ON THE SIZE OF THE ACTIVE SITE IN PROTEASES. I. PAPAIN

Israel Schechter and Arieh Berger

The Weizmann Institute of Science, Rehovoth, Israel

Received March 10, 1967

In order to investigate the size of the active site of papain, we have studied its action on 40 diastereoisomeric peptides of alanine, ranging in length from Ala_2 to Ala_6. The variations observed in the rates of hydrolysis lead to the conclusion that papain, an endopeptidase, has a large active site which extends over about 25 Å and can be divided into 7 "subsites", each accommodating one amino acid residue of the peptide substrate. The subsites are located on both sides of the catalytic site, 4 on the one side and 3 on the other. This suggests that there may be a long "groove" analogous to that found by X-ray analysis in lysozyme (Johnson and Phillips, 1965). The substrate is visualized as fitting into the groove, binding to several subsites of specific geometry (see scheme, Fig. 1).

The active site of an enzyme performs the twofold function of binding the substrate and catalyzing the reaction. The efficiency of these actions determines the

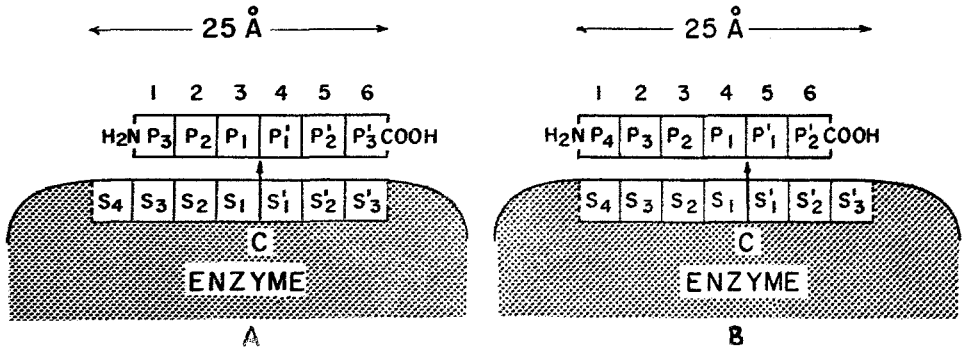

Fig. 1. Schematic representation of two possible enzyme-substrate complexes of papain with a hexapeptide. The active site of the enzyme is composed of 7 "subsites" (S_1 - S_4 and S_1' - S_3') located on both sides of the catalytic site C. The positions, P, on the hexapeptide substrate are counted from 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.

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 studying the kinetics of the enzyme's reactions with different substrates.

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 et al., 1965).

Changes in substrate composition may affect either the degree of enzyme-saturation (K_m), or the catalytic step (k_3), 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 et al., 1941). This is equal to k_3/K_m, the rate at infinite substrate dilution.

MATERIALS AND METHODS. The synthesis of the alanine peptides used as substrates has already been described (Schechter and Berger, 1966a). Papain, 2x crystallized (Worthington) was used. For reaction conditions see Fig. 3.

The compositions of the reaction mixtures at different times were determined by quantitative paper electrophoresis (Schechter and Berger, 1966b). Two examples are given in Fig. 2. When necessary, the reaction products were isolated and their stereochemical composition determined by further degradation with CP-A and/or LAP (Schechter and Berger, 1966b). 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

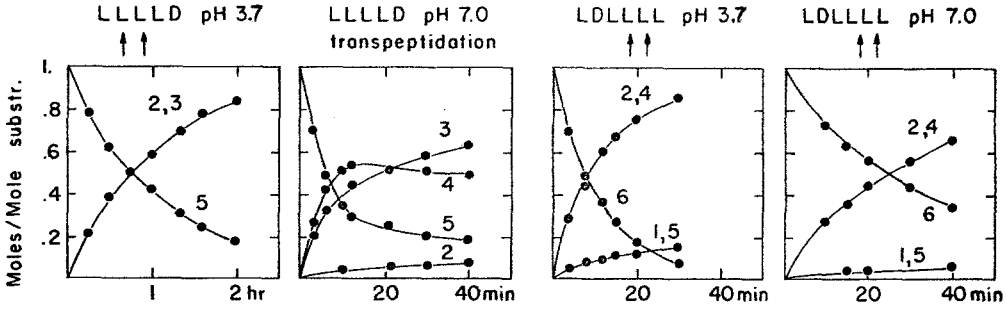

Fig. 2. The course of hydrolysis of alanine peptides by papain. Arrows indicate bonds cleaved; numbers give the size of the peptide.

Ala_2(DL). Ala_5(LL-L-LD) gave Ala_3(LLL 34%; LLD 66%) and Ala_2(LL 66%; LD 34%), showing splits at both bonds indicated, in a ratio of 2:1. In the case of Ala_5(LLLLL) it was necessary to synthesize a labelled substrate. Ala_5(L^* L-L-LL, C^{14}) gave Ala_2 (L^* L) and Ala_3(L^* LL) in a ratio of 1:2. 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.

RESULTS AND DISCUSSION. Proteolytic coefficients, C, for cleavages at pH 3.7 are given in Fig. 3 on a log scale [1]. No hydrolysis occurs at bonds involving a D-residue, or at L-L bonds in DL-L sequences (Schechter et al.,1965). We interpret the observed variations in C on the basis of the following assumption: the substrates are lined up on the enzyme in such a way that the CO-NH group being hydrolyzed always occupies the same place (the catalytic site); the amino acid residues occupy adjacent subsites, those towards the NH_2-end occupying subsites S_1, S_2 etc., those towards the COOH-end occupying subsites S_1', S_2' etc. For convenience we number the "positions" 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 Fig. 1).

Table I gives the information which can be extracted from Fig. 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. A striking feature is the cooperative effect of a double replacement on opposite sides of the susceptible bond, as in Ala_5(DLLLD) $\rightarrow$ Ala_5(LLLLL), resulting in a quotient of 100 (J/S, Fig. 3); this is the product of the values for the two single replacements (J/N = 10 and J/Q = 10).

Also in polypeptides, the decrease in hydrolysis rate due to the presence of D-

---

[1] 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 Fig. 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. Evidence for the existence of subsites S_1', S_2' and S_3' was obtained by Brubacher and Bender (1967) who found that cinnamoyl-papain was deacylated by Gly_nNH_2(n = 1 to 3) at rates increasing with n.

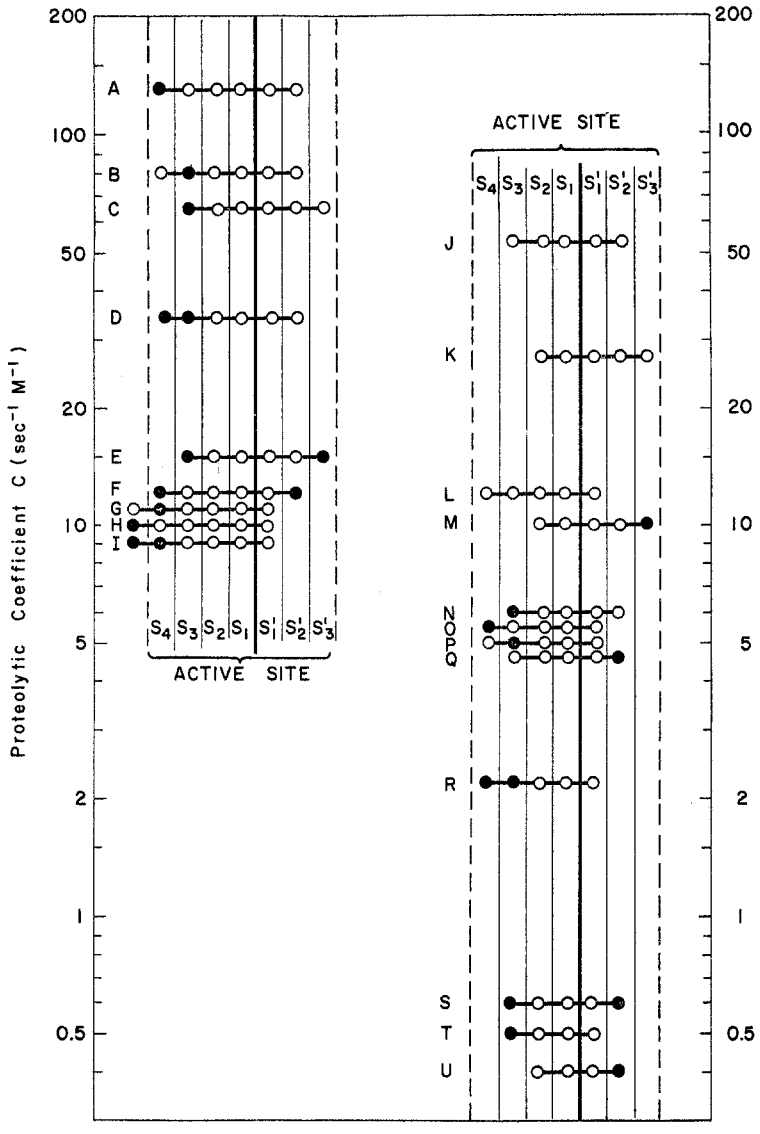

Fig. 3. The proteolytic coefficients C (in $sec^{-1}M^{-1}$) for the hydrolysis of individual bonds in a series of alanine peptides at pH 3.7, 36°C. The peptides are drawn schematically at the height corresponding to the C-value for the bond crossing the heavy vertical line (empty circles -o- indicate L-residues, full circles -●- 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. Reaction mixtures: substrate 0.02 M; papain 1.6×10^{-5} M (0.33 mg/ml); mercaptoethanol 0.01 M; EDTA 0.002 M; citric acid 0.025 M; Na_2HPO_4 0.016 M.

TABLE I 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 Fig. 3. The letters in parentheses correspond to the C-values from which the ratio is calculated.

P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'
S_4	S_3	S_2	S_1	S_1'	S_2'	S_3'
a) Diastereoisomeric replacement D → L						
2.4 (B/D)	3.8 (A/D)	$> 10^2$	$> 10^3$	$> 10^3$	12 (A/F)	4.5 (C/E)
2.2 (L/O)	10 (J/N)				10 (J/Q)	2.7 (K/M)
2.5 (P/R)	2.2 (L/P)				9 (N/S)	
	2.5 (O/R)					
	8.3 (Q/S)					
b) Elongation by L-residue						
11 (P/T)	12 (Q/U)	> 50	-	-	23 (A/O)	12 (C/N)
15 (B/N)					15 (B/P)	
					15 (D/R)	
					11 (N/T)	
c) Elongation by D-residue						
2.4 (A/J)	1.5 (S/U)	-	-	-	1.2 (S/T)	2.5 (E/N)
6.2 (D/N)	2.4 (C/K)				2.0 (F/O)	
2.2 (F/Q)	1.4 (E/M)					
4.4 (R/T)						

residues follows the pattern of Table I. We found that the DLLD sequences in the ordered copolymer $(LDL)_8$ are not digested, the DLLLD sequences in $(LDLL)_8$ are digested slowly ($C \sim 1$), and the DLLLLD sequence in $(LLDLL)_2$ quite rapidly ($C = 30$).

To sum up, from Table I it can be seen that the effect of diastereoisomeric replacement or of elongation is strongly felt over a range of seven amino acid residues, namely in 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 - S_4 and S_1' - S_3' in Fig. 1 and Table I). This size - about 25 Å, taking 3.5 Å per residue - is larger than was hitherto suspected. It is roughly as long as the binding site in lysozyme which accommodates 6 sugar units (Phillips, 1966).

If we assume that the active sites in proteases in general are large, it might

be possible to explain "unexpected" cleavages observed in proteins (Hill, 1965). It is reasonable to expect that the subsites show, in addition to their stereospecificity, specific interactions with different amino acid side-chains. We have indeed observed this in carboxypeptidase-A (Abramowitz et al., 1967). This would mean that the reactivity of a certain bond in a protein depends not only on the two residues forming this bond (which occupy S_1 and S_1' during hydrolysis) but also on the nature of the residues in its neighborhood, occupying adjacent subsites. In other words, the enzymes seem to be capable of "recognizing" quite a large portion of a peptide chain.

References

Abramowitz, N., Schechter, I., and Berger, A. (1967), to be published.
Brubacher, L.J., and Bender, M.L. (1967), to be published.
Hill, R.L. (1965), Advan. Protein Chem. 20, 37.
Irving, G.W., Fruton, J.S., and Bergmann, M. (1941), J. Biol. Chem. 138, 231.
Johnson, L.N., and Phillips, D.C. (1965), Nature, 206, 761.
Phillips, D.C. (1966), Scientific American, 215, 78.
Schechter, I., Abramowitz, N., and Berger, A. (1965), Israel J. Chem. 3, 101.
Schechter, I., and Berger, A. (1966a), Biochemistry, 5, 3362.
Schechter, I., and Berger, A. (1966b), Biochemistry, 5, 3371.