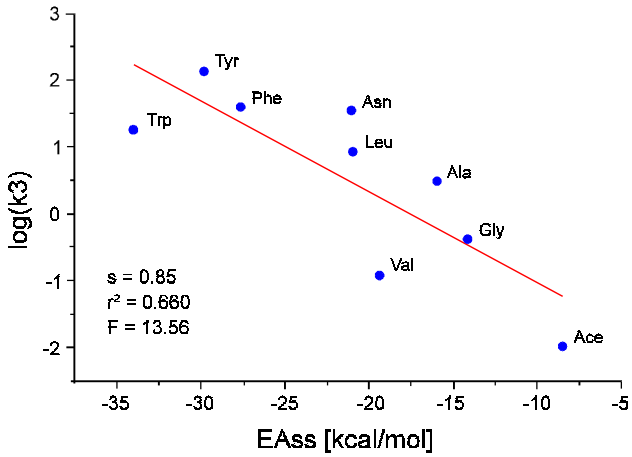


Figure 8. *Relationship between the deacylation rate constant k3 and calculated association energies (set 4).*

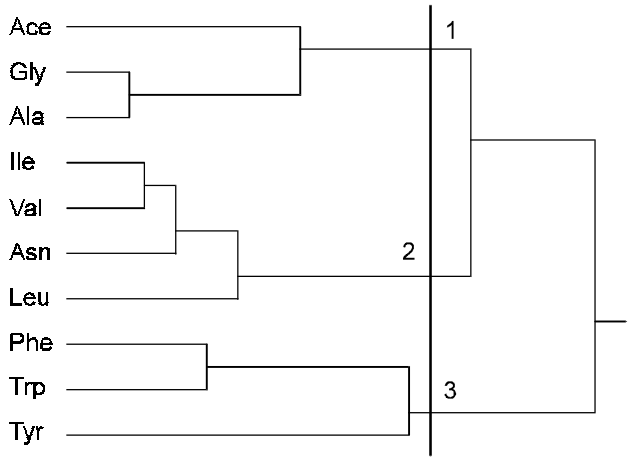


Figure 9. *Result from hierarchical cluster analysis of 32 geometry variables (set 4).*

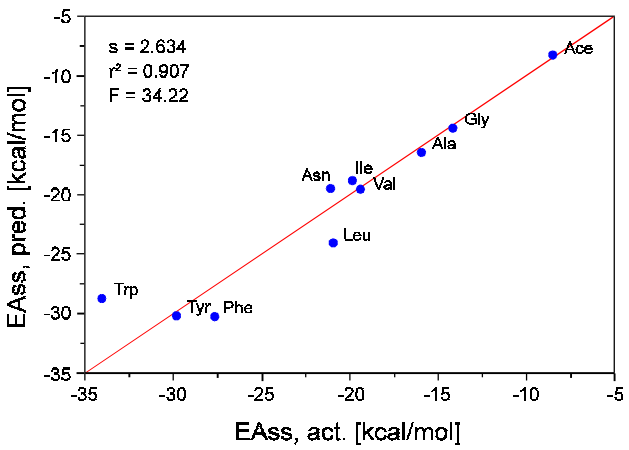


Figure 10. *Quality of the PLS fit between 32 geometry variables and the association energy (set 4).*

the reduced X-block of the two non-redundant variables have the same structure like obtained from cluster analyses of the whole data set. The obtained fit is presented in figure 10. Again set 4 is the most predictive one, although other averaged sets yield the same relationship. Interestingly, the Ac-Trp substrate is predicted by geometries to yield a smaller association energy than calculated from force field energies. This result corresponds to the measured reactivity of this substrate (see figure 8).

These analyses enable one to state a probable relationship between geometry variables and substrate reactivity. The relationship is determined by dihedral angles mostly, like shown in table 4. The only region which does not contribute to the model is the border of the binding pocket from Ser217 to Cys220. Variables from this region exhibit small variance only.

It was not possible to reduce the set of initial geometry variables to a smaller amount which enables both better interpretation and still high statistical significance. In our models, the distances d1, d2, d5 and d6 bear enough information to discuss sources of high substrate specifity. Hierarchical cluster analyses of these four distances yield clear separation of the three aromatic substrates, while the remaining structures are grouped unsystematically. The clustering is produced by a different behaviour of all four distances in the case of specific aromatic substrates (see table 5).

Relations to the reaction mechanism

We now have to explain the observed specific behaviour of aromatic substrates in terms of the reaction mechanism, which is presented in figure 11. Our current knowledge of the mechanism of deacylation yields at least three conditions necessary for this reaction.

At first, the contact between the carbonyl oxygen of the ester group to cleave and the NH-hydrogens of the protein backbones at Gly193/Ser195 is thought to yield a rate enhancement. This interaction is already present in the reactant state. It becomes stronger in the transition state involving water attack and proton transfer to His57. It reaches its maximum in the tetrahedral intermediate, because the negative charge of the carbonyl oxygen rises during the reaction. Our finding that the hydrogen bond distances d5 and d6 measured in the reactant state are related to reactivity, has to be interpreted by necessary changes of these distances through the reaction event, assuming that the found relationship is true. If it is valid, then elongation of d5 coupled with shortening of d6 occurs through water attack. The induction of these changes in the acylenzyme should yield a reactant state that is closer to the transition state. Experimental results can give some evidence for our hypothesis. In [22] it was shown by spectroscopic studies on adducts of chymotrypsin with several substrates, that one hydrogen bond to the carbonyl oxygen of specific substrates seems to be lost, whereas it remains present for unspecific substrates. Our model predicts d5 to be the broken hydrogen bond. It is the one which has an unfavourable geometry, because it requires formation of a ring with a high steric strain (see figure 2).

The second requirement for a successful deacylation reaction can be developed from the breakdown of the tetrahedral intermediate into carboxylate and reformed active site (see figure 11). During formation of the tetrahedral intermediate Nε2(His57) becomes protonated. This proton is thought to be transferred onto the Ser195 sidechain. Therefore a hydrogen bond must be formed between protonated His57 and acylated Ser195. It can only be formed, if the distance between Oγ(Ser195) and Nε2(His57) is small enough, as indicated by the d1 variable. The dihedral angle θ1 contributing to the PLS model may be responsible for this change.

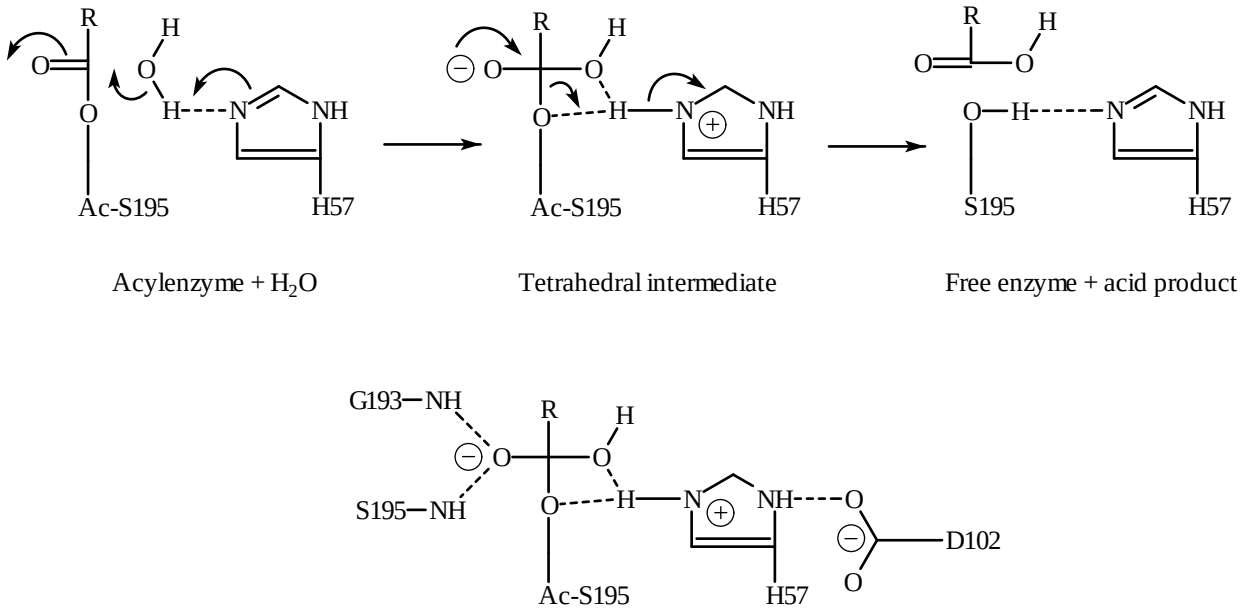


Figure 11. *Mechanism and catalysis of deacylation.*

The third statement has to include the His-Asp interaction, which is described by the d2 variable. There exists experimental evidence for the dependency of reaction rates on changes of His-Asp interactions. Kinetic studies in H_2O/D_2O mixtures enable one to measure the number of protons which can be exchanged between enzymatic residues and the solvent during a reaction. These measurements yielded the conclusion [23], that in the case of specific substrates two protons can be exchanged, whereas the reaction with unspecific substrates involves exchange of one proton only. This result was interpreted by a shortening of the His-Asp distance in the case of specific substrates. A second set of experiments was done to calculate the pKa of His57 by evaluations of pH-profiles of measured reaction rates [24,25]. It was shown that there exists a relationship between the ability of His57 to accept a proton from the solvent and the substrate structure. These results were never explained, but they should represent just another method to access features of the His-Asp dyad which are influenced by substrate binding. From this point of view our finding that the d2 variable is influenced in the case of specific substrates bears some truth. It is possible, that the interaction between Asp102 and the Ser214 sidechain is responsible for this change, because we built a hydrogen bond between these sidechains, and because the dihedral angles Ψ214/Φ214 were shown to contribute to the derived PLS model (see table 4).

Summary

The modelling of various acylenzyme structures and the extraction of representative variables enabled us to develop hypotheses for relationships between events occurring during substrate immobilization and substrate reactivity. The obtained results represent a detailed access to the „induced strain“ hypothesis. It was shown that interactions between the substrate and the active site in the binding pocket can influence the position of atoms involved in the following reaction event. The effect on structures of the reaction centre depends on substrate structure. In the case of specific substrates strong interactions lead to unique constellations of catalytic residues, which we related to possible geometry changes during the reaction. Except high reactive substrates, the obtained relationships are too complex to explain differences between low and very low reactivity by geometry variables.

In summary, is was surprising, that modelling experiments using force field simulations of reactant states yielded a hypothesis of the behaviour during the enzymatic reaction. This hypothesis is able to describe specific structure deformations as one possible source of different substrate reactivity by using QSAR techniques, which we applied to uncommon structure descriptors.

Supplementary Material Available

Tables of sets 1 and 4 as comma separated text files, including geometry variables, energies, energy differences and experimental reactivity data.

Bibliographic References

1. Fischer, E. *Ber. Dtsch. Chem. Ges.*, **1894**, *27*, 2985.
2. Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*. VCH: Weinheim 1993.
3. Pauling, L. *Nature*, **1948**, *161*, 707.
4. Menger, F. M. *Biochemistry*, **1992**, *31*, 5368.
5. Page, M. I.; Williams, A. *Enzyme Mechanisms*. Royal Society of Chemistry: London 1987.
6. DeTar, D. F. *J. Am. Chem. Soc.*, **1983**, *103*, 107.
7. Wipff, G.; Dearing, A.; Weiner, P. K.; Blaney, J. M.; Kollman, P. A. *J. Am. Chem. Soc.*, **1983**, *105*, 997.
8. Bemis, G. W.; Carlson-Golab, G.; Katzenellenbogen, J. A. *J. Am. Chem. Soc.*, **1992**, *114*, 570.
9. Dixon, M. M.; Matthews, B. W. *Biochemistry*, **1989**, *28*, 7033.
10. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T., Tasumi, M. *J. Mol. Biol.*, **1977**, *112*, 535.
11. Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In: *Crystallographic Databases - Information Content, Software Systems, Scientific Applications*; Allen, F. H.; Bergerhoff, G.; Sievers, R. (Eds.); Data Commission of the International Union of Crystallography: Bonn Cambridge Chester 1987; p. 107.
12. TRIPOS Associates, Inc.
13. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Chandra Singh, U.; Ghio, C.; Alagona, G.; Profeta, S., Weiner, P. *J. Am. Chem. Soc*, **1984**, *106*, 765.
14. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F., Stewart, J. J. P. *J. Am. Chem. Soc.*, **1985**, *107*, 3902.
15. Kallies, B.; Mitzner, R. *J. Chem. Soc., Perkin Trans. 2*, **1996**, in the press (5/08360E).
16. Atkins, P. W. *Physikalische Chemie*. 1st ed.; VCH: Weinheim 1988.
17. Hansch, C.; Grieco, C.; Silipo, C., Vittoria, A. *J. Med. Chem.*, **1977**, *20*, 1420.
18. Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. Wiley: New York 1991.
19. Dive, G.; Dehareng, D.; Ghuysen, J. M. *J. Am. Chem. Soc.*, **1994**, *116*, 2548.
20. Schellenberger, V.; Braune, K.; Hoffmann, H. J., Jakubke, H. D. *Eur. J. Biochem.*, **1991**, *199*, 623.
21. Zerner, B., Bender, M. L. *J. Am. Chem. Soc.*, **1964**, *86*, 3669.
22. Whiting, A. K., Peticolas, W. L. *Biochemistry*, **1994**, *33*, 552.
23. Elrod, J. P.; Hogg, J. L.; Quinn, D. M.; Venkatasubban, K. S., Schowen, R. L. *J. Am. Chem. Soc.*, **1980**, *102*, 3917.
24. Hirohara, H.; Philipp, M., Bender, M. L. *Biochemistry*, **1977**, *16*, 1573.
25. Béchet, J. J.; Dupaix, A., Roucous, C. *Biochemistry*, **1973**, *12*, 2566.