PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Structure of the Papain Molecule

J. Drenth, J. N. Jansonius, R. Koekoek, L. A. A. Sluyterman and B. G. Wolthers

Phil. Trans. R. Soc. Lond. B 1970 **257**, 231-236
doi: 10.1098/rstb.1970.0022

Phil. Trans. Roy. Soc. Lond. B **257**, 231–236 (1970) [231]
Printed in Great Britain

IV. CYSTEINE PROTEINASES

The structure of the papain molecule

By J. Drenth,† J. N. Jansonius,† R. Koekoek,†
L. A. A. Sluyterman‡ and B. G. Wolthers†
† *Laboratorium voor Structuurchemie, Rijksuniversiteit Groningen, The Netherlands*
‡ *Philips Research Laboratories, N.V. Philips Gloeilampenfabrieken, Eindhoven, The Netherlands*

1. Introduction

Papain (EC 3.4.4.10) is a proteolytic enzyme which is isolated from the Papaya, a common tropical tree. It is a sulphydryl enzyme and its SH group is required for enzymic activity. Papain as usually prepared (Kimmel & Smith 1954) contains only a small portion of active molecules. The majority of the molecules are inactive because their sulphydryl group is blocked. Part of

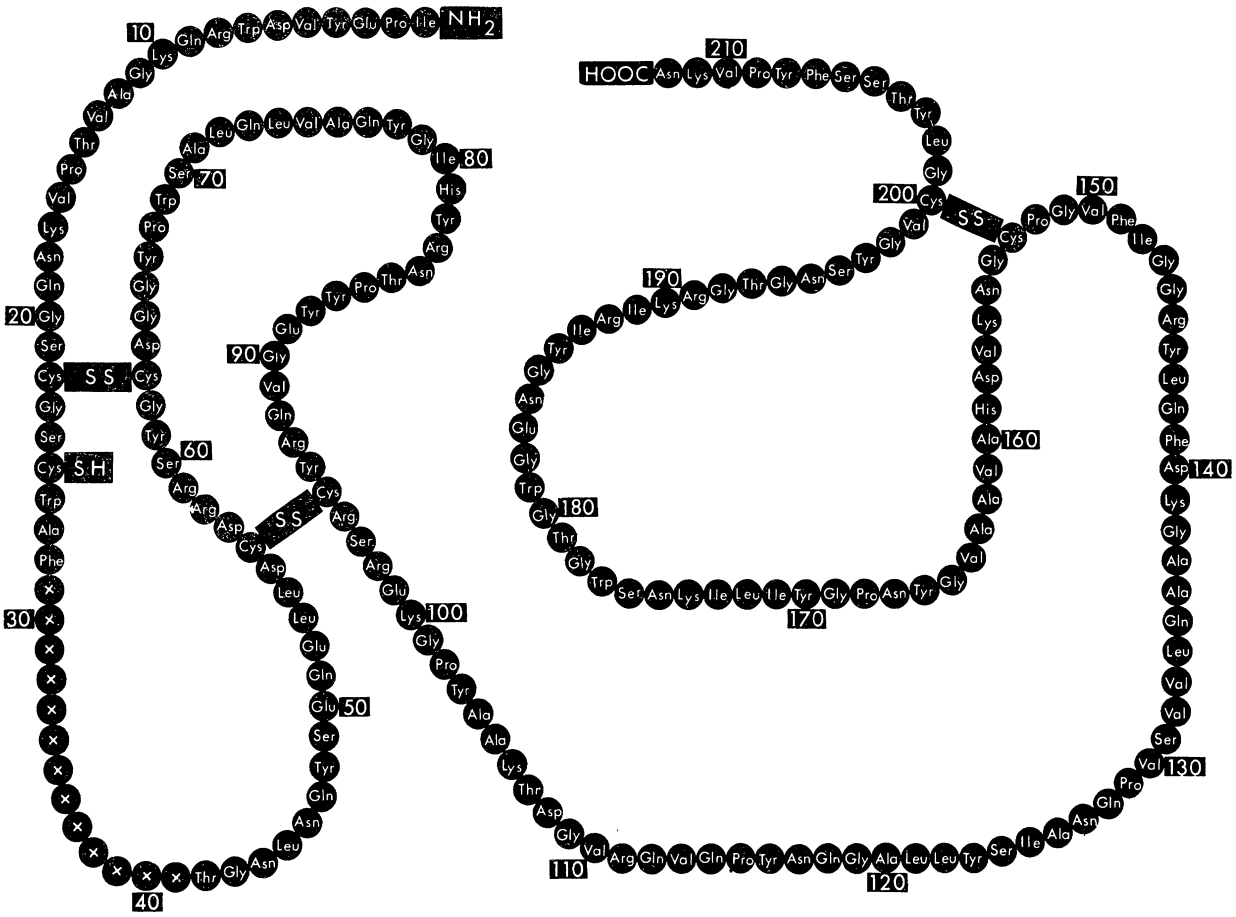

Figure 1. Amino acid residue sequence in papain. The electron density map suggested the presence of an extra valine residue in the original sequence. This has now been confirmed by Dr E. L. Smith. This residue is incorporated in the sequence as Val-130 thus increasing by one the numbers assigned to subsequent residues as reported at the meeting.

the blocking is caused by disulphide formation with cysteine (Sluyterman 1967). This disulphide can be reduced by an excess of cysteine resulting in an active enzyme preparation. The free SH content never reaches 100 % and is often not more than about 50 %, so that we must distinguish between papain molecules with a reversibly and an irreversibly blocked SH group. The chemical nature of the irreversible blocking is not yet known. It might well be due to a higher oxidation state of the sulphur which cannot be reduced by an excess of cysteine (Glazer & Smith 1965).

The papain molecule consists of a single chain of 212 amino acid residues with three disulphide bridges (figure 1), two of them in the first half of the chain and the third one in the second half. Both chemical and X-ray studies have contributed to our knowledge of the primary structure (Light, Frater, Kimmel & Smith 1964; Drenth, Jansonius, Koekoek, Swen & Wolthers 1968*b*).

2. The determination of the structure

The crystals were grown from methanol–water (2:1). They are orthorhombic with $a = 4.50$ nm, $b = 10.43$ nm, $c = 5.08$ nm, space group $P2_12_12_1$, $n = 4$. X-ray analysis by the method of isomorphous replacement together with anomalous scattering data, has resulted in an electron density map with a resolution of 0.28 nm (Drenth, Jansonius, Koekoek, Swen & Wolthers 1968*b*). The X-ray intensities were measured with a Hilger and Watts linear diffractometer (Phillips 1964; Arndt, North & Phillips 1964). Phase angles were derived from five isomorphous derivatives, containing: *p*-mercury (II) benzoate, *p*-mercury (II)-benzenesulphonate, *p*-mercury (II)-aniline, papain-SHgCl and papain-SHgCl + $HgCl_2$.

In papain-SHgCl mercury (II) chloride had reacted with the free sulphydryl group.

3. The structure of the molecule

The electron density map showed the trace of the polypeptide chain. The majority of the C=O groups and nearly all the side groups were clearly visible and this enabled us to construct a molecular model. Amino acid sequences in peptides of papain as determined previously by chemical methods (Light *et al.* 1964) were not sufficiently overlapping to give the complete sequence in the molecule. As a result of the X-ray structure determination we could put the various parts into the right order and add extra residues which were lost in the chemical work. From the size and shape of the side groups we suggest for the extra residues:

-Ser-Ala-Val-Val-Ala-Val-Arg-Ala-Ile-Ile-Arg-	Val or Thr	-Glu or Gln
29 30 31 32 33 34 35 36 37 38 39	40	41

The final word for this part of the sequence must come from chemical studies.

The molecule measures roughly $5.0 \times 3.7 \times 3.7$ nm. It is a binuclear protein with two hydrophobic cores. The two parts of the molecule contain approximately equal numbers of residues. The chain crosses from one to the other part near residue 111. The folding of the chain in one part is to a great extent independent of the folding in the other part. Figure 2 shows the α-carbon chain and the binuclear nature of the molecule can be distinguished easily. Apparently the principle behind the construction of this molecule is the following.

We assume that folding starts at the N-terminus. The chain first establishes a hydrophobic core which is then completely surrounded by other residues to form the first part of the molecule:

the process then repeats itself with the remaining 100 residues. There is one feature in this construction which could have some significance: each hydrophobic core starts with an α-helix. At residue 24 the first α-helix begins and one side of this five-turn helix is the nucleus, the starting point, for the core of part I. It goes as far as 43. A two-turn α-helix runs from 50 to 58. This one is connected to the previous α-helix by means of a link between Glu-50 and residue 35. This latter residue is one of the thirteen residues the identity of which has been tentatively derived from X-ray work. After a further twelve residues, a third α-helix of three turns completes this

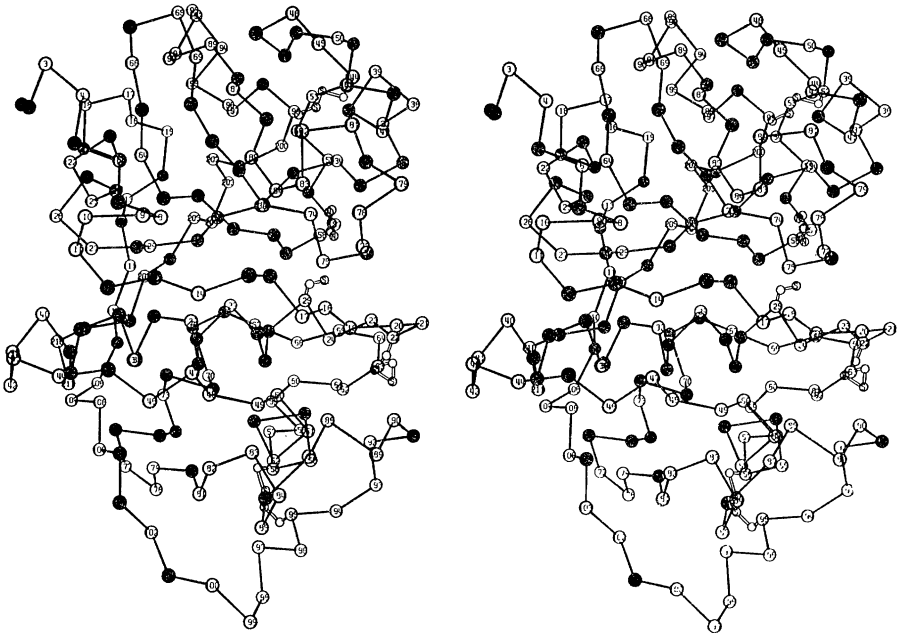

FIGURE 2. Stereoscopic drawing showing the positions of the α-carbon atoms in the papain molecule. Hydrophobic residues have a black dot. The picture was calculated by a computer, programmed by Dr R. E. Dickerson, and will appear in *The structure and action of proteins* by R. E. Dickerson & I. Geis to be published by Harper and Row. The additional residue Val-130 (mentioned in legend of figure 1) has not been incorporated.

hydrophobic core. The chain then adds (from residue 82) a mainly hydrophilic loop to complete this first part of the molecule. At residue 111 the chain moves to part II. It is still hydrophilic but then turns into another short α-helix beginning at residue 117 which is hydrophobic on one side and this is the nucleus for the core of part II of the molecule. After this helix is finished at residue 126, the chain continues as an extended chain which adds some hydrophobic residues to the core. From residue 160 it is completely hydrophobic. The residues 164 to 173 form a β-sheet of two antiparallel chains which is hydrophobic internally and hydrophilic externally. The β-sheet acts as a wall of the hydrophobic core. The last part of the chain is hydrophobic where it contributes to the core of part II of the molecule. The centre of the core of part I is surrounded by 14 methyl groups. Part II has 16 methyl groups around its core. Now the molecule is complete to fulfil its function as a proteolytic enzyme. Between the two parts is a cleft, containing the SH group of Cys-25, which constitutes the active site.

The two parts of the molecule are connected by a variety of bonds. First is the covalent bond in the main chain where it crosses from part I to part II near residue 111. Second is the binding of the first 10 residues to part II. These residues belong covalently bonded to part I but

are geographically placed against part II and attached to it mainly by two hydrogen bonds from Val-5 to the main chain in the β-sheet. Also the side group of Val-5 is directed to the hydrophobic core of part II. The third connexion between parts I and II is at the end of the chain when it leaves part II. Pro-209 has hydrophobic interactions with side groups of part I. An additional hydrophobic interaction occurs between Trp-69 and Phe-207. Altogether the connexion between parts I and II looks too strong to allow a movement of one half of the molecule with respect to the other and we have no indication that this occurs. It would be a nice experiment if we could selectively split a bond in the main chain near residue 111, separate the two parts and crystallize them as two different proteins. We then could check whether or not the chains in the two parts do fold independently of each other. The complete molecule has three disulphide bridges, that between residues 56 and 95, is right handed and those between residues 22 and 63 and 153 and 200, are left handed. Summarizing the architecture of the molecule we can say: It is constructed with two hydrophobic cores to which contribute the internal parts of four short α-helices and a small pleated sheet as well as the internal parts of pieces of extended chain.

The structure of the papain molecule as presented here has been determined with crystals grown from methanol–water (2:1). What would be the structure in water? We have no direct proof but there is strong evidence that no or at least no major conformational change takes place going from the alcoholic medium to water. The evidence is based on crystallographic results and on o.r.d. (Drenth, Hol, Visser & Sluyterman 1968*a*). They found a crystal form which was grown from a water-rich medium as well as from an alcoholic medium. O.r.d. curves down to 220 nm are exactly the same for the alcoholic medium and for water. They indicate the same low helix content as can be derived from the model, about 22 %. We may well assume that our picture of the molecule and of the active site is unchanged in water.

4. The active site

The groups which probably play a role in the catalytic mechanism are drawn in figure 3. The free SH group of Cys-25, together with the imidazole group of His-159, are responsible for the splitting of the substrate. This imidazole group is hydrogen bonded by its N3 atom to the carbonyl group of the Asn-175 side chain. Asp-158 and perhaps Asp-64 increase the affinity of the active site for positively charged groups and diminish it for negatively charged ones. The pK of their carboxyl group is about 4 and above this pH value negatively charged groups are prevented from entering the active site region. Asp-158 may well bind an arginine or lysine side chain of the substrate. This explains the slight preference of papain for substrates with basic groups.

Although the structure of this proteolytic SH enzyme is different from the serine proteinases chymotrypsin (Matthews, Sigler, Henderson & Blow 1967) and subtilisin (Schubert Wright, Alden & Kraut 1969) the enzymes have common features. Their active sites have either an SH or an OH in addition to a histidine residue which is connected to aspartic acid in chymotrypsin and to asparagine in papain.

A point which should be made perfectly clear is that the electron density map was calculated using the structure factors of crystals in which the papain molecules have their sulphydryl group blocked. This means that although the map shows the sulphur atom of Cys-25 this is not present as an SH group since, in some molecules, it is connected to another cysteine molecule in a

disulphide bridge and in others it is blocked in some other way. Electron density features resembling those of cysteine were indeed found close to the sulphur position in the active site region. No conclusion could be drawn about the irreversible blocking of the sulphydryl group. At the position of the sulphur atom of Cys-25 the map shows the highest electron density indicating that the sulphur atom has a well defined single position in the X-ray structure which is at Van der Waals distance from nitrogen N1 of the imidazole ring of His-159.

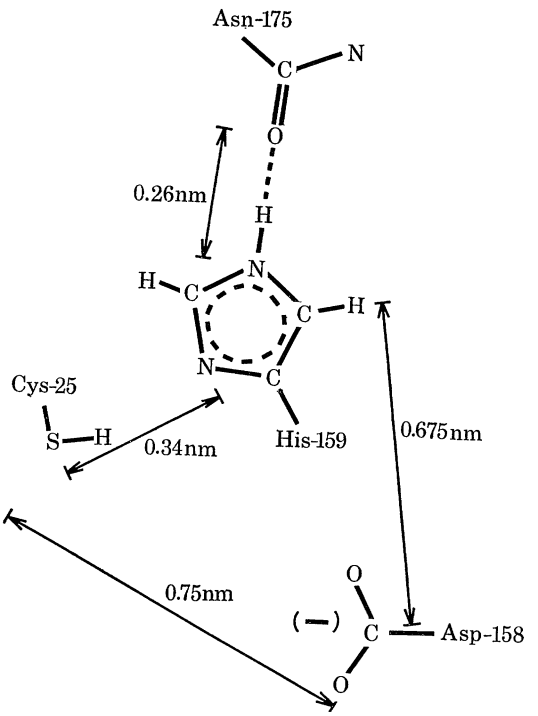

FIGURE 3. The active site of papain.

We do not know the position of the sulphur atom when the molecule contains the free sulphydryl group. If we assume that it remains in the same position when unblocked then no orbital overlap is possible between the sulphur and the imidazole ring. Any interaction should be due to hydrogen bonding, either S–H ... N or S_H ... H–N.

The first system, in which SH is the hydrogen donor, seems to be extremely weak or absent in water (Edsall 1965). The latter system is found in crystal structures (Hordvik 1963; Hamilton & Ibers 1968). Any hydrogen bond involving S seems to be very weak and thus a strong interaction between the S atom and the imidazole ring is unlikely. We do not need this interaction for giving a plausible explanation of the catalytic mechanism.

The reaction mechanism we propose (Sluyterman & Wolthers 1968) is given in figure 4. In the acylation step the reaction proceeds to the right. In deacylation water replaces the leaving group and the reaction goes to the left. The C=O bond of the substrate is polarized by the positive charge on the imidazolium ion. This allows the sulphur atom to join with the carbonyl carbon. Its proton is transferred to the oxygen of the alcohol group. The pH dependence of k_{cat} for the papain catalysed hydrolysis of *N*-α-benzoyl-L-arginine ethyl ester is governed by a group with a p*K* of about 9. This could very well be the sulphydryl group which is known to

ionize near this pH value. Acylation as well as deacylation will not take place with a deprotonated histidine. But this requires the unusually high pK value for this histidine of at least 9.5 to 10. In the model compound Gly-His-Gly the imidazole group has a pK of 7.5. However, from studying the molecular model we can give at least two reasons which could appreciably raise the pK of His-159. (*a*) The negatively charged carboxyl group of Asp-158 at a distance of 0.675 nm from the imidazole group has an electrostatic effect which will be enhanced by the

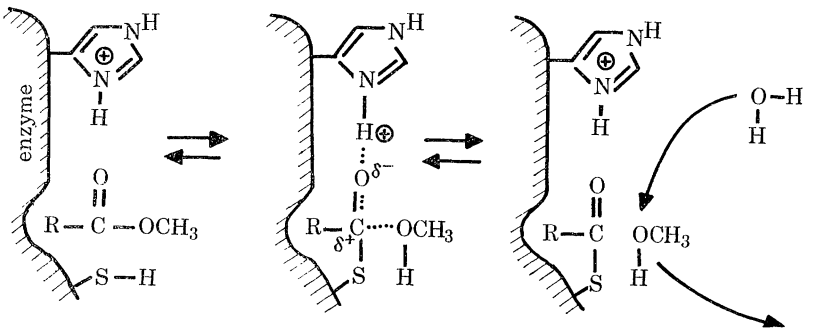

FIGURE 4. Proposed reaction mechanism of papain.

partially non-polar environment. Since $\Delta pK = 240/DR$, in which D denotes the dielectric constant and R the distance, a mean D value of 40 accounts for an increase in pK of about one unit. (*b*) Hydrogen bond formation between the side group oxygen of Asn-175 and the N3 atom of His-159 could be strengthened by the hydrophobic environment which involves the side groups of two tryptophanes, a phenylalanine and one or two leucines and a valine. Such a hydrogen bond could very well increase the pK by about two units.

Evidently these effects are quite sufficient to increase the pK of the imidazole group in the enzyme into the 9.5 to 10 range. Then, in the optimum pH range, the imidazole group is present as an imidazolium ion and is capable of a kind of acid catalysis as proposed in figure 4.

A part of the present investigation has been carried out under the auspices of the Netherlands Foundation for Chemical Research (S.O.N.) and with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

REFERENCES (Drenth *et al.*)

Arndt, U. W., North, A. C. T. & Phillips, D. C. 1964 *J. scient. Instrum.* **41**, 421.
Drenth, J., Hol, W. G. J., Visser, J. W. E. & Sluyterman, L. A. A. 1968*a* *J. molec. Biol.* **34**, 369.
Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Wolthers, B. G. 1968*b* *Nature, Lond.* **218**, 929.
Edsall, J. T. 1965 *Biochemistry* **4**, 28.
Glazer, A. N. & Smith, E. L. 1965 *J. biol. Chem.* **240**, 201.
Hamilton, W. C. & Ibers, J. A. 1968 *Hydrogen bonding in solids*, p. 167. New York and Amsterdam: Benjamin.
Hordvik, A. 1963 *Acta Chem. Scand.* **17**, 2575.
Kimmel, J. R. & Smith, E. L. 1954 *J. biol. Chem.* **207**, 515.
Light, A., Frater, R., Kimmel, J. R. & Smith, E. L. 1964 *Proc. natn. Acad. Sci. U.S.A.* **52**, 1276.
Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. 1967 *Nature, Lond.* **214**, 652.
Phillips, D. C. 1964 *J. scient. Instrum.* **41**, 123.
Schubert Wright, C., Alden, R. A. & Kraut, J. 1969 *Nature, Lond.* **221**, 235.
Sluyterman, L. A. A. 1967 *Biochim. biophys. Acta* **139**, 430.
Sluyterman, L. A. A. & Wolthers, B. G. 1968 *Proc. K. ned. akad. Wet.* (Ser. B) **72**, 14.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES