ELSEVIER

Molecular modeling of substrate–enzyme reactions for the cysteine protease papain

Yu Lin and William J. Welsh

Department of Chemistry and Center for Molecular Electronics, University of Missouri-St. Louis, St. Louis, Missouri, United States

AM1 quantum mechanical reaction coordinate (RC) calculations were run to simulate the rate-limiting deacylation (hydrolysis) reaction for a series of para-X–PhC(O)NHCH$_2$–C(Y)–S–papain intermediates, where X = OCH_3, CH_3, H, Cl, NO_2 and Y = O (thioester) or S (dithioester), for which a large body of structural, kinetic, and spectroscopic data is available. Several reaction zones, in particular the so-designated Large Zone and Small Zone, were extracted for these RC simulations from the fully solvated and energy-minimized X-ray crystal structure of papain (pdb9pap) *bound to the appropriate substrate moiety. The major structural difference between these two zones was the absence of the oxyanion hole in the latter. For both the thioester and dithioester cases, the calculated E_a value associated with the parent (X = H) acyl–enzyme intermediate was lower by ca. 10 kcal/mol for the Large Zone than for the Small Zone. The magnitude of this difference suggests that the oxyanion hole plays a functional if not essential role in stabilizing the anionic tetrahedral intermediate with the cysteine proteases. The calculated E_a value was lower by ca. 10 kcal/mol for the thioester [–C(O)–S–] than for the corresponding dithioester [–C(S)–S–], in qualitative agreement with kinetic data for this series of substrates which reveal that the specific rate constant for deacylation k_3 is ca. 60 times larger for the former. This difference is also consistent with both AM1 and 6-31G* calculations on model intermediates, which indicate that the weaker polarity of the dithioester compared with the thioester [i.e., -C(←S)–S– versus –C(→O)–S–] renders the former a much poorer site for nucleophilic attack. The anionic tetrahedral intermediate is energetically more stable for the dithioester than for the corresponding thioester, a finding that is discussed in terms of its kinetic and mechanistic implications. The mode of attack by the H_2O nucleophile is "concerted" rather than "sequential" in terms of the order of proton abstraction by His-159 and nucleophilic attack on the acyl–enzyme intermediate. While the presumably key $S_{thiol} \cdots N$ nonbonded contact remained almost constant (ca. 2.90 Å) up to formation of the [TS] structure, the substrate torsion angles ϕ and ψ rotated significantly as the hybridization around the reaction site transforms from sp^2 to sp^3 during formation of the tetrahedral intermediate. The AM1-calculated frontier molecular orbitals for model thioester and dithioester acyl–enzyme intermediates generally associate the HOMOs with the reaction site and the LUMOs with the benzamide moiety. Computer graphics images corroborate our view that, in relation to the $S_{thiol} \cdots N$ interaction, the HOMOs and LUMOs should be identified, respectively, with S_{thiol} and N rather than the reverse, as suggested by other workers.*

INTRODUCTION

The cysteine proteases are proteolytic enzymes that depend on a nucleophilic thiol group (–SH) of a cysteine residue (Cys-25 in papain) for their enzymatic activity. Like the analogous serine proteases, the cysteine proteases are functionally diverse and widely distributed in nature. Some prominent examples include the mammalian proteases cathepsin B, H, L, and S (part of the armory of lysozomes), the plant proteases papain, ficin, bromelain, and actinidin, and the bacterial proteases clostripain and streptococcal proteinase.[1–6]

The main focus of our research on the cysteine proteases is the design of mechanism-based chemotherapeutic inhibitors of cathepsin B, which has been implicated in a host of disease states including cancer, muscular dystrophy, and arthritis.[1–5] As part of this larger study,[7] we have completed a detailed exploration of the hydrolysis of several substrate–enzyme intermediates using computational chemistry in

Color Plates for this article are on pages 92 and 93.

Address reprint requests to: William J. Welsh, Department of Chemistry and Center for Molecular Electronics, University of Missouri-St. Louis, St. Louis, Missouri 63121.

Received 26 December 1996; revised 18 March 1996; accepted 26 March 1996.

655 Avenue of the Americas, New York, NY 10010

0263-7855/96/$15.00
PII S0263-7855(96)00028-8

concert with molecular computer graphics. Until the X-ray crystallographic coordinates for cathepsin B became available,[8,*] we employed papain as a surrogate model structure, based on the premise that significant insights and findings would be largely transferable between the two homologous cysteine proteases.

The 1.65-Å X-ray crystal structure of Cys-25-oxidized papain (*pdb9pap*)[9] exhibits two domains separated by a cleft that contains the active-site thiol (–SH) of Cys-25 located in close proximity to His-159 but on the opposite side. The imidazole side chain of His-159 is well positioned to act as a general base catalyst to activate the nucleophilicity of the –SH group during acylation and, later, of the attacking H_2O molecule during deacylation. The backbone NH of Cys-25 and the side chain NH_2 of Gln-19, two hydrogen bond donors, form the "oxyanion hole" for binding the substrate and promoting its reactivity to nucleophilic attack. The substrate appears to be tethered by hydrogen bonds on either side of the cleft involving the backbone CO of Asp-158 and the NH of Gly-66 (Color Plate 1). The sizable distance between Asp-158 and the substrate suggests the possibility that this hydrogen bond is mediated by an intervening H_2O molecule.[5,6,9,10]

Like most cysteine proteases, papain is capable of hydrolyzing proteins, amides, esters, and even thionoesters [i.e., R–C(S)–OR']. The reaction pathway is commonly divided into two processes, acylation and deacylation, both of which are purported to proceed through formation of an anionic tetrahedral intermediate (Figure 1). Although considerable discussion and debate persist over mechanistic details, the basic components of acylation and deacylation can be described as follows. The acylation process can be considered to begin with partial or total abstraction of the Enz–SH proton of Cys-25 by the imidazole ring of His-159 (i.e., Im $\rightarrow$ ImH^+).† This is followed by Enz–S^- attack on the carbonyl carbon of the RC(O)–NHR' substrate to form the first tetrahedral intermediate. Proton transfer from ImH^+ to –NHR' promotes amide bond cleavage, yielding NH_2R' as the first product and leaving the acylated enzyme [i.e., RC(O)–S–Enz]. The deacylation process, in which the acylated enzyme is hydrolyzed, begins with partial or total proton abstraction from the attacking H_2O by His-159 (again, Im $\rightarrow$ ImH^+), followed by nucleophilic attack by OH^- on the carbonyl carbon of RC(O)–S–Enz to form the second tetrahedral intermediate. Proton transfer from ImH^+ to –S–Enz promotes RC(O)···S–Enz bond cleavage, yielding the second product RC(O)OH and the regenerated native enzyme Enz–SH.[5,6]

Relevant to the present discussion, the acylated enzyme for ester substrates is the same thioester [RC(O)–S–Enz] as described above for amide or protein substrates. However, the acylated enzyme is the dithioester [RC(S)–S–Enz] for

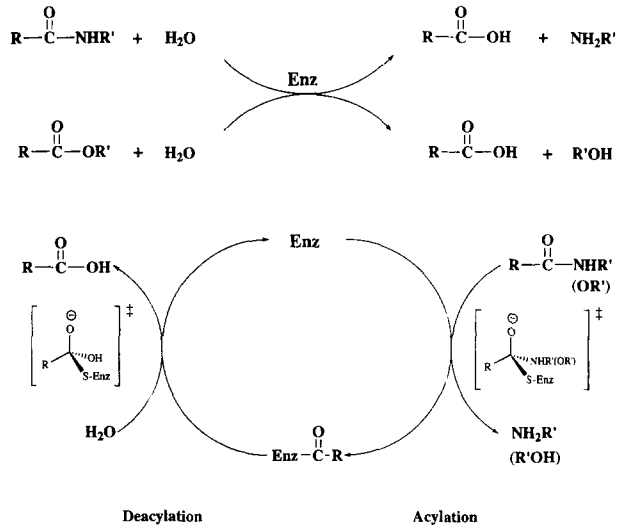


Figure 1. Illustration of the putative reaction pathway for hydrolysis of protein (amide) and ester substrates by cysteine proteases. The reaction is divided into an acylation step and an deacylation step, each of which is purported to form an anionic tetrahedral intermediate as illustrated.

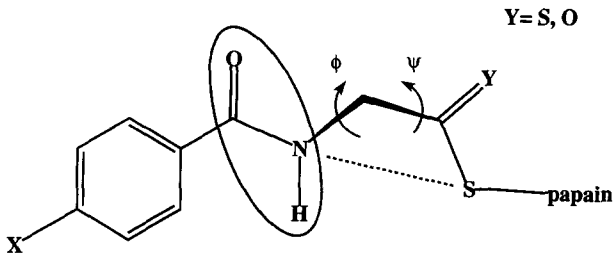
thionoester substrates. The dithioester linkage is not indigenous to the biochemistry of cysteine proteases, yet the wealth of spectral, structural, and kinetic data justified its inclusion in our study.[11–17] Kinetic evidence on the cysteine proteases indicates that the rate-limiting step is acylation for proteins and amides but is deacylation for esters and thionoesters. (Similarly for the serine proteases, the rate-limiting step is also acylation for both proteins and amides and deacylation for esters. Unlike the cysteine proteases, serine proteases cannot hydrolyze thionoesters[1–3,18,19]).

The major objectives of the present study were threefold: (1) to simulate the rate-limiting deacylation reaction for a series of model acyl–enzyme intermediates, (2) to monitor the energetics and binding interactions along the reaction pathway, especially those pertaining to the role of the oxyanion hole in stabilizing the transition state, and (3) to compare the calculated energetics with the observed kinetics for the substrate family under study in terms of the effect of substituents. The AM1 semiempirical molecular orbital method has demonstrated its utility in several applications relevant to the present study.[20–26] To this end, we present results of AM1 quantum mechanical reaction coordinate (RC) calculations on model systems for the deacylation of a series of *X*-substituted thioester (Y = O) and dithioester (Y = S) acyl–enzyme intermediates (Figure 2). The significance of this series lies in the fact that values of their deacylation rate constant k_3 for the corresponding substrates show a strong linear correlation with the Hammett σ of the *X* substituents (correlation coefficient r = 0.99 for both Y = O and Y = S).[13–15] The deacylation rate is observed to increase with increasing electron-accepting ability of the *X* substituent, a surprising result given the physical remoteness of these substituents from the reaction site.

In a previous report from this laboratory,[27] we proposed that the *X* substituent may exert its kinetic effect during reaction by modulating the extent of charge delocalization

*The X-ray crystallographic coordinates of human cathepsin B were kindly donated by R. Huber (Max-Planck Institut für Biochemie, Martinsried, Germany).

†With respect to the interaction between the His-159 and Cys-25 side chains, the "resting state" of papain (and other cysteine protease) prior to acylation has been suggested to be the ion pair $ImH^+ \cdots S^-$ rather than the alternative neutral form Im ··· HS. See van Duijnen, P.T., Thole, B.T., Broer, R., and Nieuwpoort, W.C. Active-site α-helix in papain and the stability of the ion pair $RS^- \cdots ImH^+$. *Ab initio* molecular orbital study. *Int. J. Quant. Chem.* 1980, **17**, 651–671.



X = -NO_2, -Cl, -H, -CH_3, -OCH_3

Figure 2. Series of model acyl–enzyme intermediates under study. The substrate moiety is shown in the so-called B-type binding conformation as determined by the values of the torsion angles φ and ψ. The focal nonbonded S_{thiol} · · · [C(O)NH] contact is denoted by the dotted line.

from the negatively charged reaction site onto the benzamide portion of the substrate. By facilitating this dispersal of charge, good *X* acceptors will stabilize the transition state and reduce the activation energy to accelerate deacylation. Crystallographic evidence[11–14,17] reveals a short, nonbonded S_{thiol} · · · N [or perhaps, more broadly, S_{thiol} · · · NHC(O)] contact that has been suggested[27] to couple the reaction site to the *X*-substituted benzamide system (Figure 2). This apparent S_{thiol} · · · N interaction is also explicable in terms of Lewis acid–base theory, where the S_{thiol} atom is the Lewis base and the N atom is the Lewis acid. Restated in the language of frontier molecular orbital (FMO) theory, good *X* acceptors reduce the highest occupied molecular orbital–lowest unoccupied molecular orbital (HOMO–LUMO) energy gap, thereby promoting the dispersal of electron density from the (anionic) reaction site HOMOs into the substrate benzamide LUMOs. These HOMOs and LUMOs are visually displayed later in the present study for model compounds of these acyl–enzyme intermediates.

MATERIALS AND METHODS

Solvation and acylation of papain crystal structure

The *pdb9pap* crystal structure of papain contains 212 residues plus 195 H_2O and 29 CH_3OH solvent molecules. Using the molecular modeling program *Insight,*‡ the 29 CH_3OH molecules were converted to H_2O molecules while the oxidized Cys-25 side chain was "reduced" to –SH to regenerate the native protein. Some minor alterations were also required to correct for disorder and discontinuities in several residues. The system pH was set to pH 7, thus imparting a +1 charge to the 12 arginine and 10 lysine residues and a −1 charge to 7 glutamine and 6 asparagine residues to yield a net charge of +9. The structure was solvated in a sphere of TIP3P H_2O molecules extending 5 Å beyond the protein surface; this procedure augmented the number of solvent molecules from 224 to 1 377. The solvated protein structure was subsequently energy minimized by molecular mechanics using the all-atom CVFF force field in the *Discover* program.‡ To prevent abrupt movements during geometry optimization and to ensure a physically realistic minimum-energy configuration, the system was energy minimized gradually in three stages: (1) only the 1 377 H_2O molecules, (2) only the 1 377 H_2O molecules and the side chains, and finally, (3) the entire system. Steepest-descent and conjugate gradient energy minimizations were applied consecutively until a root-mean-square gradient of 0.001 kcal/mol/Å^2 was achieved.

To mimic initial conditions for the deacylation reaction, the papain structure described above was "acylated" computer graphically to create the proper acyl–enzyme intermediate by attaching the substrate moiety *X*–PhC(O)NH–CH_2–C(*Y*)– to the S atom of Cys-25. The torsion angles φ [C(O)–NH—CH_2–C(*Y*)] and ψ [NH–CH_2—C(*Y*)–S_{thiol}] about the substrate –CH_2– unit were preset to −79° and −15.5°, respectively, such that the substrate moiety assumed a "B-type" conformation. This conformation, which spans a range of φ, ψ values encompassing these preset values, has been described elsewhere as the likely binding mode for these substrates.[11–15,17] Moreover, the B-type conformation exhibits an S_{thiol} · · · N contact of ca. 2.9 Å (based on X-ray crystallographic analysis of several model compounds), which is short compared with the value 3.35 Å for the sum of the van der Waals radii of S and N atoms.[17] Consequently, the values adopted by φ and ψ are considered crucial inasmuch as they dictate the proximity and orientation of the focal S_{thiol} · · · N [or S_{thiol} · · · NHC(O)] interaction (Figure 2). Energy minimization of the solvated acyl–enzyme intermediate was achieved again in three stages: (1) only the H_2O molecules, (2) only the H_2O molecules and the substrate moiety, and (3) the entire system. The resulting geometry-optimized structure served as the starting point for the RC calculations described below.

Extraction of enzyme–substrate reaction zone

The sheer size of the total system (i.e., enzyme, substrate moiety, and 1 377 H_2O molecules) precluded any hope of performing large-scale AM1 RC calculations. A feasible alternative is to extract an "enzyme reaction zone" containing, at the minimum, those molecular components deemed essential to the reaction mechanism. We selected two such reaction zones, appropriately designated the Small Zone and the Large Zone (Figure 3). The Small Zone, counting 69 atoms, comprised (1) the substrate moiety, (2) a Cys-25 model compound, bonded at its reactive S atom to the substrate moiety, (3) the imidazole ring of His-159 serving as the general base catalyst, and (4) an H_2O molecule, situated between the reaction site and the imidazole ring and poised to attack the C–S scissile bond. The Large Zone, counting 113 atoms, subsumed the Small Zone, then added the following components: (1) the main chain NH of Cys-25 and side-chain NH_2 of Gln-19 (i.e., the oxyanion hole), (2) the backbone CO of Asp-158 and the NH group of Gly-66, and (3) Asn-175, which forms a hydrogen bond with His-159. A conspicuous distinction between the Small Zone and the Large Zone is the inclusion of the oxyanion hole in the latter. Consequently, differences in the energetics and mechanistic aspects of deacylation between the two reaction

‡The molecular modeling programs *Insight* (version 2.3) and *Discover* (version 2.9) are products of Biosym/Molecular Simulations, Inc., since renamed MSI, Inc. (San Diego, CA).

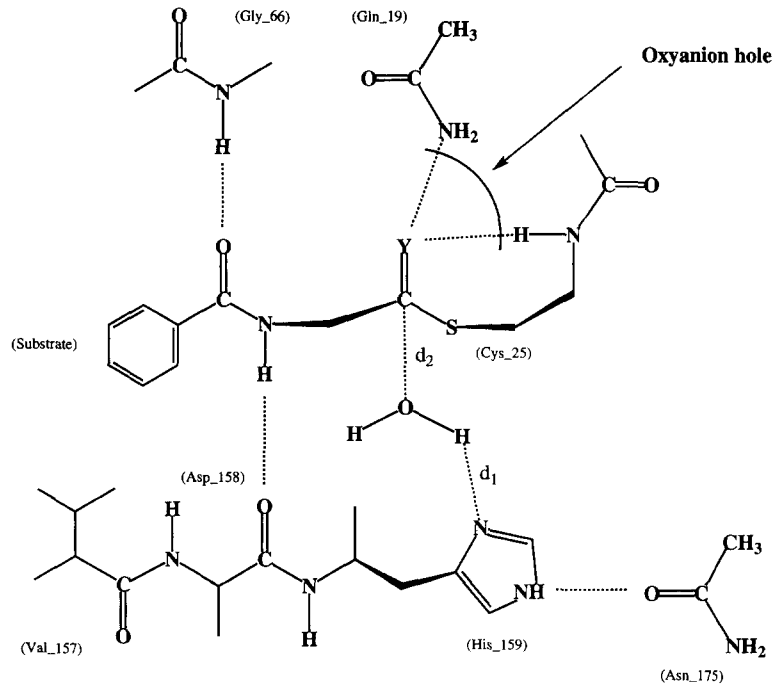


Figure 3. Illustration of the Large Zone considered in the AM1 reaction coordinate simulations of papain deacylation. The reaction coordinates d_1 *and* d_2 *are indicated.*

zones will reflect in large part its influence. Other reaction zones, differing from the Small Zone or the Large Zone by the addition or deletion of selected residues, were also investigated in less detail. The largest system for which AM1 RC calculations were performed consisted of 166 atoms.

AM1 reaction coordinate calculations

Following extraction of the enzyme reaction zone as described above, the truncated residues were capped by adding H atoms or CH_3 groups as deemed appropriate. Preceding each RC calculation, AM1 energy minimization was performed on the reaction zone. This procedure caused ϕ to rotate from $-79°$ to $-100°$ while ψ remained near $-15°$. The AM1 RC method[28] was used to simulate the reaction pathway during deacylation. In the RC method, one degree of freedom (e.g., interatomic distance) is constrained to move incrementally while all other degrees of freedom are allowed to relax via energy minimization. We adopted a two-dimensional RC scheme by analogy to Schröder and coworkers[29] in their study of the related serine proteases. The first reaction coordinate d_1 was defined as the distance between the $N^{\delta 1}$ atom of His-159 and the nearest H atom of the attacking H_2O; the second reaction coordinate d_2 was defined as the distance between the O atom of the attacking H_2O and the target C(*Y*)–S– carbon of the scissile bond (Figure 3). The deacylation reaction was then driven in the forward direction by simultaneously decreasing the values of d_1 and d_2 along their respective reaction coordinates, thereby simulating (1) proton abstraction from the attacking H_2O molecule by His-159 (via d_1) and (2) nucleophilic attack by OH^- on the –C(*Y*)–S–carbon atom (via d_2). During this process, the attacking H_2O molecule is split heterolytically.

A variety of pathways for this process was explored by conducting a systematic grid search over the span $d_1 = 2.7$

Table 1. Experimental[a] and calculated[b] data for deacylation of subject thioester and dithioester acyl–enzyme intermediates

	Large Zone						Small Zone					
	Thioester					Dithioester	Thioester	Dithioester				
Parameter	OMe	Me	H	Cl	NO_2	H	H	OMe	Me	H	Cl	NO_2
k_3 (s^{-1})	2.07	2.91	3.53	6.41	10.6	0.082	3.53	0.030	0.061	0.082	0.093	0.178
E_a (kcal/mol)	35.0	34.4	35.1	34.1	33.7	40.0	39.7	50.3	50.0	49.0	48.5	47.5
Hammett σ	−0.27	−0.17	0.00	0.23	0.78	0.00	0.00	−0.27	−0.17	0.00	0.23	0.78

[a]Taken from Ref. 15.
[b]Present work.

Å → 1.0 Å and d_2 = 3.9 Å → 1.4 Å; at each grid point, the AM1 heat of formation (ΔH_F) was calculated by geometry optimization with respect to all degrees of freedom except d_1 and d_2. Values of the relative energy $E(d_1,d_2)$ were obtained by normalizing each ΔH_F value to that for the starting structure in the simulation. The Mopac program§ employed in the present study can accommodate a single reaction coordinate (e.g., d_1) automatically but not two at a time as desired in the present application. We therefore adopted a semiautomatic procedure for generating the d_1–d_2 energy grid by running the RC calculation automatically over the variable d_2 several times, each time manually setting d_1 to a different fixed value along its reaction coordinate. In practice, each energy grid typically required 64 separate constrained AM1 energy minimizations (i.e., 8 d_1 values times 8 d_2 values).

The deacylation reaction was simulated for a variety of acyl–enzyme intermediates, some of which were treated using both the Small Zone and the Large Zone for the sake of comparison. Each choice of reaction zone confers its own advantages and disadvantages, namely, the Small Zone is computationally more tractable but skimps on active-site residues, while the Large Zone is physically more complete and realistic but is computationally tedious. The Small Zone was employed to study the deacylation reaction for the series of para-*X*-substituted dithioester [–C(S)–S–] intermediates with X = OCH_3, CH_3, H, Cl, NO_2 and for the para-*X*-substituted thioester [–C(O)–S–] intermediate with X = H. In a complementary fashion, the Large Zone was employed for the series of para-*X*-substituted thioester [–C(O)–S] intermediates with X = OCH_3, CH_3, H, Cl, NO_2 and for the para-*X*-substituted dithioester [–C(S)–S–] intermediate with X = H (Figure 3). In all cases, the X = H intermediate was considered the "parent" model system and thus served as the structural template for subsequent calculations using the other *X* substituents.

RESULTS AND DISCUSSION

Energetics of deacylation

The parent (X = H) dithioester model system is depicted in Figure 3 to illustrate the elements of a typical RC simulation. Starting at d_1 = 2.78 Å and d_2 = 3.91 Å, the d_1–d_2 energy grid was generated by driving the d_2 reaction coordinate through the values 3.91, 3.5, 3.0, 2.5, 2.0, 1.8, 1.6, and 1.4 Å several times, each time setting d_1 to the values 2.78, 2.4, 2.1, 1.8, 1.6, 1.4, 1.2, and 1.0 Å in succession. Plots of $E(d_1, d_2)$ vs. d_1 and d_2 yielded an energy surface for the deacylation reaction, where the minimum-energy trajectory from reactants to products represents the reaction pathway and the maximum-energy point along this trajectory represents the transition state [TS]. The location of the [TS] was verified by performing a normal-mode analysis with the FORCE option in Mopac to ascertain the existence on one and only one imaginary vibrational frequency indicative of a saddle point.

For each of the other *p*-*X*-substituted thioester and dithioester intermediates (i.e., X = OCH_3, CH_3, Cl, and NO_2), the general shape of the energy surface, as well as the structures of reactants, intermediates, and products yielded results similar to their X = H counterparts. Rather than display the individual energy surfaces and [TS] structures for the entire series of substrate–enzyme intermediates, the energy surface for the X = H case is depicted in Color Plate 2 while the calculated activation energy E_a and relevant structural data for the [TS] are summarized in Table 1. The value of E_a was taken as the difference in energy between the [TS] structure and the initial reactant (i.e., the acyl–enzyme intermediate). These calculated E_a values likely represent upper bounds to the true E_a value due to the inherent limitations of the molecular model. For example, the present model excludes thermal fluctuations, which would tend to lower E_a values by sampling configurations corresponding to the close approach of reactant species.[16] Furthermore, the true E_a is an activation free energy inclusive of entropy effects, which will tend to promote [TS] formation while our calculated E_a is an activation enthalpy. It is also expected that the AM1 procedure embodies some amount of imprecision. Aside from these limitations, the present methodology has been demonstrated to be useful for noting general trends and for comparing results from similar systems.[27,29–31,‖] Encouragingly, the results from the present calculations are largely consistent with experimental data on the deacylation kinetics for this family of substrates. An Arrhenius-like plot (Figure 4) of the calculated E_a values versus $\ln(k_3)$ for these five para-*X*-substituted intermediates yielded a strong linear correlation (correlation coefficient r = 0.95) for the dithioester intermediate (i.e., Y = S).

For both the thioester and dithioester cases, the calculated

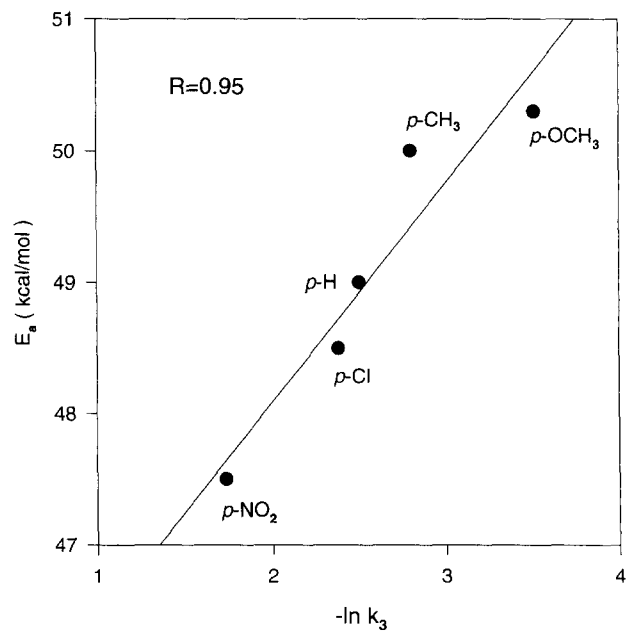


Figure 4. Arrhenius-type plot of the AM1-calculated values of the activation energy (E_a) versus $-ln\ k_3$ (Small Zone, Y = S), where k_3 is the specific rate constant for deacylation of the dithioester acyl–enzyme intermediate.

§Mopac edition 6.0 (Stewart, J.J.P., Frank J. Seiler Research Laboratory, U.S. Air Force Academy, Colorado Springs, CO 80840) was accessed through Biosym's *Insight* graphics interface (see footnote ‡).

‖For example, see Refs. 22–24 and 26.

E_a value associated with the parent (X = H) acyl–enzyme intermediate was lower by ca. 10 kcal/mol for the largest Large Zone considered (166 atoms) than for the Small Zone.** This difference is likely due to the influence of the oxyanion hole that is present in the Large Zone but absent in the Small Zone. The magnitude of this stabilizing effect is consistent with the ca. 14 kcal/mol calculated similarly by Kollman and co-workers in the case of the serine proteases.[29,30] This finding that the oxyanion hole stabilizes the tetrahedral intermediate disagrees with Asboth et al., who concluded that such stabilization is essential with the serine proteases[18] but not with the cysteine proteases.[19] These workers based their conclusion on kinetic measurements that revealed that serine proteases can hydrolyze ester [–C(O)–O–] substrates but not thionoester [–C(S)–O–] substrates, whereas cysteine proteases like papain, can hydrolyze both types. Consistent with Asboth et al., the AM1 geometric data (see below) reveal that during deacylation the $C=S_{thiono}$ bond of the dithioester penetrates more deeply into the oxyanion hole than does the corresponding C=O bond of the thioester. It is logical to conclude that the oxyanion hole of the serine proteases may be less able than that of the cysteine proteases to accommodate the greater size of the S_{thiono} atom and/or the greater extension of the $C\rightarrow S_{thiono}$ bond. However, this argument has been questioned by Arad et al.,[31] who found that the oxyanion hole of both trypsin (a serine protease) and papain are nearly superimposible and thus presumably are equally able to accommodate the steric bulk of the S_{thiono} atom.

Regarding the above issue, it is important to recognize that the acyl–enzyme (and subsequent tetrahedral) intermediate formed during the rate-limiting deacylation step is different in each case. With the serine and cysteine proteases, the relevant acyl–enzyme intermediates will be –C(O)–O– and –C(O)–S–, respectively, for ester substrates and –C(S)–O– and –C(S)–S–, respectively, for thionoester substrates. As a consequence, the fact that thionoester substrates are hydrolyzed by cysteine proteases but not by serine proteases may involve electrostatics besides just sterics, as argued by Asboth and Polgar.[18] For example, differences between –C(S)–O– and –C(S)–S– in the distribution of (negative) charge during the deacylation step would likely influence the specific interaction of the S_{thiono} atom with the oxyanion hole. These differences in charge distribution would also likely influence the stability of the tetrahedral intermediate in each case, thus affecting the reactivity of the intermediate and especially of the leaving group (realizing that, in general, –S– is a better leaving group than –O–). Such an "electrostatic" explanation would be more consistent with the superimposibility of the oxyanion holes of trypsin and papain as found by Arad et al.[31]

In their modeling study of the serine proteases, Kollman et al.[29,30] noted that the aspartate belonging to the catalytic triad might be more important than the oxyanion hole in stabilizing the tetrahedral intermediate. In papain, the far distance (5.46 Å in *pdb9pap*) between $O^{\delta 1}$ of Asp-158 and $N^{\delta 1}$ of His-159 casts doubt on the likelihood of a functioning catalytic triad involving Asp-158.[9] However, $O^{\delta 1}$ of Asn-175 and $N^{\epsilon 2}$ of His-159 are close enough (2.77 Å in *pdb9pap;* 2.99 Å from *Discover* energy minimization) to form a hydrogen bond.[6,9] This observation has raised speculation whether Asn-175 rather than Asp-158 participates in the putative catalytic triad.[29,30] While this possibility is currently under investigation in our laboratory,[7] exploratory RC simulations find scant evidence of a catalytic role for Asn-175 in terms of lowering E_a.

In another comparison, the calculated E_a value for the parent X = H acyl–enzyme intermediate was lower for the thioester [–C(O)–S–] than for the dithioester [–C(S)–S–] by ca. 10 kcal/mol in the Large Zone. This difference is qualitatively consistent with experiments on this series of substrates indicating that k_3 at 20°C is ca. 60 times larger for the thioester than for the corresponding dithioester. This difference is also consistent with both previous AM1 and 6-31G* calculations[27] and with the present AM1 calculations on the model acyl–enzyme intermediates, which find a reversed dipole for the dithioester compared with the thioester [i.e., –C(←S)–S– versus –C(→O)–S–], insinuating that the former is a much poorer site for nucleophilic attack.

Corresponding thioester and dithioester acyl–enzyme intermediates also displayed differences in their propensity toward product formation. Starting at one step beyond the [TS] along the reaction coordinate (i.e., d_1 = 1.2 Å, d_2 = 1.4 Å for Large Zone), unconstrained AM1 energy minimization of the two systems led to different final geometries. For the thioester [–C(O)–S–], the $C–S_{thiol}$ scissile bond (3.70 Å) and the water H–OH bond (3.26 Å) undergo cleavage, whereupon the tetrahedral intermediate essentially decomposes into product (i.e., RCOOH + Enz–S^-). For the dithioester [–C(S)–S–], the $C–S_{thiol}$ scissile bond (1.87 Å) and the water H–OH bond (2.30 Å) are elongated beyond their equilibrium bond lengths but remain intact. Likewise, the hybridization around the reaction site is distinctly sp^2 (product-like) for the thioester and sp^3 (TS-like) for the dithioester. The tetrahedral intermediate thus appears to be more stable for the dithioester than for the thioester, thus suggesting slower kinetics for the former. We believe that this extra stability of the dithioester [-C(=S)-S-] is derived partially from the ability of the highly polarizable S_{thiono} to promote delocalization of the anionic intermediate –1 charge over the reaction site.[27] The deacylation pathway of the corresponding thioester [-C(=O)-S-], lacking the extra stabilizing effect of S_{thiono}, proceeds directly to product with a large build-up of negative charge on S_{thiol}. In addition, deacylation of the thioester will be favored over the dithioester by the greater stability of the carboxylic acid

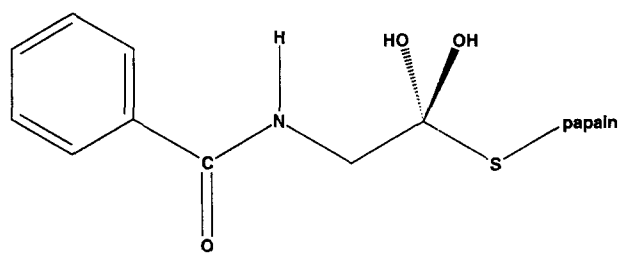


Figure 5. Illustration of the gem*-diol tetrahedral intermediate formed during some simulations of the thioester and dithioester deacylation reaction.*

**This largest Large Zone contained additional residues that effectively tethered the ImH^+ of His-159, thereby preventing it from unrealistically migrating toward the *Y* atom of $^{\delta+}C \rightarrow Y^{\delta-}$ at the reaction site. This measure increased confidence in the calculated E_a value.

Table 2. Key geometric data from AM1 reaction coordinate simulation of deacylation of dithioester and thioester model acyl–enzyme intermediates[a]

	Dithioester (Y = S)								Thioester (Y = O)						
	RC[b]								RC[b]						
Parameter	1	2	3	4	5	6	**7**	8	1	2	3	4	5	**6**	7
d_1 (Å)	2.78	2.50	2.20	2.00	1.80	1.60	**1.40**	1.20	2.62	2.30	2.00	1.80	1.60	**1.40**	1.20
d_2 (Å)	4.50	4.00	3.50	2.50	2.00	1.80	**1.60**	1.40	3.92	3.00	2.50	2.00	1.80	**1.60**	1.40
ϕ (deg)	−97.8	−99.0	−103.5	−111.5	−109.5	−110.2	**−123.6**	−135.7	−103.1	−106.3	−112.8	−117.4	−117.2	**−130.6**	−147.0
ψ (deg)	−14.59	−17.00	−19.22	−18.97	−10.91	3.06	**16.20**	44.92	−10.92	2.36	7.04	17.33	10.62	**30.54**	46.44
$S_{thiol} \cdots N^c$	2.90	2.91	2.92	2.93	2.91	2.89	**2.92**	3.03	2.92	2.91	2.90	2.91	2.93	**2.93**	3.00
$S_{thiol} \cdots C^c$	3.71	3.76	3.84	3.94	3.84	3.72	**3.81**	3.86	3.76	3.68	3.72	3.71	3.76	**3.79**	4.34
$C–S_{thiol}{}^d$	1.65	1.64	1.64	1.65	1.70	1.71	**1.79**	1.90	1.71	1.72	1.73	1.76	1.81	**1.87**	2.16
$C=Y^e$	1.57	1.58	1.58	1.58	1.60	1.63	**1.68**	1.73	1.24	1.24	1.24	1.24	1.25	**1.26**	1.29

[a]The location of the [TS] is highlighted in boldface.
[b]RC, Reaction coordinate.
[c]Interatomic distance of postulated nonbonded interaction (in angstroms).
[d]Bond length of scissile bond (in angstroms).
[e]Bond length of target C=O bond for thioester or C=S bond for dithioester (in angstroms).

[–C(O)–OH] product in the first case relative to the thionocarboxylic acid [–C(S)–OH] product in the second case.

Mechanistic findings

The RC simulations located the characteristically tetrahedral [TS] structure on the d_1–d_2 reaction surface near the grid points d_1 = 1.6 Å and d_2 = 1.4 Å for both the thioester (Y = O) and dithioester (Y = S) intermediates. Between the equilibrium state and [TS], the C–S scissile bond has elongated from 1.71 to 1.87 Å (9.4%) for the thioester and from 1.65 to 1.79 Å (8.5%) for the dithioester. Likewise, the C = *Y* bond has elongated from 1.24 to 1.26 Å (1.6%) for the thioester and from 1.57 to 1.68 Å (7%) for the dithioester. Thus, the C = S bond is longer than the corresponding C = O bond at equilibrium (by 0.33 Å) and even more so at the [TS] (by 0.42 Å). As previously proposed,[27] the small perturbation in the length of the C = O bond (1.6%) compared with the corresponding C = S bond (7%) implies that the intermediate in the deacylation of the thioester [–C(O)–S–] is much closer to severance of the C–S scissile bond and, hence, farther along the reaction pathway toward formation of products. This argument is corroborated by noting that the C–S scissile bond is more elongated at the [TS] for the thioester (9.4%) than for the dithioester (8.5%) on a percentage basis. Furthermore, more of the −1 charge at the reaction center has already transferred to the S_{thiol} leaving group for the thioester than for the dithioester.[27]

Pertaining to the attacking H_2O nucleophile, the present results indicate that proton abstraction by His-159 occurs simultaneous with, rather than prior to, nucleophilic attack on the thioester acyl–enzyme intermediate. Even at the [TS], the H · · · OH bond is elongated to 1.13 Å but has not yet broken. Only after cleavage of the C(O)–S scissile bond (d_{C-S} = 3.70 Å) does the H_2O dissociate into OH^- and H^+ (d_{HO-H} = 3.26 Å). Similar results were found for the dithioester case as well. Therefore, the "abstraction" and "attack" steps of deacylation appear to be concerted rather than sequential. In a related matter, a curious observation during the RC simulations for the thioester acyl–enzyme intermediate was the tendency for the ImH^+ moiety (representing His-159) to migrate toward the emerging anionic O atom of $C^{\delta+} \rightarrow O^{\delta-}$ in the tetrahedral intermediate. The resulting electrically neutral intermediate is a *gem*-diol (Figure 5) similar to that identified by Arad et al.[31] as a possible stable intermediate in their molecular modeling of the acylation of papain. In the present study, this *gem*-diol formed during all RC simulations in which the oxyanion hole was missing (e.g., the Small Zone). It also formed during simulations where the oxyanion hole was present and the ImH^+ was unconstrained so as to allow migration toward the carbonyl O atom. However, this *gem*-diol intermediate did not

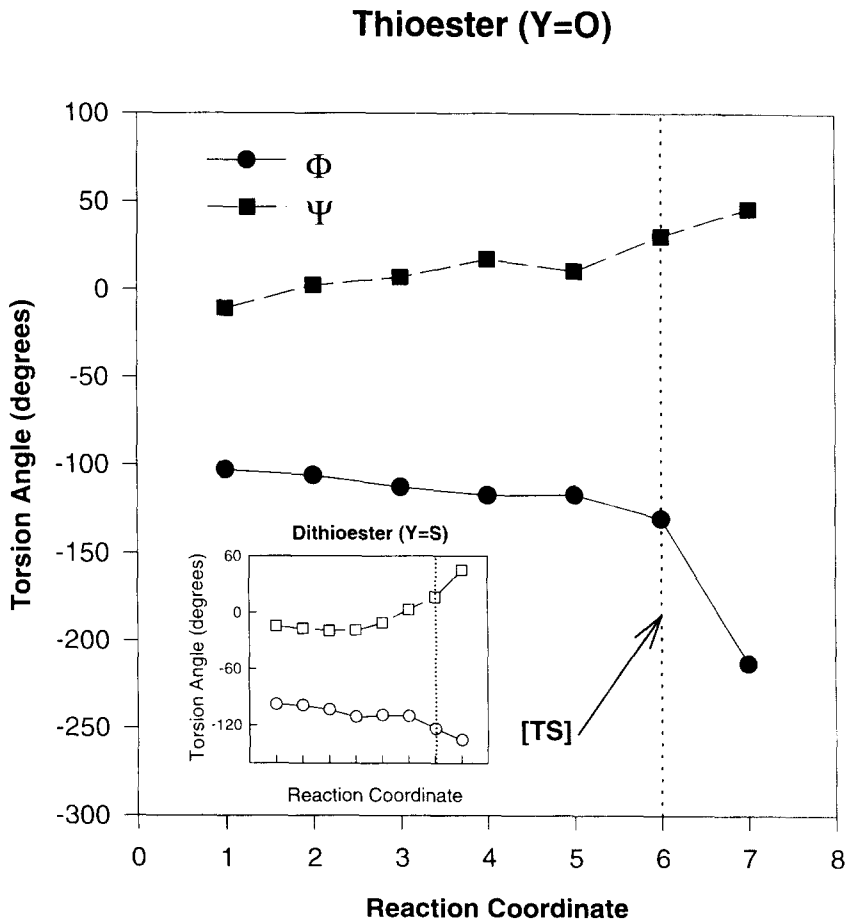


Figure 6. Plot of the AM1-calculated torsion angles ϕ and ψ along the deacylation reaction pathway for the parent (X = H) thioester acyl–enzyme intermediate.

form for the Large Zone in which the ImH^+ group was effectively tethered by attached residues. Instead, the ImH^+ remained poised in position to donate its proton to the emerging S^-_{thiol} of Cys-25 at the completion of the deacylation step. We thus regard the formation of the *gem*-diol as an artifact arising from the limited size of the reaction zone. It should be noted that an intermediate analogous to the *gem*-diol was observed less often in simulations involving the corresponding dithioester acyl–enzyme intermediate. This is consistent with two general observations from this study: (1) the polarity of the $^{\delta+}C=Y^{\delta-}$ bond is appreciably less for Y = S than for Y = O and (2) S is a poorer proton acceptor than O.

Analysis of the calculated structural data failed to provide evidence of a Cys · · · His · · · Asp catalytic triad in papain comparable to the Ser · · · His · · · Asp triad in the serine proteases. The *Discover*-optimized acyl–enzyme structure for the thioester exhibited a point of closest approach between Asp-158 and His-159 of 5.29 Å for the $O^{\delta1} \cdots N^{\delta1}$ distance. This value compares well with the 5.46 Å observed in the *pdb9pap* crystal structure of papain.[9] The corresponding $O^{\delta1} \cdots N^{\delta1}$ distance in the serine protease chymotrypsin is much closer at 2.61 Å.[32] Furthermore, the side-chain O atoms of Asp-158 form hydrogen bonds with the main-chain NH groups of both Ala-136 and Ala-137 and not with the Im side chain of His-159. These structural comparisons argue against a significant role for a Cys · · · His · · · Asp catalytic triad in papain.

Key structural data derived from the RC simulations on the parent (X = H) thioester and dithioester intermediates are summarized in Table 2 for the Large Zone. Among the most notable findings, the $S_{thiol} \cdots N$ distance remains nearly constant at ca. 2.90 Å up to formation of the [TS] structure, after which it begins to increase as the reaction proceeds to product. Since the $S_{thiol} \cdots N$ distance remains nearly invariant prior to the [TS], one might reasonably expect both ϕ and ψ to vary only slightly. Surprisingly, they both change significantly along the reaction coordinate. Initially at −100° and −15°, respectively, ϕ and ψ rotated incrementally to ca. −130° and +33° at the [TS] (Figure 6). These rotations occur as the hybridization around the target C = Y carbon transforms from sp^2 to sp^3 during formation of the anionic tetrahedral intermediate. They also appear to rotate so that the C = Y bond (which becomes $C^{\delta+} \rightarrow Y^{\delta-}$ for the tetrahedral intermediate) points more optimally toward the oxyanion hole.

The fact that the $S_{thiol} \cdots N$ distance remains constant despite the large changes in ϕ and ψ reinforces the notion that this nonbonded contact plays a special role in the deacylation mechanism.[11–17,27] At the same time, some consideration should be given to the $S_{thiol} \cdots C$ interaction [i.e., $S_{thiol} \cdots C(O)NH$] as playing a role in this unique nonbonded contact. This proposition is supported by the present calculations, which find that the $S_{thiol} \cdots C$ distance is also nearly constant (i.e., 3.70 Å) along the reaction pathway up to the [TS]. On the basis of partial atomic charges calculated by AM1 and 6-31G*, the $S^{\delta-} \cdots C^{\delta+}$ interaction appears to be electrostatically more attractive than the $S^{\delta-} \cdots N^{\delta-}$

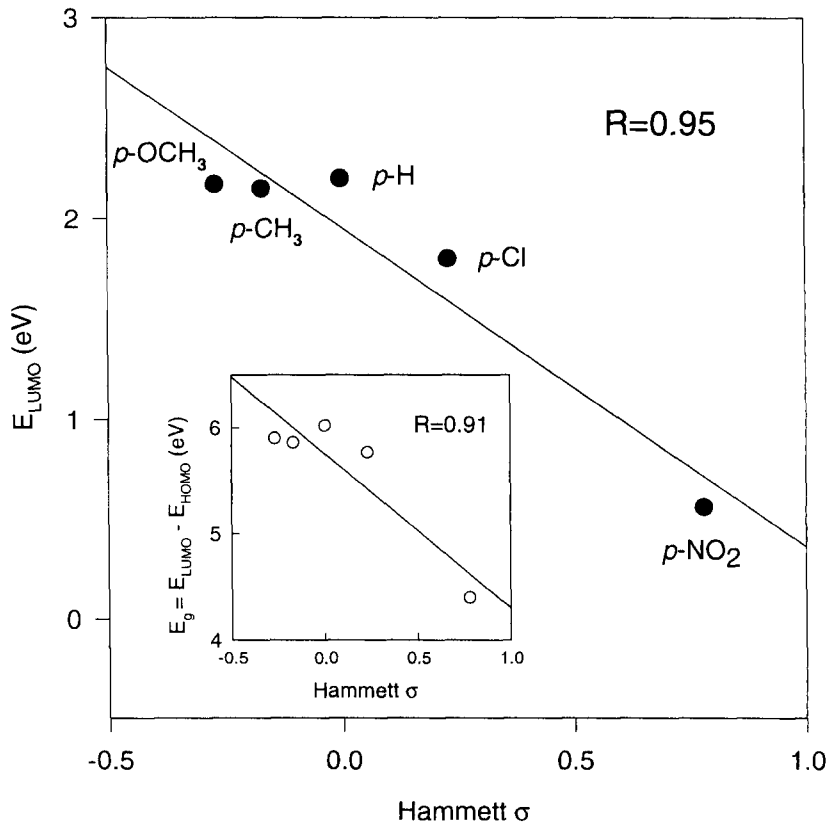


Figure 7. Linear plot of the AM1-calculated E_{LUMO} versus Hammett σ for the series of p-X-*substituted acyl–enzyme intermediates under study. Inset: Linear plot of the AM1-calculated band gap $E_g = E_{LUMO} - E_{HOMO}$ versus Hammett σ for the same series of substrate intermediates.*

interaction. Interestingly, the S · · · C contact distance is particularly short at $\phi = -105°$, $\psi = +15°$, which was deduced by Kim et al.[17] on the basis of resonance Raman spectra to be the B-type conformation adopted by the X = H acyl–papain intermediate. This evidence raises some question as to the exact nature of this key nonbonded interaction, i.e., whether it is S · · · N or S · · · C, or possibly some combination S · · · [C(O)NH].

Possible catalytic role of frontier molecular orbitals

The proposed role for the S_{thiol} · · · N [or perhaps S · · · C or S · · · C(O)NH] nonbonded interaction in the catalytic mechanism elicits questions as to its fundamental physical nature.[17,27] Two possibilities, which are not mutually exclusive, are an electrostatic interaction (e.g., $S^{\delta-} \cdots C^{\delta+}$) and an orbital overlap interaction (e.g., HOMO · · · LUMO). Regarding the orbital overlap explanation, other workers[11–17] view the S_{thiol} · · · N interaction in terms of an attraction between the N atom as the nucleophile and the S_{thiol} atom as the electrophile. They have enunciated[17] two possible qualitative interpretations from MO theory: (1) an HOMO localized on the amide N overlaps with an LUMO σ* located near the S_{thiol}—CH_2 ·⋱· papain moiety of the reaction site; and (2) the lone-pair electrons on the amide N atom interact with the *d* orbitals of the S_{thiol}. On the basis of our previous work[27] as well as the present calculations, we have deduced an opposing interpretation in which the location of the interacting HOMOs and LUMOs is the reverse of that proposed by these workers. Specifically, we locate the HOMOs at the electron-rich reaction site near S_{thiol} and the LUMOs on the (relatively speaking) electron-poor benzamide ring system [i.e., *X*–Ph–C(O)NH]. In contrast to these other workers, we view the S_{thiol} as the nucleophile and the amide N atom as the electrophile.

To substantiate our alternative view, computer graphics images are presented of the AM1-calculated HOMOs and LUMOs for the X = H model acyl–enzyme intermediate Ph–C(O)NH–CH_2–C(*Y*)–S–CH_2CH_3 for both the thioester (Y = O) and the dithioester (Y = S) cases (Color Plate 3). These images visualize the model acyl–enzyme intermediates at two points along the reaction pathway: (1) before hydrolysis, and (2) at the early stages of hydrolysis (i.e., while forming an adduct with an H_2O molecule). The torsion angles φ and ψ were fixed at −105° and +15°, respectively, in accordance with the experimental values deduced by Kim et al.[17] for the corresponding substrate–papain complex, after which AM1 geometry optimization was performed on the remaining degrees of freedom. For the thioester case, the HOMO resides near the reaction site while the LUMO resides on the benzamide moiety. For the dithioester case, the HOMO, LUMO, and LUMO+1 reside near the reaction site while the LUMO+2 resides on the benzamide moiety. The association of the LUMO and LUMO+1 with the reaction site in the dithioester case is understood by recognizing that S_{thiol} and especially S_{thiono} are electron deficient at the onset of deacylation and so assume more LUMO than HOMO character; they become electron acceptors and convert to HOMOs as the deacylation reaction proceeds.[27] For both the thioester and dithioester cases, these HOMOs and LUMOs remain situated in these positions at both points along the reaction pathway, refuting any suggestion that the presence of the H_2O nucleophile might alter or reverse the disposition of the HOMOs and LUMOs. We believe that these images provide compelling evidence in support of our assertion that the HOMOs and LUMOs should be identified, respectively, with S_{thiol} and N rather than the reverse.

To evaluate a possible catalytic role for such HOMO–LUMO interactions in the deacylation reaction, we sought evidence for a correlation between these orbital energies and the Hammett constant σ for the series of *p-X*-substituted acyl–enzyme intermediates. We reasoned that the larger k_3 values observed for stronger electron-accepting *p-X* substituents may be due to enhanced HOMO–LUMO overlap, which would facilitate charge delocalization from the reaction center onto the benzamide portion of the substrate. This view was corroborated by a linear regression plot of E_g vs. σ giving a correlation coefficient r = 0.91 (Figure 7, inset). Analysis of the HOMO energy and LUMO energy versus σ revealed that the former remained nearly constant while the latter decreased appreciably. This behavior is predictable inasmuch as the LUMOs are in closer proximity to the perturbing *p-X* substituent and, hence, should be affected more than the HOMOs located at the reaction site. A plot of the LUMO energy vs. σ (Figure 7) yielded an even stronger linear correlation (r = 0.95), supporting the hypothesis that HOMO–LUMO interactions play a significant role in the catalytic mechanism of the deacylation of papain and likely other cysteine proteases.[27]

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support for this research from the Institute of General Medicine of the National Institutes of Health and from the University of Missouri-St. Louis Graduate School for a Summer Fellowship to Y.L. The authors also thank Biosym Technologies, Inc. (San Diego, CA) for access to their programs *Insight* and *Discover*. Finally, the authors are indebted to Dr. Gregory Duncan for his collaboration in this area of research.

REFERENCES

1 Shaw, E. Cysteinyl proteinases and their selective inactivation. *Adv. Enzymol. Relat. Areas Mol. Biol.* 1990, **63,** 271–347
2 Duffy, M.J. The role of proteolytic enzymes in cancer invasion and metastasis. *Clin. Exp. Metastasis* 1992, **10,** 145–155
3 Berquin, I.M. and Sloane, B.F. Cysteine proteases and tumor progression. *Perspect. Drug Design Discovery* 1994, **2,** 371–388
4 Sloane, B.F., Moin, DK., Krepela, E., and Rozhin, J. *Cancer Metastasis Rev.* 1990, **9,** 333
5 Brocklehurst, K. Acyl group transfer-cysteine proteases. In: *Enzyme Mechanisms* (Page, M.I. and Williams, A., eds.). Royal Society of Chemistry, London, 1987, pp. 140–158
6 Lowe, G. The cysteine proteases. *Tetrahedron* 1976, **32,** 291–302

7 Lin, Y. and Welsh, W.J. Rational design of mechanism-based inhibitors of cathepsin B. In preparation (1996)
8 Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., and Bode, W. The refined 2.15 Å x-ray crystal structure of human liver cathepsin B: The structural basis for its specificity. *EMBO J* 1991, **10,** 2321–2330
9 Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A., and Drenth, J. Structure of papain at 1.65 Å resolution. *J. Mol. Biol.* 1984, **179,** 233–256
10 Kamphuis, I.G., Drenth, J., and Baker, E.N. Thiol proteases: Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *J. Mol. Biol.* 1985, **182,** 317–329
11 Huber, C.P., Ozaki, Y., Pliura, D.H., Storer, A.C., and Carey, P.R. Precise structural information for transient enzyme–substrate complexes by a combined x-ray crystallographic-resonance Raman spectroscopic approach. *Biochemistry* 1982, **21,** 3109–3115
12 Carey, P.R. and Storer, A.C. Molecular details of enzyme–substrate transients by resonance raman spectroscopy. *Acc. Chem. Res.* 1983, **16,** 455–460
13 Varughese, K.I., Storer, A.C., and Carey, P.R. Directional preference for a catalytically important N · · · S contact seen in acyl-thiolproteases. *J. Am. Chem. Soc.* 1984, **106,** 8252–8257
14 Carey, P.R., Lee, H., Ozaki, Y., and Storer, A.C. Rate–structure correlation for enzyme–substrates intermediates: Resonance Raman and kinetic studies on some N-benzoylglycine (dithioacyl) papains. *J. Am. Chem. Soc.* 1984, **106,** 8258–8262
15 Storer, A.C. and Carey, P.R. Comparison of the kinetics and mechanism of the papain-catalyzed hydrolysis of esters and thiono esters. *Biochemistry* 1985, **24,** 6808–6818
16 Tonge, P.J., Lee, H., Sans Cartier, L.R., Ruzsicska, B.P., and Carey, P.R. Striking changes observed in key acyl-enzyme linkages by resonance raman experiments near 77 K. *J. Am. Chem. Soc.* 1989, **111,** 1496
17 Kim, M., Birnbaum, G.I., Hynes, R.C., Neugebauer, W., and Carey, P.R. Fine-tuned conformation of dithioacyl-papain intermediates: Insights from resonance raman spectroscopy. *J. Am. Chem. Soc.* 1993, **115,** 6230–6237
18 Asboth, B. and Polgar, L. Transition-state at the oxyanion binding sites of serine and thiol proteinases: Hydrolysis of thiono and oxygen esters. *Biochemistry* 1983, **22,** 117–122
19 Asboth, B., Stokum, E., Khan, I.U., and Polgar, L. Mechanism of action of cysteine proteinases: Oxyanion binding site is not essential in the hydrolysis of specific substrates. *Biochemistry* 1985, **24,** 606–609
20 Williams, R.V., Kurtz, H.A., and Farley, B. The use of semiempirical energy partitioning in the study of through space (homoaromatic) interactions. *Tetrahedron* 1988, **44,** 7455–7460
21 Sotomatsu, T., Murata, Y., and Fujita, T. Correlation analysis of substituent effects on the acidity of benzoic acids by the AM1 method. *J. Comput. Chem.* 1989, **10,** 94–98
22 Dewar, M.J.S. and Yuan, Y.-C. AM1 studies of E2 reactions. 1. Mechanism and leaving group effects. *J. Am. Chem. Soc.* 1990, **112,** 2088–2094
23 Dewar, M.J.S. and Yuan, Y.-C. AM1 Studies of E2 reactions. 2. Regioselectivity, stereochemistry, kinetic isotope effects, and competition with S_N2 reactions. *J. Am. Chem. Soc.* 1990, **112,** 2095–2105
24 Merz, K.M., Jr., Hoffmann, R., and Dewar, M.J.S. Mode of action of carbonic anhydrase. *J. Am. Chem. Soc.* 1989, **111,** 5636–5649
25 Katagi, T. Theoretical studies on the alkaline hydrolysis of *N*-methylcarbamates. *J. Comput. Chem.* 1990, **11,** 524–530
26 Bacaloglu, R., Bunton, C.A., and Ortega, F. Interaction of nitroarenes with hydroxide ion. An AM1 molecular orbital treatment. *J. Am. Chem. Soc.* 1989, **111,** 1041–1047
27 Duncan, G.D., Huber, C.P., and Welsh, W.J. Molecular orbital studies of the structure and reactivity of model substrate intermediates in the deacylation of the cysteine protease papain. *J. Am. Chem. Soc.* 1992, **114,** 5784–5794
28 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 1985, **107,** 3902–3909
29 Schröder, S., Daggett, V., and Kollman, P. A comparison of the AM1 and PM3 semiempirical models for evaluating model compounds relevant to catalysis by serine proteases. *J. Am. Chem. Soc.* 1991, **113,** 8922–8925
30 Daggett, V., Schröder, S., and Kollman, P. Catalytic pathway of serine proteases: Classical and quantum mechanical calculations. *J. Am. Chem. Soc.* 1991, **113,** 8926–8935
31 Arad, D., Langridge R., and Kollman, P.A. A simulation of the sulfur attack in the catalytic pathway of papain using molecular mechanics and semiempirical quantum mechanics. *J. Am. Chem. Soc.* 1990, **112,** 491–502
32 Tsukada, H., and Blow, D.M. Structure of α-chymotrypsin refined at 1.68 Å resolution. *J. Mol. Biol.* 1985, **184,** 703