A Quantum Mechanical Study of the Active Site of Aspartic Proteinases†

A. J. Beveridge*,‡,‖ and G. C. Heywood§,⊥

Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England, and Department of Surgery, St. George's Hospital Medical School, University of London, Cranmer Terrace, London SW17 ORE, England

Received May 21, 1992; Revised Manuscript Received October 27, 1992

ABSTRACT: We have performed ab initio self-consistent field (SCF) and configuration interaction (CI) calculations on the active site of the aspartic proteinases pepsin and endothiapepsin. The active site, which carries a formal negative charge to effect hydrolysis, was modeled as a formic acid/formate anion moiety and a water molecule, and the nearest hydrogen bonding residues (Gly34, Ser35, Gly217, and Thr218, with respect to the residue numbering in endothiapepsin) were modeled as formamide and methanol molecules. Four possible binding modes for the active-site water molecule were considered. In contrast to previous theoretical studies, we predict that the most stable form has the water molecule forming a bifurcated hydrogen bond to the inner oxygens of Asp32 and -215, with Asp32 being ionized. The calculations suggest that the water molecule prefers to bind across the shortest $O_{D32}\cdots O_{D215}$ diagonal of the active-site carboxyl groups and therefore the binding mode of the water molecule for all the native aspartic proteinases can be readily predicted by measuring these distances.

The aspartic proteinases comprise a family of hydrolytic enzymes widely distributed in nature and involved in processes as diverse as milk clotting (chymosin) (Foltmann, 1966), blood pressure regulation (renin) (Burton et al., 1980), digestion (pepsin) (Schwann, 1836), and retroviral maturation (HIV-I and -II proteinases) (Kramer et al., 1986). The active sites are characterized by two aspartate carboxyl groups (D32 and D215 in endothiapepsin) containing a tightly bound water molecule (Figure 1). The pH/activity (hydrolysis) profiles are bell-shaped (Fruton, 1976, 1987), suggesting two mechanistically implicated deprotonations associated with the aspartate dyad. It is generally accepted that the active site carries a formal negative charge to effect hydrolysis and therefore a correct assignment of the associated pK_a's is essential toward understanding the hydrolytic mechanism. Furthermore, in the absence of neutron diffraction data, a detailed understanding of the hydrogen bonds responsible for binding the water molecule to the active site is crucial if one is to accurately describe the enzyme mechanism.

This problem has been addressed by two theoretical studies. Hadzi et al. (1987) have performed ab initio self-consistent field (SCF) calculations on endothiapepsin, where the active site has been modeled as a negatively charged formic acid/formate anion moiety and the position of the water molecule determined by geometry optimization. Goldblum (1988) has also studied the active site of penicillopepsin using the semiempirical quantum mechanical MNDO/H method (Goldblum, 1987). Several different models for the active site were investigated; the most sophisticated represented the two active-site aspartates by a formic acid/formate anion moiety with the neighboring residues Thr34, Gly35, Ser36, Thr214, Gly215, and Thr216 replaced by methanol and formamide molecules. The optimum position of the bound water molecule was again determined by geometry optimization. These calculations indicated that there was no clear preference for ionization of one aspartate over the other, suggesting the existence of a prototropic equilibrium.

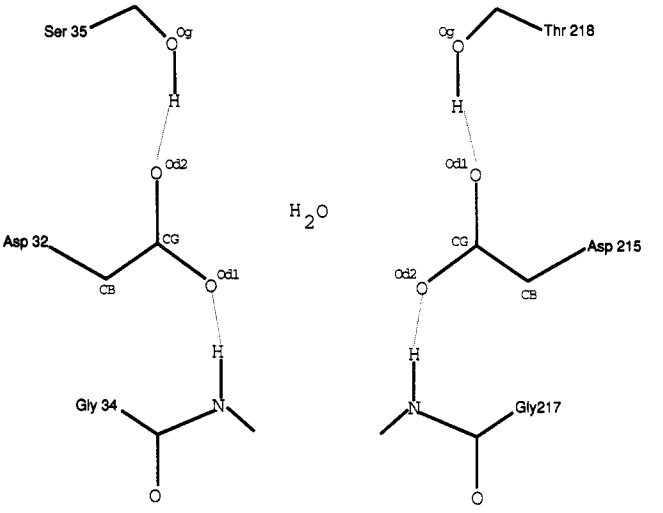

FIGURE 1: Schematic illustration of the active site of aspartic proteinases. The active site consists of the two carboxyl groups from Asp32 and -215 and a tightly bound water molecule. The active site is almost planar and carries an overall charge of −1. Protons involved in forming hydrogen bonds between the water molecule and the active site are not explicity drawn because the exact locations of these protons have yet to be determined experimentally. Gly34, Ser35, Gly217, and Thr218 form hydrogen bonds from above and below the plane of the active site to Asp32 and -215.

Both of these theoretical studies have assumed that one of the aspartate inner oxygens is protonated.[1] The water molecule then forms a hydrogen bond to the outer oxygen of the ionized aspartate. Curiously, there have been no theoretical studies to date that have considered the possibility of the water molecule forming a bifurcated hydrogen bond to the inner oxygens of the active-site carboxyl groups. The primary

† A.J.B is a recipient of a CRAY University Research Fellowship. G.C.H. acknowledges the support of the Wellcome Trust.

* Author to whom correspondence should be addressed.
‡ Birkbeck College, University of London.
‖ Present address: Research School of Chemistry, The Australian National University, GPO Box 4, Canberra ACT 2601, Australia.
§ St. George's Hospital Medical School, University of London.
⊥ Deceased.

[1] We define the inner oxygens as Od1 of Asp32 and Od2 of Asp215. The outer oxygens are Od2 of Asp32 and Od1 of Asp215 (Figure 1).

0006-2960/93/0432-3325$04.00/0

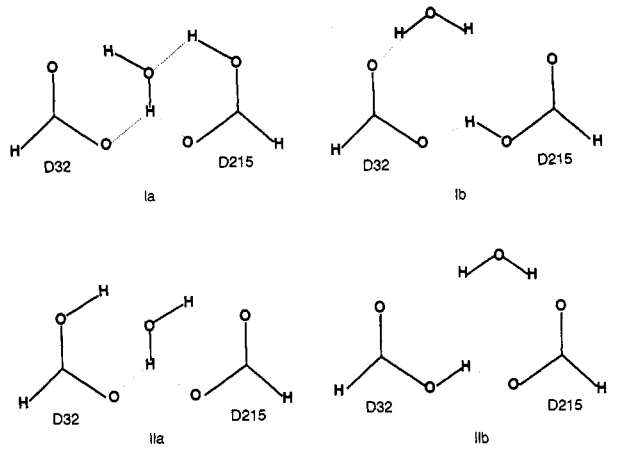

FIGURE 2: Four protomers considered in the present study for the active site of aspartic proteinases. Three models for the active site have been used. Model 1: the active site is modeled as a water molecule and a formic acid/formate anion moiety. Model 2: the four nearest hydrogen bonding residues (Gly34, Ser35, Gly217, and Thr218; Figure 1) are modeled using formamide and methanol molecules. Model 3: residues 34, 35, 217, and 218 are represented using atom-centered point charges. This simulates the electrostatic and polarization effects of the local enzyme environment. Previous theoretical studies considered only protomers Ib and IIb. The ab initio calculations on endothiapepsin by Hadzi et al. used ammonia and water molecules to mimic the hydrogen bonds from residues 34, 35, 217, and 218. The MNDO/H calculations on penicillopepsin by Goldblum used methanol and formamide molecules to represent residues Thr34, Gly35, Ser36, Thr214, Gly215, and Thr216. Thr34 and Thr214 (not indicated in Figure 1) form part of the local peptide backbone. Note: the residue numbering for penicillopepsin is slightly different.

objective of the present study has been to examine this possibility and compare it energetically to the nonbifurcated forms previously considered (Figure 2).

THEORETICAL METHODOLOGY

All ab initio SCF calculations were performed with the ab initio quantum chemistry package GAMESS (Guest, 1990). Second-order configuration interaction (CI) wavefunctions were computed using the DIRECT-CI module (Saunders & Van Lenthe, 1983). Four basis sets were used: STO-3G, 4-31G, 4-31G*, and 4-31G** (Hehre et al., 1969, 1970, 1972; Ditchfield et al., 1971). All geometry optimizations were performed with the 4-31G basis set. This was considered to be the smallest basis set which could adequately describe intermolecular hydrogen bonds (Scheiner et al., 1983). All calculations were carried out on the ULCC and ATLAS CRAY-XMP supercomputers.

THEORETICAL CALCULATIONS

We have performed calculations on two aspartic proteinases, pepsin (Sielecki et al., 1990) and endothiapepsin (Pearl & Blundell, 1984).

(*i*) *Endothiapepsin.* Proceeding in an analogous fashion to previous theoretical studies, we modeled the active site of endothiapepsin as a formic acid/formate anion moiety and a water molecule. This will subsequently be referred to as model 1. Given that it is clearly impossible, with such a simplified model, to account for the structural effects and forces exerted on the active site by the surrounding residues, we have applied certain geometric constraints to our active-site model. In all geometry optimizations described below, the CG–CG distance between D32 and D215 has been held fixed at the crystallographic distance of 5.23 Å, and the C–H directions for both the formic acid and the formate anion have been held in alignment with the CB–CG directions in the crystal structure (Figure 1) (Pearl & Blundell, 1984). In endothiapepsin, the deviation from planarity of the active-site carboxyl groups is 6°. This near-planarity of the active site is a feature common to nearly all of the crystal structures determined for the native aspartic proteinases, and we therefore decided to fix the relative planar orientation of the formic acid/formic anion moiety to 6° in all subsequent calculations.

Four possible arrangements of the active site were considered (Figure 2). The water molecule was allowed to bind in two distinct ways: either by forming a bifurcated hydrogen bond between the inner oxygens of D32 and D215 or by forming a single hydrogen bond to the outer oxygen of the ionized aspartate [this latter model is essentially the form studied previously (Hadzi et al., 1987; Goldblum, 1988)].

Geometry optimization of each structure was performed in four stages.

(i) The geometries (bond lengths and angles) of both the formic acid and the formate anion were optimized in the absence of the water molecule.

(ii) The water molecule was then positioned in either a bifurcated or a nonbifurcated binding mode and its position optimized (6 degrees of freedom). The torsion angle of the acidic hydrogen of formic acid was also optimized, yielding a total of 7 degrees of freedom.

(iii) The position of the water molecule was then held fixed, and the bond lengths and angles of the formic acid/formic anion moiety were optimized.

(iv) Finally, the position of the water molecule, the bond lengths of water, the torsion angles involving hydrogen bonds, and the bond lengths and angles of the formic acid/formate anion moiety were optimized simultaneously. All other geometric parameters (e.g., CG–CG), as described previously, were held fixed.

The energies for the 4-31G SCF optimizations are tabulated in Table I; the optimized structures are shown in Figure 3. The nonbifurcated structures, Ib and IIb, are predicted to be significantly more stable than their bifurcated counterparts. In particular, the most stable bifurcated form, Ia, is 6.80 kcal mol^{-1} less stable than Ib. The relative stability of the nonbifurcated forms can be largely attributed to the reduction in electronic repulsion between the two inner oxygens by the formation of a strong hydrogen bond orientated along the O···O axis. Removal of this hydrogen bond, as in the case of the bifurcated forms, would be expected to lead to an increase in electronic repulsion. Theoretically, the main deficiency of SCF wavefunctions is the neglect of electron correlation, and intuitively, one would expect the correlation energies for the bifurcated forms to be significantly higher than those for nonbifurcated structures.

To investigate this further, we calculated second-order CI wavefunctions, including all single and double excitations (SDCI), for each structure using the DIRECT-CI module of GAMESS. Size consistency corrections were determined using Davidson's algorithm (Langhoff & Davidson, 1974). It is impractical for computational reasons to include all of the occupied and virtual orbitals in the CI expansion. Therefore, several trial calculations were performed with structures Ia and Ib to determine which orbitals and reference configurations should be included in the CI wavefunctions.

Preliminary calculations using only the SCF ground state as a reference configuration and restricting the configurational subspace to the 17 highest valence and the 21 lowest virtual orbitals were performed on structures Ia and Ib. These

Table I: Calculated SCF Energies (au) for Endothiapepsin (Model 1), Relative Stabilities (kcal mol^{-1}) in Parentheses

structure	4-31G	4-31G*	4-31G**
Ia	−452.3397845 (6.80)	−452.560524 (7.53)	−452.583561 (7.74)
Ib	−452.3506192 (0.00)	−452.572517 (0.00)	−452.595902 (0.00)
IIa	−452.3287445 (13.73)	−452.552463 (12.58)	−452.575434 (12.84)
IIb	−452.3463307 (2.69)	−452.568162 (2.73)	−452.591779 (2.59)

Table II: Calculated CI Energies (au) for Endothiapepsin (Model 1), Relative Stabilities (kcal mol^{-1}) in Parentheses

structure	1R/17/21[a]	3R/17/21[a]	1R/22/21[a]	1R/22/21[b]
Ia	−452.587685	−452.597773	−452.654464 (3.52)	−452.853299 (1.75)
Ib	−452.593541	−452.603563	−452.660073 (0.00)	−452.856093 (0.00)
IIa			−452.644857 (9.55)	−452.844162 (7.49)
IIb			−452.654807 (3.30)	−452.849802 (3.95)

[a] 4-31G basis set. [b] 4-31G* basis set.

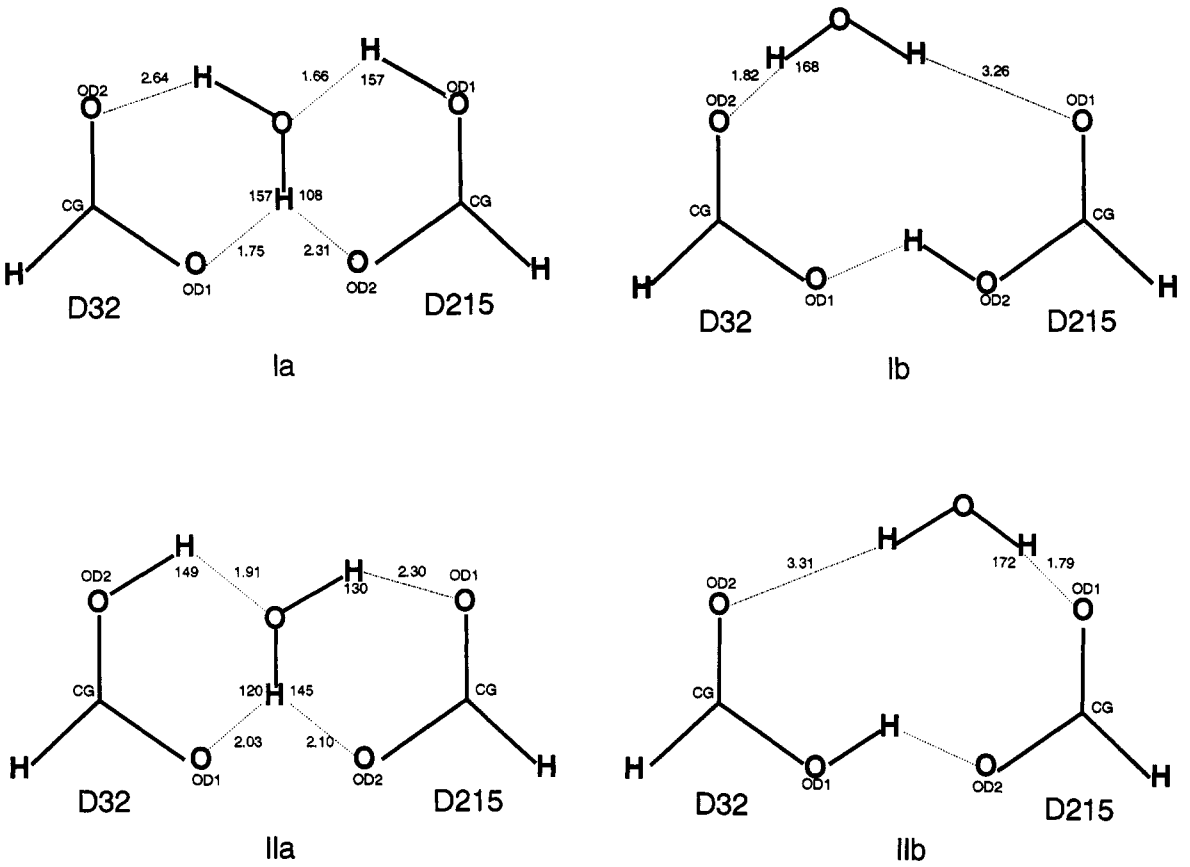

FIGURE 3: Four optimized geometries (4-31g basis set) for the active site of endothiapepsin. The water molecule lies in the plane of the two carboxyl groups.

calculations are denoted 1R/17/21 (1 reference configuration/17 valence orbitals/21 virtual orbitals). The results are tabulated in Table II. CI reduces the difference in energy between Ia and Ib from 6.80 (SCF) to 3.67 kcal mol^{-1}. Furthermore, examination of the CI expansion coefficients indicated that the next two most important configurations involved double excitations from the π_{215} and σ_{32} valence orbitals. In both of these orbitals, the electron density is centered largely on the oxygen atoms. Further calculations were performed by explicitly adding these excitations to the list of reference configurations. The results of these calculations (3R/17/21) are tabulated in Table II. The energies of both Ia and Ib were significantly lowered (by over 6 kcal mol^{-1}), but the difference in the relative energy (3.63 kcal mol^{-1}) changed only slightly.

The active space was then expanded further to include all valence orbitals (1R/22/21). Inclusion of the remaining five valence orbitals resulted in a modest change to the relative energy. It was therefore decided, for the sake of both accuracy and computational expediency, to restrict all subsequent CI calculations to the 1R/22/21 level. At this level of theory, the most stable form is again predicted to be Ib. However, the most stable bifurcated form is now only 3.52 kcal mol^{-1} less stable (Table II).

The CI calculations were then repeated with a 4-31G* basis set. The results are tabulated in Tables I and II. The 4-31G* basis set increases the difference in SCF energy between Ia and Ib; this, however, is more than compensated for by CI, which reduces the energy difference to 1.75 kcal mol^{-1}. Comparing the CI results for the 4-31G and 4-31G* basis sets indicates that the relative stabilities of both bifurcated structures are significantly improved by the addition of d-type polarization functions to the heavy atoms. This observation can be largely attributed to the effects of angular correlation

Table III: Calculated SCF Energies (au) for Endothiapepsin (Model 2), Relative Stabilities (kcal mol^{-1}) in Parentheses

structure	STO-3G	4-31G
Ia	−1007.193784 (0.0)	−1019.511100 (0.0)
Ib	−1007.180799 (8.15)	−1019.504409 (4.20)
IIa	−1007.171802 (13.79)	−1019.501458 (6.05)
IIb	−1007.174633 (12.02)	−1019.502519 (5.38)

in the 4-31G* basis set. Calculations at the SCF level have also been performed with the 4-31G** basis set. The results are tabulated in Table I. Inclusion of p-type polarization functions on the hydrogen atoms has resulted in relatively minor changes to the predicted SCF stabilities.

The results of our CI calculations suggested that the relative stabilities of the four structures might be significantly altered by the surrounding enzyme environment. This hypothesis was investigated by expanding model 1 to include two methanol and two formamide molecules representing the hydrogen bonds from residues Gly34, Ser35, Gly217, and Thr218. Gly34 and Gly217 form imido hydrogen bonds to the inner oxygens of the active-site carboxyl groups; these hydrogen bonds were mimicked using the two formamide molecules. The geometry of formamide was optimized with a 4-31G basis set and overlayed onto the crystallographic coordinates of Gly34 and Gly217. Ser35 and Thr218 form hydroxyl hydrogen bonds with the outer oxygens of D32 and D215, respectively. Methanol was optimized with a 4-31G basis set and overlayed onto Ser35 and Thr218. The hydroxyl hydrogen atoms were then orientated to produce the shortest possible hydrogen bonds to the outer oxygens of the active-site carboxyl groups. This model will be subsequently referred to as model 2. SCF calculations on all four structures (Ia, Ib, IIa, and IIb) were then performed using two different basis sets, STO-3G and 4-31G. The results are tabulated in Table III.

Structure Ia, with D32 ionized and the water binding in a bifurcated mode, is now predicted to be the most stable form. The STO3G calculations predict Ia to be more stable than Ib by 8.15 kcal mol^{-1}, while the 4-31G calculations reduce the relative stabilization energy to 4.20 kcal mol^{-1}. In general, the relative energies calculated with the STO-3G basis are twice as great as those calculated using a 4-31G basis set. Nevertheless, both basis sets predict the same ranking order for stability, i.e., Ia > Ib > IIb > IIa. This result is quite encouraging because despite the obvious limitations of the minimal STO-3G basis set, the results are qualitatively similar to the larger 4-31G basis. This suggests that once the active site has been optimized with a double ζ basis set, larger calculations, including several of the neighboring active site residues, can be performed with the smaller STO3G basis.

Perhaps the most interesting result, however, is the relative stability of Ia with respect to IIa. Both forms are bifurcated; Ia has D32 ionized while IIa has D215 ionized. The 4-31G calculations predict Ia to be more stable by 6.05 kcal mol^{-1}, and therefore one would expect pK_a1 to be associated with Asp32. The relative stability of Ia can be readily understood by comparing the optimized structures for the two bifurcated forms (Figure 3). Three strong hydrogen bonds are predicted for Ia, but the two strongest hydrogen bonds are from the innermost proton on the water molecule to $OD1_{32}$ (1.75 Å; $<OD1_{32}$–H–O_{water} = 157°) and O_{water} to H_{215} (1.66 Å; $<O_{water}$–H_{215}–$OD1_{215}$ = 157°). Therefore, the water molecule is essentially bound diagonally across the active-site carboxyl groups. Structure IIa has four hydrogen bonds between the water molecule and the carboxyl groups, but the two hydrogen bonds orientated diagonally across the carboxyl groups

Table IV: Calculated Molecular Charges for Endothiapepsin (Model 2)

structure	formamide$_{34}$[a]	methanol$_{35}$	formamide$_{217}$	methanol$_{218}$
Ia	−0.032	−0.057	−0.027	−0.025
Ib	−0.032	−0.052	−0.024	−0.025
IIa	−0.019	−0.017	−0.045	−0.052
IIb	−0.027	−0.019	−0.042	−0.043

[a] Subscripts indicate which amino acid residue is being replaced by methanol/formamide.

Table V: Calculated SCF and CI Energies (au) for Endothiapepsin Using a Point Charge Representation for the Methanol and Formamide Molecules Used in Model 2, Relative Stabilities (kcal mol^{-1}) in Parentheses

structure	SCF (4-31G)	CI (1R/22/21)
Ia	−453.362473 (2.32)	−453.674631 (0.00)
Ib	−453.366175 (0.00)	−453.672833 (1.13)
IIa	−453.453139 (9.28)	−453.665049 (6.01)
IIb	−453.359695 (2.37)	−453.665541 (5.70)

($OD2_{32}$–$OD2_{215}$) would be expected to be significantly weaker. The hydrogen bonds are both longer (1.91, 2.10 Å), and the angles are significantly smaller (149°, 145°). In the crystal structure of endothiapepsin, the $OD1_{32}$–$OD1_{215}$ distance is 4.43 Å, while the $OD2_{32}$–$OD2_{215}$ distance is 4.76 Å. Our calculations strongly suggest that the water molecule binds in a bifurcated mode across the shortest O_{32}···O_{215} diagonal. This rationale explains why D32 should be preferentially ionized. The difference between the $OD1_{32}$–$OD1_{215}$ and $OD2_{32}$–$OD2_{215}$ distances is a structural feature common to all the native enzyme crystal structures (Newman et al., 1990; Sielecki et al., 1989; Cooper et al., 1990; Lapatto et al., 1989). We therefore propose that the ionization state of the active site of all the aspartic proteinases can be readily predicted by measuring these distances.

A Mulliken population analysis (Mulliken, 1955) of the 4-31G SCF wavefunctions revealed that a significant amount of charge (0.13 e) has been transferred from the active site onto neighboring residues. The results are tabulated in Table IV. The results are fairly similar for all four structures. In general, most of the charge transfer occurs between the ionized aspartate and the two nearest hydrogen bonding residues. One would expect the predicted charge transfer to preferentially stabilize the bifurcated forms as this should help reduce the electronic repulsion between the two inner oxygens. A comparison of Tables I, II, and III supports this interpretation as both Ia and IIa have both become relatively more stable.

Electrostatic and polarization effects would also be expected to influence the relative stabilities. We have investigated the effect that the local electrostatic field has on the active site by replacing neighboring residues with partial atomic charges. We first examined the validity of such an approach by deriving atomic point charges for methanol and formamide. The program QUEST (Singh & Kollman, 1986) was used to calculate wavefunctions for both methanol and formamide with an STO-3G basis set. Partial atomic charges were then derived using the methodology developed by Singh and Kollman (1984) for the AMBER force field (Weiner et al., 1984). These point charges were then used to replace the methanol and formamide molecules used in model 2. SCF and SDCI (1R/22/21) calculations with a 4-31G basis set are tabulated in Table V. A comparison of the results in Tables III and V indicates that by modeling the formamide and methanol as point charges, qualitatively similar results to model 2 can be obtained at the CI(1R/22/21) level.

Table VI: Calculated SCF and CI Energies (au) for Endothiapepsin (Model 3), Relative Stabilities (kcal mol^{-1}) in Parentheses

structure	SCF (4-31G)	CI (1R/22/21)	correlation energy[a]
Ia	−453.498975 (1.65)	−453.812727 (0.00)	196.9
Ib	−453.501606 (0.00)	−453.808993 (2.34)	192.9
IIa	−453.486971 (9.18)	−453.802302 (6.54)	197.9
IIb	−453.494876 (4.22)	−453.801693 (6.92)	192.5

[a] Correlation energies are in kilocalories per mole.

We therefore extended the above electrostatic model by replacing the point charges for methanol and formamide with partial atomic charges for the four hydrogen bonding residues Gly34, Ser35, Gly217, and Ser218. The active site was again modeled as a formic acid/formate anion moiety and a bound water molecule; the partial atomic point charges were taken directly from the AMBER united-atom force field (Weiner et al., 1984). This will be referred to as model 3. SCF and SDCI calculations (1R/22/21) for all four tautomers are tabulated in Table VI. Comparison of the CI results in Tables V and VI suggests that extending the electrostatic field to include the four nearest residues improves the relative stability of both bifurcated forms; in particular, the ranking order has also changed, and IIa has now become more stable than IIb, yielding a new ranking order: Ia > Ib > IIa > IIb. Combining the results for models 2 and 3 suggests that the local enzyme environment preferentially stabilizes the bifurcated forms as a direct result of charge transfer, electrostatic, and polarization effects. Furthermore, Table VI indicates that the difference in correlation energy between structures Ib and IIa is 5.0 kcal mol^{-1}. If we were to correct the values in Table III (model 2) to include the difference in correlation energies, structure IIa would be predicted to be more stable than Ib, and a revised ranking order of Ia > IIa > Ib > IIb would be predicted for the relative stabilities. Overall, our calculations predict that, contrary to previous theoretical studies, the water molecule binds by forming a bifurcated hydrogen bond to the inner oxygens of D32 and D215. Asp32 is preferentially ionized and can be assigned to pK_a1.

(*ii*) *Pepsin.* Pepsin (Sielecki et al., 1990) is one of the few aspartic proteinase structures which displays a significant deviation from planarity at the active site. The active-site carboxyl groups are inclined at an angle of 35°. This is a high-resolution structure (1.8 Å), so it is not obvious why the active site is significantly nonplanar. It may possibly be due to small errors in the torsion angles of the aspartate side chains; comparison with a high-resolution pepsin structure determined at Birkbeck College (Cooper et al., 1990) reveals that a good fit to the electron density can be obtained with coplanar aspartate groups. Nevertheless, we felt calculations on an aspartic proteinase with a nonplanar active site would be potentially very interesting. The nonplanarity of the active site might well disrupt the hydrogen bonds formed by the bifurcated structures, and torsional changes in the active site might well have mechanistic implications.

We have restricted our calculations to models 1 and 3. Proceeding as before, we used model 1 to optimize the geometries of the active site. The final geometries are shown in Figure 4, and the calculated SCF and CI (1R/22/21) energies are tabulated in Table VII. Comparing these results with endothiapepsin, we find that there are two noticeable differences. The relative stability of Ia, at the SCF level, is 0.6 kcal mol^{-1} more stable for pepsin; this is undoubtedly due to the OD1···OD1 distance being 0.09 Å shorter in the crystal structure. Comparison of the optimized structures for endothiapepsin and pepsin (Figures 3 and 4) indicates that pepsin forms stronger hydrogen bonds with the water molecule across the $OD1_{32}$···$OD1_{215}$ diagonal; the hydrogen bonds are both slightly shorter and the angles slightly better. The differences in relative stability are reversed by CI. Presumably, the nonplanarity of the active site slightly weakens the hydrogen bonds. The difference in energy between the two bifurcated forms is substantially higher in pepsin. This is largely due to the nonplanarity of the active site. The water molecule in IIa forms four hydrogen bonds between the carboxyl oxygens; however, all of these bonds would be expected to be significantly weaker than the two diagonal hydrogen bonds formed by Ia. The bonds are all longer, and the angles are smaller. These results emphasize the importance of forming strong hydrogen bonds diagonally across the active site.

We then used model 3 to investigate the effect of the local electrostatic field on the active site. The SCF and CI results are tabulated in Table VII. Again Ia is predicted to be the most stable form, the relative stability being Ia > Ib > IIb > IIa.

Structure IIa is again predicted to be substantially less stable, and it appears that relatively large torsional changes by the carboxyl groups have a small effect when Asp32 is ionized, but dramatically destabilize bifurcated structures with Asp215 ionized. Such an effect may well have mechanistic implications.

DISCUSSION

The main conclusion from our calculations is that the water binds at the active site by forming a bifurcated hydrogen bond to the inner oxygens of the active site. A third hydrogen bond is formed between the water oxygen and the acidic proton of the neutral aspartate. The two strongest hydrogen bonds are formed between the inner oxygen of D32 to the innermost proton on the water molecule and between the water oxygen to the acidic proton of the neutral aspartate. These two hydrogen bonds orientate the water molecule diagonally across the active site. In both pepsin and endothiapepsin, D32 is predicted to be preferentially ionized. Our results suggest that the ionization state of the active site can be readily predicted by the following simple rule: the aspartate which has the inner oxygen on the shortest O_{32}···O_{215} diagonal is ionized at pK_a1.

The results of our 4-31G SCF calculations using model 2 for the active site predict that Ia is the most stable tautomer by 4.20 kcal mol^{-1}. However, these calculations neglect the effect of electron correlation. A comparison between the SCF and CI energies for models 1 and 3 (Tables I, II, and VI) indicates that the difference in correlation energies for structures Ia and Ib in endothiapepsin can be as high as 6 kcal mol^{-1} (4-31G* basis, model 1). If we correct the 4-31G SCF energies in model 2 by this amount, Ia would be expected to be at least 10 kcal mol^{-1} more stable than the most stable nonbifurcated form. Furthermore, expanding our models to include more of the active site would be expected to increase the predicted stability of Ia as a result of the surrounding electrostatic field.

While our calculations predict that D32 is ionized in the native enzyme, this need not be the case for the enzyme–substrate complex. In the case of endothiapepsin, IIa (D215 ionized) is predicted to be 6 kcal mol^{-1} less stable. Therefore, it is conceivable that substrate binding may well provide the energy necessary to reverse the ionization of the active site. The calculations on pepsin, on the other hand, predict IIa to

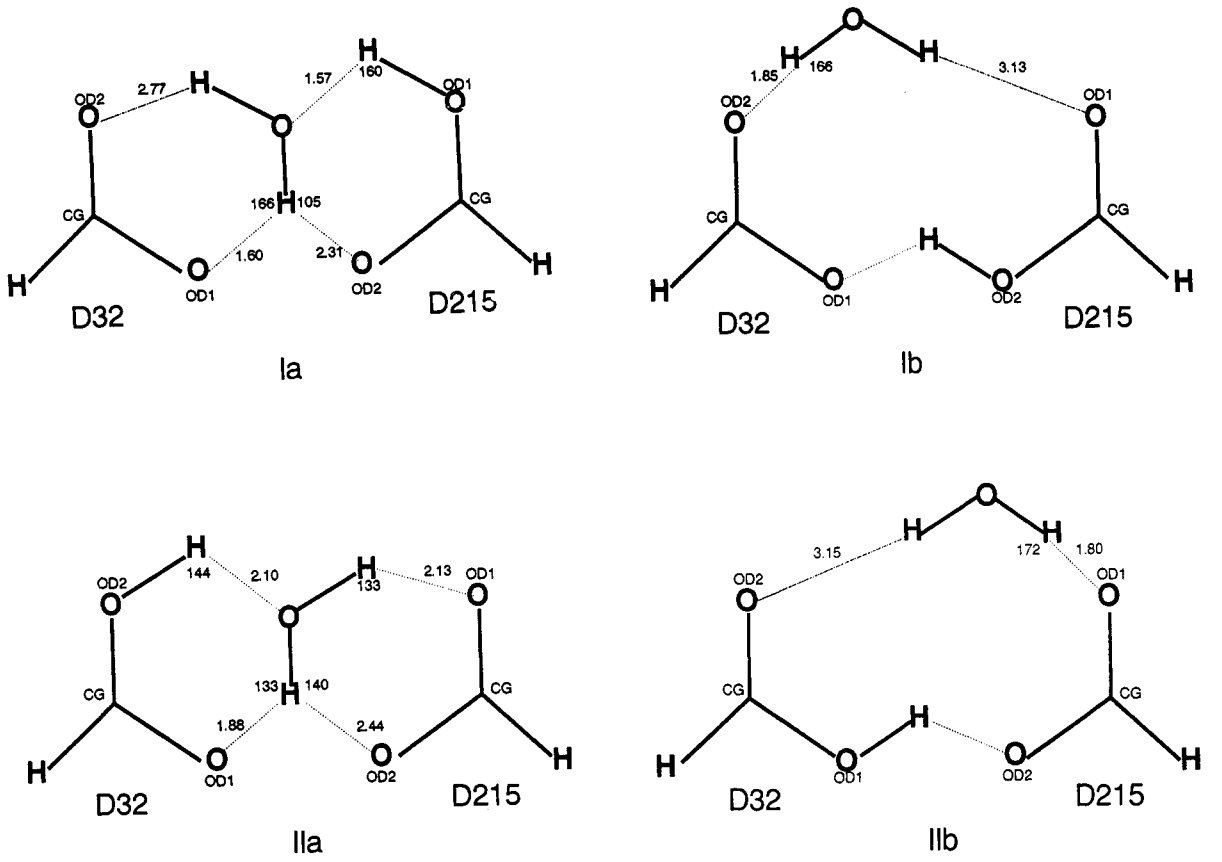

FIGURE 4: Four optimized geometries (4-31g basis set) for the active site of pepsin.

Table VII: Calculated SCF and CI Energies (au) for Pepsin (Models 1 and 3), Relative Stabilities (kcal mol^{-1}) in Parentheses

model	structure	SCF (4-31G)	CI (1R/22/21)
1	Ia	−452.343093 (6.23)	−452.656566 (4.36)
	Ib	−452.353025 (0.00)	−452.663515 (0.00)
	IIa	−452.320704 (20.28)	−452.641997 (16.26)
	IIb	−452.343629 (5.90)	−452.652167 (7.12)
3	Ia	−453.505302 (0.00)	−453.816735 (0.00)
	Ib	−453.503592 (1.07)	−453.812047 (2.94)
	IIa	−453.479567 (16.15)	−453.795158 (13.54)
	IIb	−453.492029 (8.33)	−453.798902 (11.19)

Table VIII: Distances (Å) between the Theoretically Determined Position of the Active-Site Water Molecule and the Crystal Structures

	endothiapepsin	pepsin
Ia	0.43	0.43
IIa	0.32	0.54
Ib	1.52	1.36
IIb	1.43	1.61

be significantly less stable (model 3, 13.5 kcal mol^{-1}), but this is undoubtedly due to the nonplanarity of the active site, which as discussed above may possibly be due to small experimental errors in the torsion angles of the asparate side chains.

We have also compared the optimized geometries for each structure (Ia, Ib, IIa, and IIb) with the crystal structures (Table VIII). In both pepsin and endothiapepsin, we find that the bifurcated structures predict the bound water molecule to be much nearer to the position found in the crystal structures. The average distance between the theoretically determined positions for the water molecule and the crystal structures is 0.43 Å for bifurcated structures and 1.48 Å for the nonbifurcated forms. If one considers the errors in both the crystal structures and the theoretical calculations, the value of 0.43 Å is reasonably acceptable. This, at least, provides some circumstantial evidence to support our theoretical results.

Covalent inhibition studies involving EPNP [1,2-epoxy-3-(*p*-nitrophenoxy)propane] and DAN (diazo-DL-norleucine methyl ester) also suggest that D32 is preferentially ionized at pK_a1. EPNP covalently inhibits pepsin by reacting predominantly with D32 (Chen & Tang, 1972; Hartsuck & Tang, 1972). The ring opening of 1,2-epoxides, such as EPNP, in acid solution favors products in which the attacking nucleophile reacts at the 2 position of the epoxide ring (Parker & Isaacs, 1959). Therefore, one would expect EPNP to preferentially react with the most nucleophilic carboxyl group (D32). DAN also inactivates pepsin, but reacts exclusively with D215 (Hofmann, 1974). In one of the earliest studies of the covalent inhibition of pepsin, it was demonstrated that DAN, in the presence of Cu(II), forms a positively charged organo-copper complex (Lundblad & Stein, 1969). The resulting copper complex can then bind electrostatically to the negatively charged aspartate (D32), before reacting with the protonated carboxyl group of D215. Our theoretical calculations provide a simple explanation for the observed preferential ionization of D32 at pK_a1.

Our active-site model also has important implications for substrate binding. Hadzi et al. (1988, 1987) extended their previous study and investigated possible docking modes for the substrate-bound enzyme. They modeled the active site as a formic acid/formate anion moiety and a nonbifurcated water molecule, and used *N*-methylacetamide as a model substrate. The electrostatic potential of their active-site model was strongly negative almost everywhere, and they found that only by displacing water could they bind the model peptide to the active site. Furthermore, the predicted substrate-bound structures had the peptide bond unfavorably oriented with respect to potential reaction mechanisms. The predicted

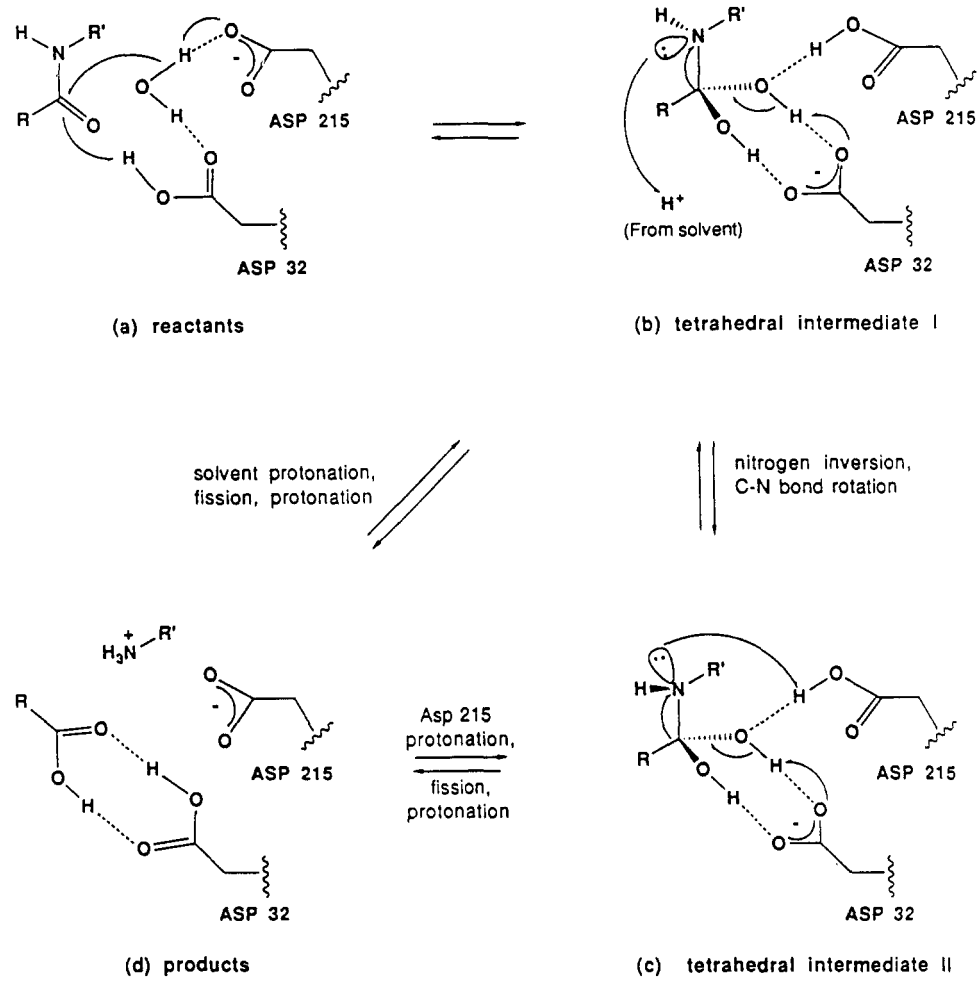

FIGURE 5: Catalytic mechanism proposed by Veerapandian et al. The mechanism is based on crystallographic analysis of a hydrated 2,2-difluorostatone inhibitor complexed with endothiapepsin. The *gem*-diol [–$C(OH)_2$–CF_2–] moiety of the inhibitor is thought to mimic the tetrahedral intermediate [–$C(OH)_2$–NH–] formed by the substrate during hydrolysis (b). Diagram reproduced with the permission of Dr. J. B. Cooper (Birkbeck College).

structures also bore no similarity to any of the inhibitor–enzyme complexes determined by crystallography (Sali et al., 1989; Cooper et al., 1987, 1989).

We feel that our active-site model would overcome these difficulties. First, the acidic proton of the neutral aspartate and the outermost proton on the water molecule both carry substantial positive charges (0.54 e and 0.42 e, respectively, [Mulliken population analysis (Mulliken, 1955)], while both outer oxygens are substantially negative (both >–0.75 e); this should produce an electrostatic potential complementary to a distorted peptide unit, thereby facilitating substrate binding. Second, substrate binding would not displace the water molecule which would remain trapped between the bound substrate and the active-site carboxyl groups. In both of our bifurcated models (Ia and IIa), the proton on the neutral aspartate and the outermost proton of the water molecule are easily accessible to an approaching substrate. Both the carbonyl oxygen and the lone pair from the nitrogen of the distorted peptide bond of the substrate can interact with these protons, enabling the substrate to reorientate the water molecule prior to nucleophilic attack.

Several crystallographic studies support this hypothesis. Recently, two crystallographic studies involving difluorostatone inhibitors have been reported (Veerapandian et al., 1992; James et al., 1992). In both of these studies, the scissile bond surrogate, an electrophilic ketone, is hydrated in the complex and forms a *gem*-diol. It has been argued that the resulting *gem*-diol, which has two hydroxyl groups on the scissile carbon, mimics the tetrahedral intermediate formed during hydrolysis (Figure 5b). The crystal structures suggest that both hydroxyl groups of the *gem*-diol form hydrogen bonds to the negatively charged carboxyl group of D32; D215 is protonated on the outer oxygen and forms a hydrogen bond to the hydroxyl group which has replaced the water molecule in the native enzyme. On the basis of these structures, both groups postulate very similar mechanisms. The substrate initially binds to an active site which has a bifurcated water molecule; D32 is protonated on the outer oxygen, and D215 is ionized. The water molecule then attacks the carbon atom of the scissile bond, producing a tetrahedral intermediate similar to that observed in inhibitor–enzyme complexes. During nucleophilic attack, the ionization of the active site is reversed, and D215 becomes protonated on the outer oxygen. This proton is then transferred to the scissile nitrogen, initiating the breakdown of the bound complex to form products (Figure 5).

In an earlier study, Suguna and co-workers examined the crystal structures of three native aspartic proteinases and postulated that the active site has a bifurcated water molecule between the carboxyl groups of D32 and D215 (D215 ionized). A model for the enzyme–substrate complex (constructed from the crystal structure of a reduced peptide inhibitor complexed with rhizopuspepsin) demonstrated that the substrate pushes the bound (bifurcated) water molecule out of the plane of the carboxyl groups. The water molecule remains hydrogen-

bonded to the active site, but has been reorientated to engage in nucleophilic attack on the carbon atom of the scissile bond (Suguna et al., 1987). On the basis of crystallographic evidence, and their model substrate–enzyme complex, Suguna proposed a mechanism which is very similar to the mechanisms proposed by Veerapandian and James.

One of the common features of all of these mechanisms is that the active site initially has D215 ionized and D32 protonated on the outer oxygen. This corresponds to model IIa. Our calculations on native endothiapepsin predict IIa (D215 ionized) to be less stable than Ia (D32 ionized) by about 6 kcal mol^{-1}. Therefore, to be consistent with the crystallagrophic studies described above, a reversal of ionization must occur during substrate binding. We anticipate that such a process should be relatively facile. Consider model Ia (Figure 3). If the acidic proton on D215 migrates onto the oxygen of the water molecule, while the outer proton of the water molecule simultaneously migrates onto D32, Ia can be readily transformed into IIa. A recent theoretical study of an intramolecular single-proton transfer in the hydrogen oxalate anion demonstrates that for relatively small distances (<1 Å) the barrier for proton transfer is small (<1 kcal mol^{-1}) (Truong & McCammon, 1991). For a concerted double-proton transfer, we anticipate that the barrier would still be relatively small. Furthermore, as described above, the acidic proton on the neutral aspartate and the outer proton of the water molecule are readily accessible to the substrate. It is therefore conceivable that the approaching substrate could assist proton transfer by orbital participation from the carbonyl oxygen and/or nitrogen atoms of the peptide bond.

It should be emphasized that our calculations have been performed only on the native enzyme and therefore do not include the effects of substrate. We are currently investigating the potential effects of substrate binding, and this will be discussed in a future publication. However, our active-site model for the native enzyme offers a more promising description for the initial stage of substrate binding than previous theoretical models (Goldblum, 1988; Hadzi et al., 1987, 1988). We have demonstrated that a bifurcated water molecule can form three or four hydrogen bonds to the active site; this should ensure that in the absence of a substrate (or inhibitor) the water molecule is tightly bound and cannot easily be displaced. As the modeling of Suguna demonstrates, the incoming substrate destabilizes the water molecule by pushing it out of the plane of the active site, and effectively primes the water molecule for nucleophilic attack. Furthermore, the nonbifurcated water molecules in models Ib and IIb form only one hydrogen bond to the active site and could be easily displaced by a substrate, and as the calculations of Hadzi (Hadzi et al., 1988) indicate, active-site models with a nonbifurcated water molecule do not appear to lead to productive binding.

Finally the results of our calculations may also have important implications for the simulation of enzyme–inhibitor complexes. A recent theoretical study (Rao & Singh, 1991) reported the results of a molecular dynamics simulation of a rhizopus pepsin–pepstatin complex. Pepstatin (Iva-Val-Val-Sta-Ala-Sta), a naturally occurring inhibitor of the aspartic proteinases, contains the unusual amino acid statine ([*S*-(*R**,*R**)]-4-amino-3-hydroxy-6-methylheptanoic acid). The 3(*S*)-hydroxyl group of the central statine residue, which displaces the active-site water molecule when pepstatin binds, is structurally related to a hydroxyl group in the tetrahedral intermediate formed during amide hydrolysis, and, therefore, it has generally been assumed that pepstatin is a good transition-state analogue (Merciniszyn, 1976). In all of the simulations reported by Rao and Singh, one of the inner oxygens of the active site was always protonated. The hydroxyl group of the central statine residue was then allowed to form hydrogen bonds between the outer oxygen of the ionized aspartate and the protonated inner oxygen of the active site. However, these simulations produced some serious inconsistencies with the crystal structure. First, they predicted the distance between the active-site carboxyl groups (D32 and D215) to be 0.5 Å less than the crystallographically determined distance, and second, geometry optimization severely distorted the active site, in some cases resulting in the two carboxyl groups becoming perpendicular to each other.

We have investigated these effects by allowing the CG_{32}–CG_{215} distances in our active-site model (model 1) to vary during the geometry optimizations (Beveridge, unpublished results). In the case of the bifurcated structures (Ia and IIa), the CG–CG distance changed only slightly (by 0.03 Å), but a shortening of 0.4 Å was observed for the nonbifurcated structures (Ib and IIb). These results suggest that inhibitors (like pepstatin) which contain a hydroxyl group may possibly bind by mimicking the water molecule which they displace. Therefore, we anticipate that the simulations reported by Rao and Singh could possibly be improved by allowing the hydroxyl group of pepstatin to form a bifurcated hydrogen bond to the inner oxygens of the active site. The hydroxyl group would then mimic the water molecule in structure Ia, and the mutual repulsion between the inner oxygens would ensure that the distance between the active-site carboxyl groups does not become too short.

ACKNOWLEDGMENT

We thank Professor J. Hermon-Taylor, Dr. J. B. Cooper, C. Dealwis, Dr. T. C. Jenkins, and Ned Slider for useful discussions.

REFERENCES

Burton, J., Cody, R. J., Herd, J. A., & Haber, E. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5476.

Chen, K. C. S., & Tang, J. (1972) *J. Biol. Chem. 247*, 2566.

Cooper, J. B., Foundling, S., Hemmings, A. M., Blundell, T. L., Jones, D. M., Hallett, A., & Selke, M. (1987) *Eur. J. Biochem. 169*, 215.

Cooper, J. B., Foundling, S. I., & Blundell, T. L. (1989) *Biochemistry 28*, 8596.

Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J., & Blundell, T. L. (1990) *J. Mol. Biol. 214*, 199.

Ditchfield, R., Hehre, W. J., & Pople, J. A. (1971) *J. Chem. Phys. 54*, 721.

Foltmann, B. (1966) *C. R. Trav. Lab. Carlsberg 35*, 143.

Fruton, J. S. (1976) *Adv. Enzymol. Relat. Areas Mol. Biol. 44*, 1.

Fruton, J. S. (1987) in *Hydrolytic Enzymes* (Neuberger, A., & Brocklehurst, K., Eds.) pp 1–37, Elsevier Science Publishers, New York.

Goldblum, A. (1987) *J. Comput. Chem. 8*, 835.

Goldblum, A. (1988) *Biochemistry 27*, 1653.

Guest, M. F. (1990) *GAMESS User Manual*, SERC, Daresbury, U.K.

Hadzi, D., Hodoscek, M., Harb, V., & Turk, D. (1987) *J. Mol. Struct. 150*, 241.

Hadzi, D., Hodoscek, M., Turk, D., & Harb, V. (1988) *J. Mol. Struct. 181*, 71.

Hartsuck, J. A., & Tang, J. (1972) *J. Biol. Chem. 247*, 2575.

Hehre, W. J., & Pople, J. A. (1972) *J. Chem. Phys. 56*, 4233.
Hehre, W. J., Stewart, R. F., & Pople, J. A. (1969) *J. Chem. Phys. 51*, 2657.
Hehre, W. J., Ditchfield, R., Stewart, R. F., & Pople, J. A. (1970) *J. Chem. Phys. 52*, 2769.
Hofmann, T. (1974) *Adv. Chem. Ser. 136*, 146.
James, M. N. G., Sielecki, A. R., Hayakawa, K., & Gelb, M. H. (1992) *Biochemistry 31*, 3872.
Kramer, R. A., Schaber, A. M., Skalka, K., Gangoly, F., Wong-Stall, & Reddy, E. P. (1986) *Science 231*, 1580.
Langhoff, S. R., & Davidson, E. R. (1974) *Int. J. Quantum Chem. 8*, 61.
Lapatto, R., Blundell, T. L., Hemmings, A., Overington, J., Wilderspin, A., Wood, S. P., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Hawrylik, S. J., Lee, S. E., Scheld, K. G., & Hobart, P. M. (1989) *Nature 342*, 299.
Lundblad, R. L., & Stein, W. H. (1969) *J. Biol. Chem. 244*, 154.
Marciniszyn, J., Hartsuck, J. A., & Tang, J. (1976) *J. Biol. Chem. 251*, 7088.
Mulliken, R. S. (1955) *J. Chem. Phys. 23*, 1833.
Newman, M., Safro, M., Frazao, C., Khan, G., Zdanov, A., Tickle, I. J., Blundell, T. L., & Andreeva, N. (1990) *J. Mol. Biol. 221*, 1295.
Parker, R. E., & Isaacs, N. S. (1959) *Chem. Rev. 59*, 737.
Pearl, L., & Blundell, T. (1984) *FEBS Lett. 174*, 96.
Rao, B. G., & Singh, U. C. (1991) *J. Am. Chem. Soc. 113*, 6735.
Sali, A., Veerapandian, B., Cooper, J. B., Foundling, S. I., Hoover, D. J., & Blundell, T. L. (1989) *EMBO J. 8*, 2179.
Saunders, V. R., & Van Lenthe, J. H. (1983) *Mol. Phys. 48*, 923.
Scheiner, S., Szczesniak, M. M., & Bigham, M. D. (1983) *Int. J. Quantum Chem. 23*, 739.
Schwann, T. (1836) *Muller's Arch. Anat. Physiol. 36*, 31.
Suguna, K., Padlan, E. A., Clark, W. S., Carlson, W. D., & Davies, D. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7009.
Sielecki, A. R., Hayakawa, K., Fujinaga, M., Murphy, M. E. P., Fraser, M., Muir, A. K., Carilli, C. T., Lewicki, J. A., Baxter, J. D., & James, M. N. G. (1989) *Science 243*, 1346.
Sielecki, A. R., Federov, A. A., Booghoo, A., Andreeva, N. J., & James, M. N. (1990) *J. Mol. Biol. 214*, 143.
Singh, U. C., & Kollman, P. A. (1984) *J. Comput. Chem. 5*, 129.
Singh, U. C., & Kollman, P. A. (1986) QUEST, University of San Francisco, San Francisco, CA.
Truong, T. N., & McCammon, J. A. (1991) *J. Am. Chem. Soc. 113*, 7508.
Veerapandian, B., Cooper, J. C., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., & Hoover, D. J. (1992) *Protein Sci. 1*, 332.
Weiner, S. J., Kollman, P. A., Case, D. A., Chandra Singh, U., Ghio, C., Alagano, C., Profeta, S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765.