Biophysical Chemistry 9 (1979) 273–280

ON THE ROLE OF THE ACTIVE SITE HELIX IN PAPAIN, AN AB INITIO MOLECULAR ORBITAL STUDY

P.Th. VAN DUIJNEN, B.Th. THOLE and W.G.J. HOL
Department of Chemistry, The University, Nyenborgh 16, 9747 AG Groningen, The Netherlands

Received 20 June 1978
Revised manuscript received 10 November 1978

On the system methanethiol/imidazole/formaldehyde (modelling the active site of papain) we performed ab initio self-consistent-field molecular orbital calculations using a rather large basis of Gaussian-type functions. A point charge representation of the long central α-helix present in the enzyme, was added in order to establish the influence of the electric field of the helix (which amounts to 10^9 V m^{-1} in the active site region) on the equilibrium: RSH...Im ⇄ RS^-...ImH^+, which is an essential step in a recently proposed mechanism for the catalytic action of papain. Our results show that the helix stabilizes the ion-pair by 15 kcal $mole^{-1}$ more than the neutral form making the two configurations energetically equivalent and lowers the energy barrier in the reaction path by 8 kcal $mole^{-1}$, thus shifting the equilibrium considerably towards the ionic situation and increasing the rate of proton transfer by several orders of magnitude. We conclude that "active site" helices, present in many enzymes, play a pertinent role in enzyme catalysis.

1. Introduction

Many enzymes have their active sites situated near the N-terminal end of an α-helix. In a recent paper [1] it was suggested that the considerable dipole field [2] of such a helix plays an important role in binding charged coenzymes and substrates. Moreover, in some enzymes, it may also enhance the rate of a basic step in the catalytic process [1]. In support of the latter suggestion, we now present quantitative evidence, based on ab initio molecular orbital calculations on models for the active site of papain.

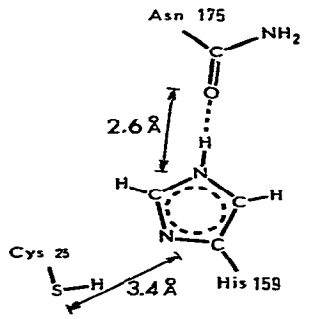

Fig. 1. Geometry of the active site of papain. The α-helix runs about perpendicular to the plane of the ring.

2. The mechanism of papain

Fig. 1 is a two dimensional representation of the active site of papain, a proteolytic enzyme, in which the essential thiol group of Cys25 is in close proximity of the imidazole ring of His159, which, in turn, is hydrogen bonded to the side chain of Asn175. In fig. 2 the mechanism as proposed by Drenth et al. [3] is schematical-

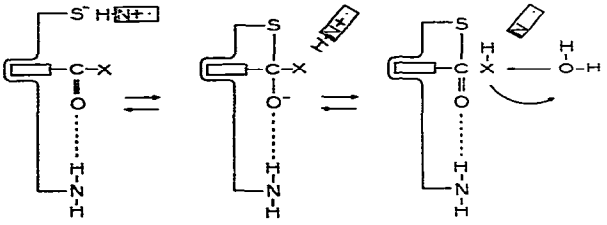

Fig. 2. Schematic representation of the mechanism proposed by Drenth et al. [3]. On the left the Michaelis–Menten complex, in the middle the tetrahedal intermediate, and on the right the acyl-enzyme with the protonated leaving group X.

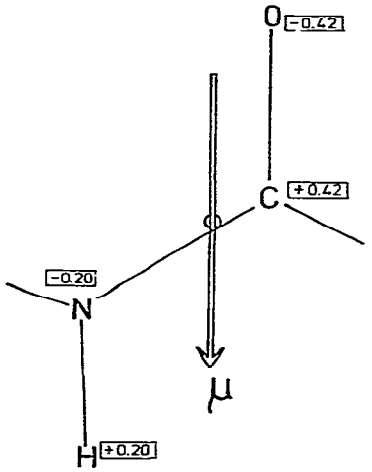

Fig. 3. Geometry and dipole moment of a single peptide unit. Numbers in boxes give the appropriate charges (in units of the elementary charge), taken from ref. [6], but slightly altered so as to preserve charge neutrality. $\mu = 3.46$ D $= 1.155 \times 10^{-29}$ Cm.

ly depicted. The features of this mechanism, relevant to the present discussion are:

i. In the free enzyme the active site contains the ion pair $S^- \ldots ImH^+$. This ion pair exists between pH = 4.2 and pH = 8.6, i.e. in the pH range of optimal activity.

ii. After formation of the Michaelis–Menten complex, the thiolate ion attacks the carbonyl carbon of the peptide bond to be broken, and a tetrahedal intermediate is formed.

iii. The imidazolium group of His159 rotates towards the leaving groupX, and protonates the latter, while the peptide bond is broken, and an acyl-enzyme is formed.

iv. In the deacylation step the leaving group is replaced by water or some other nucleophile, and the acyl-enzyme is hydrolyzed, probably by the same steps as mentioned above, but in reverse order.

Experimental evidence [4] for the first and third steps in this proposed mechanism is rather circumstantial, but ab initio MO calculations [5] on some model systems for the active site of papain support the proposed first step of the mechanism, albeit only qualitatively. Our extended calculations described here, include the effect of the "active site helix" in papain.

3. The α-helix in papain

Papain contains a long central α-helix comprising residues 24–43[‡]. The N-terminal end of this helix is situated close to the active site: the peptide bond of Cys25 is actually part of the helix backbone.

The electrical field produced by the helix backbone is due to the charge distribution of its peptide units. Fig. 3 shows a single peptide unit with formal charges [6] leading to a dipole moment of about 3.5 D, a value supported by experimental evidence [2] of various kinds. In an α-helix, these individual peptide dipoles are virtually parallel to the helix axis [2], giving rise to a considerable total dipole moment of the helix. These aligned dipoles have the effect of a line dipole with a dipole density of about 3.5 D per 1.5 Å, or 0.8×10^{-19} C. The electric field of such a line dipole is equal to that of a positive charge at the amino end, and a negative charge at the carboxyl and of the helix, each of magnitude 0.8×10^{-19} C (one half elementary charge). Due to the co-operative fashion in which the peptide units are ordered, the individual dipole moments may be increased by about 30 percent [2], resulting in a correspondingly stronger field. On the other hand, any field generated in the protein will be compensated to some extent by the reaction field induced in the surrounding solvent. By neglecting both the increase of the individual dipoles, and the reaction field, a reasonable estimate of the actual situation is probably obtained.

In order to investigate the role of the helix in the first step of the mechanism discussed above, we calculated the potentials on an essential line in the active site region, viz. the line connecting S^{γ} of Cys25 and $N^{\delta 1}$ of His159. Initially, we investigated the effect of making various choices as to which atoms contribute to the potential, using the charges of ref. [6], and atom positions from structural data on papain. The results are collected in fig. 4a, from which can be seen that indeed the helix backbone is the major cause of the field. Neither the inclusion of side chains, nor considering all the atoms of the enzyme does significantly change the slope of the potential curves. In these calculations an effective dielectric constant:

$$\epsilon = 1 + (R - R_A)/R,$$

[‡] Structual information on papain was kindly provided to us by Prof. J. Drenth.

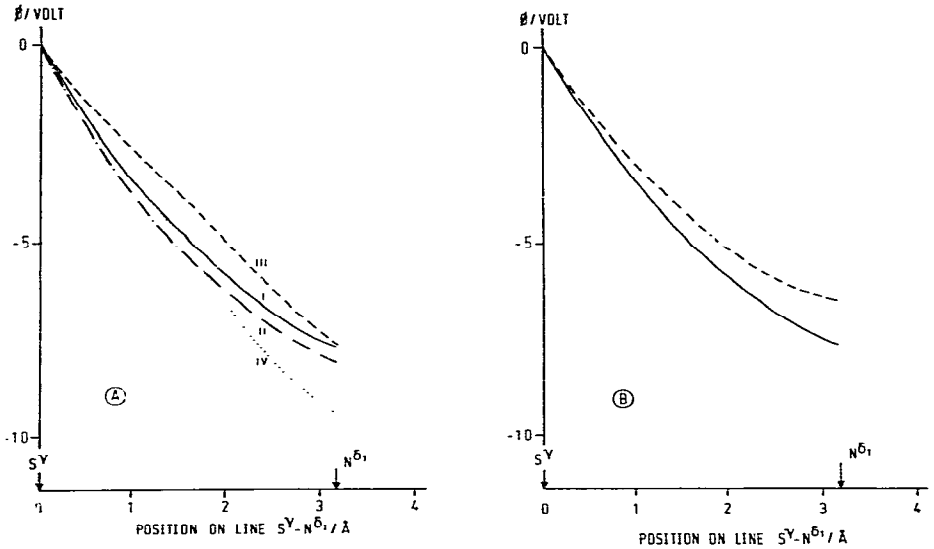

Fig. 4. Electrical potentials on the line $S^{\gamma}\ldots N^{\delta 1}$ in papain. The potential at S is taken as (arbitrary) zero point. Partial charges on the atoms in complete residues were taken from ref. [6]. For calculations using the backbone only, charges of fig. 3 were used. (A) Dielectric constant $\epsilon = 1.0 + (R - R_A)/R$, with R the distance between a point on the line $S^{\gamma}\ldots N^{\delta 1}$ and a contributing atom, R_A = 1.5 Å. For $R < R_A$, $\epsilon = 1.0$. i) backbone active site helix (residues 24–43), ii) backbone Cys25 + all other atoms of the helix, iii) backbone all residues, iv) backbone Cys25, His159, Asn175 + all other atoms. (B) —— as i) in (A), – – – – – as i) in (A), but with $\epsilon = 1.0$.

was applied, with R the distance between a contributing atom and the point in which the potential is evaluated, and R_A the radius of a sphere around the latter point, such that for $R - R_A < 0$ the dielectric constant is unity. The effect of using unit dielectric constant throughout, is displayed in fig. 4b. Here, curve "i" of fig. 4a, obtained with the effective dielectric constant, is compaired with that using $\epsilon = 1$. The latter curve is somewhat steeper, but the difference in slope is of the same order as the differences in fig. 4a. In all cases, the resulting electric field strength is about 10^9 Vm^{-1}.

When a proton is transfered from S^{γ} to $N^{\delta 1}$, negative charge will be left behind on Cys25, while the electronic charge distribution on His159 will show a polarization in the direction of $N^{\delta 1}$. The net result will be a change in dipole moment of the system. From model calculations [5,9] (see also next section), including all electrons of the relevant groups, we estimated this change in dipole moment to be of the order of 10D. The main component of this dipole moment increase lies, of course, along the line $S^{\gamma}\ldots N^{\delta 1}$. Moving positive charge towards His159 will cost energy. However, in a field of strength and direction as generated by the helix, the associated change in dipole moment will be stabilized by about 10 kcal $mole^{-1}$, relative to the situation without the helix. This is such a large number that no serious theoretical model for the active site of papain should exclude the helix.

4. Ab initio molecular orbital calculations

In our single determinant, ab initio molecular orbital calculations, the model compounds methanethiol, imidazole, and formaldehyde were used to represent Cys25, His159, and Asn175, respectively (fig. 5). Geometries of these model compounds were obtained, partly from experiment [7], and partly from earlier

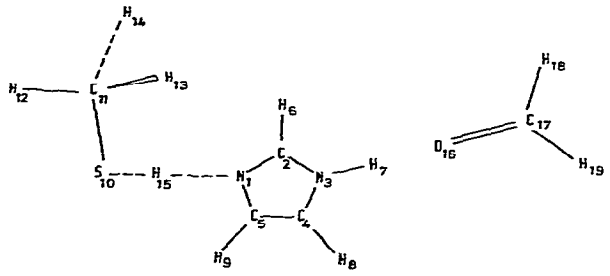

Fig. 5. Model system for the active site of papain. Only H_{13} and H_{14} are not in the plane of the ring. Imidazole geometry from ref. [7], methanethiol and formaldehyde geometries from ref. [5]. $S_{10}-N_1 = 3.35$ Å; $N_3-O_{16} = 2.90$ Å. The helix runs about perpendicular to the molecular plane, with the N-terminal end just below it. The coordinates of this model system as well as the positions of the point charges of the helix used in the calculations are given in table 1.

calculations [5]. Co-ordinates were transformed so as to give maximum coincidence with the structural data on papain, obtained from X-ray diffraction. The helix was represented by a number of point charges, taken from ref. [6], but changed into N = −0.204, H = +0.204, C = +0.422, and O = −0.422 in order to preserve charge neutrality of the backbone. These charges were given positions coinciding with those of the appropriate atoms in the enzyme. Only backbone atoms of residues 24–43 were considered, and unit dielectric constant was applied.

On these models we performed all electron, ab initio calculations, using a double zeta basis, consisting of 119 functions, contracted from 186 primitive gaussian orbitals [8]. Full technical details of these and related calculations will be given elsewhere [9]. Here, we only stipulate the necessity of using a rather large basis set. Minimal basis sets lack the diffuse functions required for an appropriate description of the electronic charge distribution in regions of the system where negative charge is accumulated. The errors introduced by using a minimal basis, are much larger for these negatively charged systems, than for neutral or positively charged systems, and therefore they tend to obscure the rather delicate energy balance associated with the type of process discussed here.

All calculations were performed on the University's CDC Cyber 74–18, with a program system newly developed [10] for this type of large scale calculations.

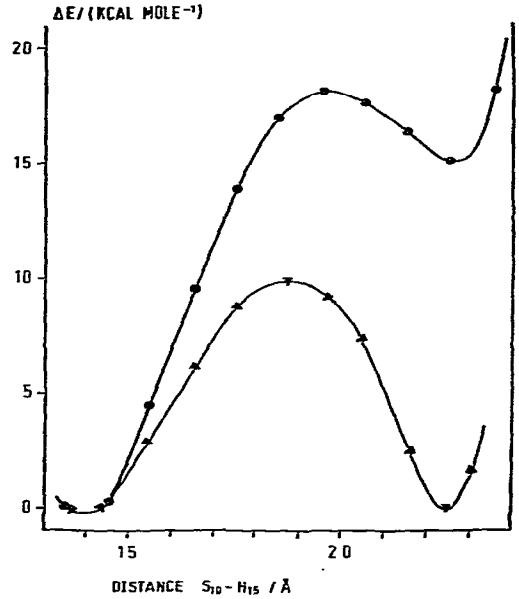

Fig. 6. Total energy of the system of fig. 5, obtained as expectation value of the molecular hamiltonian, for various positions of H_{15} on the line $S_{10}\ldots N_1$. Energy of the first point ($S_{10}-H_{15} = 1.35$ Å) is taken as (arbitrary) zero point. Top curve: no helix field. Bottom curve: helix field included.

5. Results and discussion

Fig. 6 gives the total energy of our model system (fig. 5) as a function of the position of H_{15} on the line $S_{10}\ldots N_1$ (i.e. S^{γ} of Cys25 and $N^{\delta 1}$ of His159), both with, and without the helix. The two curves clearly show that proton transfer, as part of the proposed mechanism for the catalytic process, is reasonable. The double minimum required for a (meta) stable position for H_{15} near N_1 is present. This, however, does not imply the existence of a stable ion pair $S^-\ldots ImH^+$ in free papain. Due to the neglect of surrounding solvent and of entropy effects, a MO calculation on symplifying models cannot establish the precise position of an equilibrium in solution. Nevertheless, the curves in fig. 6 indicate that the frequency with which an ion pair occurs in the active site is much larger than one would expect from the pK values of the free amino-acids. Also, this frequency is much larger than for a

Table 1
Coordinates of the point charges in the active site helix 24–43 and of the active site model methanediol/imidazole/formaldehyde. The charges on the backbone atoms were taken as in ref. [1]. In this approximation the C^{α} atoms have no charge. The coordinates of these atoms are therefore not included in the list. The atom names of the active site model are as in fig. 5

Positions of helix point charges

x	y	z	Residue number	Atom name
16.4	24.4	14.5	24	H
16.1	24.3	17.5		C
16.3	23.7	18.5		O
16.3	25.6	17.3		N
16.0	25.9	16.3	25	H
16.4	26.5	19.6		C
17.1	26.9	20.6		O
15.1	26.1	19.7		N
14.7	25.8	18.7	26	H
14.8	25.0	21.9		C
14.9	25.3	23.1		O
15.1	23.9	21.4		N
14.9	23.8	20.3	27	
17.0	22.8	22.6		C
17.4	22.3	23.7		O
17.7	23.6	21.8		N
17.2	24.0	20.9	28	H
19.3	24.9	23.2		C
20.1	24.7	24.1		O
18.5	25.9	23.0		N
17.8	25.9	22.1	29	H
18.1	26.5	25.4		C
18.8	26.8	26.4		O
17.2	25.5	25.4		N
16.8	25.4	24.4	30	H
17.8	24.0	27.3		C
18.1	24.2	28.6		O
18.4	23.1	26.6		N
18.0	23.1	25.5	31	H
20.6	23.0	27.8		C
21.2	22.5	28.7		O
20.8	24.2	27.3		N
20.1	24.5	26.5	32	H
21.3	25.7	29.2		C
22.0	25.9	30.2		O
20.0	26.0	29.2		N
19.5	25.7	28.3	33	H
19.3	25.6	31.5		C
20.0	25.8	32.5		O
18.6	24.5	31.3		N
18.1	24.4	30.4	34	H
20.0	23.1	32.7		C
20.4	23.1	33.9		O
20.9	22.8	31.7		N
20.4	22.9	30.8	35	H
22.9	23.4	33.0		C
23.4	22.9	34.0		O
22.9	24.7	32.7		N
22.4	24.9	31.7	36	H
23.1	25.5	35.1		C
24.0	25.4	35.9		O
21.8	25.5	35.3		N
21.2	25.8	34.4	37	H
21.7	24.0	37.3		C
22.1	23.9	38.5		O
21.7	23.0	36.4		N
21.3	23.2	35.4	38	H
23.6	21.5	37.4		C
23.9	20.7	38.2		O
24.4	22.5	36.9		N
23.9	23.1	36.2	39	H
26.0	23.6	38.4		C
27.0	23.6	39.1		O
25.0	24.5	38.6		N
24.2	23.4	37.8	40	H
24.9	25.1	41.1		C
25.4	25.9	42.0		O
24.3	24.0	41.3		N
24.0	23.5	40.3	41	H
24.9	22.1	42.9		C
25.7	22.1	43.8		O
24.7	21.2	42.0		N
24.0	21.4	41.2	42	H
27.0	19.9	41.9		C
27.7	19.2	42.5		O
27.4	20.9	41.0		N
26.5	21.4	30.5	43	H
29.2	20.1	39.6		C
30.4	19.6	39.6		O

Coordinates of active site model

N_1	17.13	30.96	16.24
C_2	18.46	30.96	16.25
N_3	18.95	32.33	16.24
C_4	17.88	33.07	16.24
C_5	16.77	32.29	16.24
H_6	19.08	30.06	16.24
H_7	19.94	32.48	16.24
H_8	18.08	34.11	16.24
H_9	15.72	32.48	16.24

Table 1 (ctd)

S_{10}	15.14	28.26	16.24
C_{11}	16.54	26.94	16.24
H_{12}	16.10	25.96	17.12
H_{13}	17.15	27.07	15.36
H_{14}	17.14	27.07	17.12
H_{15}	15.94	29.35	16.24
O_{16}	21.75	32.98	16.24
C_{17}	22.92	33.29	16.24
H_{18}	23.71	32.55	16.24
H_{19}	23.23	34.32	16.24

system containing the thiol/imidazole hydrogen bond, but lacking the helix.

Table 2 contains the results of a Mulliken population analysis [11], as well as the computed dipole moment and total energies for various positions of H_{15}. In examining the table, one should realize that a population analysis, being just one of the infinite number of ways to split up the total number of electrons, has only qualitative value. This is particularly true for a large basis, like the one used here, because the diffuse functions contained in such a basis, sometimes extend over several nuclei, thus making it impossible to assign the charge described by one such function uniquely to a single atom. Furthermore, computed dipole moments tend to be too large (in comparison with experment) because of the absence of polarization functions. The trends reflected in the table are beyond doubt, however.

As to be expected, the overall effect of proton transfer is accumulation of negative charge on Cys25, in particular on S^{γ}, rendering it more nucleophilic. Introduction of the helix has only minor effects on the

Table 2
Some results from ab initio LCAO-SCF-MO calculations on the model of fig. 5. Top entries: no helix. Bottom entries: helix included

Distance	Mulliken gross charges [a]					Total dipole moment (D)	Total energy (Hartrees) [b]
S_{10}–H_{15} (Å)	CH_3	S	H_{15}	Im	Form		
1.35	−0.06	−0.20	+0.21	+0.01	+0.05	11.4	−775.739
	−0.17	−0.16	+0.24	+0.03	+0.05	12.6	−774.783
1.45	−0.10	−0.23	+0.23	+0.04	+0.05	11.8	−775.738
	−0.17	−0.19	+0.26	+0.07	+0.05	13.1	−774.782
1.55	−0.10	−0.25	+0.24	0.08	+0.05	12.4	−775.732
	−0.18	−0.22	+0.27	+0.08	+0.05	13.8	−774.778
1.65	−0.10	−0.27	+0.25	+0.09	+0.05	13.1	−775.724
	−0.19	−0.25	+0.28	+0.11	+0.05	14.4	−774.773
1.75	−0.12	−0.31	+0.27	+0.12	+0.05	14.0	−775.717
	−0.22	−0.28	+0.30	+0.15	+0.05	15.6	−774.769
1.85	−0.15	−0.35	+0.29	+0.15	+0.05	15.1	−775.712
	−0.23	−0.32	+0.33	+0.17	+0.05	1.68	−774.767
1.95	−0.17	−0.38	+0.33	+0.19	+0.05	16.4	−775.710
	−0.25	−0.37	+0.36	+0.22	+0.04	18.2	−774.768
2.05	−0.19	−0.44	+0.36	+0.22	+0.04	17.8	−775.711
	−0.27	−0.41	+0.39	+0.26	+0.04	19.5	−774.772
2.15	−0.22	−0.49	+0.39	+0.26	+0.04	19.2	−775.713
	−0.30	−0.46	+0.41	+0.28	+0.04	20.9	−774.779
2.25	−0.24	−0.53	+0.40	+0.30	+0.04	20.2	−775.715
	−0.31	−0.49	+0.42	+0.33	+0.05	21.9	−774.782
2.35	−0.25	−0.56	+0.39	+0.36	+0.05	21.2	−775.710
	−0.33	−0.51	+0.41	+0.38	+0.05	22.7	−774.780

[a] Corrected for nuclear charges. [b] 1 Hartree (atomic unit of energy) = 27.2 eV = 627 kcal mole^{-1}.

charge distribution. (The relatively strong displacement of electrons towards CH_3 is an artifact, caused by using a point charge representation for the helix backbone. The sum of charges for CH_3 and S is altered much less by introducing the helix.) The large dipole moment increase of about 10 D may be thought of as resulting from two contributions. First, the proton is moved in a *static* distribution of electrons, which adds about 5 D to the dipole moment. Second, relaxation of the electronic charge distribution adds another 5 D. Rather than using the estimated overall dipole moment and approximate field strength for discussing the energy effects, we return to fig. 6, which contains total energies obtained as expectation values of the molecular hamiltonian.

The top curve, i.e. the energy effect of transferring H_{15} without the helix field, does not agree, in a quantative sense, with the proposed mechanism. If we look at the proton transfer as a dynamical process, taking place prior to, or concomitant with the attack of S, the net enthalpy change is about 15 kcal mole^{-1}, with a barrier of 18 kcal mole^{-1}, relative to the non-ionic situation. This would result in a rather endothermic process at a low rate. Introduction of the helix changes this picture drastically. The minimum near the imidazole ring is lowered by no less than 15 kcal mole^{-1}, thus making the two minima practically equivalent. The only obstacle against proton transfer is the barrier, which is lowered by about 8 kcal mole^{-1} to a value of about 10 kcal mole^{-1}, which in any case would increase the computed rate of proton transfer by a factor of 6×10^4, a number easily associated with rate increases by enzyme catalysis. Also, since the two stable proton positions are equivalent, proton tunneling should be considered as a possibility, depending on barrier height and width. In both cases the barrier height is probably overestimated by several kcal mole^{-1}, because of the neglect of electron correlation, typical for Hartree-Fock calculations.

The effect of the helix field, as given here, is probably a lower limit to the actual effect. First, the coincidence of our model system with the actual structural parameters is such, that the helix is somewhat farther removed from the active site than in practice. Second, in representing (for practical reasons) Asn175 by formaldehyde instead of, say, formamide (which has a larger polarizability) we restricted the change in dipole moment associated with the proton transfer. Moreover, we did not move H_7 (fig. 5), which again may have the effect of restricting the dipole moment increase. Third, the co-operative effect of the typical alignment of the individual peptide dipoles in the helix was neglected, which may have been compensated, however, by also neglecting the reaction field of the surrounding solvent, molecules of which may actually be present quite close to the active site.

In conclusion, we suggest that, among other things, the α-helices, frequently encountered near the active sites of enzymes, play a pertinent role in the catalytic process. In addition, the helix field is very likely also important in stabilizing the tetrahedal intermediate. In papain, for example, the oxygen of this intermediate forms a hydrogen bond with the NH-group of Cys25 [12]. Hence, the helix and its field should be included in the discussion of proteins and their properties.

In a number of enzymes the reaction mechanism appear to be quite similar to that of papain, although different groups may be involved. Even in cases where no active site helix is present, preliminary calculations indicate that, nevertheless, the protein backbone produces a potential which resembles those given in fig. 4. Studies on these systems will be presented in forthcoming papers.

Finally, we like to discuss a question put forward by the referee: Why ab initio calculations on systems of this size? In our opinion computational chemistry should not primarily be concerned with reproducing experimental numbers to the last decimal place. Many semi-empirical schemes have been developed to do just that. However, their successful application requires in many cases a thorough understanding of the chemistry of the system under investigation and a considerable experience in the use of the various semi-empirical methods available. In contrast to this, ab initio calculations are straightforward in the sense that after the choice of the basis set and the geometry, no further parameters can or need be chosen. The effects of the limitations like reducing basis sets, or not taking correlation into account, are well understood and can often be estimated rather accurately. Moreover, when applied in fields where the chemistry is less well understood, like in enzymatic catalysis, considerable insight might be obtained from the trends revealed by, even inaccurate, non-empirical calculations.

Acknowledgement

We thank Professor J. Drenth for many stimulating discussions. We are greatly indebted to the management of the University's Computing Centre for making available large amounts of computer time.

References

[1] W.G.J. Hol, P.Th. van Duijnen and H.J.C. Berendsen, Nature 273 (1978) 443.
[2] A. Wada, Adv. Biophys. 9 (1976) 1.
[3] J. Drenth, H.M. Swen, W. Hoogenstraaten and L.A.AE. Sluyterman, Proc. Koninkl. Akad. Wetensch. C78 (1975) 104.
[4] L. Polgar, FEBS Letters 38 (1974a) 187; 47 (1974) 15; G. Lowe, Phil. Trans. Roy. Soc. B257 (1970) 237.
[5] Ria Broer, P.Th. van Duijnen and W.C. Nieuwpoort, Chem. Phys. Letters 42 (1976) 525.
[6] D. Poland and H.A. Sheraga, Biochemistry 6 (1967) 3791.
[7] S. Martinez-Carera, Acta Cryst. 20 (1966) 783.
[8] B. Roos and P. Siegbahn, Theoret. Chim. Acta 17 (1970) 209.
[9] P.Th. van Duijnen, B.Th. Thole, Ria Broer and W.C. Nieuwpoort, to be published.
[10] P.Th. van Duijnen and B.Th. Thole, unpublished.
[11] R. Mulliken, J. Chem. Phys. 23 (1955) 1833.
[12] J. Drenth, K.H. Kalk and H.M. Swen, Biochemistry 15 (1976) 3731.