

Structure of Actinidin, after Refinement at 1·7 Å Resolution

E. N. BAKER

*Department of Chemistry, Biochemistry and Biophysics
Massey University, Palmerston North
New Zealand*

(Received 18 February 1980)

The structure of the sulphhydryl protease, actinidin, after refinement at 1·7 Å resolution, is described. The positions of most of the 1666 atoms have been determined with an accuracy better than 0·1 Å; only two residues (219 and 220) and the side-chain of a third (87) cannot be seen. In addition, the model contains 272 solvent molecules, all taken as water, except one which may be an ammonium ion. Atomic *B* values give a good indication of the mobility of different parts of the structure. Actinidin has a double domain structure, with one domain mostly helical in its secondary structure, and the other domain built around a twisted β -sheet. The geometry of hydrogen bonds in helices, β -structure and turns has been analysed. All are significantly non-linear, with the angle N–H...O $\sim 160^\circ$. Carbonyl groups are tilted outwards from the axis of each helix, the tilting apparently unaffected by whether or not additional hydrogen bonds are made (e.g. to water or side-chain atoms). Each domain is folded round a substantial core of non-polar side-chains, but the interface between domains is mostly polar. Interactions across this interface involve a network of eight buried water molecules, the buried carboxyl and amino groups of Glu35, Glu50, Lys181 and Lys17, other polar side-chains and a few hydrophobic groups. One other internal charged side-chain, that of Glu52, is adjacent to a buried solvent molecule, probably an ammonium ion. Other side-chain environments are described. One proline residue has a *cis* configuration. The sulphhydryl group is oxidized, probably to SO₂[−], with one oxygen atom clearly visible but the other somewhat less certain. The active site geometry is otherwise compatible with the mechanism proposed by Drenth *et al.* (1975, 1976) for papain. The positions of the 272 solvent molecules are described. The best-ordered water molecules are those that are internal (total of 17), in surface pockets, or in the intermolecular contact regions. These generally form three or four hydrogen bonds, two to proton acceptors and one or two to proton donors. Other water molecules make water bridges on the surface, sometimes covering the exposed edges of non-polar groups. Intermolecular contacts involve few protein atoms, but many water molecules.

1. Introduction

Actinidin is a proteolytic enzyme obtained from the fruit of the Chinese gooseberry or Kiwifruit, *Actinidia chinensis*. It has a free sulphhydryl group, which is essential for activity (Arcus, 1959; McDowall, 1970), and has therefore been grouped with other sulphhydryl proteases, such as the plant enzymes papain, ficin and bromelain,

the bacterial enzyme streptococcal proteinase, and the mammalian lysosomal enzymes, cathepsins B1 and B2 (Glazer & Smith, 1971; Lowe, 1976).

The three-dimensional structure of actinidin was determined from an X-ray crystallographic analysis at 2.8 Å resolution (Baker, 1977). This showed that the polypeptide chain conformation was virtually identical to that of papain (Drenth *et al.*, 1971), at least within the accuracy of the analysis. The structure could not be reported in full detail, however, because complete amino acid sequence information was not then available. The amino acid sequence determination has now been completed (Carne & Moore, 1978), and has revealed that, allowing for a small number of insertions and deletions, about 48% of the residues are the same as the corresponding ones in papain (Mitchel *et al.*, 1970). Furthermore, the X-ray analysis has been extended to 1.7 Å resolution, and crystallographic refinement (Baker & Dodson, 1980) has led to a very accurate description of the protein structure (accuracy of atomic positions <0.1 Å) and identification of many water molecules associated with the protein.

The purpose of the present paper is to describe details of the structure of particular interest, paying special attention to (1) the general organization of the molecule and the interactions that contribute to its stability and/or activity, (2) the hydrogen-bonding geometry within elements of secondary structure (since only after refinement are hydrogen bond distances and angles reliable enough), and (3) the roles played by solvent molecules in their interactions with the protein and with each other. A detailed comparison of actinidin and papain will be made later, when refinement of the structure of papain, at present in progress, is complete.

2. Experimental

(a) Crystals and data collection

Actinidin was prepared as described by Carne & Moore (1978), but crystals were not grown as for the 2.8 Å analysis, where actinidin was crystallized as the tetrathionate-inhibited enzyme, and the inhibitor then removed by washing crystals in dithioerythritol (Baker, 1976, 1977). Instead, in order to ensure that no tetrathionate remained bound, the inhibitor was removed in solution, before crystallization, and the free enzyme crystallized. Crystallization conditions were otherwise identical to those used previously (Baker, 1973), and unit cell data were unchanged.

For the 1.7 Å analysis and refinement a new set of diffractometer data (24,157 independent reflections, including "unobserved" reflections) was collected, with most reflections being measured twice. A total of 5 crystals were used. None of the data used for the 2.8 Å analysis was incorporated in this data set.

(b) Refinement

Most of the technical details of the refinement of the structure have been described elsewhere (Baker & Dodson, 1980). The methods used, and some of the results, must briefly be restated, however, as they are important to a proper evaluation of the structure.

The refinement method used was that of fast Fourier least-squares refinement, developed by Agarwal (1978) and first applied by Isaacs & Agarwal (1978) in the refinement of insulin. Atomic positions (x , y and z co-ordinates) were refined independently, and after the first 7 cycles of refinement, when the crystallographic R value† had been reduced to 0.29, individual isotropic temperature factors (B values) were also refined. The protein

† $R = \Sigma |F_o| - |F_c| / \Sigma |F_o|$.

structure was regularized at intervals (i.e. constrained close to standard geometry) using the model-fitting routines of Dodson *et al.* (1976). Non-bonded contacts were not constrained in any way, however, and no energy calculations were incorporated.

Solvent molecules (all regarded as water) were added at intervals, as they appeared in difference maps, but only following rather selective criteria, namely that they should have persisted through several difference maps, be remote from parts of the protein structure still in doubt, and be within approximate hydrogen-bonding distance of suitable protein atoms or other solvent molecules.,

The final R value for a regularized structure (root-mean-square deviations 0.01 Å from standard bond lengths and 3.0° from standard bond angles) was 0.171 for all 23,990 reflections between 10 Å and 1.7 Å resolution (or 0.152 if reflections with $I < 2\sigma_I$ were omitted). This model consisted of 1657 protein atoms (24 of partial occupancy) and 163 water molecules.

More recently, hydrogen atoms have been added, in their calculated positions, and this has reduced R to 0.168 (0.149 with a 2σ cut-off of data). Furthermore, a close examination of the final difference electron density map, and of the existing water structure, suggested that further water molecules should be added. While "water molecules" could be added almost indefinitely (and probably often erroneously) to account for low-level density, it was clear that many of the low-level peaks (with heights 0.3 to 0.4 e/Å³) were in positions where they made the water structure more complete, hydrogen-bonding to, and often bridging between, existing water molecules and/or protein atoms. Thus a further 109 water molecules have now been added; they have been given B values of 50 Å², but have not been refined.

The final model† consists of 1657 protein atoms (24 of partial occupancy) and 272 water molecules (163 refined, 109 unrefined). The R value is 0.165 (0.146 after reflections with $I < 2\sigma_I$ are omitted).

(c) Accuracy

Standard deviations in atomic positions have been derived from the diagonal elements of the inverse least-squares matrix and the spread of values is shown in Table 1, as a function of the atomic B values. For most of the protein structure (~1350 atoms) B values are less than 15 Å², and the overall estimated standard deviation is 0.055 Å. This agrees well with a value of ~0.07 Å, estimated from the size of shifts in least-squares cycles near the end of refinement, and with a *maximum* value of ~0.10 Å derived from the variation of R as a function of resolution (Luzzatti, 1952). For the refined water molecules (W1 to W163), with B values 7 to 40 Å², positional standard deviations will be slightly higher, ~0.05 to 0.15 Å.

Since the final protein model has been constrained close to standard geometry, the standard deviations in bond lengths and angles given in Table 1 reflect mainly the tightness of constraints; they are derived from the spread of values after regularization. Waterpaugh *et al.* (1979), following their experience in refinement of rubredoxin, emphasize the need for caution in using idealized bond-angle parameters in model refinement. The restraints used for actinidin, however, do allow for reasonable deviations from standard values. The r.m.s. deviation of 3.2° is of the same order as is found commonly in small molecule structures, and individual deviations of up to 10 to 12° occur. The only parts of the protein structure not regularized are the 3 disulphide bridges, which therefore offer an independent check on the validity of the refinement. Encouragingly, then, these are found to have very similar geometry to small organic disulphide molecules (Shefter, 1970). S—S bond lengths are 1.97, 2.01 and 1.96 Å, C—S—S angles range from 101.0° to 108.3°, and the dihedral angles (C—S—S—C) are 90.8°, 84.5° and 96.0°.

Non-bonded contacts have not been constrained during refinement. The error in these is therefore derived from the standard deviations in atomic positions, and varies from

† Co-ordinates and B values have been deposited with the Protein Data Bank, Brookhaven National Laboratory, Upton, N.Y. 11973, U.S.A.

TABLE 1
Standard deviations

A. *In atomic positions (Å)*

	$B =$	0-5	5-10	10-15	15-20	$> 20 \text{ Å}^2$	Overall
From inverse least-squares matrix		0.041	0.050	0.063	0.073	0.081	0.055
From variation of R with resolution						maximum 0.10	

B. *In bond lengths and angles (from spread of values after regularization)*

	Average $B =$	0-5	5-10	10-15	15-20	$> 20 \text{ Å}^2$	Overall
C—C bonds		0.007	0.007	0.008	0.010	0.012	0.008 Å
C—N bonds		0.005	0.006	0.007	0.009	—	0.006 Å
C—O bonds		0.004	0.004	0.004	0.004	—	0.004 Å
Bond angles		2.8	3.0	3.7	4.6	5.0	3.2°

C. *In non-bonded contacts*

Between internal atoms ($B = 5-15 \text{ Å}^2$)	$\sim 0.1 \text{ Å}$
Between external protein atoms ($B = 10-30 \text{ Å}^2$)	$\sim 0.14 \text{ Å}$
Between water molecules ($B = 10-50 \text{ Å}^2$)	$\sim 0.1-0.3 \text{ Å}$

0.1 Å (e.g. for hydrogen bonds or contacts between internal atoms) to $\sim 0.3 \text{ Å}$ (e.g. for distances involving the less well-defined water molecules). Similarly errors in hydrogen bond angles and torsional angles (ϕ, ψ) range from 5° to a maximum of $\sim 10^\circ$.

While the above standard deviations apply to most of the structure, there is also an *uncertainty* associated with some residues. Some side-chains (e.g. those of Ser3, Tyr4, Glu21, Asp97, Tyr130, Asp142, Lys145, Gln146, Glu172 and Glu191, all external, projecting into the surrounding liquid) have very weak density and correspondingly high B values (see Fig. 1). They are clearly very flexible and the orientation seen is probably only the major one of a variety of possible orientations. Static disorder is seen for 5 side-chains (Ser9, Ser42, Ser44, Arg58 and Thr129) and probably exists in a less well-defined way for others (e.g. Met211). B values (see Fig. 1) probably give the best guide to the reliability of different parts of the structure. All unrefined water molecules (W164 to W272) should be treated with caution.

For one side-chain (that of Glu87), no density is apparent beyond C_β , and the two C-terminal residues, Asn219 and Asn220, are not included in the model at all. Some very weak density was seen beyond residue 218 in the final difference map but it was not considered reliable enough and these 2 residues must be regarded as disordered.

(d) *Positional changes resulting from refinement*

The average shift in main-chain atom positions (from those obtained from the 2.8 Å map) was 0.50 Å (median 0.45 Å). Although no peptide units had to be turned around completely (180°), 3 were twisted through more than 60° (those of 8 to 9, 95 to 96 and 151

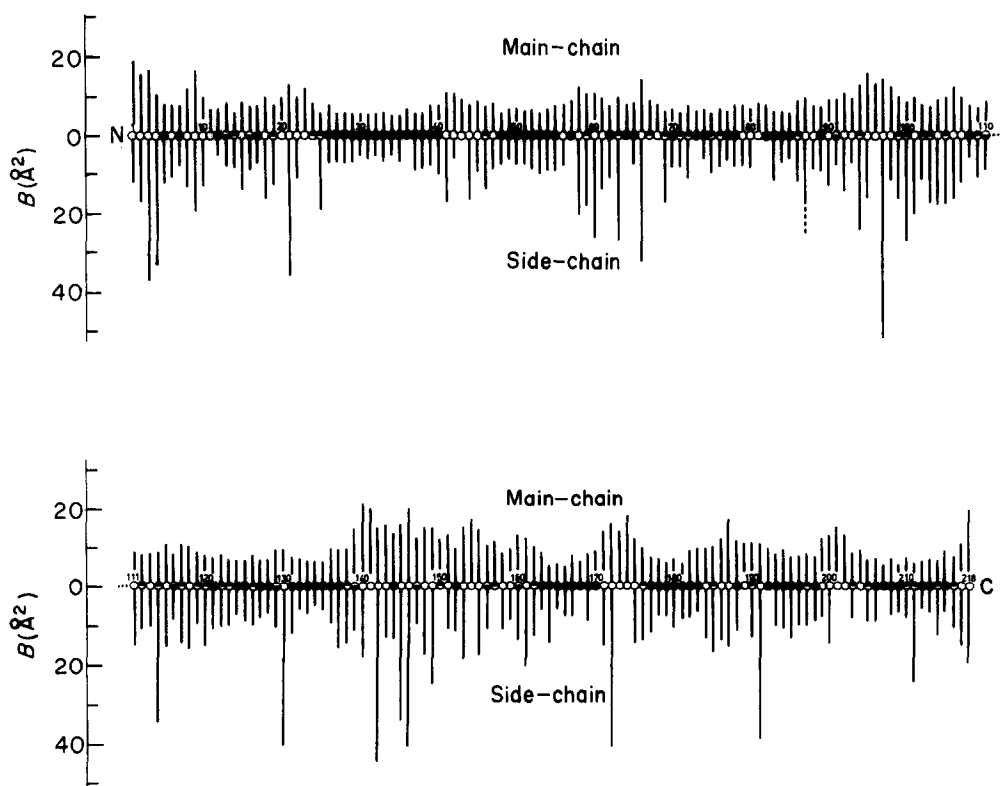


Fig. 1. A plot of B values for all residues, showing the flexibility of different parts of the structure. Average values for main-chain atoms (N, C $_{\alpha}$, C, O) and side-chain atoms (C $_{\beta}$ and beyond) are plotted for each residue. Residues totally exposed to the external solution are shown (○), those with only side-chain exposed (◐), those with only main-chain atoms exposed (◑), and those totally buried (main-chain and side-chain) (●). (Note that the mean amplitude of vibration of an atom, $\bar{U}$, is related to the B value by $B = 8\pi^2\bar{U}^2$, so that $B = 20 \text{ Å}^2$ corresponds to a mean amplitude of vibration of 0.5 Å .)

to 152), while others were moved sufficiently to change the hydrogen-bonding pattern originally deduced.

The average shift for side-chain atoms was 0.84 Å (median 0.54 Å) including the contributions of 19 side-chains which had to be manually rebuilt into quite new orientations (Glu21, Thr33, Ile37, Lys39, Thr59, Arg63, Asp66, Gln75, Ile78, Gln94, Val100, Ile110, Glu114, Gln125, Asp138, Ile166, Thr171, Ile208 and Met211). These gross changes arose from uncertainties or misinterpretations of the original 2.8 Å map. In addition, a number of atoms were shifted quite large distances ($>1 \text{ Å}$) during the automated least-squares cycles.

(e) Amino acid sequence changes

The refinement has required one definite change to the amino acid sequence; residue 86, identified chemically as Asp (Carne & Moore, 1978), proved to be clearly one carbon atom longer, fitting perfectly as Gln or Glu. It has been taken as Glu, to preserve the same charge as in the chemical sequencing. If this is so, one of the carboxyl oxygens, O $_{\epsilon 2}$ is only 2.5 Å from O $_{\epsilon 2}$ of Glu50, suggesting that they may perhaps share a proton, with one Glu having

a high pK value. (The pH of solution is 5.9.) This is found to occur in lysozyme where 2 adjacent Glu residues result in a pK value of 5.9 for Glu35 (Parsons & Raftery, 1970).

Three other residues could also represent sequence errors or variations (see Table 2). Ser42 appears to be disordered, with 2 alternative positions for O_γ , but residual positive density in the positions of both O_γ and O'_γ (despite low B values) suggests it could actually be Thr. (Note that in papain residue 42 is Thr.) Ala101 had density beyond C_β in the 2.8 Å map. During refinement this extra density weakened but it was later found that a water molecule (W92) had moved to be only 1.8 Å from C_β . On removal of W92, the density persisted. Either residue 101 should be Ser, or there could be genetic variations at this point with some molecules containing Ala101 and some Ser101 (note that this is the most variable part of the sequence, compared with papain). Finally, Ile164 could possibly be Val. Throughout refinement, $C_{\gamma 1}$ and $C_{\delta 1}$ tended to move together, B values are high for an internal side-chain, and $C_{\delta 1}$ makes several close contacts (3.1 to 3.2 Å) with other side-chains. No alternative Ile conformation could be found, and a more satisfactory fit may be as Val.

TABLE 2

Possible sequence changes suggested by refinement

Residue number	Chemically determined identity	Identity from X-ray work	Comments
42	Ser	Thr	Extra peak adjacent to C_β . Either disordered Ser (with 2 positions for O_γ) or Thr
86	Asp	Glu	Side-chain definitely 1 carbon atom longer. Fits perfectly as Glu
101	Ala	Ser	Weak peak 2 Å from C_β . Could be genetic variation here?
164	Ile	Val	Refines poorly. $C_{\delta 1}$ moves too close to $C_{\gamma 1}$, but no alternative conformation. $C_{\delta 1}$ some close contacts.

(f) B values

Several authors (Frauenfelder *et al.*, 1979; Artymiuk *et al.*, 1979) have commented recently on the significance of atomic temperature factors (B values) in indicating the flexibility of parts of a protein molecule. The refinement method used for actinidin is particularly suitable for refining B values, and although the latter have not yet been extensively analysed it is clear that they hold considerable physical significance.

To provide an objective assessment, Table 3 lists the B values of all Lys and Phe side-chains. For the latter, the individual B values are reasonably consistent within each group. The Lys side-chains show the effects of different environments. Lys17 and Lys181 are internal, constrained by hydrogen bonds and salt linkages, and their B values are low. Lys39 is confined in a surface pocket, and forms 3 or 4 H bonds (through $N\zeta^\dagger$); its B values are also low. Lys106 and Lys217 are more exposed (although in each case $N\zeta$ forms 3 H bonds) and B values are consistently higher, while Lys145 projects right out from the surface of the molecule and has high B values, increasing to ~ 50 Å² at the end of the side-chain.

† The IUPAC-IUB (1970) nomenclature conventions are generally used throughout this paper. In Tables, and occasionally in the text, the Latin equivalents of the Greek subscripts are used for clarity (e.g. $N\zeta$ for N_ζ , CE for C_ϵ). Main-chain carbonyl groups are designated CO(123) etc., and main-chain carbonyl oxygen atoms 123 O or O(123). Main-chain amide nitrogens are identified NH(123) or 123 N.

TABLE 3
B values for all Lys and Phe side-chains

Lys	CB	CG	CD	CE	NZ		
17	6.6	7.8	6.0	5.8	7.6		
39	8.0	5.2	8.7	7.1	9.2		
106	10.2	12.4	13.4	24.2	21.7		
145	16.3	22.8	33.8	50.0	48.3		
181	4.5	7.2	9.8	12.8	8.0		
217	12.7	16.4	18.7	15.4	14.0		
Phe	CB	CG	CD1	CE1	CZ	CE2	CD2
28	6.8	5.4	8.4	9.5	8.5	6.1	6.3
74	5.0	5.8	5.2	8.7	6.3	5.3	8.5
76	4.8	6.7	9.1	9.0	7.0	9.2	9.6
144	9.9	9.0	13.2	15.6	13.2	18.9	13.7
152	12.5	10.4	10.0	9.3	15.2	12.8	11.3

The average *B* values for the main-chain and side-chain atoms of each residue are plotted in Fig. 1. Main-chain atoms involved in helices (e.g. 24 to 43, 69 to 81, 50 to 57 and 120 to 130) and strands of β -sheet have low *B* values, while more loosely held regions have higher values (see e.g. the external loops 92 to 100 and 139 to 160, and the exposed turn, 171 to 175, between strands C and D of the β -sheet). Internal side-chains have much lower *B* values than external ones (except where the latter are involved in specific interactions such as H bonds or ion pairs). In many peptide units the carbonyl oxygen atom has a higher *B* value than the other atoms, indicating some torsional flexibility.

3. The Structure

(a) General organization

The overall polypeptide chain conformation of actinidin is shown in Figure 2, where α -carbon positions are plotted. The molecule has a clearly defined double domain structure, with the active site lying between the two domains on the top surface. Domain I (residues 19 to 115 and 214 to 218) consists principally of residues from the first half of the chain and domain II (1 to 18 and 116 to 213) residues from the second half. The N-terminal and C-terminal ends act as "straps" crossing over

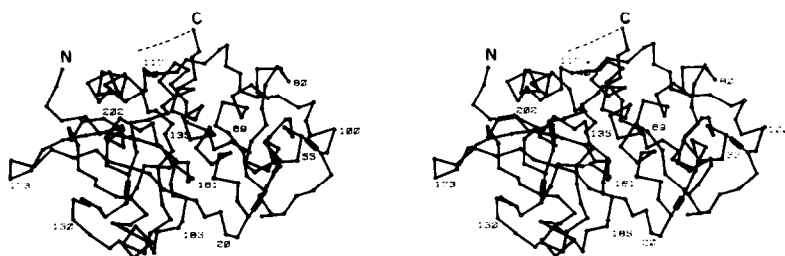


FIG. 2. A plot of α -carbon atom positions for actinidin. The N-terminal and C-terminal ends are shown. The broken line represents the position of weak density which could be the 2 disordered C-terminal residues, 219 and 220. The α -carbon atoms of Cys25 and His162 are shown as filled circles. The 3 disulphide bridges, 22-65, 56-98 and 156-206 are shown (—■—).

from one domain to the other to help hold the two halves of the molecule together. Interactions involving these straps are detailed in Table 4. The N-terminal strap is wrapped around the surface of domain II, while the C-terminal strap is inserted between residues 113 to 116 and the rest of domain I (so that Pro215 projects into the non-polar core of domain I while hydrogen bonds are made with residues 111 to 114 on the outside).

TABLE 4
Interactions involving the N-terminal and C-terminal straps

<i>A. Between residues 1 to 17 and the rest of domain II</i>				
H-bonds:	2 O...169 OH	3.20 Å	5 N...169 O	2.90 Å
	7 NE1...128 O	2.82 Å	5 O...169 N	2.85 Å
	14 O...181 NZ	2.70 Å	7 N...167 O	2.83 Å
			17 O...185 N	3.11 Å
Hydrophobic interactions:	Trp7...Pro132			(3.5-4.0 Å)
	Val13...Lys181 (CD, CE)			(3.8-4.2 Å)
	Ile16...Tyr193			(3.7-4.0 Å)
<i>B. Between residues 214 to 218 and the rest of domain I</i>				
H-bonds:	215 O...131 ND2	2.88 Å	214 N...114 O	3.00 Å
	217 NZ...109 O	2.91 Å	214 O...114 N	2.85 Å
	218 N...111 OD1	2.81 Å	216 N...112 O	2.82 Å
			216 O...111 N	2.78 Å
Hydrophobic interactions:	Pro215...Val34	(~4 Å)		
	Pro215...Phe74	(3.7-4.0 Å)		
	Pro215...Ile110	(~4.5 Å)		

Domains I and II could clearly fold up independently of each other. Residues 1 to 17 are probably laid over the surface of domain II *after* the latter has assumed its final folded structure; several internal water molecules lie between 1 to 17 and the rest of domain II. Probably these bind to domain II during the folding process and are then trapped when the N-terminal section is laid over the top. Residues 214 to 218 must, however, become associated with domain I *before* it assumes its final folding, since they must tuck inside residues 113 to 116. Residues 113 to 116 are part of a piece of extended chain (105 to 120) and it seems likely that this remains somewhat "floppy" until the C-terminal end has associated with the rest of domain I.

Both domains have a congregation of non-polar side-chains in their interior (see later), but the interface between the two, lying below the active site, is largely polar, made up of mostly polar side-chains and a network of buried water molecules. (This is also discussed later.)

There are three disulphide bridges (22 to 65, 56 to 98 and 156 to 206); all have normal geometry (see above, Experimental, section (c)).

(b) *Hydrogen bonding*

Hydrogen bonds involving protein atoms are shown schematically in Figure 3 (for main-chain atoms) and listed in Table 5 (for side-chain atoms). In both cases hydrogen bonds are assumed if the donor-acceptor distance is less than 3.2 Å and bond angles are reasonable. A few approaches of 3.2 to 3.4 Å which may represent weak attractions are also noted.

"Lost" hydrogen bonds (where main-chain amido or carbonyl groups neither interact with solvent nor form hydrogen bonds) have been noted before, e.g. in

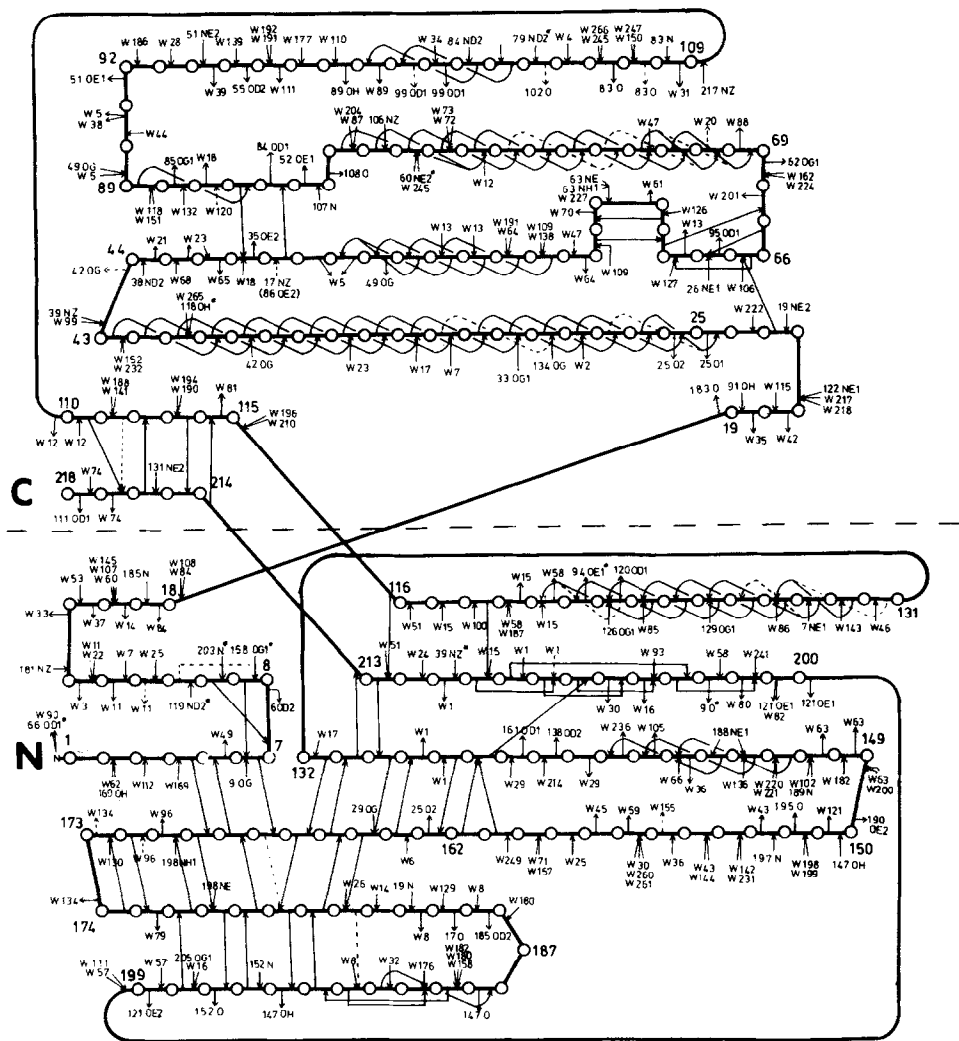


FIG. 3. A schematic diagram showing all hydrogen bonds formed by main-chain atoms. Possible hydrogen bonds for which the distance is rather too long (3.3 to 3.5 Å) are shown as broken lines. Intermolecular hydrogen bonds are designated *.

TABLE 5

Hydrogen bonds made by polar side-chains

Ser3	OG	Gln75	OE1...W34, (79 OD1)	Gln131	OE1...W181
Tyr4	OH...6 OD1, W116	Asn79	NE2...W40, W135	Ser134	NE2...215 O, W17
Asp6	OD1...4 OH, 8 NH2, W122	Asp80	OD1...(75 OE1)	Asp138	OG...29 O, 29 OG, 33 OG1
Trp7	OD2...8 N, 8 NE	Asn84	ND2...W97, 104 O	Asp142	OD1...W71, W59, W264
Arg8	NE1...128 O	Thr85	OD1...(W4)	Lys145	OD2...140 N, W59, W71
	NE...6 OD2	Glu86	OD2...W91, W244	Gln146	OD1
	NH1...W176	Glu87	OD1...85 N, 88 ND2		OD2...W236, W237
Ser9	NH2...6 OD1, W156	Asn88	ND2...102 O, 105 OE1		NZ
Asp15	OG...6 O, W164	Tyr89	OE1...88 N, W54		NE2
	OD1...W22, W264	Thr91	OE1...W5, W38	Tyr147	OH...150 O, 195 N
	OD2...W71, W250, W264	Thr92	OE2...17 NZ, 48 O, 50 OE2	Ser149	OG...W237
Lys17	NZ...35 OE2, 47 OG, 50 OE1, 86 OE2	Gln94	OE1...63 NH2†, W138	Thr153	OG1...W43
Gln19	OE1...183 OG	Asp95	ND2...61 OD1†, 84 OD1	Thr158	OG1...8 O†, W11
	NE2...22 O, 25 O1	Asp97	OH...55 OD1, 89 N	Asp161	OD1...139 N
Glu21	OE1...W219	Asp99	OD1...65 N		OD2...W71
	OE2	Gln103	OD2...W27, W127	His162	ND1...25 SG
Cys25	O1...25 N, 19 NE2, W216, W223	Asp104	OD1...W131		NE2...182 OD1
	O2...163 N, W212, W222	Gln105	OD2...W111, W131	Tyr169	OH...129 OG1, 2 O
Trp26	NE1...64 O		OD1...102 N	Thr171	OG1...176 OD1, W112
Ser29	OG...134 OG, 164 O		OD2...W175	Glu172	OE1
Thr33	OG1...30 O, 134 OG		OE1		OE2...(195 NH2)
Glu35	OE1...W3, W9		NE2	Asp176	OD1...171 OG1, W206
	OE2...17 NZ, 48 N		NE2	Tyr177	OH...W116
Asn38	OD1...W31		NE2	Trp178	NE1...121 OE2
	ND2...W23, W48, 44 O		NE2	Lys181	NZ...14 O, W3, W9
Lys39	NZ...W21, 43 O, 210 O†, 210 OG1†		NE2	Asn182	OD1...162 NE2
Thr41	OG1		NE2		ND2
Ser42	OG...38 O, 44 N, W48		NE2	Ser183	OG...19 OE1, W6
Ser44	OG		NE2	Trp184	NE1...W77
Ser47	OG...17 NZ, W55		NE2		
Ser49	OG...52 N, 89 O, W5		NE2		

Glu50	OE1...17 NZ, W2	Gln105	NE2...107 OH	Asp185	OD1...W183, W185
	OE2...86 OE2, W76	Lys106	NZ...60 OE1, 80 O, W56, (59 OG1†)		OD2...186 N, W107
Gln51	OE1...92 N	Tyr107	OH...105 NE2, W73	Thr186	OG1...W163
Glu52	NE2...55 OD2, 94 O, W146	Thr109	OG1	Thr187	OG1...W129, W183
	OE1...84 N	Asp111	OD1...218 N, W173	Trp188	NE1...144 O, 145 O
	OE2...W4		OD2...W228	Glu190	OE1...195 NE, W83
Asp55	OD1...W78, 89 OH	Thr112	OG1...W149, W190	Glu191	OE2...150 N, 195 NH1
	OD2...51 NE2, 96 N	Tyr113	OH...W40, W207, W209		OE1
Arg58	NE	Glu114	OE1...W196		OE2
	NH1...W192	Asn115	OD1...W52	Tyr193	OH...W41
	NH2...W127		ND2	Arg195	NE...190 OE1
Thr59	OG1...W87, 106 NZ†	Tyr118	OH...W123, 40 O†		NH1...190 OE2
Gln60	OE1...106 NZ†	Asn119	OD1...210 OG1		NH2...(172 OE2)
	NE2...79 O†		ND2...10 O†, W255, W256	Arg198	NE...177 O
Asn61	OD1...88 ND2†, 63 NH1	Asn120	OD1...W28, 94 NE2†		NH1...170 O, W112
	ND2...W114		ND2...W98	Asn199	NH2...W10, W103
Thr62	OG1...69 N, 72 OD2	Glu121	OE1...200 N, W82		OD1...W10, W39
Arg63	NE...60 O, W227		OE2...178 NE1, 199 N, W16		ND2...176 OD2, W27
	NH1...60 O, 61 OD1, W226	Trp122	NE1...21 O†	Thr205	OG1...197 O, W67
	NH2...88 OD1†, W50, W128	Gln125	OE1...W10, W27	Thr210	OG1...39 NZ†, 119 OD1
Asp66	OD1...1 N†		NE2...W62	Ser213	OG...W24
	OD2...W62, W127	Thr126	OG1...W101, 122 O	Tyr214	OH...W189
Tyr69	OH...W68, W125	Thr129	OG1...125 O, 169 OH	Lys217	NZ...W31, W167, 109 O
Thr71	OG1...W69, W75	Tyr130	OH	Tyr218	OH
Asp72	OD1...W56, W69				
	OD2...W61, 62 OG1				

† Intermolecular contact.

elastase (Sawyer *et al.*, 1978) and carp parvalbumin (Moews & Kretsinger, 1975), and in the latter case are considered to have some functional significance. There are very few of such lost hydrogen bonds in actinidin. In all, only 14 carbonyl and 17 amido groups do not form definite hydrogen bonds, and of these 10 carbonyl and 12 amido groups are accessible to solvent (and most do, in fact, have weak density adjacent). Of the four internal carbonyl groups which are not hydrogen bonded, three (those of residues 82, 131 and 157) are probably stabilized by other interactions with neighbouring polar groups (CO(131) is antiparallel with CO(127) and CO(128); CO(157) is antiparallel with CO_{δ1}(138); and O(82) is 2.96 Å from O_{δ1} of Asp80). The fourth, CO(49) is directed towards a non-polar region; this is discussed below (section (d),(i)). Of the five internal amido groups not hydrogen bonded, four are directed towards potential hydrogen bond acceptors 3.5 to 4 Å away. Thus very little hydrogen-bonding potential is not utilized.

The same is true for potentially hydrogen-bonding side-chain atoms. Of 168 such atoms in actinidin, only 22 (13%) do not form identified hydrogen bonds (see Table 5), and of these only one (Asn182 N_{δ2}) is not exposed to the external solution. A total of 93 (55%) hydrogen-bond to other protein atoms and the remaining 53 (32%) to water only.

(c) *Main-chain conformational angles*

The torsion angles ϕ , ψ along the polypeptide backbone are listed in Table 6, and shown graphically in Figure 4, as the usual Ramachandran plot (with "allowed" regions outlined according to Ramakrishnan & Ramachandran (1965)). Apart from Gly residues, only seven residues have conformations with ϕ in the range 50 to 100°. Five of these (Ala93, Asn119, Glu191, Asn199 and Cys206) are in the region for a left-handed α -helix, although none is part of a turn, or near-turn, of left-handed helix, and only one of these five (Cys206) is found in a "reverse turn" (Crawford *et al.*, 1973). Both of the remaining two residues (Asp185 and Gln60) are in external loops where an unfavourable conformation seems unlikely to be "forced". Gln60, however, is one of two residues (60 and 61) that are insertions at this point in actinidin, relative to papain (Baker, 1977), and that are accommodated in a type II' reverse turn (Venkatachalam, 1968) involving residues 59 to 62. The unusual conformation for Gln60 may then result from the need to insert the two extra residues without disturbing the rest of the structure.

Other conformational angles, τ and ω , have been restrained during the model refinement (by regularization). The looseness of constraints allows for some variation, however. The angles N-C α -C, for example, average $\sim 108^\circ$ in pieces of extended chain (the strands of β -sheet, and the long extended section 92 to 120) but $\sim 112^\circ$ in helices and up to 117 to 119° in some turns. This is qualitatively the same as is seen in rubredoxin (Watenpaugh *et al.*, 1979) although the deviations in the latter are larger.

(d) *Secondary structure*

The secondary structure of the molecule is shown schematically in Figure 3. Actinidin falls into the category of $\alpha + \beta$ proteins (Levitt & Chothia, 1976), in that

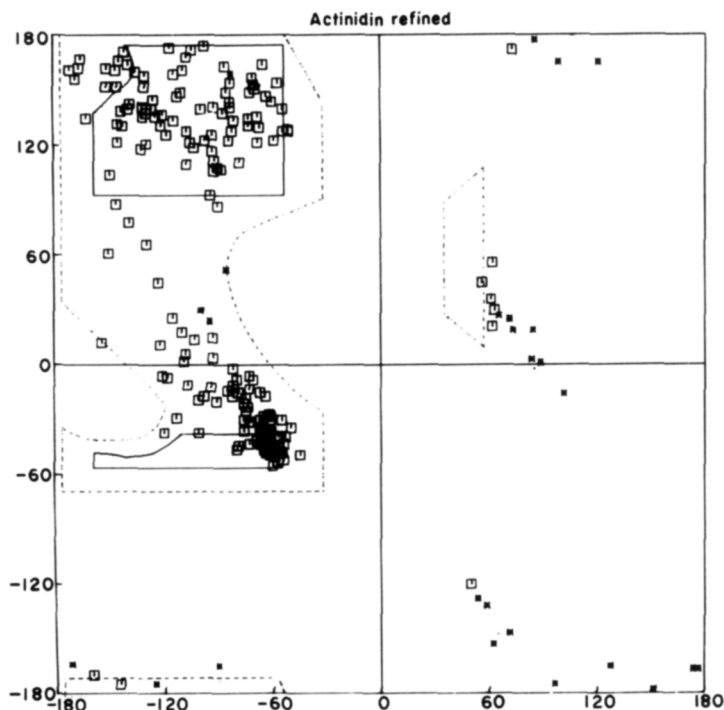


FIG. 4. Plot of ϕ, ψ values for actinidin. "Allowed" regions are according to Ramakrishnan & Ramachandran (1965). Glycine residues are denoted *, all others $\square$.

α -helix and β -sheet regions are segregated along the polypeptide chain. As in others of this type, one domain (domain I) is largely helical, while the other is built around a somewhat irregular twisted β -sheet.

(i) *Helices*

Helices have been identified from the ϕ, ψ values of the residues which comprise them, and from patterns of hydrogen bonding. Taking a helix to extend from the first hydrogen-bond donor in a helical region to the last hydrogen-bond acceptor, there are three major helices in actinidin (residues 24 to 43, 69 to 81 and 120 to 130) and four short ones (50 to 57, 85 to 89, 99 to 104 and 141 to 147).

As can be seen from Table 6, the ϕ, ψ values within the helices (ignoring the first residue and the last one or two) approximate those of an ideal α -helix ($-57, -47$), although the mean values (Table 7) tend towards ($-63, -40$). Similar average values have been found in other refined protein structures, e.g. ($-67, -41$) in bovine pancreatic trypsin inhibitor (Deisenhofer & Steigemann, 1975), and ($-64, -40$) in erythrocrucorin (Steigemann & Weber, 1979). This is consistent with the observation (see below) that *all* carbonyl groups tend to tilt outwards, away from the helix axis.

In every case except one, the first hydrogen bond of a helix is formed by a residue with non-helical (ϕ, ψ) values. The exception is 50 to 57, where residue 49 might have been expected to form the first hydrogen bond, but instead has its carbonyl oxygen

Phe28	-54	-52	Ile83	-145	152	Asp138	-92	106	Gly192	100	-18
Ser29	-65	-46	Asn84	-97	174	Ala139	-110	18	Tyr193	-107	168
Ala30	-64	-43	Thr85	-86	163	Ala140	-75	-18	Met194	-138	143
Ile31	-61	-48	Glx86	-62	-44	Gly141	-84	157	Ile195	-107	128
Ala32	-54	-43	Glu87	-61	-43	Asp142	-65	-28	Arg196	-131	141
Thr33	-72	-31	Asn88	-73	-31	Ala143	-61	-44	Leu197	-51	129
Val34	-69	-41	Tyr89	-139	78	Phe144	-64	-48	Arg198	-114	159
Glu35	-56	-47	Pro90	-71	157	Lys145	-61	-36	Asn199	62	56
Gly36	-59	-51	Tyr91	-81	134	Gln146	-82	-11	Val200	-107	-11
Ile37	-66	-35	Thr92	-118	-7	Tyr147	-51	128	Gly201	97	163
Asn38	-62	-43	Ala93	63	30	Ala148	-120	-37	Gly202	57	-134
Lys39	-62	-49	Gln94	-135	160	Ser149	-165	167	Ala203	-71	-8
Ile40	-55	-48	Asp95	-84	123	Gly150	95	-177	Gly204	84	175
Thr41	-80	-46	Gly96	-127	-177	Ile151	-92	112	Thr205	-68	122
Ser42	-91	-20	Asp97	-73	131	Phe152	-73	135	Cys206	62	21
Gly43	87	-1	Cys98	-59	123	Thr153	-121	-6	Gly207	72	17
Ser44	-121	131	Asp99	-90	108	Gly154	126	-167	Ile208	-50	-34
Leu45	-97	123	Val100	-63	-35	Pro155	-72	149	Ala209	-93	4
Ile46	-131	136	Ala101	-64	-41	Cys156	-168	156	Thr210	-76	-21
Ser47	-82	128	Leu102	-76	-30	Gly157	-92	-167	Met211	-150	61
Leu48	-109	161	Gln103	-67	-30	Thr158	-122	11	Pro212	-85	149
Ser49	-78	111	Asp104	-79	-15	Ala159	-94	93	Ser213	-151	162
Glu50	-76	-36	Gln105	-88	107	Val160	-60	144	Tyr214	-151	152
Gln51	-57	-39	Lys106	-118	126	Asp161	-154	12	Pro215	-83	154
Glu52	-62	-39	Tyr107	-113	-29	His162	-145	122	Val216	-124	136
Leu53	-67	-43	Val108	-131	140	Ala163	-92	141	Lys217	-134	112
Ile54	-60	-50	Thr109	-130	158	Ile164	-145	-175	Tyr218	-79	-
Asp55	-79	-44	Ile110	-135	160						

TABLE 7

Dihedral angles† within helices‡

Residues	$\langle\phi\rangle$	$\langle\psi\rangle$
(24)-25-41-(42-43)	-62 (8)	-43 (6)
50-55-(56-57)	-67 (8)	-42 (6)
(69)-70-79-(80-81)	-61 (7)	-42 (10)
(85)-86-88-(89)	-65 (5)	-39 (6)
(99)-100-103-(104)	-67 (5)	-35 (5)
(120)-121-129-(130)	-63 (3)	-38 (7)
(141)-142-145-(146-147)	-63 (2)	-39 (8)

† Standard deviations (from spread of values) in parentheses.

‡ Residues in parentheses are omitted from calculation of the means, but are included in the hydrogen-bonding in the helix.

directed away from NH(53) and NH(54) (3.7 Å and 3.9 Å distant, respectively) towards the non-polar core of domain I. The "loss" of this hydrogen bond is probably attributable to the hydrogen bonds made by NH(50) with the carboxyl group of Glu50 and an internal water molecule, W5.

The hydrogen bonding tends to be irregular at the start of each α -helix (see Table 8, where hydrogen-bonding details are given). Helix 24 to 43 begins with a 4 $\rightarrow$ 1 hydrogen bond (the angles are more suitable for a 5 $\rightarrow$ 1 interaction, but the distance, 3.31 Å, is too long), 69 to 81 begins with two 4 $\rightarrow$ 1 interactions (again the angles are more suitable for a 5 $\rightarrow$ 1 bond, but the distances, 3.37 Å and 3.23 Å, are too long), and both 50 to 57 and 120 to 130 begin with bifurcated (4 $\rightarrow$ 1 and 5 $\rightarrow$ 1) bonds.

Each helix ends with one or two residues whose ϕ, ψ values correspond more to those of an α_{II} helix (-93, -18). This phenomenon has been noted before (Nemethy *et al.*, 1967) and has the result that the helix tends to end with one or two bifurcated or irregular hydrogen bonds.

The hydrogen bond lengths (O...N) vary considerably, 2.7 to 3.3 Å, with a mean value of 2.98 to 2.99 Å (Table 8), rather longer than the value of 2.87 Å for a standard α -helix (Arnott & Wonacott, 1966).

Each of the principal α -helices has a polar side and a non-polar side. The central α -helix, 24 to 43, is completely buried, but its polar side-chains form part of the distinctly polar interface between domains I and II. Thus, although buried, one side of this helix is still associated with a number of water molecules and other polar groups.

(ii) β -Structure

The twisted β -sheet around which domain II is built is shown in Figure 5. It consists basically of four strands, all antiparallel; strand B, which is a split strand, comprising residues 5 to 7 and 133 to 137; strand C, residues 162 to 172; strand D, residues 175 to 182; and strand E, residues 192 to 198. Two additional hydrogen bonds at each edge are provided by residues 213 (antiparallel with strand B) and 152 (parallel to strand E) and these could possibly be regarded as "strands" A and F.

The sheet includes one classic β -bulge (Richardson *et al.*, 1978), residues 167 to

TABLE 8

Hydrogen bonds in principal α -helices(A) *Residues 24-43*

H bond	Distance (Å)	C—O...N	C—O...H	N—H...O
24 O...27 N†	2.92	129	117	135
(24 O...28 N†	3.31	151	145	161)
25 O...29 N	2.89	162	159	172
26 O...30 N	3.09	143	137	158
27 O...31 N	2.89	154	149	163
28 O...32 N	2.91	155	150	165
29 O...33 N	3.27	153	150	165
30 O...34 N	3.12	157	149	153
31 O...35 N	2.84	169	165	170
32 O...36 N	3.07	149	142	159
33 O...37 N	2.94	151	148	164
34 O...38 N	2.94	158	150	157
35 O...39 N	3.09	152	145	156
36 O...40 N	3.01	157	153	169
37 O...41 N	2.88	156	150	160
38 O...42 N	2.77	156	148	158
39 O...43 N	2.71	136	125	135
<i>Mean</i>	<u>2.98 ± 0.13</u>	<u>154 ± 7</u>	<u>150 ± 6</u>	<u>162 ± 6</u>

Other H bonds:	28 O...W2	2.75 Å	C—O...W = 113°	
	29 O...134 OG	3.01 Å	C—O...OG = 117°	O...OG—CB = 96°
	30 O...33 OG1	3.22 Å	C—O...OG1 = 118°	O...OG1—CB = 123°
	32 O...W7	2.84 Å	C—O...W = 133°	
	33 O...W17	2.84 Å	C—O...W = 124°	
	35 O...W23	2.87 Å	C—O...W = 116°	
	38 O...42 OG	2.75 Å	C—O...OG = 135°	O...OG—CB = 99°

(B) *Residues 69-81*

H bond	Distance (Å)	C—O...N	C—O...H	N—H...O
69 O...72 N†	2.95	125	113	137
(69 O...73 N†	3.37	151	142	145)
70 O...73 N†	2.84	119	103	125
70 O...74 N	3.23	153	147	156
71 O...75 N	2.87	147	139	157
72 O...76 N	2.96	145	138	152
73 O...77 N	3.29	154	152	168
74 O...78 N	2.98	161	157	169
75 O...79 N	2.81	156	152	168
76 O...80 N	2.78	164	157	157
77 O...81 N†	3.02	148	136	125
77 O...82 N†	2.99	150	145	131
(78 O...81 N†	3.27	108	97	140)
<i>Mean</i>	<u>2.99 ± 0.18</u>	<u>154 ± 6</u>	<u>149 ± 7</u>	<u>161 ± 6</u>

Other H bonds:	72 O...W47	2.82 Å	C—O...W = 124°
	77 O...W12	3.06 Å	C—O...W = 115°

TABLE 8 *continued*(C) *Residues 120–129*

H bond	Distance (Å)	C—O...N	C—O...H	N—H...O
120 O...123 N†	3.13	117	102	118
120 O...124 N	3.05	157	150	157
121 O...125 N	2.79	150	142	155
122 O...126 N	3.03	150	145	161
123 O...127 N	3.05	149	143	159
124 O...128 N	2.94	162	157	165
125 O...129 N	3.05	155	145	141
(126 O...129 N†	3.24	107	94	127)
127 O...130 N†	2.96	121	111	147
<i>Mean</i>	<u>2.98±0.10</u>	<u>154±5</u>	<u>147±5</u>	<u>156±8</u>

Other H bonds:	120 O...W15	C—O...W = 100°	
	122 O...126 OG1	C—O...OG1 = 147°	O...OG1—CB = 104°
	123 O...W85	C—O...W = 123°	
	125 O...129 OG1	C—O...OG1 = 140°	O...OG1—CB = 119°
	127 O...W86	C—O...W = 109°	

† Those H bonds omitted from calculation of means.

168 opposing residue 179. The bulge allows strand C to transfer its hydrogen bonding from one half to the other of the split strand B, while at the same time maintaining its hydrogen bonding to strand D.

Omitting four residues with non-standard angles (residues 7 and 192, at the ends of strands, and 167 to 168) the average dihedral angles for residues comprising the β -sheet are $\langle\phi\rangle = -121 \pm 28^\circ$ and $\langle\psi\rangle = 143 \pm 19^\circ$. These values agree well with those expected theoretically for an antiparallel β -sheet with a right-handed twist, namely $\phi = -120^\circ$, $\psi = 150^\circ$ (Chothia, 1973), and with those found in other

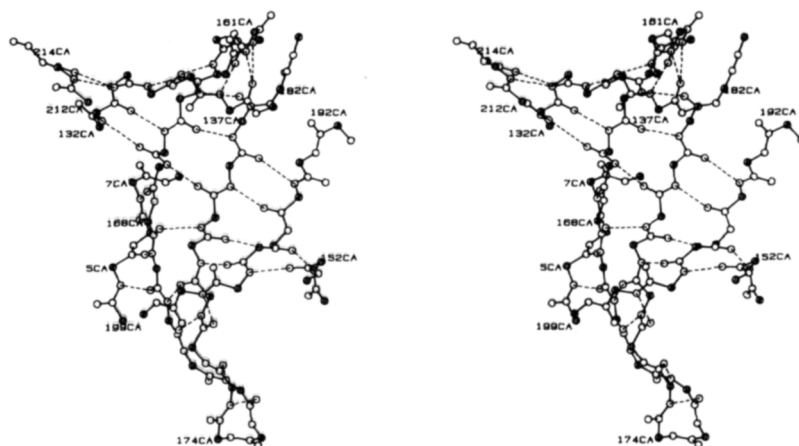


FIG. 5. The twisted β -sheet in domain II of actinidin. α -Carbon (CA) atoms are shown cross-hatched. Broken lines represent hydrogen bonds.

refined protein structures, e.g. prealbumin (Blake *et al.*, 1978), $\phi = -119$ to -126° , $\psi = 134$ to 136° , and bovine pancreatic trypsin inhibitor (Deisenhofer & Steigemann, 1975), $\phi = -114^\circ$, $\psi = 139^\circ$. As in the latter proteins, the standard deviations, obtained from the spread of values, indicate considerable variability. The distribution of ϕ values appears to be bimodal, with about 40% of residues having ϕ clustered around -95° (-80 to -110) and the remainder centred on $\phi = -140^\circ$. The distribution of ψ values is unimodal. Thus many residues have ϕ, ψ values like those for a straight β -strand ($-139, 135^\circ$), while others correspond to a more extremely twisted strand, with $\phi \sim -95^\circ$.

Details of hydrogen bonding in the β -sheet, and in one other short piece of anti-parallel β -structure (residues 111 to 116 and 212 to 216) are given in Table 9. The

TABLE 9
Hydrogen bonds within β -structure

H-bond	Distance (Å)	C—O...N	C—O...H	O...H—N	
5 N...169 O	2.91	156	159	161	
5 O...169 N	2.85	146	140	154	
7 N...167 O	2.83	170	169	172	
111 N...216 O	2.78	160	152	154	
112 O...216 N	2.82	138	144	162	
114 N...214 O	2.85	169	164	164	
114 O...214 N	3.00	165	158	157	
116 N...212 O	2.96	135	130	161	
133 N...166 O	3.22	151	147	162	
133 O...166 N	2.73	168	171	167	
134 N...213 O	2.86	174	174	148	
134 O...213 N	2.92	154	149	161	
135 N...164 O	2.99	150	150	178	
135 O...164 N	2.76	163	161	173	
137 N...162 O	2.81	152	148	162	
137 O...162 N†	2.92	165	150	129	Bifurcated H-bond
137 O...161 N†	2.88	125	124	172	
165 N...181 O	3.16	150	147	164	
165 O...181 N	2.77	164	164	164	
167 N...179 O	2.81	160	152	156	
(168 O...179 N†	2.91	122	124	175)	β -bulge
170 N...177 O	3.07	144	140	163	
170 O...177 N	2.99	152	153	173	
172 N...175 O	2.87	141	139	159	
(172 O...175 N†	2.96	117	112	162)	β -bend
176 O...198 N	2.86	123	127	166	
178 N...196 O	2.84	153	145	167	
178 O...196 N	3.00	154	147	160	
180 N...194 O	2.76	161	158	165	
180 O...194 N	2.97	151	144	155	
(182 N...192 O†	3.35	155	147	152)	Long (end of strand)
152 N...195 O	2.73	144	146	172	
152 O...197 N	2.88	163	166	168	
Mean	<u>2.89 ± 0.11</u>	<u>154 ± 12</u>	<u>152 ± 11</u>	<u>163 ± 7</u>	

† Those omitted from total.

average hydrogen bond length is shorter than that seen in the α -helices (2.89 ± 0.11 Å compared with 2.98 ± 0.10 Å) and agrees well with normal N-H...O hydrogen bond lengths. Several longer bonds are found at the ends of strands (but are included in Table 9 because their geometry is otherwise favourable). Only one hydrogen bond is bifurcated (137 O...161 N, 162 N at the end of strands B and C).

The four central strands of the β -sheet are largely buried in the interior of domain II. Not surprisingly there is a fairly non-polar distribution of amino acids in the β -sheet, namely Val₇Ile₄Gly₃Tyr₃Leu₂Ala₂Trp₂Arg₂Asp₂Met₁Thr₁Ser₁Asn₁His₁Glu₁Lys₁. Polar residues mostly occur at the ends of strands. Only three of the amido and carbonyl groups buried in these strands do not have hydrogen-bonding partners in adjacent strands; NH(136), CO(136) and CO(163) hydrogen bond to internal water molecules (see Fig. 3). The main-chain hydrogen bonding of the sheet is also relatively uninterrupted by additional hydrogen bonds from polar side-chains or water molecules. Only four such interactions occur (163 O...Ser29 O _{γ} ; 170 O...Arg198 N _{η 1}; 177 O...Arg198 N _{ϵ} ; 181 O...W26).

(iii) Turns

Reverse turns, as described by Venkatachalam (1968) and by Crawford *et al.* (1973) are listed in Table 10. All those for which C _{α} ¹...C _{α} ⁴ is less than 6 Å are given; all have 4 $\rightarrow$ 1 hydrogen bonds. Turns expected to have only Gly at certain positions include types I' and III' (Gly at positions 2 and 3), type II' (Gly at 2) and type II (Gly at 3). The restrictions are followed by 172 to 175 (I'), 63 to 66 (II'), 201 to 294 (II') but not by 190 to 193 (I', Glu at position 2), 205 to 208 (III', Cys at position 2), 204 to 207 (II, Cys at position 3) and 59 to 62 (II', Gln at position 2). Even in these cases, however, the 4 $\rightarrow$ 1 hydrogen bond remains.

Residues 201 to 210 form a very tight piece of structure on the surface of the molecule, including four Gly residues, and encompassing four turns as well as two other hydrogen bonds (see Fig. 6). No part of this structure could be described as helical, and residues 205 to 210 form a more or less planar S-shaped piece of structure, made up of two interlocking turns (205 to 208 and 207 to 210). It appears to be similar to a conformation previously noted in concanavalin A (Reeke *et al.*, 1975).

Several other chain reversals occur where the angles ϕ_2, ψ_2 and ϕ_3, ψ_3 roughly correspond to reverse-turn configurations (e.g. 19 to 22 and 117 to 120), but where the distance C _{α} ¹...C _{α} ⁴ is greater than 6 Å, and no 4 $\rightarrow$ 1 hydrogen bond exists. In the case of 117 to 120 a tightly bound water molecule links residues 1 and 4, i.e. 117 O...W15...120 N.

(iv) Linearity of hydrogen bonds in secondary structure

Hydrogen bond angles are listed in Tables 8 to 10. Since all the atoms involved are well-ordered, with $B = 5$ to 10 Å², standard deviations are ~ 0.1 Å in individual hydrogen bond lengths and $\sim 5^\circ$ in individual hydrogen bond angles. For an ideal linear hydrogen bond, the angle N-H...O = 180° , and for ideal interaction with a trigonal oxygen atom, C-O...H should be $\sim 120^\circ$ †.

In helices a consistent pattern emerges. None of the hydrogen bonds are linear.

† C-O...H refers to the angle (in degrees) at the atom O. Similarly for other angles, N-H...O (angle at H atom), etc.

TABLE 10
Geometry of turns

Residues	Sequence	ϕ_2	ψ_2	ϕ_3	ψ_3	Type	$C_\alpha^1 \dots C_\alpha^4$ (Å)	$O^1 \dots N_4$ (Å)	C-O...N	C-O...H	O...H-N
6-9	Asp-Trp-Arg-Ser	-64,	-17	-67	-34	III	5.72	2.99	123	113	140
7-10	Trp-Arg-Ser-Ala	-67,	34	-73,	-6	III	5.15	3.12	107	99	152
8-11	Arg-Ser-Ala-Gly	-73,	-6	-108,	6	I	5.39	3.27	113	111	164
59-62	Thr-Gln-Asn-Thr	50,	-120	-93,	15	II'	5.29	2.72	146	141	160
63-66	Arg-Gly-Cys-Asp	52,	-130	-67,	-15	II'	5.40	3.04	142	145	141
64-67	Gly-Cys-Asp-Gly	-67,	-15	-73,	-13	III	5.93	3.07	123	121	169
172-175	Glu-Gly-Gly-Val	64,	25	82,	1	I'	5.31	2.96	117	112	162
190-193	Glu-Glu-Gly-Tyr	61,	36	100,	-18	I'	4.98	3.18	105	102	160
201-204	Gly-Gly-Ala-Gly	57,	-134	-71,	-8	II'	5.48	2.97	136	137	148
204-207	Gly-Thr-Cys-Gly	-68,	122	62,	21	II	4.90	3.10	117	122	150
205-208	Thr-Cys-Gly-Ile	62,	21	72,	17	III'	5.55	3.17	121	115	161
207-210	Gly-Ile-Ala-Thr	-50,	-34	-93,	4	I	5.30	2.90	127	124	167
							123 ± 12		120 ± 14		156 ± 9

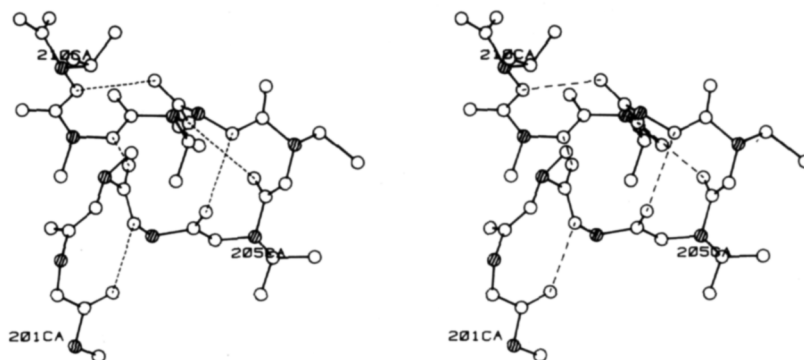


FIG. 6. The conformation of residues 201 to 210, a tightly folded piece of structure, encompassing 4 turns (201 to 204, 204 to 207, 205 to 208 and 207 to 210), and including 4 glycine residues (201, 202, 204 and 207). Atoms etc. as in Fig. 5, and for Figs 6, 7, 8, 9, 10, 11, 12, 14, 18, 19 and 20.

The average angle $N-\hat{H}\dots O$ is $161 \pm 7^\circ$ and in only one case, out of 29 regular $5 \rightarrow 1$ hydrogen bonds is it $> 170^\circ$. The mean value of the angle $C-\hat{O}\dots N$ is $154 \pm 6^\circ$, and that of $C-\hat{O}\dots H$ is $149 \pm 7^\circ$. (Note that $C-\hat{O}\dots H$ is always less than $C-\hat{O}\dots N$.) These angles agree nicely with the average $C-\hat{O}\dots N$ angle of 156° found in the helices of the highly refined structure of erythrocrucorin (Steigemann & Weber, 1979). They are consistent with a tilting of the carbonyl groups out from the helix axis (at an angle of $\sim 12^\circ$). It is of interest that the tilting of the carbonyl groups is apparently unaffected by any additional hydrogen bonds with adjacent side-chains (e.g. Ser) or water molecules, i.e. *all* carbonyl groups are tilted this way. Where additional hydrogen bonds do occur (see Table 8), the angle $C-\hat{O}\dots D$ ($D = \text{donor, e.g. water molecule}$) is usually $\sim 120^\circ$, as expected.

The hydrogen bonds in β -sheets are also non-linear, with the angle $N-\hat{H}\dots O$ averaging $163 \pm 7^\circ$ and with only 7 out of 33 having this angle $> 170^\circ$. Non-linearity

TABLE 11

Residues forming non-polar cores of actinidin

Domain I		Domain II	
Trp26	Phe74	(Val116)	Ile166
Ala30	(Phe76)	Leu124	Trp178
Ile31	Ile77	Val128	Val180
Val34	(Ile78)	Val133	(Trp184)
(Ile37)	Ile83	Val135	Trp188
(Ile46)	Val108	Leu137	Met194
Leu53	Pro215	(Tyr147)	Ile208
		(Phe152)	Ala209
		Ile164	(Pro212)

Residues in parentheses lie in the surface of the molecule, in contact both with the non-polar core and with the external solution.

of hydrogen bonds within β -sheets has been commented on before (Chothia, 1973), and while the reasons may be associated with the twisting of the sheet, it is interesting to note the similarity between the hydrogen bond angles in both helices and β -structure. The angles $C-\hat{O}\dots N$ (average $154\pm 12^\circ$) and $C-\hat{O}\dots H$ (average $152\pm 12^\circ$) are more variable in the β -sheet, however, and there is no clear pattern to their values.

The hydrogen bonds in reverse turns, or 3_{10} bends, are sometimes referred to as bent because the angle $C-\hat{O}\dots N$ is $\sim 120^\circ$ (e.g. in rubredoxin, in eight 3_{10} bends, $C-\hat{O}\dots N = 110$ to 136° (Watenpaugh *et al.*, 1979)). The same is true of the turns in actinidin, namely $C-\hat{O}\dots N = 105$ to 146° (average $123\pm 12^\circ$). However, the hydrogen bonds are no more non-linear than those in helices and β -structure, as judged by the angles $N-\hat{H}\dots O$ (average $156\pm 9^\circ$). They may be more favourable, since $C-\hat{O}\dots H$ is $\sim 120^\circ$ and an oxygen atom lone pair then points directly towards the hydrogen atom.

(e) Side-chain environments

(i) Non-polar side-chains

Each domain has a core of buried non-polar side-chains (see Table 11). These are mostly Val, Leu or Ile, and although a number of aromatic groups contribute, they do not, with the possible exception of Phe74, intrude into the centre (see e.g. Fig. 7). Instead, as has been also noted in elastase (Sawyer *et al.*, 1978), they tend to lie either on the periphery, just inside the surface of the molecule, or right on the surface. Residues in the latter category (Phe76, Tyr147, Phe152) often have one edge of their aromatic ring in contact with the core, and the other edge accessible to the external solution.

None of the aromatic side-chains (6 Trp, 5 Phe, 14 Tyr) are packed plane-to-plane in a manner indicative of π - π interactions. (The only likely interaction of this sort

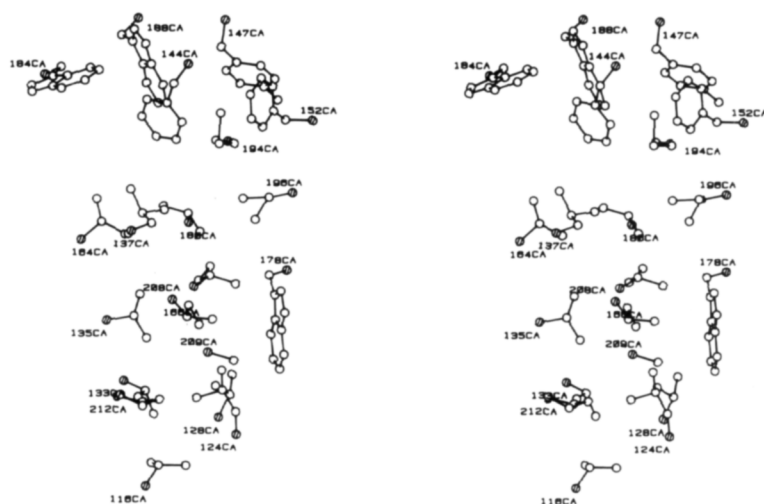


FIG. 7. The hydrophobic core of domain II showing the packing of non-polar side-chains. Some aromatic side-chains are seen, all at the periphery of the core. The aromatic cluster (comprising Phe144, Tyr147, Phe152, Trp184 and Trp188) is seen at the top. Note the "edge-on" packing.

is between His162 and Trp184 in the active site, see below.) The only sizeable cluster of aromatic side-chains is that comprising Phe144, Tyr147, Phe152, Trp184 and Trp188 (see Fig. 7); these pack roughly in an "edge-on" manner (i.e. the edge of one ring approaching the plane of another) as is often seen in crystals of aromatic small molecules.

Few of the non-polar side-chains in actinidin are fully exposed to the solvent. There are a small number of scattered non-polar groups on the surface (e.g. Val5, Val100, Leu102 and Val216) and just one sizeable hydrophobic "patch", comprising Ile151, Val175, Leu197, Val200 and C β —C γ of Thr153. This has no known function.

(ii) *Proline residues*

Four out of seven Pro residues are external. One of them, Pro155 (Fig. 8), is a *cis*-proline (as it also is in papain), but although it comes immediately before Cys156, which forms **one** of the three disulphide bridges in actinidin, there is no obvious

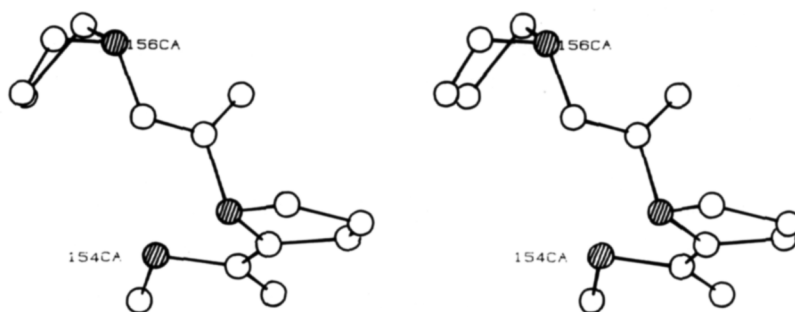


FIG. 8. The *cis*-proline residue, 155. Gly154 could be joined to Cys156 relatively easily, suggesting Pro155 could have been an earlier evolutionary insertion, then being required to adopt the *cis* configuration.

structural requirement for a *cis*-proline in this position. On the contrary, it makes a small loop projecting from the surface and could perhaps have been an insertion earlier in the evolution of the sulphydryl proteases (since Gly154 could be joined to Cys156 with little disturbance to the structure).

Of the other three external Pro residues, Pro90 comes between Tyr89 and Tyr91, both involved in structurally important hydrogen bonds, but the other two, Pro2 and Pro117, are in pieces of extended chain and do not appear particularly important to the structure. The three remaining Pro residues are Pro132 (in the domain-domain interface), Pro212 (in the surface of domain II) and Pro215 (in the non-polar core of domain I).

(iii) *Tyr and Trp side-chains*

Of the six Trp side-chains, three (26, 178, 188) are completely buried, and in each case the indole nitrogen, N ϵ_1 , is hydrogen bonded (twice to carbonyl oxygen atoms, and once to a carboxyl group) (see Table 5). Two (7, 184) lie in the molecular surface and both are hydrogen bonded (184 to water, 7 to a carbonyl oxygen). The sixth,

Trp122, projects from the surface, but is involved in intermolecular interactions (hydrogen bonded to a neighbouring carbonyl group).

Only one of the 14 Tyr side-chains (169) is completely buried; its —OH group makes two hydrogen bonds, helping to bind the N-terminus of the polypeptide chain to the rest of the molecule by linking residues 2 and 129. Three of the five partially buried Tyr residues have similarly important structural roles. The hydroxyl groups of Tyr89 and Tyr147 both form two hydrogen bonds, helping to stabilize loosely bound external loops. Tyr89 links residues 55 and 99 (see Fig. 9), while Tyr147 links

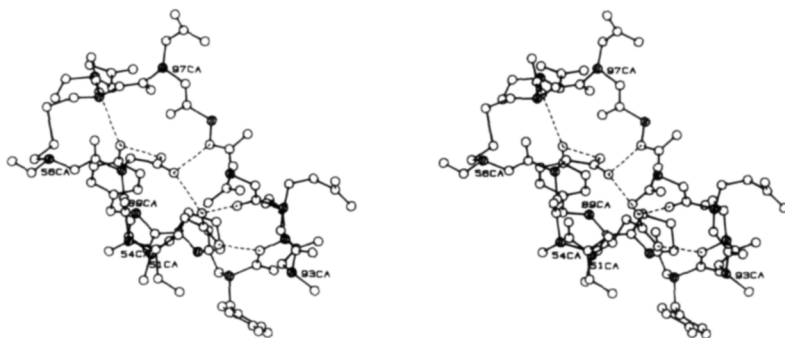


FIG. 9. Diagram showing how polar side-chains (Gln51, Asp55 and Tyr89) hold a loosely bound external loop (92 to 100) to the rest of the molecule, through a network of hydrogen bonds.

residues 150 and 195, helping to hold the external loop, 146 to 152, to the β -sheet of domain II. Tyr91 hydrogen bonds to the carbonyl oxygen of residue 19 at the domain-domain interface. The remaining two partially buried Tyr residues (177 and 193) have their —OH groups projecting into solution, where they hydrogen bond to water molecules.

The remaining eight Tyr residues are all external, generally hydrogen bonded to water.

(iv) *Charged side-chains*

Acidic and basic side-chains, which should normally be charged at pH 6, are mostly distributed over the surface, as expected. Although there is a large preponderance of acidic residues over basic ones (27 to 11), consistent with the low isoelectric point, 3.1 (McDowall, 1970), there is no obvious concentration of charge on any one region of the surface.

Of particular interest in the structure of actinidin is the presence of one buried basic group (Lys181) and no fewer than three buried acidic groups (Glu35, Glu50 and Glu52). None of these has any known functional role. Glu35, Glu50 and Lys181 all lie within the interface between domains I and II, along with Lys17, partially buried at the end of the interface, and beyond it Glu86. Thus the charge on the two buried acid groups, Glu35 and Glu50, is partly neutralized by Lys17, and further neutralized by interaction with Lys181, which is linked to Glu35, not directly but through two water molecules. This complex arrangement is shown in Figure 10. Finney (1977)

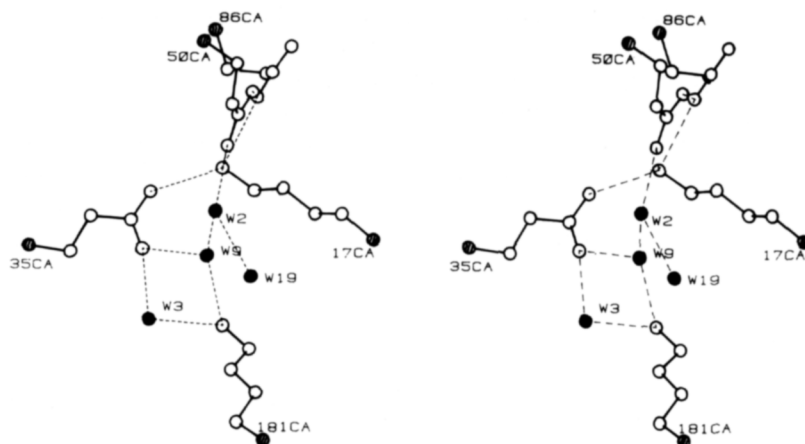


FIG. 10. The complex of charged residues, Glu35, Glu50, Lys17 and Lys181 buried in the domain I-domain II interface. Lys17 interacts directly with Glu35 and Glu50; Lys181 interacts with Glu35 *via* 2 water molecules (●). Residue 86, beyond Glu50 and Lys17, is either Glu or Gln.

has suggested that water molecules can have a charge-spreading role, and the interaction between Glu35 and Lys181 seems to be a good example.

The other buried acid group, the carboxyl group of Glu52, has no obvious positively charged partner. It lies in a partly hydrophobic environment, adjacent to the side-chains of Ile83 and Phe76. One carboxyl oxygen receives a hydrogen bond from the main-chain amido group of residue 84, while the other is 2.66 Å from a buried solvent molecule or ion, W4 (see Fig. 11). The latter was originally taken as a water molecule, but there is good reason to suppose it may actually be an ammonium ion. It has three hydrogen-bonding neighbours, within 3 Å (52 O_{ε2}, 103 O, 105 O) but all are hydrogen-bond acceptors. Two more distant interactions (3.60 Å and 3.70 Å) are made by the carboxyl oxygen, O_{δ1}, of Asp80, and the main-chain carbonyl oxygen of residue 82 (Fig. 11). A buried ammonium ion in this position, lying between

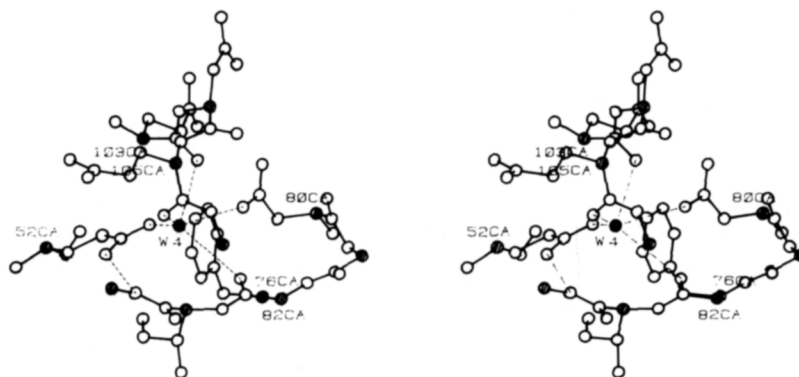


FIG. 11. Surroundings of the buried carboxyl group of Glu52 and the postulated ammonium ion, W4. The latter makes short approaches to Glu52 O_{ε2} (2.66 Å), 103 O (2.79 Å) and 105 O (2.80 Å) and longer ones to Asp80 O_{δ1} (3.60 Å) and 82 O (3.70 Å).

Asp80 and Glu52, would neutralize the charge on the buried Glu52 nicely, and satisfy the hydrogen-bonding requirements of the groups around it.

Most of the remaining charged side-chains are external, interacting with the solvent only, but structurally important interactions include those made by the side-chains of (1) Glu121, partly buried in a deep surface pocket, and hydrogen-bonded to NH(199) and NH(200) from an external loop, N_{ε1} of the buried Trp178, and an internal water molecule, W16; (2) Asp55, hydrogen-bonded to Gln51, Tyr89 and NH(96) (see Fig. 9); and (3) Arg195, which links Glu190 (2.9 Å away on one side) and Glu172 (3.3 Å away on the other side), giving a "double" ion pair interaction Glu172...Arg195...Glu190. There is only one other surface ion pair, Asp6...Arg8.

(v) *Other polar side chains*

Very few are found completely buried in actinidin. Four (Ser29, Thr33, Ser134 and Ser183) are situated in the domain I-domain II interface (see below). In domain II, Asn182 is buried, surrounded by the side-chains of Trp184, Trp188, Phe144 and Ile164, and hydrogen-bonded to the catalytically important His162. In domain I only Ser49 is buried (hydrogen-bonded to CO(89)).

Other polar side-chains, e.g. Gln51, lie in the molecular surface where, like Tyr89, Tyr147, Asp55 and Glu121, mentioned earlier, they have an important structural role, binding external sections of the polypeptide chain to the main part of the molecule (see e.g. Fig. 9).

(f) *The domain I-domain II interface*

The polypeptide chain crosses between domains I and II in only three places, residues 16 to 22 crossing on one side of the molecule and 115 to 119 and 212 to 214 crossing on the other side (see Fig. 2). In between, there is an extensive interface between the two domains. The surprising feature of this interface is that its interior is mostly polar in nature, rather than non-polar. Only three non-polar side-chains are buried within it; Ala32 from domain I and Val165 from domain II, which oppose each other across the interface (in papain these are Val32 and Ala162), and Pro132 from domain II which is adjacent to Gly36 from domain I. Non-polar groups also appear on the periphery of the interface (the side-chains of Ala163 and Ile70 on the top surface, Val14 and Leu45 on the bottom, Phe28 and the aliphatic chain of Lys17 at one end, and Ile40, Ala12, Trp7 and part of the side-chain of Lys39 at the other), apparently shielding the polar *interior* from the external solution.

Within the interface are the hydrogen-bonded side-chains of Ser29, Thr33 and Ser134, the charged side-chains of Lys17, Lys181, Glu35 and Glu50 (see previous section e(iv)) and a network of eight linked water molecules, hydrogen-bonded to each other and to protein atoms. These lie below the active site and extend along the interface from Ser183, approximately parallel to the central α -helix (24 to 43). Interactions within the interface are listed in Table 12. The water network, and its associated protein atoms, is shown in Figure 12 (a stereo diagram) and schematically in Figure 13. The internal water structure is also discussed in more detail below.

There are several possible reasons for this distinctly polar interface. It is possible that the water molecules may just be trapped during folding, i.e. that the two domains

TABLE 12
Domain I-domain II interface

A. Residues contributing

From domain I:	Phe28, Ser29, Ala32, Thr33, Glu35, Gly36, Lys39, Ile40, Leu45, Ser47, Glu50, Ile70, Glu86
From domain II:	Trp7, Ala12, Val14, Lys17, Gln131, Pro132, Ser134, Ala163, Val165, Lys181, Ser183

B. Interactions across interface

Hydrophobic:

Phe28 CD1...Lys17 CG, CD	~4.3 Å	Leu45 CD1...Val14 CG1	3.9 Å
Ala32 CB...Val165 CG1	3.3 Å	Ile70 CD1...Ala163 CB	3.8 Å
Thr33 CG2...Val165 CG2	3.6 Å	Gly36 CA...Pro132 CG, CD	~4.3 Å
Ile40 CD1...Trp7 CZ3, CE3	~4.0 Å	Lys39 CG...Ala12 CB	3.8 Å

H-bonds, ion pairs:

Glu35 OE2...Lys17 NZ	2.88 Å	Ser29 O...Ser134 OG	3.01 Å
Glu50 OE1...Lys17 NZ	2.78 Å	Ser29 OG...Ser134 OG	3.08 Å
Glu86 OE2...Lys17 NZ	2.97 Å	Ser29 OG...Ile164 O	2.86 Å
Ser47 OG...Lys17 NZ	3.07 Å	Thr33 OG1...Ser134 OG	3.02 Å
Gln19 OE1...Ser183 OG	2.66 Å	Pro215 O...Gln131 NE2	2.88 Å

Involving water molecules:

W2...Phe28 O	2.75 Å	W3...Val14 N	2.89 Å
W2...Glu50 OE1	2.67 Å	W3...Glu35 OE1	2.63 Å
W2...W9	2.83 Å	W3...Lys181 NZ	2.84 Å
W2...W19	2.89 Å	W3...W7	2.98 Å
W9...Glu35 OE1	2.66 Å	W7...Ala12 O	2.88 Å
W9...Lys181 NZ	2.71 Å	W7...Ala32 O	2.84 Å
W6...Ala163 O	2.80 Å	W14...Lys17 N	2.85 Å
W6...Ser183 OG	2.89 Å	W14...Asn182 O	2.81 Å
W6...W26	2.62 Å	W14...W19	2.76 Å
W26...Lys181 O	3.05 Å	W19...W26	2.88 Å
W17...Thr33 O	2.84 Å	W17...Gln131 NE2	3.08 Å
W17...Pro132 O	2.69 Å		

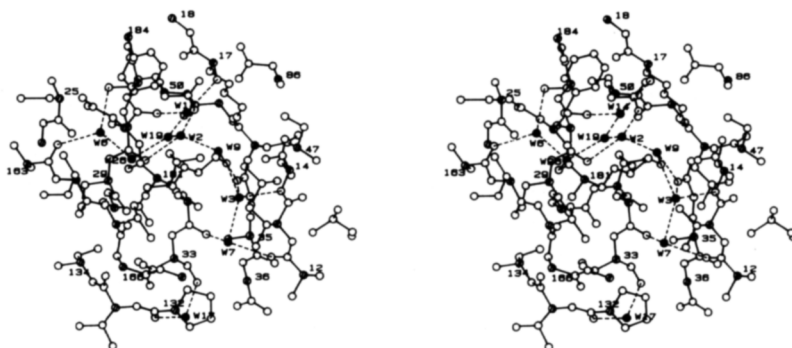


FIG. 12. The network of water molecules (●) in the domain-domain interface of actinidin. Most of the protein atoms which contribute to the interface are shown (C_{α} atoms, hatched circles). Refer to Table 12 for details.

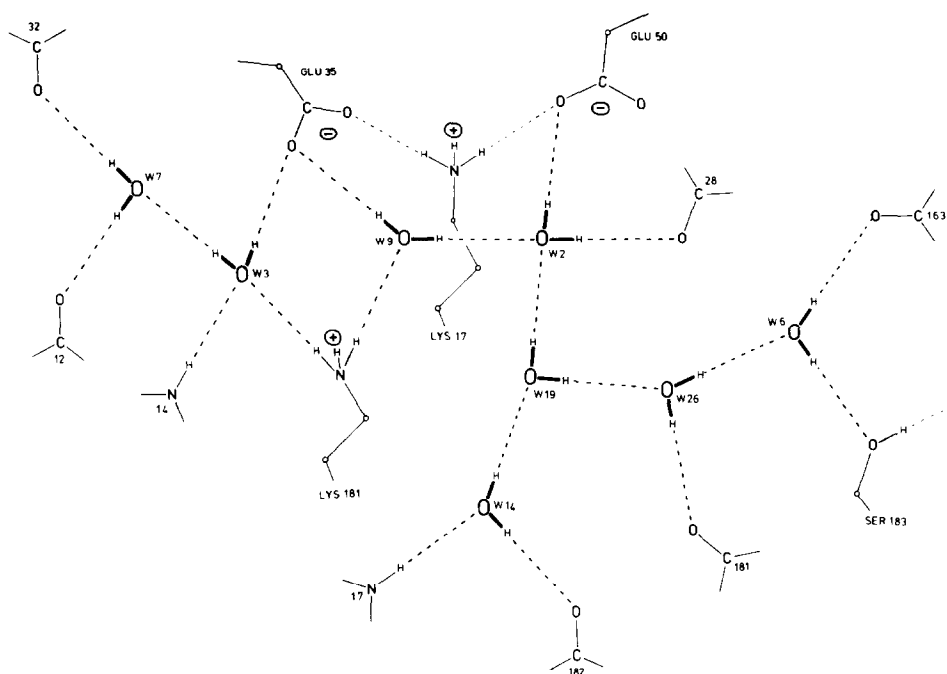


FIG. 13. A schematic representation of the network of 8 internal water molecules in the domain-domain interface. Each water molecule forms hydrogen bonds with 2 proton acceptors and either 1 or 2 proton donors; all water protons are thus utilized. The 4 charged side-chains, of Lys17, Lys181, Glu35 and Glu50, are also shown.

fold up independently, each with a polar exterior (one side of the helix 24 to 43 would then be on the outside of domain I, and side-chains such as Lys181, Glu35, Glu50 etc. would also be external), with associated water molecules; and that the water remains trapped, when the two domains mesh together. It may be, however, that the polar interior (and in particular the water molecules) has a functional importance. Lowe & Yuthavong (1971) have suggested that substrate binding should be accompanied by a widening of the active site cleft in papain, and this widening has been confirmed by crystallographic inhibitor binding studies (Drenth *et al.*, 1975). Since papain contains a similar network of internal water molecules in the domain-domain interface, below the active site, it is possible that the more flexible, malleable structure provided by the water molecules allows the active site above it to be more elastic (perhaps to "breathe", oscillating between a wider and a narrower structure). It is interesting to note that in the serine proteases, elastase and chymotrypsin, most of the internal water molecules lie between the two β -barrels (Sawyer *et al.*, 1978; Birktoft & Blow, 1972), a somewhat similar situation to that seen here.

(g) *The active site*

Although the active site remains qualitatively as originally described (Baker, 1977), shifts of 0.3 to 0.6 Å during refinement have changed the detailed interpretation

significantly; the Cys25 S_γ ...His162 $N_{\delta 1}$ distance has been reduced by 0.6 Å, the angle between the ring planes of His162 and Trp184 has changed by 20°, and other distances and angles in the active site have been more actively described.

The sulphur atom of the essential Cys25 lies close (3.30 Å) to one nitrogen, $N_{\delta 1}$ of His162, while the other nitrogen, $N_{\epsilon 2}$, is hydrogen-bonded (2.79 Å) to $O_{\delta 1}$ of Asn182 (see Fig. 14). Asn182 is buried, in a non-polar environment, and the hydrogen bond 162 $N_{\epsilon 2}$...182 $O_{\delta 1}$ is shielded from the external solution by the indole ring of Trp184. The $-NH_2$ group of Asn182 is left without a hydrogen-bonding partner, and is directed inwards towards the plane of the aromatic ring of Trp188.

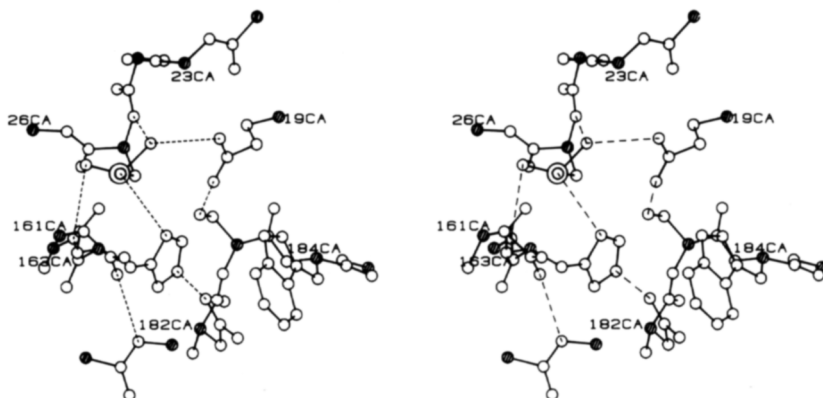


FIG. 14. The catalytic site in actinidin. The sulphur atom (double circle) of Cys25 is 3.30 Å from $N_{\delta 1}$ of His162, while $N_{\epsilon 2}$ is hydrogen-bonded (2.79 Å) to $O_{\delta 1}$ of Asn182. The hydrogen bond is covered by the ring of Trp184. The 2 oxygen atoms of the oxidized sulphhydryl group are shown, one of them hydrogen bonded to NH(25) and Gln19 $N_{\epsilon 2}$ (i.e. in the "oxyanion hole"), the other hydrogen-bonded to NH(163). The carboxyl group of Asp161 is turned away from the catalytic site, to hydrogen bond with NH(139).

The active site geometry is complicated by oxidation of the sulphhydryl group, probably to SO_2^- as in papain (Drenth *et al.*, 1975), during the time taken for crystallization and data collection (less than 10% of activity remains afterwards). One oxygen atom, O_1 , is clearly visible, attached to S_γ of Cys25 (it was visible also in the original 2.8 Å map as an elongation of the density for S_γ). It is hydrogen bonded to the main-chain amide of residue 25 (2.75 Å away) and the side-chain $-NH_2$ group of Gln19 (3.24 Å away), i.e. it occupies the "oxyanion hole" which forms the binding site for the carbonyl oxygen of a substrate molecule (Drenth *et al.*, 1976). The same site is occupied by one of the oxygen atoms in oxidized papain (Drenth *et al.*, 1976), and the presence of this stabilizing site, adjacent to S_γ , no doubt explains the very ready oxidation of sulphhydryl proteases.

The existence of a second oxygen atom, O_2 attached to S_γ , is less certain. In the present model, O_2 has been placed in a weak extension of the density (see Fig. 15), in a reasonable position (3.13 Å from NH(163), and with the angle $O_1-S_\gamma-O_2 = 117^\circ$). Difference maps, however, suggest slight disorder around the position of S_γ , and the placement of this second oxygen atom is somewhat uncertain as a result.

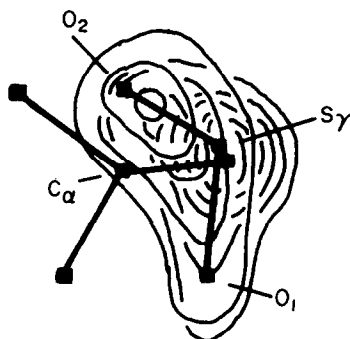


FIG. 15. Electron density for the side-chain of Cys25 (from a difference map calculated after its omission from the model). The positions of S_γ and the 2 oxygen atoms attached to it are shown.

Attempts to obtain crystals of actinidin in the fully reduced form have not so far been successful. However, there is evidence that little change in structure accompanies oxidation. The original 2.8 Å multiple isomorphous replacement map was obtained from crystals with ~50% activity, yet there was no sign of significant disorder, nor any significant difference in protein structure from that described here. Drenth *et al.* (1975, 1976) have compared fully oxidized papain with papain that is a mixture of oxidized and reduced (with cysteine bound) forms; the only major difference was a rotation of the imidazole group of His159 by 30° about its C_β — C_γ bond. The slight disorder in the position of Cys25 S_γ in actinidin still only gives a maximum uncertainty in position of 0.2 to 0.4 Å.

The arrangement of groups in the active site of actinidin generally supports the mechanism proposed by Drenth *et al.* (1975, 1976). The plane of the imidazole ring of His162 makes an angle of 40° with the indole ring of Trp184. In this orientation, $N_{\delta 1}$ points towards the position that would be occupied by the nitrogen atom of the leaving group during catalysis, and is therefore ideally arranged for proton transfer. The sulphur atom, S_γ , of Cys25 lies 2.7 Å out of the plane of the imidazole ring, so that no hydrogen bond would be possible even if the sulphur was not oxidized.

A rotation of 35° about C_β — C_γ of His162 would cause Cys25 S_γ to be coplanar with the imidazole ring, allowing a good S_γ ... $N_{\delta 1}$ hydrogen bond (distance ~3.0 Å). Furthermore, the rings of His162 and Trp184 would then be approximately parallel, about 3.5 Å apart, giving additional stabilization. This is the structure proposed for free, reduced papain, in the absence of substrate (Drenth *et al.*, 1976). Such a rotation would not disrupt the orientational His162...Asn182 hydrogen bond.

Other features of the active site remain unchanged by refinement. The side-chain of Gln19 is fixed by a hydrogen bond to the side-chain of Ser183 (19 $O_{\epsilon 1}$...183 O_γ , 2.66 Å), leaving the —NH₂ group to project into the active site (see above). The carboxyl group of Asp161, sometimes implicated in kinetic mechanisms (e.g. Zannis & Kirsch, 1978; Angelides & Fink, 1978), is turned away from the active site, as in papain, so that it hydrogen bonds with the main-chain amido group of residue 139 (161 $O_{\delta 1}$...139 N, 2.93 Å). In this position it is ~5.6 Å from $N_{\delta 1}$ of His162, and ~7.1 Å from S_γ of Cys25. Rotation about C_α — C_β of Asp161 would not bring it appreciably closer.

TABLE 13

*Environment of water molecules**A. Internal (17)*

W1...136 N; 136 O; 208 O; 211 N; (207 O)	W7...12 O; 32 O; W3	W16...121 OE2; 197 O; 205 N
W2...28 O; 50 OE1; W9; W19	W8...184 N; 185 O; 192 O	W17...33 O; 131 NE2; 132 O
W3...14 N; 35 OE1; 181 NZ; W7	W9...35 OE1; 181 NZ; W2	W19...W2; W14; W26
W4...52 OE2; 103 O; 105 O	W13...53 O; 54 O; 64 N	W23...35 O; 38 ND2; 46 N; 46 O
W5...50 N; 51 N; 86 OE1; 89 O; (49 OG)	W14...17 N; 182 O; W19	W26...181 O; W6; W19
W6...163 O; 183 OG; W26	W15...117 O; 120 N; 120 O; 209 O	

B. In active site cleft (18)

W20...W24; W75; W88	W115...20 O; W77; W217	W215...W214
W24...211 O; 213 OG; W20; W52	W126...61 O; W125; W224	W216...25 O1; W77; W223
W51...116 O; 212 O; W52; W99	W162...68 O; W88; W212	W217...21 O; W115; W124
W52...115 OD1; W24; W51; W160	W201...68 N; W222; W224	W222...23 O; 25 O2; W201; W212
W77...184 NE1; W115; W216	W212...25 O2; W162; W222	W223...25 O1; W216
W88...70 N; W20; W162	W214...139 O; W215	W224...68 O; 68 N; W126; W201

C. In surface pockets or depressions (34)

W10...125 OE1; 198 NH1; 199 OD1	W48...38 ND2; 42 OG; W150; W243	W105...143 N; W235; W239
W12...77 O; 110 N; 110 O; W72; (82 N)	W57...198 O; 199 O; W79; W131	W112...3 O; 171 OG1; 198 NH1; W103
W18...47 O; 87 N; W253	W59...138 OD2; 157 N; W29; W238	W116...4 OH; 177 OH; W156
W29...138 O; 141 N; W36; W59	W61...61 N; 72 OD2	W127...58 NH2; 63 O; 66 OD2; 95 OD2
W30...156 O; 206 N; W142; W240	W67...205 OG1; W142; W231	W142...153 O; W30; W67; W155; W231
W31...38 OD1; 109 N; 217 NZ; W150; W167	W76...50 OE2; 86 OE1; W84; (86 OE2)	W150...107 O; W31; W48
W32...191 O; W37; W172; W272	W79...176 N; W57	W155...W142; (156 N)
W35...18 OG; 20 N	W83...190 OE1; W156; W176	W156...8 NH2; W83; W116
W36...144 N; 155 O; W29	W84...18 N; 18 O; W76	W173...111 OD1; 217 NZ
W37...16 N; W32; W41	W86...127 O; 131 OE1; W189	W176...8 NH1; 190 O; W83
W41...193 OH; W22; W37; W133	W103...198 NH2; W27; W62; W112	
W47...57 O; 72 O; W94; W154		

D. *Between molecules I and I' (22)*

W11...13 N; 13 O; 158 OG1†	W65...47 N; W68	W128...61 OD1†; 63 NH2†; 88 OD1; W54; (W225)
W21...39 NZ; 45 N; W25	W68...45 O; 69 OH†; W65	W147...W256; W271; (9 OG)
W22...13 O; 15 OD1; W41; W45	W71...15 OD2; 138 OD2†; 159 O†; W157	W153...W33; W249
W25...11 O; 158 O†; W21	W93...204 O†; W240; W241	W157...159 O†; W33; W71; W250; (15 OD2)
W33...15 N; W153; W157	W99...43 O; W51; W152	W249...160 O†; W153; (45 O)
W45...158 N†; W22; W140	W123...118 OH†; W166; W255	W255...119 ND2†; W123; W256
W50...63 NH2†; 105 OE1; W118	W125...69 OH†; W54; W126	W264...15 OD1; 15 OD2; 138 OD1†; W71
W54...61 OD1†; 85 OG1; W125; W128		

† Protein atoms belonging to molecules I' at ($x, y, -1 + z$)

E. *Between molecules I and II' (26)*

W27...95 OD2†; 125 OE1; 199 ND2; W103	W101...126 OG1; W124; W218	W146...51 NE2†; 55 OD2†; W78; W119
W28...93 O†; 120 OD1; W42	W106...65 O†; W90; W218	W161...W42; W108
W39...94 OE1†; 95 N†; 199 OD1	W111...97 N†; 97 OD2†; 199 O	W186...92 O†; W187
W42...21 N†; W28; W161	W118...88 O†; W50; W151; W257; W268	W187...119 O; 202 N; W58; W80; W186
W58...121 N; 202 O; W82; W187; (119 O)	W119...W139; W146	W192...58 NH1†; 58 NE†; 96 O†; W195; W191
W62...2 O; 66 OD2†; 125 NE2; W103	W131...97 OD1†; 97 OD2†; W57	W195...176 OD2; W192
W80...202 N; W187	W132...87 O†; W137	W218...21 O†; W101; W106
W82...121 OE1; 201 N; W58; W139	W137...W132; W151; W257; W267	W257...W118; W137
W90...1 N; W106	W139...95 O†; W82; W119	

† Protein atoms belonging to molecule II' at ($-\frac{1}{2} + x, \frac{1}{2} - y, 1 - z$)

F. *Between molecules I and III (16)*

W34...75 OE1; 101 O†; W135; W149; W248	W73...78 O; 107 OH†; W70; W242	W159...104 OD2; W40; W56; W91
W40...75 NE2; 104 OD1†; 113 OH; W69;	W87...59 OG1; 81 O†; W202; W204; W272	W204...81 O; W87; W97
W159	W91...80 OD2; 104 OD2; W159	W242...W73; W243; (W141)
W56...72 OD1; 106 NZ†; W94; W159	W104...105 OE1; W248	W245...79 O; 106 O†; W117; W247; W266
W69...71 OG1; 72 OD1; W40	W135...75 NE2; W34; W149	W248...W34; W104; W188; W242
W70...60 N; W72; W73; W202		
W72...78 O; W70; W141; (W12)		

† Protein atoms belonging to molecule III at ($3/2 - x, -y, -\frac{1}{2} + z$)

TABLE 13 *continued*G. *In other intermolecular contact areas (3)*

W237...142 OD2; 149 OG† between molecule I and molecule III' at $\dagger (3/2 - x, 1 - y; -\frac{1}{2} + z)$
 W228...111 OD2; W229 between molecule I and molecule IV' at $(1 - x, -\frac{1}{2} + y, 3/2 - z)$
 W229...W134, W228

H. *External water molecules (135)*

W38...86 OE1; 91 N; W120	W145...16 O; W53; W252; W270	W209...113 OH; W160
W43...153 N; 153 OG1; 154 O; W199	W148...130 O	W210...115 O
W44...90 O; W78; W174	W149...112 OG1; W34; W135; W188	W211...W230
W46...130 O; W166	W151...88 O; W118; W137; W267; W268	W213...W63; W66; W200
W49...6 N; W96; W164; W169	W152...42 O; W99; W160	W219...21 OE1; W85; W124
W53...15 O; W55; W145; W251	W154...W47; W94; W97; W109	W220...146 O
W55...47 OG; W53; W168; W251	W158...189 O; W180; W182	W221...146 O
W60...16 O; W107	W160...W52; W75; W152; W209	W225...W128
W63...148 N; 149 N; 149 O; W213	W163...186 OG1	W226...60 O; 63 NH1; W227
W64...55 O; 56 O; 58 N; W205	W164...9 OG; 9 OG'; W49; W271; (W165)	W227...60 O; 63 NE; W236
W66...143 O; W213	W165...9 OG; (W164)	W230...97 OD1; W57; W211; W233; W234
W74...217 N; 217 O; W181	W166...W46; W123	W231...153 O; W67; W142
W75...71 OG1; W20; W160	W167...217 NZ; W31	W232...42 O
W78...55 OD1; W146; W177; W44	W168...W55; W269	W233...W230
W81...115 N	W169...4 O; W49	W234...W230
W85...123 O; W113; W197; (W219)	W170...W122; W133; W140; W171	W235...W105
W89...100 N; W179	W171...W170	W236...142 N; 142 OD2
W94...W47; W56; W97; W154	W172...W32; W133	W238...W59; W239; W263
W95...W49; W143	W174...W44; W175	W239...W105; W238; W262
W96...171 N; (171 O)	W175...99 OD2; W174; W177	W240...W30; W93
W97...79 ND2; W94; W154; W204; (W266)	W177...97 O; W78; W175	W241...201 O; W93
W98...120 ND2; W100	W178...W179	W243...W48
W100...118 O; W98	W179...W89; W178; (99 OD2)	W244...80 OD2; W117; W246
W102...147 O;	W180...186 O; 189 O; W158	W246...W117; W138; W244
W107...16 O; 185 OD2; W60	W181...131 OE1; W74	W247...107 O; W245
W108...18 O; W161; W184	W182...148 O; 189 O; W158	W250...15 OD2; W157; W251; W252
W109...56 O; 58 O; W154	W183...185 OD1; 187 OG1; W185	W251...W53; W55; W250

W110...98 O	W184...18 OG; W108; W185	W252...W145; W250
W113...W85; W196	W185...18 OG; 185 OD1; W183; W184	W253...W18; W254
W114...61 ND2	W188...111 O; W149; W248	W254...W120; W253
W117...W244; W245; W246; W266	W189...214 OH; W86; W197	W256...119 ND2; W147; W255
W120...W38; W254; (86 O)	W190...112 OG1; 113 O; 114 OE2; W194	W258...W259
W121...151 N; W198	W191...55 O; 96 O; W192	W259...W258
W122...6 OD1; W170	W193...W129; W136	W260...156 O; W140; W171; W261
W124...W101; W217; W219	W194...113 O; W190; W207	W261...156 O; W260; W262
W129...184 O; W193; (187 OG1)	W196...114 OE1; 115 O; W113	W262...W239; W261
W130...172 O	W197...W85; W189	W263...W238
W133...W41; W170; W172	W198...151 O; W121; W199	W265...40 O
W134...174 N; W229; (173 N)	W199...151 O; W43; W198; W200	W266...106 O; W117; W245; (W97)
W136...145 O; W193	W200...149 O; W199; W213	W267...W137; W151
W138...56 O; W246	W202...W70; W87; W203	W268...W118; W151
W140...W45; W170; W260	W203...W202; W205	W269...W168; W270
W141...111 O; W72; (W242)	W205...W64; W203	W270...W145; W269
W143...129 O; W95	W306...176 OD1	W271...9 OG; W147; W164
W144...154 O	W207...113 OH; W194; W208	W272...W32

Beyond the catalytic site (see Fig. 16) is the non-polar binding site for a substrate side-chain (the S_2 site, following the terminology of Berger & Schechter (1970)). In actinidin this is lined by the side-chains of Tyr69, Ile70, Ala136, Val160, Ala163 and Met211. The side-chain of Met211 stretches across the end of this pocket, making it noticeably shorter than in papain.

There are a number of water molecules in the active site cleft (see Fig. 16 and Table 13). However, with the exception of W77, attached to $N_{\epsilon 1}$ of Trp184, the only well-ordered water molecules are those in and beyond the S_2 site. Otherwise only poorly ordered, weakly bound solvent is found in the catalytic site, suggesting that it does not have any important *structural* role in the catalytic mechanisms.

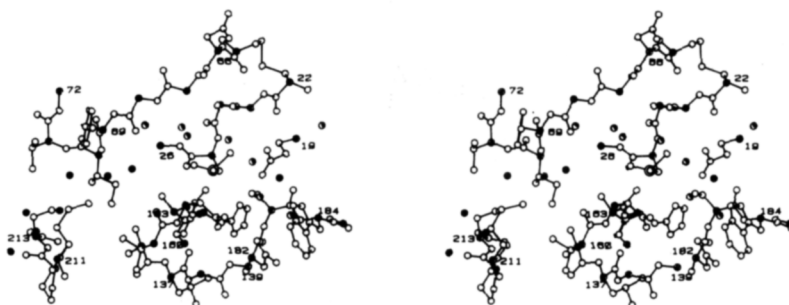


FIG. 16. The entire active site cleft in actinidin. The catalytic site lies towards the right (the sulphur atom of Cys25 shown as a double circle). Towards the left is a non-polar pocket (the S_2 site) flanked by the side-chains of Tyr69, Ile70, Ala136, Val160, Ala163 and Met211. Well ordered water molecules are shown as filled circles, weakly bound, poorly ordered water molecules as half-filled circles.

(h) Solvent structure

All solvent molecules or ions have been taken as water molecules except for W4 (probably an ammonium ion; see section (e)(iv), above). The 163 refined water molecules have been sorted, and numbered according to their B values, so that for water molecules W1 to W163 the serial number gives an indication of their reliability (W1 is the best-ordered, with $B = 7.2 \text{ \AA}^2$). As the distribution in Figure 17 shows, many are very well-ordered, with $B < 20 \text{ \AA}^2$, while the highest B value is only $40 \text{ \AA}^2$, less than those of the more mobile amino acid side-chains. The remaining 109 water molecules (W164 to W272) have not been refined and there is no significance in their serial numbers, except that all are less-ordered and less-reliable than those which have been refined.

Water molecules are listed in Table 13, according to their environments in the structure. Also listed are hydrogen-bonding partners for all water molecules, hydrogen bonds being assumed where the contact distance is less than $3.3 \text{ \AA}$, and angles (e.g. in the case of protein NH groups) are reasonable. The best-ordered water molecules are those which are internal (all 17 internal waters have $B < 15 \text{ \AA}^2$) but other well-ordered water molecules are found either in crevices in the protein surface or at the contact areas between molecules in the crystal. Of the best 65 water molecules, with $B < 25 \text{ \AA}^2$, 17 are internal, 20 are in surface pockets or crevices, 18 lie in inter-

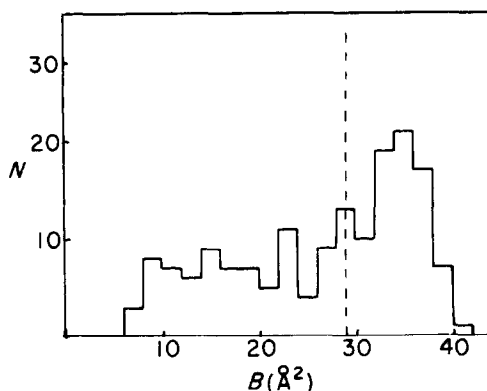


FIG. 17. Distribution of B values ($\text{\AA}^2$) for the 162 refined water molecules. Many are firmly bound, and well-ordered (46 of them have $B < 20 \text{ \AA}^2$). N , number of molecules.

molecular contact areas, and only 10 are classed as external (i.e. water molecules on or outside the protein surface, in contact with the main solvent areas in the crystal). Most of the external water molecules have fewer protein ligands and are significantly less-ordered than those bound in crevices, intermolecular sites, etc.

Internal water molecules, having no contact with the external solution, have been found in other refined protein structures, e.g. there are 25 such in trypsin (Bode & Schwager, 1975), and present indications are that they may be a common feature of proteins of this size. In actinidin the internal water molecules fall into two categories. Some (W1, W4, W5, W8, W13, W15, W16, W17, W23) occupy discrete sites, in which a single water molecule occupies a small cavity and is bound exclusively to protein atom ligands. These water molecules fill otherwise empty spaces, and in doing so probably impart some extra stability to the structure by hydrogen-bonding with internal polar groups which would otherwise have no hydrogen-bonding partner (in each case at least one, and usually two, of the protein ligand atoms is in this situation). All nine of these discrete internal water molecules have either three or four protein ligands, two of which are invariably proton acceptors (16 out of 18 are carbonyl oxygens), the third a proton donor (usually backbone $-\text{NH}$), and the fourth, where present, is usually also a proton donor. The only exceptions are the ammonium ion W4 (3 oxygen ligands, all acceptors), and W15 (3 oxygen acceptors, 1 NH donor).

Two typical discrete internal water sites are shown in Figure 18. One (W8) is three-co-ordinate. The other (W1) has four ligands arranged approximately tetrahedrally, but has a fifth possible ligand (207 O) somewhat further away ($3.33 \text{ \AA}$). Inspection of the geometry suggests that a bifurcated hydrogen may exist, one water proton being directed between 136 O and 207 O. More "regular" five-co-ordinate water molecules (i.e. with 5 close neighbours) are not unusual (as has previously been noted in the crystal structure of rubredoxin (Watenpaugh *et al.*, 1978)). Eleven such are listed in Table 13, and there are a number of four-co-ordinate waters which, like W1, have a fifth neighbour only slightly more distant.

The remaining eight internal water molecules are interconnected in a cluster at the centre of the domain-domain interface (Figs 12 and 13). Again, each has either

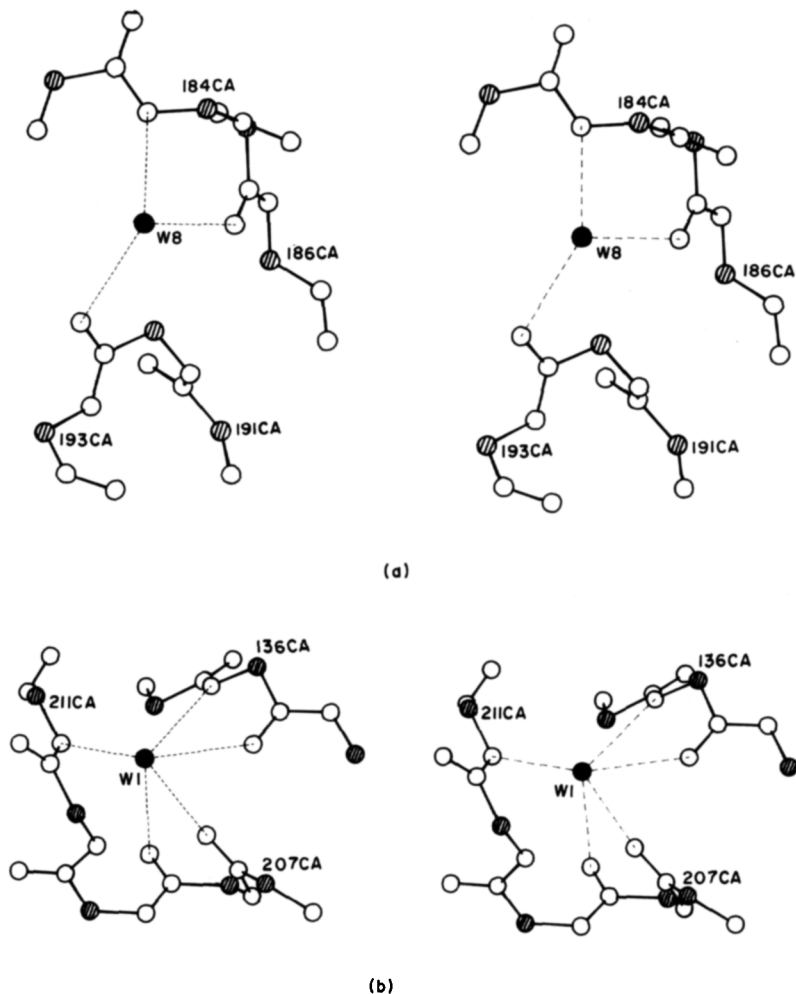


FIG. 18. Two typical binding sites for discrete internal water molecules (●). In (a) W8 has 3 ligands in an approximately trigonal arrangement, 2 proton acceptors (185 O 2.88 Å; 192 O 2.91 Å) and 1 proton donor (184 N 2.83 Å). In (b) W1 has 4 close neighbours, 2 proton acceptors (208 O 2.74 Å; 136 O 2.87 Å) and 2 proton donors (211 N 2.87 Å; 136 N 3.14 Å), as well as a 5th neighbour somewhat more distant (207 O 3.33 Å). Three of the ligands (211 N, 208 O, 136 N) are tetrahedrally disposed and the 4th tetrahedral position lies between 136 O and 207 O, suggesting a bifurcated hydrogen bond (the proton directed between 136 O and 207 O).

three or four hydrogen-bonding partners, two being proton acceptors and the remaining one or two proton donors. Thus all water protons are utilized. Those waters with only three ligands generally have the fourth position blocked by a non-hydrogen-bonding group. Most of the water molecules in this network have two or three protein ligands, but one (W19) in the centre of the cluster, is bound only to other waters.

All the internal water molecules can be considered an integral part of the protein structure, present also in solution. The same may also be true of some of the water

molecules bound in pockets on the surface. For example, W29 and W36 occupy a deep crevice between two antiparallel sections of polypeptide chain, 138 to 143 and 155 to 157 (see Fig. 19) and are probably essential to the structure. It will be interesting to see whether these and other water molecules occupying pockets formed mainly between main-chain atoms (e.g. W12, W18, W47) are found in papain as well (where the main-chain conformation is the same but many side-chains are different). Other tightly bound water molecules, e.g. W10 (see Fig. 19), W31 and W48, occupy pockets formed by polar side-chains. These help to stabilize the surface conformation of the molecule.

A monolayer of water molecules covers most of the surface of the protein. Seldom is it more than one molecule thick, however, except at or near the points of intermolecular contact. This simply reflects the fact that only the better-ordered water molecules can be identified as such, and where the protein surface is adjacent to a large body of liquid, the water structure rapidly becomes less-ordered. Of the 163

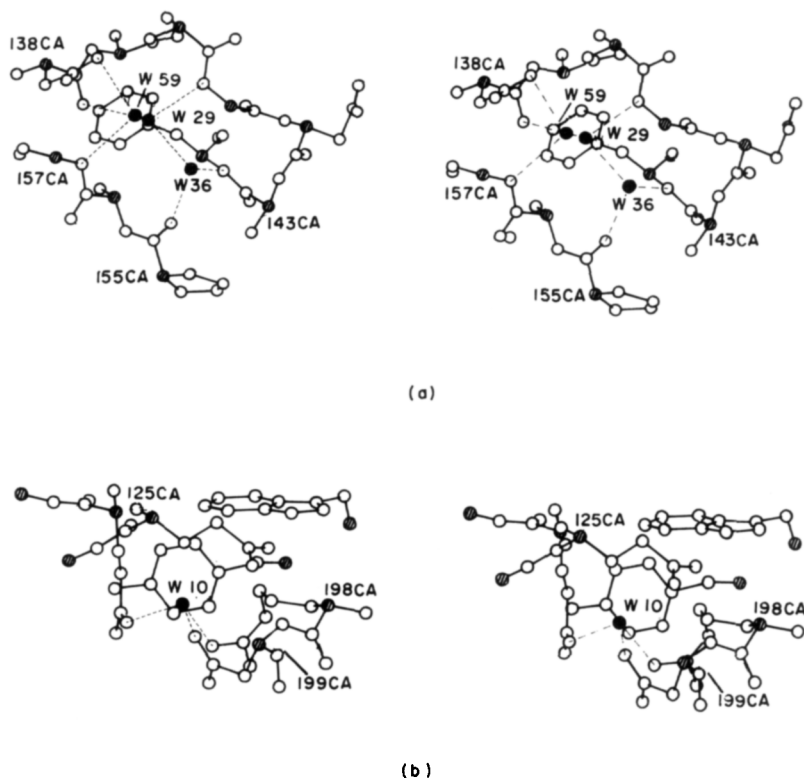


FIG. 19. Water molecules (●) tightly bound in surface pockets. In (a) 2 water molecules, W29 and W36, occupy a deep crevice between 2 sections of the main-chain. A 3rd water molecule, W59, is also shown, somewhat above the crevice. In (b) W10 occupies a deep pocket formed by the side-chains of Gln125, Arg198 and Asn199. Only 3 hydrogen bonds are formed (125 OE1 2.88 Å; 198 NH2 2.96 Å; 199 OD1 2.70 Å), the potential fourth site blocked by the side-chain of Trp178. In both (a) and (b) the water molecules must be considered an integral part of the protein structure.

“best” water molecules (those that have been refined) only 13 are not hydrogen-bonded to a protein atom. Of the 109 more poorly ordered waters (not refined) 42 have no protein ligand. Further, weaker, peaks can be seen in difference maps but are below the error level.

As has been observed in other proteins (Finney, 1977) the more firmly bound surface water molecules sometimes act as “anchoring points” for further water molecules extending out from the protein surface. Many bridges occur between polar groups on the surface; the side-chains of Arg8 and Glu190 are linked by W83 and W176 (a water-mediated salt bridge?); W83, W156, and W116 form a bridge between Glu190 and Tyr177; and W76 and W84 form a bridge across the end of the domain I-domain II interface, covering the exposed edge of Phe28 in the process. More extensive networks fill shallow depressions in the protein surface. In several places networks of water molecules cover the protruding edges of non-polar side-chains. In domain II the edges of Phe152 and Tyr147 project out from the centre of a large loop of the polypeptide chain (residues 143 to 154) and are covered by a monolayer of eight water molecules (see Fig. 20). Similar clusters are seen covering the exposed

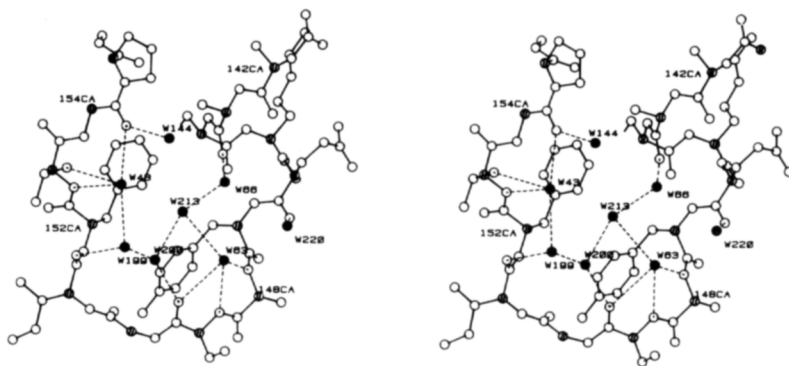


FIG. 20. A layer of 8 water molecules (●) covering the exposed edges of the aromatic rings of Tyr147 and Phe152. The network is “anchored” by attachment to atoms of the surrounding polypeptide chain (residues 143 to 154).

edges of the aromatic rings of Phe76 and Tyr89. Where non-polar groups project more fully from the surface, water molecules are sometimes found to be arranged around them, but only when there are adjacent polar groups to which any such clusters can be anchored. Even then the water is not well-ordered. The largest hydrophobic patch on the surface (the side-chains of Val200, Val175, Leu197 and Ile151) is not associated with any ordered water; clearly the water structure is much more mobile around such regions.

As in other protein structures, e.g. rubredoxin (Watenpaugh *et al.*, 1978), bovine pancreatic trypsin inhibitor (Deisenhofer & Steigemann, 1975) and erythrocrucorin (Steigemann & Weber, 1979), extensive networks of water molecules are found at the regions of contact between molecules in the crystal (see Table 13). Many of these water molecules may be present only as a consequence of the crystal packing, but some, interacting with several protein ligands, would certainly remain despite it.

Many good hydrogen bonds are made by these intermolecular water molecules, and undoubtedly contribute greatly to the stability of the crystal lattice. The number of intermolecular interactions involving water far outweigh direct interactions between protein atoms (see below).

In general, the degree to which water molecules are ordered can be correlated with the number of hydrogen bonds they form with protein atoms. The "best" 54 water molecules (W1 to W54) have an average 2.2 protein ligands; the next 54 (W55 to W108) average 1.6; the next 55 (W109 to W163) average 1.1; while the remaining 109 average 0.8 protein ligands. Carbonyl oxygen atoms are by far the most common protein ligands for water molecules (41%), followed by main-chain NH groups (18%). Amongst side-chains, carboxyl groups hydrogen bond most often to water (39% of side-chain interactions involve carboxyl oxygens) but this partly reflects the large proportion of acidic side-chains in actinidin.

It must be emphasized that the solvent structure described here only represents about 30% of the total water content of the crystals. More peaks can be seen (in difference maps), particularly further out from the protein surface; some may be real

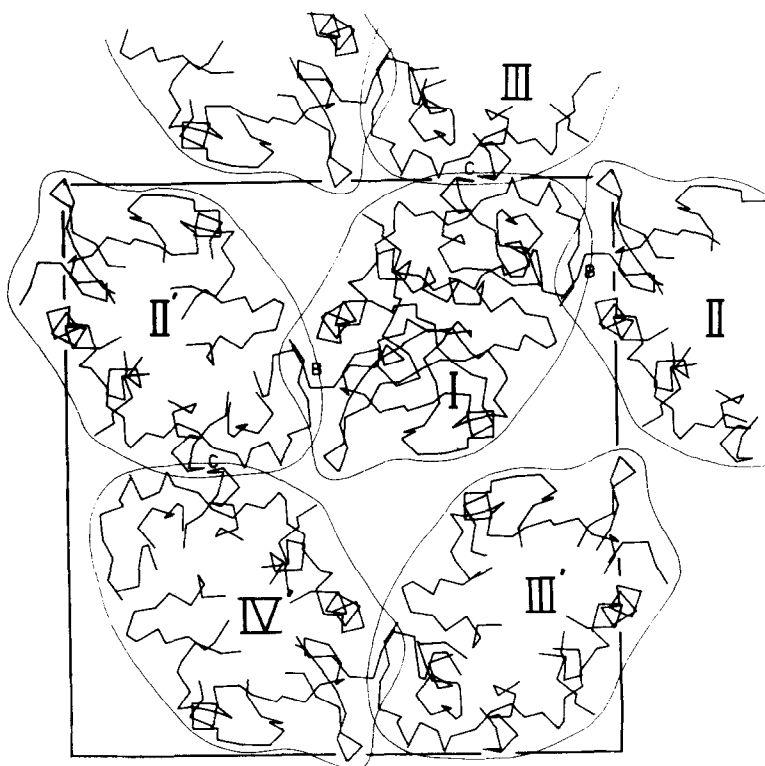


FIG. 21. The packing of the molecules of actinidin within the crystal. The reference molecule I (x, y, z) makes contact with molecule II ($\frac{1}{2} + x, \frac{1}{2} - y, 1 - z$) at region B, molecule III ($\frac{3}{2} - x, -y, -\frac{1}{2} + z$) at region C, and molecule I' ($x, y, 1 + z$) at region A (not shown). Intermolecular interactions at these points of contact are listed in Table 14. Water molecules in these regions are listed in Table 13.

(if poorly ordered, or partly occupied) water sites, but others may represent only noise.

(i) *Packing of the molecules in the crystal*

The packing of the molecules is shown diagrammatically in Figure 21, and intermolecular contacts are listed in Table 14.

The three main regions of contact are region A, involving the top of one molecule (residues 60 to 63, 69, 118, 158 to 159, 202 to 203, 210) and the bottom of the next

TABLE 14

Intermolecular contacts

A. *Region A* (between molecules I and I'; see text)

H-bonds:	8 O...158 OG1	2.40 Å	9 O...203 N	2.94 Å
	10 O...119 ND2	2.86 Å	39 NZ...210 O	2.91 Å
	40 O...118 OH	2.68 Å	39 NZ...210 OG1	2.81 Å
	88 ND2...61 OD1	3.00 Å	88 OD1...63 NH2	3.09 Å
van der Waals':	Leu45 CD1...Ala159 CB	(4.0 Å)		
	Gly43 CA...Tyr118 ring	(3.6 Å)		
	Gly11 CA...Thr158 OG1	(3.5 Å)		
Water bridges:	13 N, 13 O...W11...158 OG1		11 O...W25...158 O	
	105 OE1...W50...63 NH2		85 OG1...W54...61 OD1	
	45 O...W68...69 OH		15 OD2...W71...159 O, 138 OD2	
	88 OD1...W128...61 OD1			

B. *Region B* (between molecules I and II; see Fig. 21)

H-bonds:	21 O...122 NE1	2.99 Å	66 OD1...1 N	2.47 Å
	97 OD2...199 O	3.14 Å	94 OE1...122 N	2.91 Å
			94 NE2...120 OD1	3.04 Å
van der Waals':	Cys65 C,O...Trp122 ring	(~4 Å)		
Water bridges:	95 OD2...W27...125 OE1, 199 ND2		93 O...W28...120 OD1	
	95 N, 94 OE1...W39...199 OD1		97 N...W111...199 O	
	66 OD2...W62...125 NE2, 2 O			

C. *Region C* (between molecules I and III; see Fig. 21)

H-bonds:	59 OG1...106 NZ	3.23 Å	60 NE2...79 O	3.14 Å
	60 OE1...106 NZ	3.18 Å	79 ND2...104 O	3.06 Å
Water bridges:	75 OE1...W34...101 O		78 O...W73...107 OH	
	72 OD1...W56...106 NZ		59 OG1...W87...81 O	
	104 OD1...W40...113 OH, 75 NE2			

D. *Others*

Between molecules I and IV (Fig. 21)	Tyr218 ring...173-174 peptide	3.5 Å
Between molecules I and III' (Fig. 21)	149 OG...W237...142 OD2	

along the z direction (residues 8 to 13, 15, 39 to 45, 88); region B, between residues 1 to 2, 120 to 125, and 198 to 202 of one molecule and 21, 66 and 92 to 99 of a symmetry-related molecule; and region C, involving residues 59 to 61, 75 to 80 and 101 to 107 of one molecule and the same residues (101 to 107, 75 to 80, 59 to 61) of a symmetry-related neighbour.

Although actinidin has only a moderate solvent content (42% by vol.), only 20 side-chains make direct contacts (<4 Å) with neighbouring molecules. There are several "patches" of exposed non-polar side-chains on the surface, but none of these are involved in the intermolecular contacts; almost all the interactions are hydrogen bonds between polar side-chains or water bridges. Only a few of the water bridges (those involving single water molecules) are listed in Table 14. Most form complex networks.

In general, intermolecular contacts seem likely to have little effect on the overall molecular structure, consistent with solution studies on other proteins (Rupley, 1969) and evidence from other crystal structures (Drenth, 1972). Nine carbonyl oxygens form intermolecular hydrogen bonds, but a preliminary comparison with the structure of papain shows that most of these have very similar orientations in the two proteins, despite different crystal packing. The active site is well removed from neighbouring molecules and should permit binding studies once crystals of the fully reduced enzyme can be prepared. The only intermolecular contact near this region involves one side-chain (Tyr69) which is part of the S_2 site and is linked through a water molecule to a neighbouring molecule.

I am very grateful to the many colleagues who assisted and encouraged my work on actinidin; in particular Dr Bryan Anderson for his help and advice during the preparation of this manuscript.

REFERENCES

- Agarwal, R. C. (1978). *Acta Crystallogr. sect. A*, **34**, 791–809.
 Angelides, K. J. & Fink, A. L. (1978). *Biochemistry*, **17**, 2659–2668.
 Arcus, A. C. (1959). *Biochim. Biophys. Acta*, **33**, 242–244.
 Arnott, S. & Wonacott, A. J. (1966). *J. Mol. Biol.* **21**, 371–383.
 Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C. & Sternberg, M. J. E. (1979). *Nature (London)*, **280**, 563–568.
 Baker, E. N. (1973). *J. Mol. Biol.* **74**, 411–412.
 Baker, E. N. (1976). *J. Mol. Biol.* **101**, 185–196.
 Baker, E. N. (1977). *J. Mol. Biol.* **115**, 263–277.
 Baker, E. N. & Dodson, E. J. (1980). *Acta Crystallogr.*, in the press.
 Berger, A. & Schechter, I. (1970). *Phil. Trans. Roy. Soc. ser. B*, **257**, 249–264.
 Birktoft, J. J. & Blow, D. M. (1972). *J. Mol. Biol.* **68**, 187–240.
 Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rérat, B. & Rérat, C. (1978). *J. Mol. Biol.* **121**, 339–356.
 Bode, W. & Schwager, P. (1975). *J. Mol. Biol.* **98**, 693–717.
 Carne, A. & Moore, C. H. (1978). *Biochem. J.* **173**, 73–83.
 Chothia, C. (1973). *J. Mol. Biol.* **75**, 295–302.
 Crawford, J. L., Lipscomb, W. N. & Schellman, C. G. (1973). *Proc. Nat. Acad. Sci., U.S.A.* **70**, 538–542.
 Deisenhofer, J. & Steigemann, W. (1975). *Acta Crystallogr. sect. B*, **31**, 238–250.
 Dodson, E. J., Isaacs, N. W. & Rollett, J. S. (1976). *Acta Crystallogr. sect. A*, **32**, 311–315.

- Drenth, J. (1972). *Fed. Eur. Biochem. Soc. Meet. Proc.* **29**, 1-7.
- Drenth, J., Jansonius, J. N., Koekoek, R. & Wolthers, B. G. (1971). *Advan. Protein Chem.* **25**, 79-115.
- Drenth, J., Swen, H. M., Hoogenstraaten, W. & Sluyterman, L. A. A. E. (1975). *Proc. Kon. Ned. Akad. v. Wetensch., Amsterdam, ser. C*, **78**, 104-110.
- Drenth, J., Kalk, K. H. & Swen, H. M. (1976). *Biochemistry*, **15**, 3731-3738.
- Finney, J. L. (1977). *Phil. Trans. Roy. Soc. ser. B*, **278**, 3-32.
- Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. (1979). *Nature (London)*, **280**, 558-563.
- Glazer, A. N. & Smith, E. L. (1971). In *The Enzymes* (Boyer, P. D., ed.), vol. 3, pp. 501-546, Academic Press, New York.
- Isaacs, N. W. & Agarwal, R. C. (1978). *Acta Crystallogr. sect. A*, **34**, 782-791.
- IUPAC-IUB Commission on Biochemical Nomenclature (1970). *J. Mol. Biol.* **52**, 1-17.
- Levitt, M. & Chothia, C. (1976). *Nature (London)*, **261**, 552-558.
- Lowe, G. (1976). *Tetrahedron*, **32**, 291-302.
- Lowe, G. & Yuthavong, Y. (1971). *Biochem J.* **124**, 107-115.
- Luzzatti, V. (1952). *Acta Crystallogr.* **5**, 802-810.
- McDowall, M. A. (1970). *Eur. J. Biochem.* **14**, 214-221.
- Mitchel, R. E. J., Chaiken, I. M. & Smith, E. L. (1970). *J. Biol. Chem.* **245**, 3485-3492.
- Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201-228.
- Nemethy, G., Phillips, D. C., Leach, S. J. & Scheraga, H. A. (1967). *Nature (London)*, **214**, 363-365.
- Parsons, S. M. & Raftery, M. A. (1970). *Biochem. Biophys. Res. Commun.* **41**, 45-49.
- Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909-933.
- Reeke, G. N., Becker, J. W. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1525-1547.
- Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 2574-2578.
- Rupley, J. A. (1969). In *The Structure and Stability of Biological Macromolecules* (Timasheff, S. N. & Fasman, G. D., eds), pp. 291-352, Marcel Dekker, New York.
- Sawyer, L., Shotton, D. M., Campbell, J. W., Wendell, P. L., Muirhead, H., Watson, H. C., Diamond, R. & Ladner, R. C. (1978). *J. Mol. Biol.* **118**, 137-208.
- Shefter, E. (1970). *J. Chem. Soc. (B)*, 903-906.
- Steigemann, W. & Weber, E. (1979). *J. Mol. Biol.* **127**, 309-338.
- Venkatachalam, C. M. (1968). *Biopolymers*, **6**, 1425-1436.
- Watenpaugh, K. D., Margulis, T. N., Sieker, L. C. & Jensen, L. H. (1978). *J. Mol. Biol.* **122**, 175-190.
- Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1979). *J. Mol. Biol.* **131**, 509-522.
- Zannis, V. I. & Kirsch, J. F. (1978). *Biochemistry*, **17**, 2669-2674.