PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Mapping the Active Site of Papain with the Aid of Peptide Substrates and Inhibitors

A. Berger and I. Schechter

Phil. Trans. R. Soc. Lond. B 1970 **257**, 249-264
doi: 10.1098/rstb.1970.0024

Phil. Trans. Roy. Soc. Lond. B **257**, 249–264 (1970) [249]
Printed in Great Britain

Mapping the active site of papain with the aid of peptide substrates and inhibitors

By A. Berger and I. Schechter
Departments of Biophysics and Chemical Immunology,
The Weizmann Institute of Science, Rehovot, Israel

Introduction

The active site of an enzyme performs the twofold function of binding a substrate and catalysing a reaction. The efficiency of these actions determines the overall activity of the enzyme towards the particular substrate, i.e. determines the specificity of the enzyme. It is therefore possible to obtain information on the active site by the kinetics of the enzyme's reactions with different substrates and inhibitors.

An important feature of the active site is its size. It should be possible to 'measure' this by using substrates or inhibitors large enough to show up the interactions of the furthermost parts of the binding site. In the present series of investigations on proteolytic enzymes, our approach is to compare the activity of the enzyme towards (*a*) peptides of increasing length, (*b*) diastereoisomeric pairs of peptides in which a particular amino acid residue has been replaced by its antipode, and (*c*) pairs of substrates in which a particular side chain (say a methyl group) has been replaced by another (say an aromatic group). The influence of these changes on reaction rates as a function of distance from the point of cleavage indicates the extent of the active site (Schechter, Abramowitz & Berger 1965; Abramowitz, Schechter & Berger 1967).

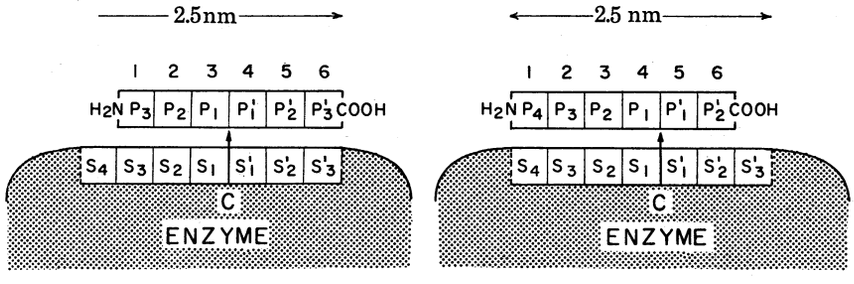


Figure 1. Schematic representation of two possible enzyme-substrate complexes of papain with a hexapeptide. The active site of the enzyme is composed of seven 'subsites' (S_1–S_4 and S'_1–S'_3) located on both sides of the catalytic site C. The positions, P, on the hexapeptide substrate, are counted from the point of cleavage and thus have the same numbering as the subsites they occupy. Complex A will yield as products two molecules of tripeptide, B one molecule of tetrapeptide and one of dipeptide.

Since the active site of proteolytic enzymes was found to be rather large (Schechter & Berger 1966*b*; Morihara & Oka 1968), capable of combining with a number of amino acid residues (seven in the case of papain, Schechter & Berger 1967), it is convenient to subdivide the binding site into subsites. A subsite is defined as the region on the enzyme surface which interacts with one amino acid residue of the substrate. Subsites are located on both sides of the catalytic site (see figure 1), being numbered S_1, S_2, etc. and S'_1, S'_2, etc. In the enzyme–substrate complex the substrate is lined up on the enzyme in such a way that the CO–NH group being hydrolysed

always occupies the same place (the catalytic site); the amino acid residues occupy adjacent subsites, those towards the amino end occupying subsites S_1, S_2, etc., those towards the carboxyl end occupying subsites S_1', S_2', etc.

Knowing the size of the active site one can proceed to map it out in terms of the particular interactions which the various subsites can exhibit (Abramowitz *et al.* 1967; Schechter & Berger 1968). These include binding forces (hydrogen bonds, electrostatic and nonbonding interactions) as well as geometrical factors such as stereospecificity and space limitations.

Changes in substrate composition may affect either the degree of enzyme saturation (K_m), or the rate of the catalytic step (k_{cat}), or both. When these parameters cannot be measured separately (e.g. K_m is too large) the overall efficiency of the enzyme action can be expressed in terms of the proteolytic coefficient C (Irving, Fruton & Bergmann 1941). This is equal to k_{cat}/K_m, the rate of reaction at infinite substrate dilution. In the mapping procedure competitive inhibitors can be even more useful than substrates. Unlike $\bar{K}_m$ ($= 1/K_m$) values which are not equilibrium but steady state constants, inhibition constants, $\bar{K}_i$ ($= 1/K_i$), are true equilibrium constants and thus directly related to the free energy of binding.

Materials and Methods

The synthesis of the alanine peptides used as substrates has already been described (Schechter & Berger 1966*a*). Other peptides were prepared by the hydroxysuccinimide method (Anderson, Zimmerman & Callahan 1964). Sequence polymers were prepared by polycondensation of the corresponding tri- and tetrapeptide hydroxysuccinimide esters in dimethylformamide or water and fractionated on Sephadex G-25 in 10 mmol l^{-1} of HCl (Schechter *et al.* 1967).

Papain, twice crystallized (Worthington), was used.

Reaction conditions for the hydrolysis of alanine peptides listed in figure 3 are given in figure 2. Hydrolysis rates were obtained as follows. The compositions of the reaction mixtures at different times were determined by quantitative paper electrophoresis (Schechter & Berger 1966*b*). Two examples are given in figure 2. When necessary, the reaction products were isolated and their stereochemical composition determined by further degradation with carboxypeptidase A and/or leucine aminopeptidase (Schechter & Berger 1966*b*). For example: the split in Ala_5 (DLL-LL) was shown to have occurred at the point indicated, since it gave Ala_2 (LL) only, and no Ala_2 (DL). Ala_5 (LL-L-LD) gave Ala_3 (LLL 34 %; LLD 66 %) and Ala_2 (LL 66 %; LD 34 %),

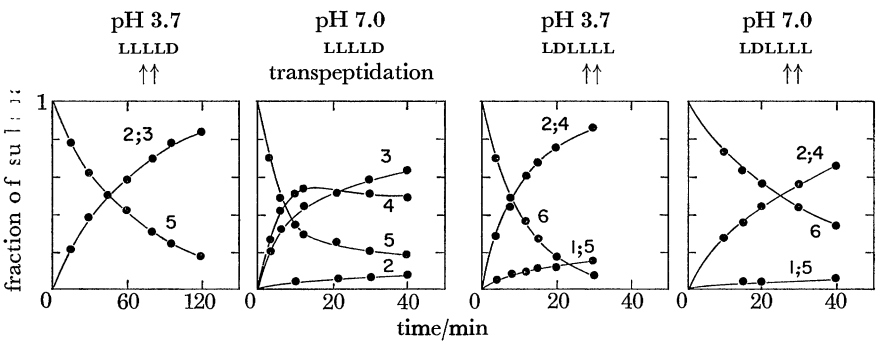


FIGURE 2. The course of hydrolysis of alanine peptides by papain at 36 °C. Arrows indicate bonds cleaved; numbers give the size of the peptide. Reaction mixture concentrations (mol l^{-1}): substrate 0.02; papain 1.6×10^{-5}; mercaptoethanol 0.01; EDTA 0.002; for pH 3.7 citric acid 0.025, Na_2HPO_4 0.016; for pH 7.0 citric acid 0.005, Na_2HPO_4 0.058.

showing splits at both bonds indicated, in a ratio of 2:1. In the case of Ala_5 (LLLLL) and Ala_6 (LLLLLL) it was necessary to synthesize a labelled substrate. Ala_5 (L*L-L-LL, ^{14}C) gave Ala_2 (L*L) and Ala_3 (L*LL) in a ratio of 1:2. Disappearance of substrate followed first order kinetics. From the data obtained, rate constants for the splitting of individual bonds could be assigned unequivocally, the sum of the individual constants in a substrate being equal to the rate of its disappearance.

In the determination of inhibition constants listed in figure 7, benzoylarginine ethyl ester served as substrate. Initial hydrolysis rates (2 to 4 % reaction) were obtained in a recording pH-stat in a medium containing KCl (0.33 mol l^{-1}), ethylenediamine tetra-acetic acid (2 mmol l^{-1}) and 2-mercaptoethanol (0.01 mol l^{-1}) at 36 °C. A $\bar{K}_m$ of 45 l mol^{-1} was determined for the substrate at pH 4.3. $\bar{K}_i$ values were determined from plots of v_0/v against $[I]$ (at constant substrate concentration $[S]$), where v_0 is the rate of the uninhibited hydrolysis and v the rate in the presence of the inhibitor at concentration $[I]$. The highest inhibitor concentration used in a run was chosen to give v_0/v in the range of 2 to 5. Plots were linear and inhibition constants were calculated from the expression $\bar{K}_i$ = slope $(1 + [S]\bar{K}_m)$. In a number of cases Lineweaver–Burk plots at constant $[I]$ were constructed and showed that the inhibition was competitive.

Determination of the size of the binding site of papain

In order to investigate the size of the active site of papain, its action on over 40 diastereoisomeric peptides of alanine, ranging in length from Ala_2 to Ala_{16} was investigated. Ala_2 (L_2) and Ala_3 (L_3) were hardly hydrolysed under the conditions used. At pH 3.7 Ala_4 (L_4), Ala_5 (L_5) and Ala_6 (L_6) were hydrolysed at all their peptide bonds except those adjacent to the *N*-terminal, at rates increasing with chain length as shown in figure 3. Peptides of L-alanine larger than Ala_6 are insoluble in water and cannot be investigated. The presence of D residues (which in general increases solubility markedly) always reduces the rate of hydrolysis. No hydrolysis occurred at bonds involving a D residue (L–D or D–L), and only trace hydrolysis of L–L bonds in D-L-L sequences. Thus the following peptides were resistant to hydrolysis: Ala_4 (LDLL, LLDL, and DLLD), Ala_5 (LLDLL), Ala_6 (LDLLDL and DLLDLL), Ala_9 (LDLLDLLDL and DLLDLLDLL), Ala_{12} (LDLLDLLDLL-DL), Ala_{15} (LDLLDLLDLLDLLDL). Further examples of the above stereospecificity rule can be found in figure 3 (e.g. Ala_5 (LDLLL) is split only at the C-terminal bond). This shows that subsites S_1, S_1' and S_2 show very strong stereospecificity.

Alanine peptides containing bonds susceptible to hydrolysis by papain are cleaved at rates which cover a wide range. Proteolytic coefficients, C (l mol^{-1} s^{-1}), for cleavages at pH 3.7 are given in figure 3 on a logarithmic scale. We interpret the observed variations in C on the basis of the assumed structure of the enzyme–substrate complex as described above: the CO–NH group being hydrolysed always occupies the same place (the catalytic site); the amino acid residues occupy adjacent subsites, those towards the amino end occupying subsites S_1, S_2, etc., those towards the carboxyl end occupying subsites S_1', S_2', etc. For convenience we number the 'positions' P of the residues in the peptide according to the subsites they occupy, and thus the numbering in a given peptide depends on which bond is split. For example, in the pair Ala_6 (LLLLLL) and Ala_6 (DLLLLL) the diastereoisomeric replacement of Ala-1 is said to be at P_3 when the bond between Ala-3 and Ala-4 is cleaved—but at P_4 when the bond between Ala-4 and Ala-5 is cleaved (see figure 1).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

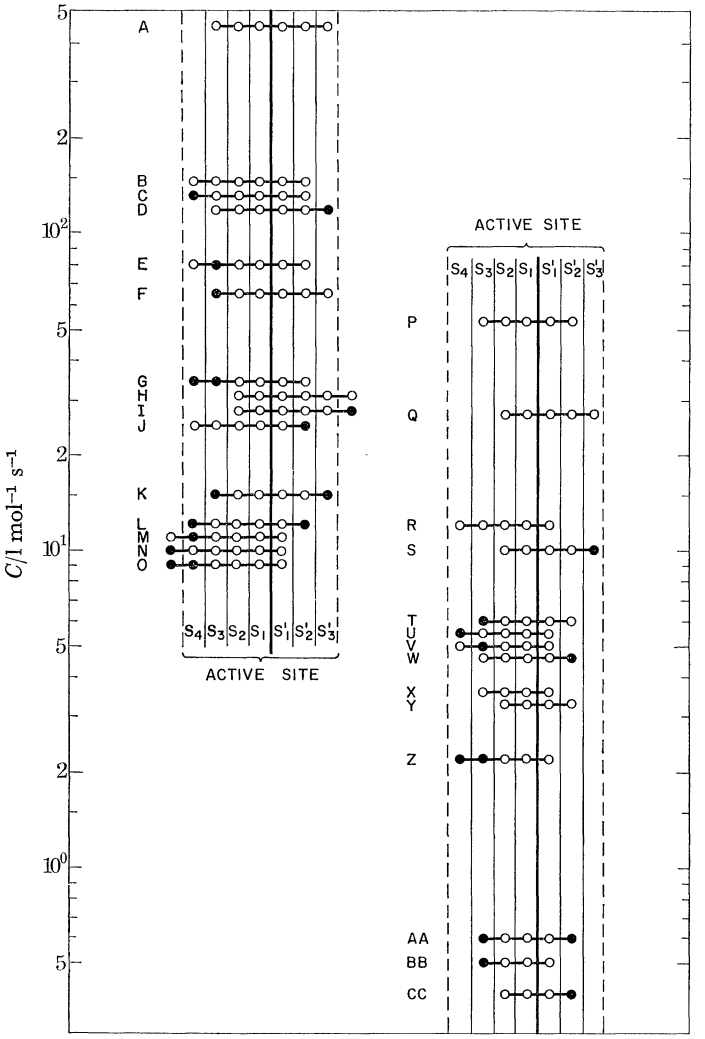


FIGURE 3. The proteolytic coefficients C (l mol^{-1} s^{-1}) for the hydrolysis of individual bonds in a series of alanine peptides at pH 3.7, 36 °C. The peptides are drawn schematically at the height corresponding to the C value for the bond crossing the heavy vertical line (○, L residues; ●, D residues; the carboxyl is to the right). The subsites (S_1, S_2 etc.) occupied by the various residues during the splitting of the bond are indicated.

Table 1 gives the information which can be extracted from figure 3 regarding the effect of diastereoisomeric replacements and of elongations by L or D residues. It is seen that the inhibiting effect of D residues is strongest in S_1 and S'_1, decreasing with distance from the catalytic point. Subsites on the amino terminal side seem to be more sensitive to replacement (at equal distance) than on the carboxyl side.

It is instructive to analyse for a given peptide the relative probabilities of the different possible modes of binding to the active site. This is illustrated for eight peptides in table 2. It is seen that in the absence of interference by D residues there is a pronounced tendency for optimal cleavage to occur in the middle of the peptide (see L_6). A D residue 'pushes' the point of attack away from itself. Thus (at pH 3.7) in L_6 the bond Ala-3–Ala-4 is the most susceptible, whereas in DL_5 it is the bond Ala-4–Ala-5. Similar changes in susceptibility can be seen in the pairs L_5–L_4D etc. A striking feature is the combined effect of a double replacement on opposite sides

TABLE 1. THE EFFECT OF CHANGES IN SUBSTRATE STRUCTURE ON RATES OF HYDROLYSIS

The numbers are the proteolytic quotients (Irving *et al.* 1941) obtained from the proteolytic coefficients in figure 3. The letters in parentheses correspond to the *C* values from which the ratio is calculated.

P_4 S_4	P_3 S_3	P_2 S_2	P_1 S_1	P'_1 S'_1	P'_2 S'_2	P'_3 S'_3
(*a*) diastereoisomeric replacement D→L						
3 (V-Z)	10 (P-T)	> 10^2	> 10^3	> 10^3	12 (C-L)	5 (F-K)
2 (E-G)	9 (D-K)				10 (P-W)	3 (Q-S)
2 (R-U)	8 (W-AA)				9 (T-AA)	3 (A-D)
2 (J-L)	7 (X-BB)				9 (Y-CC)	
1 (B-C)	7 (A-F)				6 (B-J)	
	4 (C-G)					
	3 (U-Z)					
	2 (R-V)					
	2 (B-E)					
(*b*) elongation by L residue						
15 (E-T)	16 (A-Q)	> 50	—	—	23 (C-U)	12 (F-T)
11 (V-BB)	15 (P-Y)				15 (E-V)	9 (A-P)
4 (R-X)	13 (D-S)				15 (G-Z)	8 (Q-Y)
3 (B-P)	12 (W-CC)				15 (P-X)	
					12 (B-R)	
					11 (T-BB)	
(*c*) elongation by D residue						
6 (G-T)	2 (F-Q)	—	—	—	2 (L-U)	3 (K-T)
4 (Z-BB)	2 (T-Y)				2 (J-K)	3 (S-Y)
2 (C-P)	2 (AA-CC)				1 (W-X)	3 (D-P)
2 (L-W)	1 (K-S)				1 (AA-BB)	
2 (U-X)						

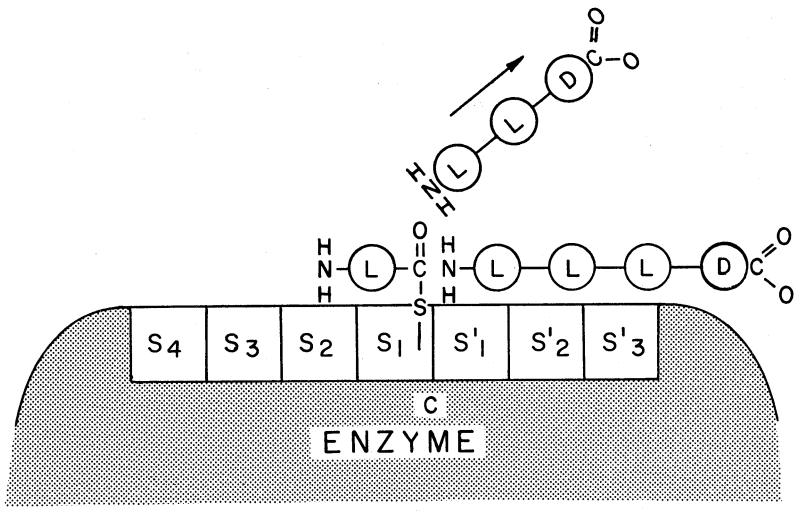


FIGURE 4. Schematic representation of the stereospecific deacylation of the acyl-papain intermediate, Pap–S–$COCHCH_3NH_2$(L) by the substrate Ala_4(LLD).

of the susceptible bond, as in Ala_5(DLLLD) → Ala_5(LLLLL), resulting in this case in a quotient of 100 (P/AA, figure 3); this is the product of the values for the two single replacements (P/T = 10 and P/W = 10).

TABLE 2. COMPARISON OF THE SUSCEPTIBILITY TO HYDROLYSIS BY PAPAIN OF THE VARIOUS BONDS IN AN ALANINE PEPTIDE

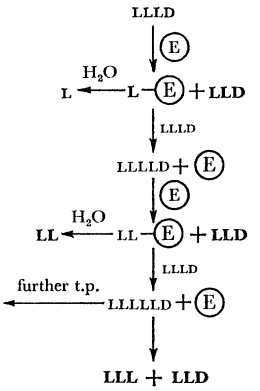
○, L residues, ●, D residues; the carboxyl is to the right. The dotted line indicates the bond cleaved, the figures giving the corresponding value for C/l mol^{-1} s^{-1}. t.p.: transpeptidation.

peptide	pH 3.7	pH 7	peptide	pH 3.7	pH 7
	10	t.p.		10	t.p.
	130			50	
	440			30	
	30		L_5	0	
L_6	0				
	10	2		5	2
	130	70		5	6
	60	80		0	
	0		DL_4	0	
DL_5	0				
	0	t.p.		0	t.p.
	25			5	
	120			10	
	30		L_4D	0	
L_5D	0				
	0			0	
	11	4		0.6	0
	14	20		0	
	0		DL_3D	0	
DL_4D	0				
	10	1.5		5	0.8
	80	25		0	
	0	1		0	
	0	0	LDL_3	0	
LDL_4	0	0			

In polypeptides, too, the decrease in hydrolysis rate due to the presence of D residues follows the pattern of table 1. We found that the DLLD sequences in the ordered copolymer $(LDL)_8$ are not digested, the DLLLD sequence in $(LDLL)_8$ are digested slowly, and the DLLLLD sequence in $(LLDLL)_2$ quite rapidly. These data are consistent with the earlier observation (Schechter & Sela 1962) that poly-DL-alanine is digested to an extent of about 10 %. About half of the bonds are split quite rapidly the other half much more slowly. This pattern of hydrolysis is to be expected if D and L residues are randomly distributed in the polymer and only runs of three or more L residues are susceptible to attack by papain.

At pH 7 extensive transpeptidation occurred with peptides which did not contain a D residue in the N-terminal or next to it. Substrates such as $Ala_6(DL_5)$ or $Ala_6(LDL_4)$ showed no transpeptidation (see figure 2). This can be explained by assuming that the substrate, acting as the acceptor, must occupy the strongly stereospecific subsites S_1' and S_2' with its two N-terminal residues during the deacylation step, and will therefore fit only if both are of the L-configuration.

A schematic representation of this process is given in figure 4. A possible mechanism for the case of Ala_4 (L_3D) is the following:



This scheme accounts for the products observed—mainly Ala_3, much less Ala and Ala_2 (figure 5). When substrate concentration is lowered, the specific rate of substrate disappearance decreases and transpeptidation is reduced, i.e. less trimer and relatively more monomer and dimer are formed. This demonstrates that the transpeptidation path of the reaction is of an order higher than one—in accordance with the proposed scheme.

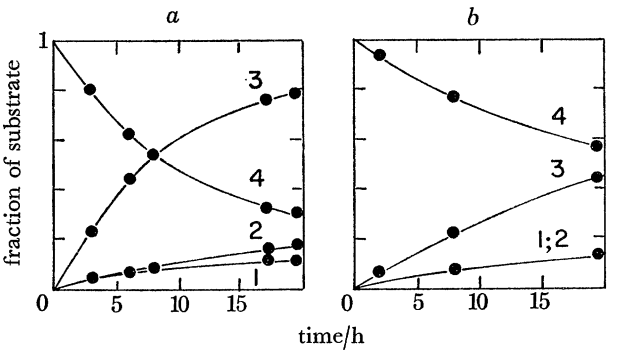


FIGURE 5. The course of degradation of Ala_4(L_3D) by papain at pH 7.0 at two substrate concentrations (*a*) 20 mmol l^{-1} and (*b*) 7 mmol l^{-1}. Reaction mixture as in figure 2.

From measurements at pH 7 it is also seen (in cases where there is no transpeptidation, e.g. DL_5, LDL_4, DL_4, LDL_3, so that individual splits can be measured) that ionization of the terminal carboxyl has the effect of greatly reducing the susceptibility of the C-terminal bond (see table 2). The same is true to a lesser extent for the penultimate bond (DL_5, DL_4D, LDL_4).

To sum up, table 1 shows that the effect of diastereoisomeric replacement or of elongation is strongly felt *over a range of seven amino acid residues*, namely P_1 to P_4 and P'_1 to P'_3. This means that the active site of papain corresponds in size to at least seven amino acid residues (subsites S_1 to S_4 and S'_1 to S'_3 in figure 1 and table 1). This size—about 2.5 nm, taking 0.35 nm per residue—is larger than was hitherto suspected. It is roughly as long as the binding site in lysozyme which accommodates six sugar units (Phillips 1966). This dimension need not necessarily be taken in a straight line; it refers only to the length of the peptide chain of the substrate when stretched.

Mapping of the active site

Hydrolysis

In order to map the active site, a number of peptides containing other L amino acids (Phe, Lys, Val, Gly) in addition to L-alanine were checked for hydrolysis by papain. Points of cleavage as well as approximate rates are given in table 3.

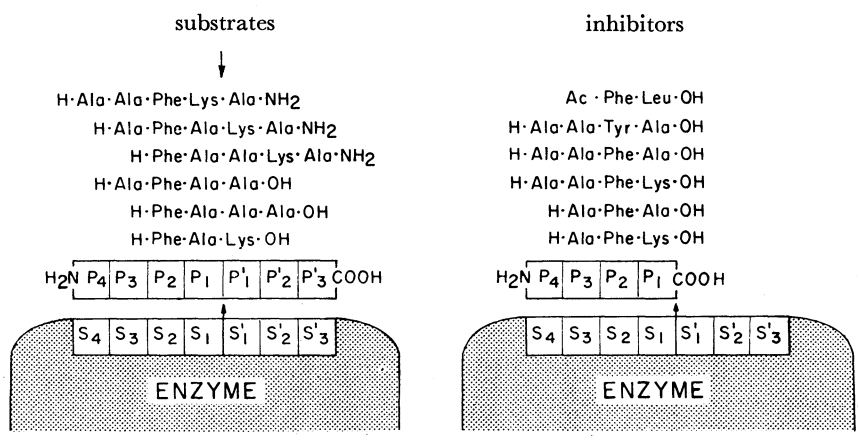
Table 3. Specificity of papain in the hydrolysis of peptides

Reaction mixtures (36 °C) contained peptide (0.02 mol l^{-1}), citrate-phosphate buffer pH 3.7 (citric acid 0.025 mol l^{-1}; Na_2HPO_4 0.016 mol l^{-1}), EDTA (0.002 mol l^{-1}), 2-mercaptoethanol (0.01 mol l^{-1}). Arrows indicate the bond split. t.p.: transpeptidation.

papain	peptide	% hydrolysis (approx.)
330 μg/ml (7 h)	Ala-Ala-Ala	40 t.p.
	Phe-Ala↓-Ala	60
	Ala-Ala-**Phe**	30 t.p.
330 μg/ml (1.5 h)	Ala-Ala-Ala-Ala	70 t.p.
	Phe-Ala↓-Ala-Ala	95
	Ala-**Phe**-Ala↓-Ala	95
	Ala-**Phe**-Gly↓-Ala	100
	Ala-Ala-Ala-**Phe**†	t.p.
	Val-Ala↓-Ala-Ala	95
	Ala-**Val**-Ala↓-Ala	95
	Ala-**Val**-Gly↓-Ala	100
	Val-**Phe**-Ala↓-Ala	80
30 μg/ml (3 h)	**Phe**-Ala↓-Lys-Ala-NH_2	5
	Ala-**Phe**-Lys↓-Ala-NH_2	30
30 μg/ml (1 h)	**Phe**-Ala↓-Ala-Lys-Ala-NH_1	60
	Ala-**Phe**-Ala↓-Lys-Ala-NH_2	60
	Ala-Ala-**Phe**-Lys↓-Ala-NH_2	60

† Substrate sparingly soluble.

It is seen that peptides in which Phe is the third or further residue from the C-terminal end are good substrates for papain (peptides containing Phe as the second residue from the C-terminal are hardly attacked). The bond split is always the next-but-one to the Phe residue in the direction of the C-terminal. In our terminology: whenever a split occurs the Phe residue is 'P_2' of the substrate, i.e. it always occupies subsite S_2 on the enzyme (see figure 6). This means



Figuee 6. Mode of binding of substrates and inhibitors in the active site of papain. The Phe residue always occupies S_2. The arrow indicates the point of cleavage.

that S_2 has a definite preference for the Phe side chain. Also the fact that the introduction of Phe into an appropriate position (third or further from the C-terminal) markedly increases the susceptibility of a given substrate to hydrolysis supports this assumption. It is interesting to note that the 'specific' splits are always purely hydrolytic, whereas in the peptides Ala_3, Ala_4, Ala_2Phe and Ala_3Phe transpeptidation occurs. These findings define a new kind of specificity, namely –Phe–X↓–Y– where the bond between X and Y is split specifically if X is preceded by phenylalanine. In our peptides containing both phenylalanine and lysine the Phe specificity as defined above dominates the pattern of hydrolysis. No 'lysine specificity' (Hill 1965) could be detected.

Valine residues behave in a manner similar to Phe residues (see table 3) and thus also show strong affinity for subsite S_2. Indeed, in the summary by Hill (1965) on hydrolysis of proteins and peptides of known sequences, it can be seen that bonds of the type Val–X↓–Y and Leu–X↓–Y are frequently attacked by papain. This type of specificity is similar to that described by Gerwin, Stein & Moore (1966) for the action of streptococcal proteinase, a sulphhydryl enzyme, on the reduced carboxymethylated B chain of insulin. Recently, Johansen & Ottesen (1968) pointed out that papain hydrolyses the oxidized B chain of insulin in a similar manner. It can thus be concluded that subsite S_2 binds preferentially hydrophobic amino acid side chains of small peptides as well as of long peptide chains.

Inhibition

The specificity of papain described above, giving evidence for strong interaction of Phe in subsite S_2, raised the possibility that peptides with Phe as the penultimate residue, while not undergoing cleavage, might bind strongly to subsites S_1, S_2 and S_3 on the one side of the catalytic sulphhydryl group (Lowe & Williams 1965; Bender & Brubacher 1966; Lowe, this volume, p. 237) of the enzyme (see figure 6), and by blocking part of the active site they would act as competitive inhibitors. A large number of peptides (tri- and tetrapeptides as well as amino-blocked di- and tripeptides) containing aromatic amino acid residues in the penultimate position were synthesized and tested for inhibition using BAEE as the substrate at pH 4.3, 36 °C. Under the experimental conditions of the inhibition experiments (1 μmol l^{-1} papain, 10–30 min) none of the compounds listed underwent hydrolysis. Even under more drastic conditions (10 μmol l^{-1} papain, 10 h) only trace digestion was observed.

Figure 7 gives the observed inhibition constants $\bar{K}_i$ ($= 1/K_i$), ranging from 10 ($pK_i = 1$) to 10^6 ($pK_i = 6$). Tripeptides such as Ala–Phe–Ala and Ala–Phe–Lys have binding constants of 1 and 3 l $mmol^{-1}$ respectively, whereas the corresponding tripeptides without Phe (Ala_3 and Ala_2Lys) have $\bar{K}_i$ values of 10 and 20 l mol^{-1} respectively.

In principle, inhibition measurements can be used to evaluate association equilibrium constants and hence thermodynamic quantities such as the free energy of binding. Work on free energies of binding of oligosaccharide inhibitors to lysozyme has been published (Sharon 1967; Sharon & Chipman 1969). However, for a detailed analysis the concentrations of the active species of the reactants must be known. It was shown by Sluyterman (1964) that in the inhibition of papain by benzoylarginine ($\bar{K}_i = 33$ l mol^{-1} at pH 4.3) the observed $\bar{K}_i$ decreases rapidly on raising the pH. He concluded that it is mainly the protonated species of the inhibitor which reacts with the enzyme. We too observed a strong pH dependence of the $\bar{K}_i$ values in the range pH 4 to pH 7.

If only the protonated species of the inhibitor molecule is in equilibrium with the enzyme,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

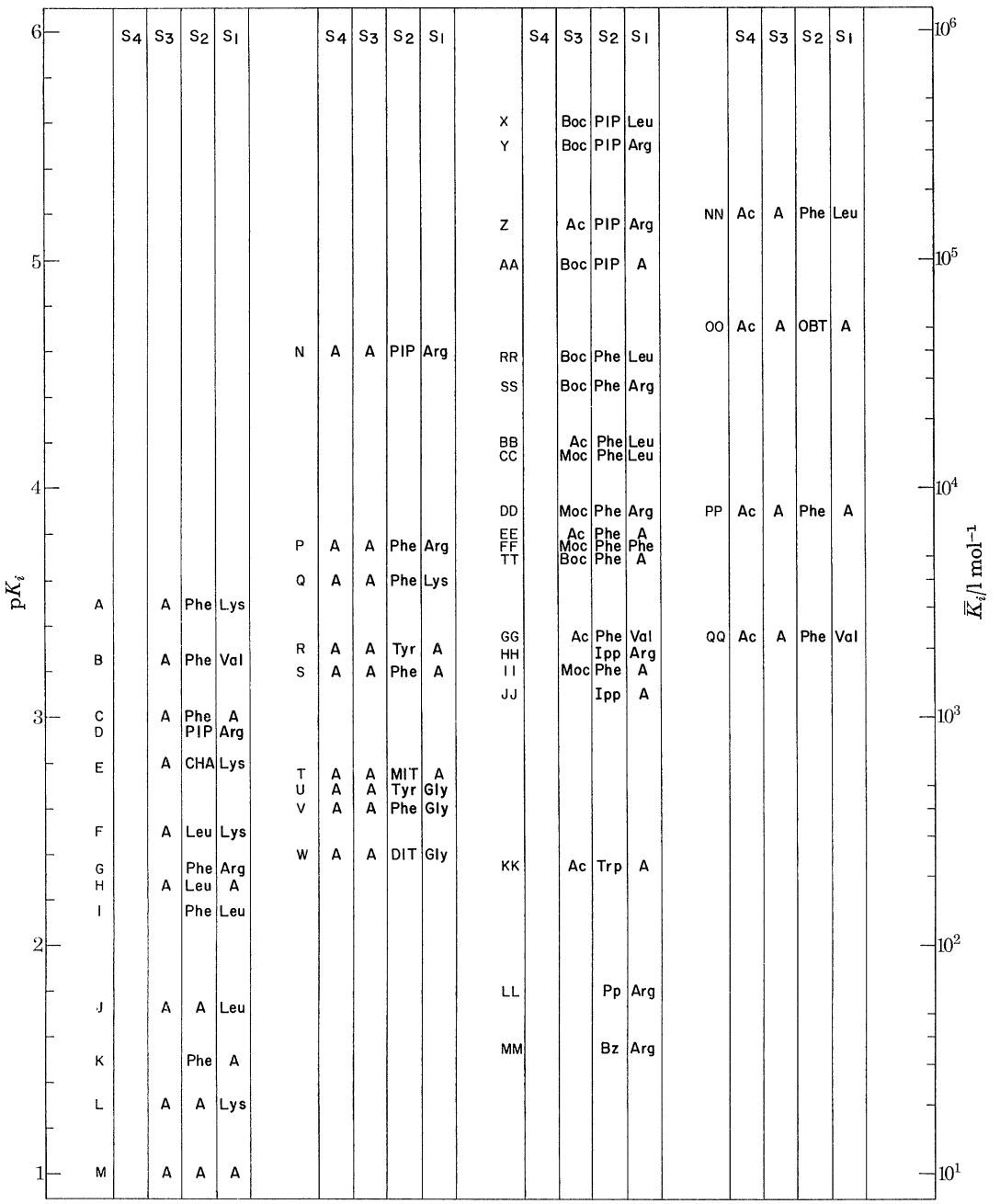


FIGURE 7. Inhibition of papain by various peptides. The inhibitors are drawn at the heights corresponding to their pK_i values (pK_i = $-\log K_i = \log \bar{K}_i$). Abbreviations (all amino acid residues are of the L-configuration): A: alanine; PIP: *p*-iodophenylalanine; MIT: *o*-monoiodotyrosine; DIT: *o*,*o*′-diiodotyrosine; OBT: *O*-benzyltyrosine; Ac: acetyl; Bz: benzoyl; Moc: methyloxycarbonyl; Boc: *tert*-butylyoxycarbonyl; Pp: *β*-phenylpropionyl; Ipp: *p*-iodo-*β*-phenylpropionyl. Substrate: Benzoyl arginine ethyl ester. For reaction conditions see Methods.

then the effective inhibitor concentration is (1-α)[I], where α is the degree of ionization of the carboxyl group. Therefore $\bar{K}_i = (1\text{-}\alpha)\bar{K}_i^0$ where $\bar{K}_i$ is the measured binding constant and $\bar{K}_i^0$ the 'true' binding constant for the protonated inhibitor species. Since for a weak acid

$$\lg \alpha - \lg(1\text{-}\alpha) = \mathrm{pH} - \mathrm{p}K_a,$$

we get for α values close to unity $\lg(1\text{-}\alpha) = -(\mathrm{pH} - \mathrm{p}K_a)$ and hence

$$\lg \bar{K}_i = \lg \bar{K}_i^0 - (\mathrm{pH} - \mathrm{p}K_a).$$

For *tert*-Boc-*p*-iodo-Phe–Arg ($\mathrm{p}K_a = 3.3$) we found $\mathrm{p}K_i (= \lg \bar{K}_i) = 5.55$ at pH 4.3; $\mathrm{p}K_i = 5.13$ at pH 5; $\mathrm{p}K_i = 4.14$ at pH 6; and $\mathrm{p}K_i = 3.26$ at pH 7, in agreement with the above assumption. From these values one obtains $\mathrm{p}K_i^0 = 6.8$ or $\bar{K}_i^0 = 6 \times 10^6$. This corresponds to a standard free energy of binding $-\Delta F^0 = 40.6$ kJ mol^{-1} (9.7 kcal mol^{-1}). This evaluation depends, however, entirely on the above interpretation of the pH dependence of $\bar{K}_i$ and seems not entirely justified at present, since at least part of the decrease of the $\bar{K}_i$ values with increasing pH may be due to changes in the properties of the enzyme.

Let us assume now that all the inhibitors listed occupy the binding site as shown in figure 6 (i.e. the C-terminal residue in S_1). This assumption is supported by the observation that Ala-Phe-Ala (L-D-L) is not an inhibitor ($\bar{K}_i < 5$ l mol^{-1}), i.e. S_2 shows almost absolute stereospecificity in inhibition similar to that observed in hydrolysis (see above). One can then try to analyze the interactions of the various side chains in the four subsites. By comparing pairs of inhibitors in which only one amino acid residue (or blocking group) is replaced by another, the affinity of a subsite towards a given side chain can be evaluated. Thus comparing for example Ala-Phe-Ala with Ala-Phe-Lys one obtains the difference in the affinity of S_1 for the lysine and alanine side chains.

Since the standard free energy of binding is related to the binding constant through the expression $-\Delta F^0 = RT \ln K$, the difference in the free energy of binding of two peptides is given by $-\Delta^2 F^0 = RT\Delta \ln \bar{K}_i$ provided that the two $\bar{K}_i$ values are determined at the same pH and the inhibitors have the same $\mathrm{p}K_a$. In this case differences in $\mathrm{p}K_i$ are proportional to differences in the free energy of binding.

At the present stage we prefer presenting the measured $\Delta\mathrm{p}K_i$ values rather than calculated ΔF^0 values because of the above reservations and also because precise $\mathrm{p}K_a$ values are not available. However, it seems that the estimated $\Delta^2 F^0$ values will be only slightly affected by differences in $\mathrm{p}K_a$ (at most 1.25 kJ mol^{-1}) since these are quite small. We found $\mathrm{p}K_a = 3.3$ for *tert*-Boc-*p*-iodo-Phe-Arg and $\mathrm{p}K_a = 3.5$ for *tert*-Boc-*p*-iodo-Phe-Ala, and this seems to be one of the largest differences expected since the former molecule contains a positively charged guanido group near its carboxyl.

Table 4 gives the $\Delta\mathrm{p}K_i$ values for the replacement of an alanine residue by another residue. In the case of Leu instead of Ala this difference is 0.8 (average). From this quantity one obtains $\Delta^2 F^0$ in kJ mol^{-1} by multiplication by the factor 5.9, i.e. the ΔF^0 for replacing an alanine by a leucine residue in S_1 is on the average 4.6 kJ mol^{-1}. Interpreted in this manner the $\Delta\mathrm{p}K_i$ values can be used for mapping the active site.

Subsite S_1 seems to bind alanine considerably better than glycine (by about 3.3 kJ mol^{-1}) and the larger side chains of lysine, arginine, leucine and phenylalanine better than alanine (by about 2.9 kJ mol^{-1}). The present data thus do not show that a positive charge increases binding in S_1. From the series Gly < Ala < Leu one might rather conclude that binding in S_1 is predominantly hydrophobic.

The case of valine, which despite its hydrophobic side chain binds weakly in S_1, seems to be a special one due to the branching on its β-carbon. It seems possible that due to space limitations at S_1 the isopropyl group would not be easily accommodated. If this is so, one might expect that in peptide substrates containing valine, crowding may be so severe (due to occupancy of S_1')

TABLE 4. ΔpK_i VALUES FOR THE REPLACEMENT OF AN ALANINE RESIDUE BY ANOTHER AMINO ACID (IN S_1 AND S_2) AND FOR ELONGATION (IN S_3 AND S_4)

S_1 (replacement of Ala)

Lys		Arg		Leu		Val		Gly		Phe	
A-C	0.5	G-K	0.8	I-K	0.7	B-C	0.3	U-R	−0.6	FF-II	0.6
F-H	0.3	P-S	0.6	J-M	0.8	GG-EE	−0.4	V-S	−0.6		
L-M	0.2	Y-AA	0.5	X-AA	0.6	QQ-PP	−0.5				
Q-S	0.4	DD-II	0.7	BB-EE	0.4						
		HH-JJ	0.2	CC-II	1.0						
		SS-TT	0.7	NN-PP	1.3						
				RR-TT	0.9						
av.	+0.4		+0.6		+0.8		−0.2		−0.6		+0.6

S_2 (replacement of Ala)

Phe		Tyr		Leu		Trp		CHA		PIP	
A-L	2.2	R-S	2.2	F-L	1.2	(KK-EE)†	0.6	E-L	1.5	(D-G)†	2.7
C-M	2.0	U-V	2.2	H-M	1.3					(N-S)†	3.0
										(AA-TT)†	3.4
										(Y-SS)†	3.2
										(X-RR)†	3.1
										av.	+3.1

S_3 (elongation)

Ala		Ac		Moc		Boc	
C-K	1.5	EE-K	2.3	II-K	1.7	Y-D	2.6
(A-G)†	1.4	BB-I	2.1	DD-G	1.6	SS-G	2.2
		Z-D	2.2	CC-I	2.0	RR-I	2.3
						TT-K	2.2
av.	+1.5		+2.2		+1.8		+2.3

S_4 (elongation)

Ala		Ac	
S-C	0.2	PP-C	0.9
Q-A	0.1	QQ-B	0.1

† Corrected by adding 2.1 (= Phe-Ala)

as not to allow valine to occupy S_1. This means that substrates would not be hydrolysed at the carbonyl group of valine. A comparison of the following splits observed supports this point:

Ala-Val-Ala↓-Ala	fast
Phe-Ala↓-Ala-Ala	fast
Phe-Val-Ala↓-Ala	fast
Phe-Val-Gly↓-Ala	fast
Phe-Ala↓-Ala	slow
Val-Phe↓-Ala	very slow
Phe-Val-Ala	none

It is seen that cuts at the Val residue do not occur even in cases where the strong binding of Phe in S_2 would favour this kind of hydrolysis. In other words, the spatial restriction preventing the Val side chain from occupying S_1 demolishes the Phe-X↓-Y specificity.

Subsite S_2 has a strong affinity for the phenylalanine side chain, the addition of a phenyl group to the alanine moiety increasing the free energy of binding by about 12.5 kJ mol^{-1} (av. ΔpK_i 2.1). The tyrosine side chain has the same effect. Leucine is also strongly bound, but

less than phenylalanine (7.5 kJ mol^{-1} for the extra isopropyl group). Comparing these values with the ones for S_1 it is seen that S_2 interacts with the side chains considerably more strongly than S_1. Moreover, substitution in the phenyl ring in the para position can further increase the binding energy in S_2: an iodine contributes about 4.2 kJ mol^{-1} (8.4 kJ mol^{-1} in the case of the phenylpropionyl blocking group); benzylation of the tyrosine oxygen increases the binding by 4.2 kJ mol^{-1}. Substitution in the meta position decreases binding. The finding that the cyclohexyl ring of cyclohexylalanine is bound less strongly than a phenyl ring (by about 3.3 kJ mol^{-1}) may indicate aromatic interaction in S_2. On the other hand the interaction in S_2 may be mainly hydrophobic, the difference observed being due to the differences in the geometry of the two rings.

Discussion

The mapping procedure described here for the case of papain seems to be applicable to other enzymes which act on chain molecules. Its prime purpose is to obtain an imprint, as it were, of the active site as a whole. On the basis of the 'map' obtained, admittedly a rather abstract one to begin with, one may expect to predict the behaviour of the binding site towards unknown molecules. As a matter of fact, in the course of this work—tacitly assuming additivity of the binding forces due to individual groups—we were able to 'improve' the binding properties of our inhibitors, as we went along, by a factor of 10^4—without yet having tackled the possibilities of subsites S_3 and S_4. Encouraged by this success we tried, in cooperation with the Groningen group of Drenth, to predict the way in which a heptapeptide, covering all the seven subsites, would bind to the papain molecule. With the help of their wire model (Drenth, Jansonius, Koekoek, Swen & Walters 1968; Drenth *et al.*, this volume, p. 231) and of a space filling model of part of the molecule we came up with a proposal which accounted for most of the features described above: stereospecificity, special interactions, space limitations.

In constructing the model, it seemed reasonable to start from the assumption that the peptide bond to be cleaved should be located so as to allow interaction with the His-159 (e.g. through binding of the peptide oxygen to the protonated imidazole nitrogen), with the possibility of subsequent interaction with the active thiol of Cys-25 (Lowe, this volume, p. 237). The second criterion was to obtain good binding, preferably by hydrogen bonds, for each of the additional amino acid residues (P_2, P_3, P_4, P_2' and P_3'). Formation of these hydrogen bonds was considered essential for gaining the necessary free energy of binding to account for the increased rates of hydrolysis observed on elongation of the substrates.† Thirdly, the mode of binding had to meet the stereospecific requirements observed in substrates and inhibitors, as well as the requirement of specific aromatic interaction in S_2 for the phenylalanine residue. The solution arrived at is illustrated in figure 8.

The use of space filling models (CPK) greatly facilitated the task of fitting the substrate to the active site, since 'close contacts' are avoided, and the permissible lengths and directions of the

† Actually, the formation of a hydrogen bond (C=O ⋯ H–N) in aqueous medium

$$C{=}O\ldots HOH + N{-}H\ldots OH_2 \rightleftharpoons C{=}O\ldots H{-}N + H{-}O{-}H\ldots OH_2$$

results in little net change of free energy, because of the concurrent dehydration of the polar groups; but without hydrogen bond formation there would be an actual loss of energy due to the loss of hydration energy involved. Therefore once all the polar groups are taken care of by hydrogen bonding, any further non-bonded interactions ('hydrophobic', dipole–dipole, aromatic, charge transfer, etc.) are a source of free energy binding. This very much resembles the situation in the helix–random coil equilibria of polypeptides in water (Lotan, Yaron & Berger 1966) where the hydrogen bonded helical structure requires for its stabilization considerable side chain interaction.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

hydrogen bonds are provided for. It was found that in order to accommodate the peptide chain on both sides of the peptide bond to be split, the 'planar' carbonyl carbon of P_1 and 'planar' amide nitrogen of P_1' had to be replaced by tetrahedral carbon and nitrogen atoms. A tetrahedral carbon has been postulated as the transition state in hydrolysis by serine enzymes, leading to the acylation of the serine hydroxyl. In the present case the tetrahedral carbon would react with the nearby sulphur atom of Cys-25 to yield the thioester intermediate, with the generation of a free amino terminal in P_1'. This 'activation' of the otherwise coplanar resonance structure of the peptide bond may be brought about by the binding energy. Thus, the process of binding would also be the initial step of catalysis (as has been postulated in the case of lysozyme by Phillips 1966).

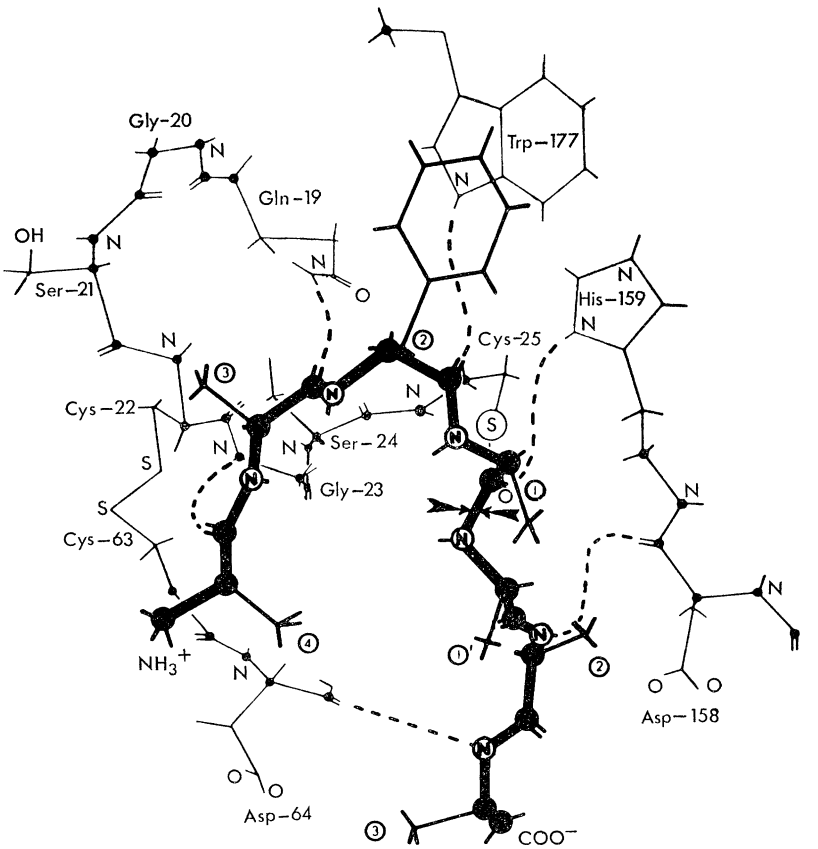


FIGURE 8. Projection of part of the papain molecule and of the heptapeptide $Ala_2PheAla_4$ in a proposed mode of binding. Hydrogen-bonded atoms are schematically connected by dotted lines. ⟹, carbonyl group.

In the proposed complex (see figure 8), the carbonyl oxygen of P_1 (O–P1) is hydrogen bonded to the protonated imidazole N of His-159; the O–P2 is hydrogen bonded to the indole N of Trp-177; the O–P 3 is hydrogen bonded to the amide N of Gln-19; the O–P4 is hydrogen bonded to the peptide N of Gly-23. There are no directed bonds involving P_1'. The N–P 2′ is bonded to the peptide O of Asp-158 and N–P 3′ to the peptide O of Asp-64.

Placing the heptapeptide thus in position, it is seen that it just fills a 'pocket' of the enzyme, producing a flush surface. This seems to be a favourable situation from the point of view of 'hydrophobic binding' particularly so since no hydratable groups on the enzyme are blocked off without compensation by hydrogen bonding. The phenyl ring covers the exposed indole

group of Trp-177, and it seems that interaction between the two aromatic systems is the reason for the observed specificity of the Phe residue both in hydrolysis and inhibition. The pronounced enhancement of binding on introducing iodine in the *para*-position may stem from electron transfer (from the indole to the phenyl ring), improved dipole interaction or hydrophobic bonding. The fact that *o*,*o*-di-iodotyrosine binds less well than tyrosine is explained by interference of one of the iodine atoms with the side chain of Gln-142 which prevents alinement of the two ring systems. The aliphatic ring in cyclohexylalanine points away from the tryptophan and thus interacts less strongly than the aromatic one. However, we could not deduce from the model why monoiodo tyrosine or tryptophan are bound less well than phenylalanine.

The strict stereospecificity of S_1, S_2 and S_1' is accounted for by the model. With the three main hydrogen bonds in place, the α-hydrogens of P_1, P_2 and P_1' (in the L-form) point towards the enzyme, the side chains away from it. Since the situation would be reversed for D residues, steric interference would not allow the hydrogen bonds to be formed.

Examination of the enzyme surface in the vicinity of the substrates in their proposed modes of binding led to a number of predictions, which could be verified experimentally. (*a*) It was seen that in benzoyl arginine there is only limited interaction between the phenyl ring and Trp-177. Phenylpropionyl arginine in which the phenyl ring can approach the indole much closer, was found to inhibit twice as well, *p*-iodophenylpropionylarginine thirty times better. (*b*) Locating Tyr in S_2 in the proposed way it was seen that a groove extending away from the phenolic group might accommodate another phenyl ring (especially in the form of a benzyl ether) which would fill out the depression. Examination of Ac-Ala-*O*-benzyl-Tyr-Ala showed that the benzyl ether group increased inhibition by a factor of almost ten. (*c*) It was observed that the methyl side chain of Ala-136 restricts the space available for P_1 and especially that a residue containing a β-branched side chain in P_1 could not be fitted in the space filling models. As discussed above, it seems that indeed Val-X bonds are not hydrolysed by papain (cf. Johansen & Ottesen 1968).

Having described this mental, and in part, experimental effort rather in detail, it may be something of an anticlimax to state that this model may have to be discarded. There are indications from preliminary X-ray work with two inhibitors containing a *p*-iodophenyl group expected to bind in S_2, that the iodine atom is found near Trp-69 rather than near Trp-177 (J. Drenth, private communication) although with low occupancy. Judging by its distance from the sulphhydryl this area is a good candidate for S_2, but a detailed proposal for this mode of binding has not yet been worked out. There is, of course, always the disturbing thought at the back of one's mind that a 'static' complex observed in the crystal is not necessarily the same as the 'dynamic' one in kinetic experiments. While the evidence presented here strongly indicates that our inhibitors indeed occupy subsites S_1, S_2, etc., when inhibiting hydrolysis of a substrate, there may be another binding area on the enzyme which an inhibitor can occupy without affecting enzyme kinetics. This site may be more accessible in the crystal than the active site itself.

Clarification of these points by three dimensional X-ray work has been hampered by the fact that the original crystals of papain and its heavy metal derivatives on which the structure was worked out are stable in high methanol concentration (above 50 %) and at high pH, whereas the inhibitors tested are effective in water at low pH. However, the situation is somewhat more promising now since some of our more recently synthesized inhibitors do bind considerably at high pH and were found to inhibit even in the presence of 50 % alcohol. One may hope that

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

complete X-ray analysis of a number of enzyme inhibitor complexes will give the answers to the points raised here.

There is one additional aspect of the problem which might be mentioned in conclusion. The research presented here was initially motivated by the wish to answer questions like: what is the nature of the binding site of the enzyme molecule, what is the source of enzyme specificity, how does substrate binding lead to catalysis and what is the nature of catalysis. However, it seems that one might attempt, using enzymes as convenient tools provided by nature, to answer questions of a more general nature, outside the realm of enzyme chemistry. The enzyme–inhibitor system is a system of two molecules, a large and a small one, which by virtue of special geometric relations show strong affinity to each other. Moreover, a family of inhibitors will be bound to the enzyme surface in a similar way. If one were to obtain from X-ray analysis the precise geometry of a series of such complexes and from physico-chemical studies their enthalpies and entropies of binding one could start to break down the measured quantities into the sum of individual interactions of groups (or single atoms) of the inhibitor with the known surface of the enzyme molecule, and to evaluate the role of the surrounding solvent on these interactions. It is unnecessary to state of what value such information would be in the understanding of related interactions such as those of antigen with antibody, the assembly of subunits or the formation of multi-enzyme systems and membranes. The state of X-ray crystallography today and the possibilities of synthesizing inhibitor molecules designed to answer specific questions of interaction seem to make it worthwhile to approach this formidable task.

The authors acknowledge the cooperation of Miss Hinga Benderly who measured inhibition constants as part of her Ph.D. thesis work, and of Mr I. Jacobson in the synthesis of peptides.

References (Berger & Schechter)

Abramowitz, N., Schechter, I. & Berger, A. 1967 *Biochem. biophys. Res. Commun.* **29**, 862.
Anderson, G. W., Zimmerman, J. E. & Callahan, F. M. 1964 *J. Am. chem. Soc.* **86**, 1839.
Bender, M. L. & Brubacher, L. J. 1966 *J. Am. chem. Soc.* **88**, 5880.
Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Walters, B. G. 1968 *Nature, Lond.* **218**, 929.
Gerwin, B. I., Stein, W. H. & Moore, S. 1966 *J. biol. Chem.* **241**, 3331.
Hill, R. L. 1965 *Adv. prot. Chem.* **20**, 37.
Irving, G. W., Fruton, J. S. & Bergmann, M. 1941 *J. biol. Chem.* **138**, 231.
Johansen, J. T. & Ottesen, M. 1968 *C. r. Trav. Lab. Carlsberg* **36**, 265.
Lotan, N., Yaron, A. & Berger, A. 1966 *Biopolymers* **4**, 365.
Lowe, G. & Williams, A. 1965 *Biochem. J.* **96**, 199.
Morihara, K. & Oka, T. 1968 *Biochem. biophys. Res. Commun.* **30**, 625.
Phillips, D. C. 1966 *Scient. Am.* **215**, 78.
Schechter, I. & Sela, M. 1962 Unpublished results quoted in E. Katchalski, M. Sela, H. J. Silman and A. Berger, 1964, in *The proteins*, vol. II (ed. H. Neurath), p. 405. New York: Academic Press.
Schechter, I., Abramowitz, N. & Berger, A. 1965 *Israel J. Chem.* **3**, 101.
Schechter, I., Benderly, H., Berger, A., Lotan, N. & Scheraga, H. A. 1967 *Abstr.* 154 *Meeting, Am. Chem. Soc., Chicago*, C-171.
Schechter, I. & Berger, A. 1966*a* *Biochemistry* **5**, 3362.
Schechter, I. & Berger, A. 1966*b* *Biochemistry* **5**, 3371.
Schechter, I. & Berger, A. 1967 *Biochem. biophys. Res. Commun.* **27**, 157.
Schechter, I. & Berger, A. 1968 *Biochem. biophys Res. Commun.* **32**, 898.
Sharon, N. 1967 *Proc. Roy. Soc.* B **167**, 402.
Sharon, N. & Chipman, D. 1969 *Science, N.Y.* (in the Press).
Sluyterman, L. A. A. 1964 *Biochim. biophys. Acta* **85**, 316.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES