Phil. Trans. R. Soc. Lond. B **319**, 369–456 (1988) [369]
Printed in Great Britain

THE STRUCTURE OF 2Zn PIG INSULIN CRYSTALS AT 1.5 Å RESOLUTION

BY EDWARD N. BAKER[1], THOMAS L. BLUNDELL, F.R.S.[2], JOHN F. CUTFIELD[3], SUSAN M. CUTFIELD[3], ELEANOR J. DODSON[4], GUY G. DODSON[4], DOROTHY M. CROWFOOT HODGKIN, F.R.S.[5], RODERICK E. HUBBARD[4], NEIL W. ISAACS[6], COLIN D. REYNOLDS[7], KIWAKO SAKABE[8], NORIOSHI SAKABE[8] AND NUMMINATE M. VIJAYAN[9]

[1] *Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand*
[2] *Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, U.K.*
[3] *Department of Biochemistry, University of Otago, Box* 56, *Dunedin, New Zealand*
[4] *Department of Chemistry, University of York, Heslington, York YO1 5DD, U.K.*
[5] *Chemical Crystallography Laboratory, University of Oxford,* 9 *Parks Road, Oxford OX1 3PD, U.K.*
[6] *St Vincents Institute of Medical Research, Fitzroy, Victoria, Australia* 3065
[7] *Department of Biophysics, Liverpool Polytechnic, Liverpool L3 3AF, U.K.*
[8] *National Laboratory for High Energy Physics, Oho-machi, Tsukuba-gun, Ibaraki-Ken,* 305, *Japan*
[9] *Biophysics Unit, Indian Institute of Science, Bangalore* 560012, *India*

(*Received* 16 *February* 1987)

[Plates 1–5]

CONTENTS

 [Published 6 July 1988

The paper describes the arrangement of the atoms within rhombohedral crystals of 2Zn pig insulin as seen in electron density maps calculated from X-ray data extending to 1.5 Å (1 Å = 10^{-10} m = 10^{-1} nm) at room temperature and refined to $R = 0.153$. The unit cell contains 2 zinc ions, 6 insulin molecules and about 3×283 water molecules. The atoms in the protein molecules appear well defined, 7 of the 102 side chains in the asymmetric unit have been assigned alternative disordered positions. The electron density over the water molecules has been interpreted in terms of 349 sites, 217 weighted 1.0, 126 weighted 0.5, 5 at 0.33 and 1 at 0.25 giving *ca.* 282 molecules.

The positions and contacts of all the residues belonging to the two A and B chains of the asymmetric unit are shown first and then details of their arrangement in the two insulin molecules, 1 and 2, which are different. The formation from these molecules of a compact dimer and the further aggregation of three dimers to form a hexamer around two zinc ions, follows. It appears that in the packing of the hexamers in the crystal there are conflicting influences; too-close contacts between histidine B5 residues in neighbouring hexamers are probably responsible for movements of atoms at the beginning of the A chain of one of the two molecules of the dimer that initiate movements in other parts, particularly near the end of the B chain. At every stage of the building of the protein structure, residues to chains of definite conformation, molecules, dimers, hexamers and crystals, we can trace the effect of the packing of like groups to like, aliphatic groups together, aromatic groups together, hydrogen-bonded structures, positive and negative ions. Between the protein molecules, the water is distributed in cavities and channels that are continuous throughout the crystals. More than half the water molecules appear directly hydrogen bonded to protein atoms. These are generally in contact with other water molecules in chains and rings of increasing disorder, corresponding with their movement through the crystals.

Within the established crystal structure we survey next the distribution of hydrogen bonds within the protein molecules and between water and protein and water and water; all but eight of the active atoms in the protein form at least one hydrogen bond.

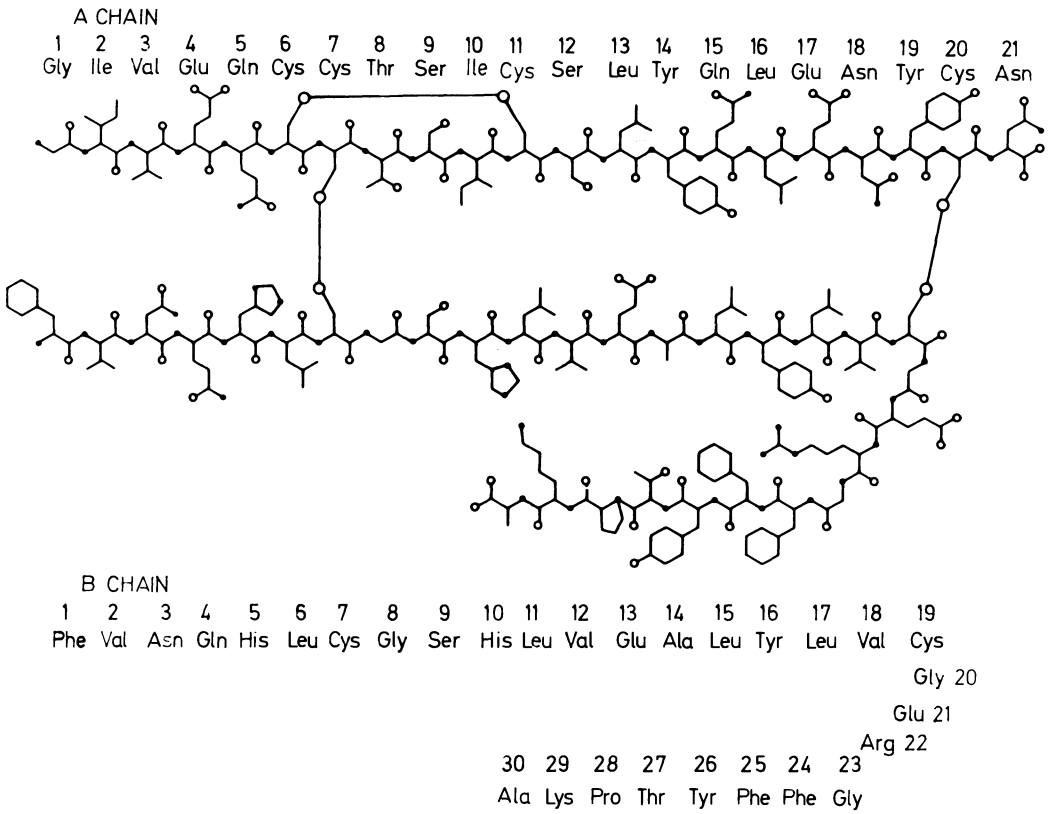
We follow with a discussion of the effect of different contacts on the observed thermal parameters and the possibility of correlating these with movements of the monomer, dimer or hexamer as a whole. The correlation seems best for molecule 1 in the dimer.

Finally we examine the relation of the crystal structure as a whole to the biological activity of insulin. The large size of the insulin receptor makes it likely that when it combines to form the receptor complex, it makes a large number of contacts with the surface of the insulin molecule. Some of these points of contact, such as, for example, B24 and B25 phenylalanine, are suggested by the changes in biological activity observed when these residues are modified. The conformational changes in the insulin chains produced by crystal packing can be seen as a model for possible changes induced by insulin contacts with the receptor that eventually we may hope to discover if the insulin-receptor complex is crystallized.

0. Introduction

The rhombohedral form of crystalline insulin first obtained by J. J. Abel in 1925 (Abel 1926) has now been explored in several different investigations. The size and shape of the unit cell was defined originally for air dry and wet cattle insulin crystals by X-ray diffraction in 1935 (Crowfoot 1935, 1937) and 1937 (Crowfoot & Riley 1938). Later work has, however, concentrated on pig insulin crystals, purified and grown by methods developed by

J. Schlichtkrull (1958). The essential structure of these crystals was solved independently in Oxford (Adams *et al.* 1969; Blundell *et al.* 1972) and in Peking (Peking Insulin Structure Group 1971), at first at resolutions of 2.8 and 2.5 Å† respectively; both solutions have been further refined since and calculations continue, particularly on new data collected at lower temperatures in China by the Peking Insulin Structure Research Group and in Japan by N. and K. Sakabe and K. Sasaki. Here we describe the conclusions we have reached through measurements at room temperature (*ca.* 20 °C) of X-ray data extending to 1.5 Å spacing following the procedures of structure solution and refinement (to be described in Hodgkin *et al.* 1988*a*, *b* (in preparation)).



FORMULA I.

The chemical structure of pig insulin (Brown *et al.* 1955) is given in formula 1.

Throughout this account we shall keep in mind the relation between the structure of the molecule and its biological activity.

1. CHARACTERISTICS OF THE CRYSTALS AND THE X-RAY ANALYSIS

1.1. *Preparation and crystal data*

The crystals were grown by dissolving 0.05 g pig insulin in 5 ml 0.02 M hydrochloric acid and then adding the following components in the order given: 0.5 ml 0.12 M zinc sulphate; 2.5 ml 0.2 M trisodium citrate; 1.5 ml acetone; 0.5 ml water. The mixture was filtered through a

† 1 Å = 10^{-10} m = 10^{-1} nm.

Millipore H A filter or Whatman no. 50 filter paper and left to stand at room temperature for several days. The pH was 6.2. The crystals were flat rhombohedra up to 0.3 mm across in early preparations and larger as the methods of purifying insulin, due to Schlichtkrull, improved.

Early analyses by J. Schlichtkrull (letter dated 10 December 1957) give the composition of pig insulin crystals grown in this way as follows.

Batch No. 111157 14.42% N, 5.1% H_2O, 0.37% Zn, 0.1% Na, < 0.02% Cl, 1.7% citric acid as sodium citrate – various batches.

The formal crystal data may now be given as follows: 2zinc insulin (pig); rhombohedral {100} hexagonal {10$\bar{1}$1}, $a = 49.0$ Å, $\alpha_R = 114.8°$, $z = 3$ (rhombohedral setting); $a = 82.5$ Å, $c = 34.0$ Å, $z = 9$ (hexagonal setting), $D_m = 1.245$ g cm^{-3} by flotation in density columns prepared from toluene and bromobenzene. V, asymmetric unit = 22250 Å^3; M, asymmetric unit = 16690. $F(000) = 9017$, $F(000)/V = 0.41$. The rhombohedral unit cell contains six molecules of pig insulin of the composition of formula 1, C256 N65 O76 S6 H388 and 2Zn. In addition, the chemical analysis above suggests the presence of 2 Na (calc. 0.122%, obs. 0.10%) and three molecules of citric acid, $C_6O_7H_6$, (calc. 1.55%, obs. 1.7%). These have not been defined for certain in our present crystal structure in the presence of water molecules, often disordered, of equivalent scattering power (later analyses give rather smaller percentages both of sodium and citrate ions in recent crystal batches). The rhombohedral unit cell contents may accordingly be given as 2Zn, 2 Na, 3|C512 N130 O152 S12 H776, insulin| 3|$C_6O_7H_6$, citric acid| and 3|270 H_2O| or 2Zn 2 Na 3|2 insulin 281 H_2O| or 2Zn 3|2 insulin 283 H_2O|. Correspondingly, the asymmetric unit contains two insulin molecules, 1 and 2, $\frac{2}{3}$ zinc atom and some 283 other atoms, mainly water.

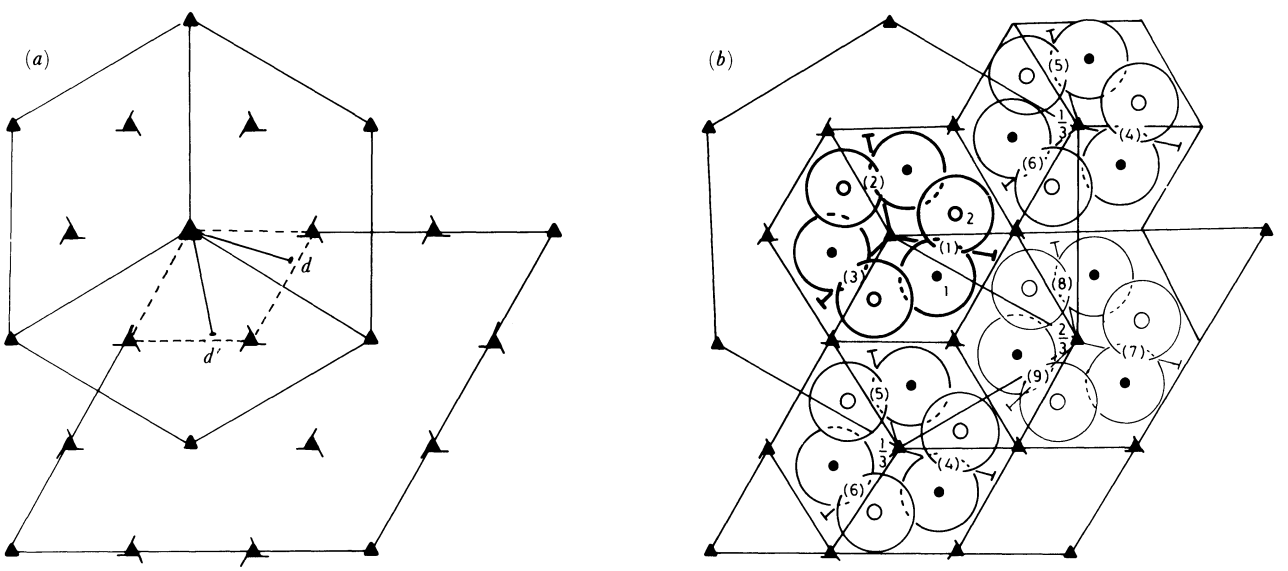


FIGURE 1.1. (*a*) The relation between the rhombohedral and hexagonal unit cells. The general coordinates of the nine units required by the symmetry elements of the hexagonal cell are

$0\,0\,0+(1)\,xyz$ (2) $\bar{y}, x-y, z$ (3) $\bar{x}+y, x, z$,
$\frac{2}{3}\,\frac{1}{3}\,\frac{1}{3}+(4)\,xyz$ (5) $\bar{y}, x-y, z$ (6) $\bar{x}+y, x, z$,
$\frac{1}{3}\,\frac{2}{3}\,\frac{2}{3}+(7)\,xyz$ (8) $\bar{y}, x-y, z$ (9) $\bar{x}+y, x, z$.

The dotted line shows the asymmetric unit employed in calculations. Within this are marked the local twofold axes, *d* and *d'*. (*b*) Diagram to illustrate the packing of the insulin molecules 1 and 2, belonging to each of the units in the insulin unit cells. They are represented by spheres centred on the approximate mass centres of the molecules as observed.

TABLE 1.1. ATOMIC PARAMETERS

(Coordinates in orthogonal ångströms.)

residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å^2	residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å^2
A1.1	Gly	N	−0.64	19.96	−14.07	1.0	26	A11.1	Cys	SG	7.32	13.97	−9.30	1.0	23
A1.1	Gly	CA	−0.39	20.03	−12.61	1.0	31	A11.1	Cys	CB	8.28	14.93	−10.50	1.0	24
A1.1	Gly	C	0.45	18.83	−12.18	1.0	34	A11.1	Cys	CA	9.49	14.18	−11.06	1.0	14
A1.1	Gly	O	1.22	18.31	−13.01	1.0	21	A11.1	Cys	C	10.30	13.73	−9.87	1.0	24
A2.1	Ile	N	0.24	18.43	−10.94	1.0	24	A11.1	Cys	O	10.24	12.52	−9.60	1.0	28
A2.1	Ile	CD1	1.08	15.83	−6.72	1.0	20	A12.1	Ser	N	10.94	14.68	−9.24	1.0	21
A2.1	Ile	CG1	1.54	16.14	−8.16	1.0	16	A12.1	Ser	OG	12.28	16.74	−7.89	1.0	26
A2.1	Ile	CB	0.49	17.01	−8.93	1.0	16	A12.1	Ser	CB	12.90	15.48	−7.91	1.0	22
A2.1	Ile	CG2	−0.97	16.53	−8.86	1.0	17	A12.1	Ser	CA	11.74	14.43	−8.10	1.0	20
A2.1	Ile	CA	1.00	17.29	−10.39	1.0	15	A12.1	Ser	C	10.98	14.61	−6.78	1.0	26
A2.1	Ile	C	0.95	16.02	−11.19	1.0	14	A12.1	Ser	O	9.93	15.24	−6.84	1.0	18
A2.1	Ile	O	1.97	15.36	−11.28	1.0	17	A13.1	Leu	N	11.52	14.16	−5.65	1.0	21
A3.1	Val	N	−0.18	15.65	−11.79	1.0	14	A13.1	Leu	CD2	12.39	11.59	−2.27	1.0	43
A3.1	Val	CG2	−2.57	13.99	−11.66	1.0	17	A13.1	Leu	CD1	10.21	11.56	−3.63	1.0	33
A3.1	Val	CG1	−1.81	12.87	−13.74	1.0	31	A13.1	Leu	CG	11.62	12.14	−3.47	1.0	38
A3.1	Val	CB	−1.71	14.14	−12.91	1.0	26	A13.1	Leu	CB	11.45	13.62	−3.26	1.0	25
A3.1	Val	CA	−0.28	14.45	−12.59	1.0	18	A13.1	Leu	CA	10.83	14.39	−4.42	1.0	19
A3.1	Val	C	0.59	14.45	−13.88	1.0	21	A13.1	Leu	C	10.81	15.86	−4.10	1.0	18
A3.1	Val	O	1.25	13.45	−14.20	1.0	19	A13.1	Leu	O	9.86	16.31	−3.43	1.0	18
A4.1	Glu	N	0.62	15.59	−14.50	1.0	24	A14.1	Tyr	N	11.76	16.64	−4.54	1.0	19
A4.1	Glu	OE2	−1.53	17.97	−15.18	1.0	28	A14.1	Tyr	OH	17.33	17.07	−1.47	1.0	40
A4.1	Glu	OE1	−2.43	16.17	−16.21	1.0	37	A14.1	Tyr	CD2	15.06	17.29	−4.32	1.0	45
A4.1	Glu	CD	−1.54	17.02	−16.03	1.0	28	A14.1	Tyr	CE2	16.14	16.84	−3.56	1.0	39
A4.1	Glu	CG	−0.35	16.90	−16.96	1.0	20	A14.1	Tyr	CZ	16.31	17.44	−2.30	1.0	43
A4.1	Glu	CB	1.06	17.06	−16.36	1.0	24	A14.1	Tyr	CE1	15.47	18.45	−1.81	1.0	35
A4.1	Glu	CA	1.48	15.74	−15.69	1.0	23	A14.1	Tyr	CD1	14.41	18.93	−2.64	1.0	37
A4.1	Glu	C	2.97	15.74	−15.34	1.0	20	A14.1	Tyr	CG	14.24	18.33	−3.88	1.0	23
A4.1	Glu	O	3.81	15.20	−16.06	1.0	29	A14.1	Tyr	CB	13.06	18.79	−4.70	1.0	23
A5.1	Gln	N	3.29	16.39	−14.24	1.0	18	A14.1	Tyr	CA	11.76	18.05	−4.25	1.0	16
A5.1	Gln	NE2	4.98	19.94	−11.15	1.0	56	A14.1	Tyr	C	10.63	18.67	−4.98	1.0	16
A5.1	Gln	OE1	6.48	19.98	−12.95	1.0	50	A14.1	Tyr	O	10.10	19.67	−4.52	1.0	24
A5.1	Gln	CD	5.41	19.75	−12.40	1.0	60	A15.1	Gln	N	10.29	18.19	−6.16	1.0	16
A5.1	Gln	CG	4.37	18.97	−13.20	1.0	33	A15.1	Gln	NE2	9.46	18.89	−11.76	1.0	35
A5.1	Gln	CB	4.82	17.55	−12.78	1.0	23	A15.1	Gln	OE1	10.46	17.06	−11.04	1.0	43
A5.1	Gln	CA	4.70	16.44	−13.80	1.0	18	A15.1	Gln	CD	10.00	18.17	−10.77	1.0	46
A5.1	Gln	C	5.21	15.10	−13.31	1.0	24	A15.1	Gln	CG	10.05	18.69	−9.37	1.0	28
A5.1	Gln	O	6.33	14.68	−13.58	1.0	26	A15.1	Gln	CB	9.02	18.20	−8.38	1.0	20
A6.1	Cys	N	4.39	14.41	−12.51	1.0	20	A15.1	Gln	CA	9.18	18.71	−6.91	1.0	14
A6.1	Cys	SG	5.42	14.72	−9.51	1.0	20	A15.1	Gln	C	7.94	18.40	−6.11	1.0	16
A6.1	Cys	CB	4.73	13.26	−10.35	1.0	17	A15.1	Gln	O	7.04	19.23	−6.05	1.0	18
A6.1	Cys	CA	4.95	13.17	−11.92	1.0	21	A16.1	Leu	N	7.84	17.25	−5.49	1.0	18
A6.1	Cys	C	4.47	11.88	−12.44	1.0	19	A16.1	Leu	CD2	5.12	14.49	−5.99	1.0	18
A6.1	Cys	O	5.13	10.85	−12.29	1.0	18	A16.1	Leu	CD1	6.48	13.23	−4.59	1.0	23
A7.1	Cys	N	3.35	11.80	−13.06	1.0	16	A16.1	Leu	CG	6.50	14.52	−5.34	1.0	18
A7.1	Cys	SG	0.60	8.90	−13.98	1.0	19	A16.1	Leu	CB	6.54	15.53	−4.23	1.0	17
A7.1	Cys	CB	1.35	10.33	−13.33	1.0	18	A16.1	Leu	CA	6.58	16.98	−4.78	1.0	17
A7.1	Cys	CA	2.86	10.52	−13.61	1.0	10	A16.1	Leu	C	6.31	17.90	−3.59	1.0	21
A7.1	Cys	C	3.09	10.42	−15.11	1.0	27	A16.1	Leu	O	5.20	18.07	−3.08	1.0	17
A7.1	Cys	O	3.60	9.45	−15.60	1.0	25	A17.1	Glu	N	7.39	18.43	−3.04	1.0	16
A8.1	Thr	N	2.69	11.40	−15.90	1.0	22	A17.1	Glu	OE2	10.36	19.29	1.47	1.0	31
A8.1	Thr	CG2	2.54	12.80	−19.37	1.0	39	A17.1	Glu	OE1	11.27	20.01	−0.34	1.0	26
A8.1	Thr	OG1	0.59	12.07	−17.87	1.0	42	A17.1	Glu	CD	10.40	19.52	0.27	1.0	26
A8.1	Thr	CB	1.99	12.49	−18.00	1.0	29	A17.1	Glu	CG	9.18	18.98	−0.41	1.0	23
A8.1	Thr	CA	2.91	11.36	−17.38	1.0	20	A17.1	Glu	CB	8.75	19.80	−1.58	1.0	17
A8.1	Thr	C	4.41	11.51	−17.64	1.0	29	A17.1	Glu	CA	7.35	19.32	−1.96	1.0	11
A8.1	Thr	O	5.05	10.72	−18.36	1.0	25	A17.1	Glu	C	6.61	20.60	−2.28	1.0	11
A9.1	Ser	N	4.96	12.55	−16.97	1.0	23	A17.1	Glu	O	6.18	21.31	−1.39	1.0	14
A9.1	Ser	OG	8.27	14.52	−17.32	1.0	31	A18.1	Asn	N	6.47	20.84	−3.58	1.0	13
A9.1	Ser	CB	6.84	14.25	−17.26	1.0	23	A18.1	Asn	ND2	7.84	23.54	−5.00	1.0	27
A9.1	Ser	CA	6.44	12.78	−17.03	1.0	24	A18.1	Asn	OD1	7.76	22.38	−6.93	1.0	28
A9.1	Ser	C	7.04	12.10	−15.80	1.0	23	A18.1	Asn	CG	7.29	22.75	−5.88	1.0	21
A9.1	Ser	O	6.51	11.08	−15.35	1.0	28	A18.1	Asn	CB	5.89	22.28	−5.51	1.0	17
A10.1	Ile	N	8.16	12.53	−15.27	1.0	18	A18.1	Asn	CA	5.71	22.01	−4.01	1.0	13
A10.1	Ile	CD1	12.45	11.25	−15.18	1.0	29	A18.1	Asn	C	4.24	21.88	−3.68	1.0	15
A10.1	Ile	CG1	11.17	12.06	−14.94	1.0	17	A18.1	Asn	O	3.51	22.89	−3.70	1.0	18
A10.1	Ile	CB	10.08	11.22	−14.30	1.0	14	A19.1	Tyr	N	3.68	20.72	−3.42	1.0	10
A10.1	Ile	CG2	9.77	9.90	−15.07	1.0	23	A19.1	Tyr	OH	2.93	20.20	−9.44	1.0	28
A10.1	Ile	CA	8.82	12.07	−14.12	1.0	14	A19.1	Tyr	CD2	3.10	19.02	−6.00	1.0	20
A10.1	Ile	C	9.16	13.27	−13.22	1.0	22	A19.1	Tyr	CE2	3.43	19.26	−7.34	1.0	21
A10.1	Ile	O	9.40	14.34	−13.81	1.0	21	A19.1	Tyr	CZ	2.60	20.08	−8.11	1.0	16
A11.1	Cys	N	9.16	13.00	−11.90	1.0	18	A19.1	Tyr	CE1	1.48	20.59	−7.53	1.0	16

TABLE 1.1. (*cont.*)

residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å²
A19.1	Tyr	CD1	1.18	20.36	−6.18	1.0	16
A19.1	Tyr	CG	2.00	19.58	−5.39	1.0	13
A19.1	Tyr	CB	1.75	19.34	−3.94	1.0	14
A19.1	Tyr	CA	2.32	20.49	−3.11	1.0	11
A19.1	Tyr	C	2.04	20.28	−1.64	1.0	19
A19.1	Tyr	O	0.88	20.01	−1.33	1.0	15
A20.1	Cys	N	3.00	20.52	−0.78	1.0	16
A20.1	Cys	SG	5.11	18.92	1.08	1.0	14
A20.1	Cys	CB	4.12	20.33	1.47	1.0	12
A20.1	Cys	CA	2.82	20.51	0.66	1.0	13
A20.1	Cys	C	2.17	21.85	1.08	1.0	18
A20.1	Cys	O	2.42	22.84	0.45	1.0	27
A21.1	Asn	N	1.40	21.90	2.17	1.0	17
A21.1	Asn	ND2	−2.24	21.28	3.59	1.0	33
A21.1	Asn	OD1	−2.06	22.47	1.65	1.0	34
A21.1	Asn	CG	−1.64	22.17	2.74	1.0	29
A21.1	Asn	CB	−0.43	22.76	3.39	1.0	22
A21.1	Asn	CA	0.81	23.15	2.68	1.0	22
A21.1	Asn	C	1.88	23.65	3.68	1.0	31
A21.1	Asn	O	2.94	22.99	3.84	1.0	45
A21.1	Asn	OE	1.58	24.58	4.40	1.0	52
B1.1	Phe	N	18.33	11.82	−3.89	1.0	62
B1.1	Phe	CD2	16.42	13.63	−2.07	1.0	31
B1.1	Phe	CE2	16.71	13.72	−0.71	1.0	54
B1.1	Phe	CZ	16.61	12.61	0.12	1.0	37
B1.1	Phe	CE1	16.22	11.38	−0.41	1.0	32
B1.1	Phe	CD1	15.98	11.29	−1.75	1.0	40
B1.1	Phe	CG	16.03	12.42	−2.57	1.0	32
B1.1	Phe	CB	15.80	12.16	−4.04	1.0	30
B1.1	Phe	CA	17.05	11.27	−4.37	1.0	29
B1.1	Phe	C	17.17	10.89	−5.85	1.0	46
B1.1	Phe	O	18.17	11.14	−6.55	1.0	50
B2.1	Val	N	16.12	10.17	−6.27	1.0	34
B2.1	Val	CG2	17.53	7.56	−7.88	1.0	86
B2.1	Val	CG1	15.24	7.33	−6.98	1.0	36
B2.1	Val	CB	16.13	8.14	−7.91	1.0	28
B2.1	Val	CA	16.01	9.67	−7.68	1.0	24
B2.1	Val	C	14.65	10.22	−8.24	1.0	34
B2.1	Val	O	13.73	10.59	−7.49	1.0	22
B3.1	Asn	N	14.56	10.31	−9.56	1.0	20
B3.1	Asn	ND2	14.11	13.12	−9.79	1.0	70
B3.1	Asn	OD1	14.43	13.81	−11.83	1.0	72
B3.1	Asn	CG	14.21	12.87	−11.10	1.0	37
B3.1	Asn	CB	13.97	11.43	−11.57	1.0	28
B3.1	Asn	CA	13.38	10.69	−10.33	1.0	34
B3.1	Asn	C	12.45	9.52	−10.62	1.0	27
B3.1	Asn	O	12.95	8.48	−11.06	1.0	35
B4.1	Gln	N	11.13	9.65	−10.49	1.0	39
B4.1	Gln	NE2	10.70	7.38	−7.00	1.0	30
B4.1	Gln	OE1	10.49	5.14	−7.27	1.0	48
B4.1	Gln	CD	10.52	6.25	−7.68	1.0	31
B4.1	Gln	CG	10.35	6.43	−9.19	1.0	41
B4.1	Gln	CB	9.98	7.92	−9.38	1.0	39
B4.1	Gln	CA	10.22	8.52	−10.78	1.0	30
B4.1	Gln	C	8.83	8.86	−11.35	1.0	14
B4.1	Gln	O	8.36	10.03	−11.20	1.0	26
B5.1	His	N	8.29	7.80	−12.00	1.0	14
B5.1	His	CD2	7.87	6.97	−15.82	1.0	32
B5.1	His	NE2	7.84	7.59	−17.01	1.0	18
B5.1	His	CE1	6.87	8.46	−17.07	1.0	29
B5.1	His	ND1	6.28	8.34	−15.88	1.0	32
B5.1	His	CG	6.86	7.45	−15.05	1.0	33
B5.1	His	CB	6.43	7.09	−13.66	1.0	25
B5.1	His	CA	6.90	7.92	−12.48	1.0	16
B5.1	His	C	6.03	7.53	−11.25	1.0	17
B5.1	His	O	6.11	6.36	−10.82	1.0	22
B6.1	Leu	N	5.26	8.45	−10.79	1.0	11
B6.1	Leu	CD2	6.09	7.59	−7.09	1.0	18
B6.1	Leu	CD1	6.83	10.00	−7.27	1.0	17
B6.1	Leu	CG	6.25	8.83	−7.99	1.0	19
B6.1	Leu	CB	4.82	9.07	−8.49	1.0	17
B6.1	Leu	CA	4.30	8.23	−9.64	1.0	23
B6.1	Leu	C	2.85	8.57	−10.05	1.0	12
B6.1	Leu	O	2.51	9.70	−10.34	1.0	14
B7.1	Cys	N	2.12	7.49	−10.11	1.0	14
B7.1	Cys	SG	1.63	7.35	−13.23	1.0	20
B7.1	Cys	CB	0.61	6.72	−11.87	1.0	31
B7.1	Cys	CA	0.75	7.57	−10.58	1.0	19
B7.1	Cys	C	−0.29	7.06	−9.56	1.0	18
B7.1	Cys	O	0.02	6.14	−8.85	1.0	19
B8.1	Gly	N	−1.44	7.62	−9.64	1.0	20
B8.1	Gly	CA	−2.58	7.16	−8.78	1.0	18
B8.1	Gly	C	−2.31	7.24	−7.33	1.0	9
B8.1	Gly	O	−1.91	8.25	−6.76	1.0	14
B9.1	Ser	N	−2.66	6.10	−6.69	1.0	12
B9.1	Ser	OG	−2.76	3.67	−5.01	1.0	21
B9.1	Ser	CB	−3.26	4.87	−4.57	1.0	15
B9.1	Ser	CA	−2.48	6.00	−5.23	1.0	12
B9.1	Ser	C	−1.03	6.12	−4.87	1.0	9
B9.1	Ser	O	−0.76	6.58	−3.76	1.0	10
B10.1	His	N	−0.17	5.70	−5.80	1.0	9
B10.1	His	CD2	1.06	2.92	−7.56	1.0	11
B10.1	His	NE2	0.95	1.63	−7.16	1.0	12
B10.1	His	CE1	1.54	1.53	−6.03	1.0	9
B10.1	His	ND1	1.95	2.70	−5.61	1.0	14
B10.1	His	CG	1.71	3.60	−6.61	1.0	9
B10.1	His	CB	2.09	5.05	−6.48	1.0	9
B10.1	His	CA	1.27	5.80	−5.45	1.0	9
B10.1	His	C	1.75	7.23	−5.32	1.0	8
B10.1	His	O	2.66	7.48	−4.53	1.0	11
B11.1	Leu	N	1.14	8.11	−6.07	1.0	9
B11.1	Leu	CD2	0.40	12.23	−8.56	1.0	16
B11.1	Leu	CD1	2.53	12.12	−7.27	1.0	12
B11.1	Leu	CG	1.11	11.70	−7.29	1.0	13
B11.1	Leu	CB	0.81	10.20	−7.25	1.0	11
B11.1	Leu	CA	1.43	9.54	−6.05	1.0	10
B11.1	Leu	C	1.03	10.15	−4.73	1.0	6
B11.1	Leu	O	1.82	10.85	−4.08	1.0	9
B12.1	Val	N	−0.19	9.82	−4.30	1.0	7
B12.1	Val	CG2	−2.64	9.01	−2.74	1.0	14
B12.1	Val	CG1	−2.69	11.27	−3.90	1.0	17
B12.1	Val	CB	−2.07	10.35	−2.79	1.0	14
B12.1	Val	CA	−0.53	10.40	−2.94	1.0	7
B12.1	Val	C	0.21	9.76	−1.79	1.0	6
B12.1	Val	O	0.47	10.48	−0.76	1.0	9
B13.1	Glu	N	0.68	8.54	−1.89	1.0	9
B13.1	Glu	OE2	−0.43	4.26	0.73	1.0	33
B13.1	Glu	OE1	1.50	5.09	1.17	1.0	28
B13.1	Glu	CD	0.54	4.94	0.49	1.0	62
B13.1	Glu	CG	0.46	5.63	−0.82	1.0	28
B13.1	Glu	CB	1.77	6.43	−1.05	1.0	12
B13.1	Glu	CA	1.51	7.90	−0.87	1.0	7
B13.1	Glu	C	2.82	8.63	−0.80	1.0	7
B13.1	Glu	O	3.40	8.91	0.27	1.0	12
B14.1	Ala	N	3.39	9.05	−1.89	1.0	10
B14.1	Ala	CB	5.15	9.88	−3.37	1.0	9
B14.1	Ala	CA	4.64	9.84	−1.99	1.0	10
B14.1	Ala	C	4.49	11.22	−1.38	1.0	10
B14.1	Ala	O	5.37	11.72	−0.68	1.0	12
B15.1	Leu	N	3.34	11.84	−1.61	1.0	9
B15.1	Leu	CD2	2.49	15.66	−2.80	1.0	16
B15.1	Leu	CD1	0.40	14.48	−3.56	1.0	13
B15.1	Leu	CG	1.77	14.34	−2.98	1.0	10
B15.1	Leu	CB	1.72	13.65	−1.64	1.0	9
B15.1	Leu	CA	3.01	13.12	−1.05	1.0	10
B15.1	Leu	C	2.97	13.02	0.45	1.0	16
B15.1	Leu	O	3.46	13.91	1.14	1.0	12
B16.1	Tyr	N	2.31	11.97	0.94	1.0	12
B16.1	Tyr	OH	1.66	9.08	8.18	1.0	18
B16.1	Tyr	CD2	2.23	9.16	4.53	1.0	10
B16.1	Tyr	CE2	2.28	8.85	5.89	1.0	18
B16.1	Tyr	CZ	1.52	9.55	6.84	1.0	13
B16.1	Tyr	CE1	0.66	10.55	6.43	1.0	12
B16.1	Tyr	CD1	0.61	10.88	5.09	1.0	13
B16.1	Tyr	CG	1.41	10.20	4.15	1.0	10

TABLE 1.1. (*cont.*)

residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å²	residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å²
B16.1	Tyr	CB	1.41	10.55	2.66	1.0	9	B25.1	Phe	CD1	−2.43	21.55	−3.85	1.0	20
B16.1	Tyr	CA	2.25	11.79	2.37	1.0	9	B25.1	Phe	CG	−2.57	21.45	−2.48	1.0	14
B16.1	Tyr	C	3.66	11.71	3.01	1.0	11	B25.1	Phe	CB	−3.69	20.61	−1.92	1.0	10
B16.1	Tyr	O	3.92	12.34	4.01	1.0	11	B25.1	Phe	CA	−3.29	19.18	−1.47	1.0	10
B17.1	Leu	N	4.48	10.91	2.35	1.0	10	B25.1	Phe	C	−3.02	18.33	−2.72	1.0	10
B17.1	Leu	CD2	8.53	8.58	1.43	1.0	36	B25.1	Phe	O	−1.92	18.36	−3.22	1.0	12
B17.1	Leu	CD1	7.27	8.08	3.61	1.0	28	B26.1	Tyr	N	−3.96	17.55	−3.19	1.0	11
B17.1	Leu	CG	7.78	9.14	2.61	1.0	27	B26.1	Tyr	OH	−4.41	11.85	−8.51	1.0	20
B17.1	Leu	CB	6.57	9.68	1.93	1.0	12	B26.1	Tyr	CD2	−3.14	14.23	−6.11	1.0	13
B17.1	Leu	CA	5.83	10.72	2.82	1.0	9	B26.1	Tyr	CE2	−3.29	13.40	−7.22	1.0	15
B17.1	Leu	C	6.67	11.99	2.88	1.0	13	B26.1	Tyr	CZ	−4.39	12.68	−7.39	1.0	17
B17.1	Leu	O	7.39	12.30	3.85	1.0	14	B26.1	Tyr	CE1	−5.49	12.77	−6.58	1.0	18
B18.1	Val	N	6.59	12.77	1.78	1.0	10	B26.1	Tyr	CD1	−5.38	13.60	−5.45	1.0	13
B18.1	Val	CG2	8.32	13.58	−0.64	1.0	16	B26.1	Tyr	CG	−4.16	14.29	−5.20	1.0	13
B18.1	Val	CG1	7.68	15.90	0.01	1.0	18	B26.1	Tyr	CB	−4.09	15.23	−4.01	1.0	13
B18.1	Val	CB	7.36	14.41	0.18	1.0	8	B26.1	Tyr	CA	−3.91	16.72	−4.37	1.0	8
B18.1	Val	CA	7.35	14.03	1.66	1.0	8	B26.1	Tyr	C	−5.01	17.22	−5.27	1.0	12
B18.1	Val	C	6.79	15.10	2.51	1.0	8	B26.1	Tyr	O	−6.21	17.04	−5.01	1.0	12
B18.1	Val	O	7.56	15.91	3.06	1.0	14	B27.1	Thr	N	−4.64	17.93	−6.34	1.0	10
B19.1	Cys	N	5.47	15.22	2.57	1.0	9	B27.1	Thr	CG2	−6.18	20.41	−5.71	1.0	17
B19.1	Cys	SG	3.85	17.36	0.91	1.0	11	B27.1	Thr	OG1	−4.18	20.58	−6.97	1.0	20
B19.1	Cys	CB	3.54	16.78	2.55	1.0	12	B27.1	Thr	CB	−5.54	20.11	−7.05	1.0	16
B19.1	Cys	CA	4.79	16.33	3.28	1.0	11	B27.1	Thr	CA	−5.56	18.56	−7.26	1.0	15
B19.1	Cys	C	4.55	16.08	4.76	1.0	20	B27.1	Thr	C	−5.25	18.12	−8.70	1.0	16
B19.1	Cys	O	4.77	17.06	5.51	1.0	17	B27.1	Thr	O	−4.55	18.82	−9.40	1.0	18
B20.1	Gly	N	4.11	14.86	5.08	1.0	15	B28.1	Pro	N	−5.77	17.03	−9.15	1.0	12
B20.1	Gly	CA	3.84	14.55	6.47	1.0	18	B28.1	Pro	CG	−6.81	14.87	−9.14	1.0	19
B20.1	Gly	C	2.82	15.48	7.18	1.0	19	B28.1	Pro	CD	−6.69	16.11	−8.34	1.0	12
B20.1	Gly	O	1.67	15.72	6.77	1.0	19	B28.1	Pro	CB	−6.19	15.16	−10.48	1.0	22
B21.1	Glu	N	3.39	15.98	8.28	1.0	20	B28.1	Pro	CA	−5.57	16.54	−10.48	1.0	14
B21.1	Glu	OE2	7.30	17.57	10.65	1.0	73	B28.1	Pro	C	−6.11	17.45	−11.59	1.0	22
B21.1	Glu	OE1	5.78	18.56	11.86	1.0	73	B28.1	Pro	O	−5.47	17.45	−12.64	1.0	24
B21.1	Glu	CD	6.10	17.83	10.94	1.0	79	B29.1	Lys	N	−7.10	18.25	−11.35	1.0	19
B21.1	Glu	CG	5.07	17.19	10.05	1.0	54	B29.1D1	Lys	NZ	−13.65	18.11	−11.18	0.5	39
B21.1	Glu	CB	3.58	17.12	10.48	1.0	23	B29.1D1	Lys	CE	−12.31	17.61	−11.67	0.5	22
B21.1	Glu	CA	2.66	16.85	9.24	1.0	24	B29.1D1	Lys	CD	−11.26	18.66	−11.37	0.5	26
B21.1	Glu	C	2.24	18.14	8.64	1.0	19	B29.1D1	Lys	CG	−10.12	18.68	−12.36	0.5	26
B21.1	Glu	O	1.25	18.73	9.11	1.0	24	B29.1D2	Lys	NZ	−13.28	22.35	−12.73	0.5	39
B22.1	Arg	N	2.95	18.56	7.57	1.0	16	B29.1D2	Lys	CE	−12.08	21.72	−13.35	0.5	41
B22.1D1	Arg	NH2	7.19	21.77	2.93	0.5	21	B29.1D2	Lys	CD	−11.09	21.21	−12.32	0.5	27
B22.1D1	Arg	NH1	5.82	23.55	2.85	0.5	18	B29.1D2	Lys	CG	−9.87	20.53	−12.93	0.5	52
B22.1D1	Arg	CZ	6.16	22.43	3.48	0.5	35	B29.1	Lys	CB	−9.02	19.69	−11.97	1.0	38
B22.1D1	Arg	NE	5.58	21.91	4.55	0.5	24	B29.1	Lys	CA	−7.65	19.15	−12.40	1.0	22
B22.1D2	Arg	NH2	8.20	18.65	3.32	0.5	20	B29.1	Lys	C	−6.78	20.31	−12.72	1.0	25
B22.1D2	Arg	NH1	7.55	20.85	3.12	0.5	20	B29.1	Lys	O	−6.94	20.99	−13.74	1.0	29
B22.1D2	Arg	CZ	7.48	19.68	3.78	0.5	12	B30.1	Ala	N	−5.87	20.55	−11.81	1.0	18
B22.1D2	Arg	NE	6.66	19.52	4.83	0.5	25	B30.1	Ala	CB	−4.39	21.98	−10.53	1.0	33
B22.1	Arg	CD	5.95	20.71	5.31	1.0	33	B30.1	Ala	CA	−5.02	21.71	−11.88	1.0	16
B22.1	Arg	CG	5.02	20.33	6.45	1.0	17	B30.1	Ala	C	−3.94	21.53	−12.91	1.0	35
B22.1	Arg	CB	3.61	20.25	5.97	1.0	30	B30.1	Ala	O	−3.35	20.45	−13.02	1.0	37
B22.1	Arg	CA	2.55	19.83	7.00	1.0	21	B30.1	Ala	OE	−3.82	22.60	−13.54	1.0	43
B22.1	Arg	C	1.23	19.78	6.23	1.0	31								
B22.1	Arg	O	0.56	20.74	5.95	1.0	22	A1.2	Gly	N	−8.86	16.94	14.29	1.0	22
B23.1	Gly	N	0.91	18.56	5.84	1.0	18	A1.2	Gly	CA	−9.93	17.03	13.24	1.0	23
B23.1	Gly	CA	−0.24	18.30	4.99	1.0	13	A1.2	Gly	C	−10.05	15.63	12.62	1.0	44
B23.1	Gly	C	0.09	18.61	3.51	1.0	11	A1.2	Gly	O	−9.78	14.73	13.41	1.0	25
B23.1	Gly	O	1.14	19.10	3.20	1.0	12	A2.2	Ile	N	−10.33	15.53	11.33	1.0	26
B24.1	Phe	N	−0.90	18.29	2.67	1.0	12	A2.2	Ile	CD1	−12.81	13.03	9.64	1.0	27
B24.1	Phe	CD2	−1.50	15.66	−0.39	1.0	12	A2.2	Ile	CG1	−11.58	13.15	8.70	1.0	37
B24.1	Phe	CE2	−2.05	14.39	−0.37	1.0	13	A2.2	Ile	CB	−10.88	14.49	9.10	1.0	40
B24.1	Phe	CZ	−1.74	13.49	0.69	1.0	14	A2.2	Ile	CG2	−9.74	14.79	8.14	1.0	23
B24.1	Phe	CE1	−0.86	13.90	1.71	1.0	19	A2.2	Ile	CA	−10.49	14.27	10.60	1.0	21
B24.1	Phe	CD1	−0.28	15.18	1.65	1.0	13	A2.2	Ile	C	−9.37	13.30	10.66	1.0	12
B24.1	Phe	CG	−0.60	16.05	0.60	1.0	10	A2.2	Ile	O	−9.58	12.09	10.97	1.0	20
B24.1	Phe	CB	0.00	17.40	0.58	1.0	9	A3.2	Val	N	−8.13	13.76	10.48	1.0	17
B24.1	Phe	CA	−0.77	18.50	1.21	1.0	10	A3.2	Val	CG2	−5.84	14.21	8.78	1.0	26
B24.1	Phe	C	−2.13	18.69	0.61	1.0	10	A3.2	Val	CG1	−4.38	12.96	10.45	1.0	32
B24.1	Phe	O	−3.14	18.47	1.20	1.0	12	A3.2	Val	CB	−5.70	13.71	10.23	1.0	21
B25.1	Phe	N	−2.07	19.07	−0.69	1.0	9	A3.2	Val	CA	−6.97	12.90	10.58	1.0	16
B25.1	Phe	CD2	−1.65	22.07	−1.65	1.0	15	A3.2	Val	C	−6.89	12.16	11.92	1.0	22
B25.1	Phe	CE2	−0.61	22.77	−2.27	1.0	20	A3.2	Val	O	−6.55	10.99	12.04	1.0	25
B25.1	Phe	CZ	−0.56	22.88	−3.66	1.0	14	A4.2	Glu	N	−7.04	13.02	12.94	1.0	17
B25.1	Phe	CE1	−1.47	22.32	−4.48	1.0	14	A4.2	Glu	OE2	−4.57	15.57	13.18	1.0	25

TABLE 1.1. (*cont.*)

residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å^2	residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å^2
A4.2	Glu	OE1	−6.76	15.96	13.38	1.0	23	A14.2	Tyr	CZ	−22.74	6.86	2.56	1.0	15
A4.2	Glu	CD	−5.70	15.46	13.73	1.0	22	A14.2	Tyr	CE1	−21.77	8.17	2.09	1.0	18
A4.2	Glu	CG	−5.61	14.56	14.95	1.0	21	A14.2	Tyr	CD1	−22.11	9.12	2.89	1.0	19
A4.2	Glu	CB	−6.81	13.69	15.27	1.0	17	A14.2	Tyr	CG	−21.46	8.76	4.08	1.0	18
A4.2	Glu	CA	−6.89	12.47	14.30	1.0	15	A14.2	Tyr	CB	−20.75	9.90	4.80	1.0	13
A4.2	Glu	C	−8.00	11.56	14.61	1.0	17	A14.2	Tyr	CA	−19.28	10.03	4.37	1.0	11
A4.2	Glu	O	−7.89	10.47	15.13	1.0	23	A14.2	Tyr	C	−18.64	11.23	4.96	1.0	13
A5.2	Gln	N	−9.20	12.05	14.36	1.0	16	A14.2	Tyr	O	−18.71	12.30	4.34	1.0	15
A5.2	Gln	OE1	−13.91	13.03	13.17	1.0	55	A15.2	Gln	N	−17.94	11.10	6.09	1.0	10
A5.2	Gln	NE2	−14.94	11.03	13.35	1.0	27	A15.2	Gln	NE2	−17.99	11.65	11.62	1.0	30
A5.2	Gln	CD	−13.95	11.90	13.63	1.0	42	A15.2	Gln	OE1	−16.86	10.03	10.60	1.0	20
A5.2	Gln	CG	−12.86	11.37	14.56	1.0	22	A15.2	Gln	CD	−17.52	11.06	10.52	1.0	28
A5.2	Gln	CB	−11.59	12.13	14.15	1.0	21	A15.2	Gln	CG	−17.89	11.67	9.21	1.0	18
A5.2	Gln	CA	−10.41	11.30	14.63	1.0	12	A15.2	Gln	CB	−16.77	11.84	8.20	1.0	14
A5.2	Gln	C	−10.43	9.94	13.98	1.0	20	A15.2	Gln	CA	−17.18	12.14	6.77	1.0	9
A5.2	Gln	O	−10.81	8.93	14.54	1.0	17	A15.2	Gln	C	−16.01	12.54	5.90	1.0	11
A6.2	Cys	N	−10.03	9.81	12.69	1.0	13	A15.2	Gln	O	−15.61	13.72	5.72	1.0	14
A6.2	Cys	SG	−10.93	9.46	9.58	1.0	13	A16.2	Leu	N	−15.35	11.52	5.32	1.0	13
A6.2	Cys	CB	−9.66	8.67	10.56	1.0	13	A16.2	Leu	CD2	−12.11	8.57	4.60	1.0	17
A6.2	Cys	CA	−10.05	8.52	12.06	1.0	13	A16.2	Leu	CD1	−12.04	10.56	6.03	1.0	17
A6.2	Cys	C	−9.10	7.52	12.67	1.0	10	A16.2	Leu	CG	−12.95	9.72	5.18	1.0	13
A6.2	Cys	O	−9.40	6.29	12.67	1.0	14	A16.2	Leu	CB	−13.59	10.52	4.06	1.0	13
A7.2	Cys	N	−8.02	7.99	13.17	1.0	11	A16.2	Leu	CA	−14.19	11.83	4.47	1.0	11
A7.2	Cys	SG	−4.18	6.82	14.13	1.0	14	A16.2	Leu	C	−14.60	12.63	3.25	1.0	16
A7.2	Cys	CB	−5.58	7.83	13.66	1.0	20	A16.2	Leu	O	−13.77	13.40	2.74	1.0	16
A7.2	Cys	CA	−6.96	7.19	13.81	1.0	17	A17.2	Glu	N	−15.78	12.42	2.76	1.0	18
A7.2	Cys	C	−7.24	6.95	15.36	1.0	14	A17.2	Glu	OE2	−19.13	11.87	−1.64	1.0	35
A7.2	Cys	O	−7.06	5.78	15.77	1.0	19	A17.2	Glu	OE1	−20.20	11.47	0.28	1.0	30
A8.2	Thr	N	−7.66	7.94	16.06	1.0	13	A17.2	Glu	CD	−19.23	11.67	−0.39	1.0	27
A8.2	Thr	CG2	−6.33	9.70	18.20	1.0	27	A17.2	Glu	CG	−17.89	11.63	0.36	1.0	28
A8.2	Thr	OG1	−8.89	9.92	18.12	1.0	27	A17.2	Glu	CB	−17.65	12.86	1.22	1.0	17
A8.2	Thr	CB	−7.73	9.06	18.39	1.0	21	A17.2	Glu	CA	−16.22	13.18	1.59	1.0	18
A8.2	Thr	CA	−7.86	7.73	17.52	1.0	20	A17.2	Glu	C	−16.17	14.69	1.81	1.0	19
A8.2	Thr	C	−9.14	7.00	17.87	1.0	26	A17.2	Glu	O	−16.12	15.47	0.80	1.0	18
A8.2	Thr	O	−9.19	6.16	18.80	1.0	25	A18.2	Asn	N	−16.09	15.07	3.10	1.0	15
A9.2	Ser	N	−10.17	7.35	17.06	1.0	20	A18.2	Asn	OD1	−18.81	16.05	4.38	1.0	34
A9.2	Ser	OG	−13.67	7.77	17.65	1.0	32	A18.2	Asn	ND2	−17.99	16.17	6.45	1.0	36
A9.2	Ser	CB	−12.33	8.07	17.44	1.0	20	A18.2	Asn	CG	−17.87	16.34	5.14	1.0	30
A9.2	Ser	CA	−11.51	6.80	17.12	1.0	17	A18.2	Asn	CB	−16.49	16.93	4.74	1.0	21
A9.2	Ser	C	−11.80	5.98	15.86	1.0	13	A18.2	Asn	CA	−16.03	16.53	3.33	1.0	19
A9.2	Ser	O	−11.14	5.01	15.47	1.0	18	A18.2	Asn	C	−14.70	17.13	2.95	1.0	18
A10.2	Ile	N	−12.88	6.38	15.16	1.0	15	A18.2	Asn	O	−14.55	18.38	2.83	1.0	20
A10.2	Ile	CD1	−16.70	3.72	15.01	1.0	42	A19.2	Tyr	N	−13.70	16.33	2.74	1.0	16
A10.2	Ile	CG1	−15.70	4.87	14.74	1.0	22	A19.2	Tyr	OH	−12.23	17.85	8.40	1.0	38
A10.2	Ile	CB	−14.37	4.52	14.05	1.0	19	A19.2	Tyr	CD2	−12.56	15.70	5.45	1.0	19
A10.2	Ile	CG2	−13.71	3.30	14.72	1.0	23	A19.2	Tyr	CE2	−12.83	16.21	6.71	1.0	24
A10.2	Ile	CA	−13.35	5.72	13.93	1.0	20	A19.2	Tyr	CZ	−12.08	17.27	7.19	1.0	34
A10.2	Ile	C	−13.97	6.90	13.11	1.0	18	A19.2	Tyr	CE1	−11.08	17.80	6.39	1.0	28
A10.2	Ile	O	−14.35	7.92	13.62	1.0	17	A19.2	Tyr	CD1	−10.85	17.30	5.13	1.0	27
A11.2	Cys	N	−14.08	6.68	11.77	1.0	12	A19.2	Tyr	CG	−11.59	16.22	4.63	1.0	21
A11.2	Cys	SG	−12.38	8.02	9.39	1.0	14	A19.2	Tyr	CB	−11.36	15.84	3.18	1.0	14
A11.2	Cys	CB	−13.69	8.70	10.42	1.0	13	A19.2	Tyr	CA	−12.36	16.72	2.38	1.0	14
A11.2	Cys	CA	−14.66	7.68	10.88	1.0	11	A19.2	Tyr	C	−12.15	16.69	0.90	1.0	13
A11.2	Cys	C	−15.30	6.88	9.77	1.0	12	A19.2	Tyr	O	−11.01	17.04	0.48	1.0	16
A11.2	Cys	O	−14.96	5.69	9.53	1.0	21	A20.2	Cys	N	−13.06	16.31	0.08	1.0	13
A12.2	Ser	N	−16.23	7.56	9.10	1.0	11	A20.2	Cys	SG	−13.92	13.76	−1.58	1.0	13
A12.2	Ser	OG	−18.87	8.54	7.88	1.0	17	A20.2	Cys	CB	−13.85	15.49	−2.16	1.0	16
A12.2	Ser	CB	−18.52	7.18	8.08	1.0	17	A20.2	Cys	CA	−12.84	16.31	−1.39	1.0	19
A12.2	Ser	CA	−17.00	6.98	8.01	1.0	10	A20.2	Cys	C	−12.98	17.80	−1.80	1.0	19
A12.2	Ser	C	−16.56	7.64	6.73	1.0	7	A20.2	Cys	O	−13.59	18.58	−1.08	1.0	19
A12.2	Ser	O	−15.97	8.75	6.71	1.0	10	A21.2	Asn	N	−12.38	18.06	−2.91	1.0	18
A13.2	Leu	N	−16.85	6.91	5.61	1.0	11	A21.2	Asn	ND2	−9.00	19.67	−3.81	1.0	40
A13.2	Leu	CD2	−16.05	4.20	2.02	1.0	19	A21.2	Asn	OD1	−10.07	21.38	−5.08	1.0	68
A13.2	Leu	CD1	−14.47	5.32	3.54	1.0	17	A21.2	Asn	CG	−10.02	20.28	−4.46	1.0	41
A13.2	Leu	CG	−15.94	5.05	3.32	1.0	18	A21.2	Asn	CB	−11.23	19.39	−4.52	1.0	19
A13.2	Leu	CB	−16.77	6.35	3.23	1.0	12	A21.2	Asn	CA	−12.40	19.40	−3.61	1.0	25
A13.2	Leu	CA	−16.53	7.44	4.26	1.0	11	A21.2	Asn	C	−13.64	19.70	−4.45	1.0	35
A13.2	Leu	C	−17.32	8.72	4.03	1.0	13	A21.2	Asn	OE	−13.88	20.89	−4.60	1.0	42
A13.2	Leu	O	−16.83	9.52	3.23	1.0	12	A21.2	Asn	O	−14.15	18.70	−4.96	1.0	31
A14.2	Tyr	N	−18.49	8.79	4.63	1.0	11								
A14.2	Tyr	OH	−23.44	5.93	1.72	1.0	25	B1.2	Phe	N	−21.77	1.13	3.58	1.0	26
A14.2	Tyr	CD2	−21.46	7.44	4.47	1.0	15	B1.2	Phe	CD2	−19.71	1.87	1.26	1.0	21
A14.2	Tyr	CE2	−22.12	6.44	3.68	1.0	14	B1.2	Phe	CE2	−19.82	2.07	−0.12	1.0	51

TABLE 1.1. (*cont.*)

residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å²	residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å²
B1.2	Phe	CZ	−20.01	3.33	−0.63	1.0	43	B10.2	His	ND1	−3.04	1.31	5.42	1.0	16
B1.2	Phe	CE1	−20.09	4.44	0.24	1.0	53	B10.2	His	CG	−3.19	2.27	6.36	1.0	10
B1.2	Phe	CD1	−20.07	4.20	1.62	1.0	36	B10.2	His	CB	−4.23	3.32	6.23	1.0	10
B1.2	Phe	CG	−19.92	2.92	2.13	1.0	17	B10.2	His	CA	−3.94	4.38	5.19	1.0	8
B1.2	Phe	CB	−19.81	2.72	3.62	1.0	23	B10.2	His	C	−5.08	5.43	5.07	1.0	10
B1.2	Phe	CA	−20.37	1.37	4.05	1.0	24	B10.2	His	O	−6.02	5.16	4.29	1.0	11
B1.2	Phe	C	−20.34	1.14	5.59	1.0	40	B11.2	Leu	N	−5.02	6.50	5.81	1.0	9
B1.2	Phe	O	−21.42	1.14	6.17	1.0	38	B11.2	Leu	CD2	−8.38	9.23	6.63	1.0	18
B2.2	Val	N	−19.10	0.90	6.03	1.0	21	B11.2	Leu	CD1	−6.74	10.31	8.11	1.0	21
B2.2	Val	CG2	−18.99	−1.95	7.61	1.0	6	B11.2	Leu	CG	−6.95	9.61	6.78	1.0	15
B2.2	Val	CG1	−16.78	−0.92	6.80	1.0	22	B11.2	Leu	CB	−5.86	8.54	6.74	1.0	7
B2.2	Val	CB	−18.05	−0.76	7.64	1.0	31	B11.2	Leu	CA	−6.07	7.52	5.62	1.0	10
B2.2	Val	CA	−18.75	0.60	7.41	1.0	37	B11.2	Leu	C	−5.97	8.18	4.28	1.0	8
B2.2	Val	C	−17.78	1.66	7.97	1.0	24	B11.2	Leu	O	−6.97	8.46	3.67	1.0	10
B2.2	Val	O	−17.10	2.33	7.20	1.0	20	B12.2	Val	N	−4.75	8.45	3.80	1.0	10
B3.2	Asn	N	−17.74	1.75	9.28	1.0	13	B12.2D1	Val	CG2	−2.82	9.79	0.80	0.5	5
B3.2	Asn	ND2	−19.33	3.41	12.47	1.0	35	B12.2D2	Val	CG2	−1.96	8.65	2.12	0.5	11
B3.2	Asn	OD1	−18.66	4.77	10.72	1.0	35	B12.2	Val	CG1	−2.81	10.67	3.15	1.0	18
B3.2	Asn	CG	−18.51	3.84	11.51	1.0	49	B12.2	Val	CB	−3.15	9.54	2.23	1.0	12
B3.2	Asn	CB	−17:27	3.01	11.38	1.0	25	B12.2	Val	CA	−4.58	9.06	2.49	1.0	8
B3.2	Asn	CA	−16.87	2.69	9.95	1.0	14	B12.2	Val	C	−5.05	8.13	1.37	1.0	8
B3.2	Asn	C	−15.46	2.10	9.88	1.0	15	B12.2	Val	O	−5.59	8.65	0.40	1.0	12
B3.2	Asn	O	−15.31	0.86	9.93	1.0	25	B13.2	Glu	N	−4.91	6.88	1.50	1.0	6
B4.2	Gln	N	−14.51	3.03	9.77	1.0	13	B13.2	Glu	OE1	−4.24	3.64	−1.50	1.0	34
B4.2D1	Gln	NE2	−10.90	0.62	6.95	0.6	29	B13.2	Glu	OE2	−2.11	3.61	−1.13	1.0	47
B4.2D1	Gln	OE1	−12.81	−0.61	6.49	0.6	18	B13.2	Glu	CD	−3.25	3.86	−0.79	1.0	54
B4.2D1	Gln	CD	−12.27	0.52	6.83	0.6	12	B13.2	Glu	CG	−3.50	4.44	0.58	1.0	16
B4.2D1	Gln	CG	−13.01	1.73	7.03	0.6	11	B13.2	Glu	CB	−5.00	4.51	0.85	1.0	13
B4.2D2	Gln	NE2	−15.11	2.18	6.16	0.4	36	B13.2	Glu	CA	−5.43	5.95	0.54	1.0	9
B4.2D2	Gln	OE1	−14.06	3.96	7.11	0.4	40	B13.2	Glu	C	−6.97	6.01	0.47	1.0	12
B4.2D2	Gln	CD	−14.07	2.78	6.75	0.4	33	B13.2	Glu	O	−7.58	6.03	−0.61	1.0	11
B4.2D2	Gln	CG	−12.98	1.76	7.00	0.4	37	B14.2	Ala	N	−7.66	6.00	1.64	1.0	7
B4.2	Gln	CB	−12.67	2.76	8.12	1.0	39	B14.2	Ala	CB	−9.60	6.04	3.11	1.0	12
B4.2	Gln	CA	−13.14	2.54	9.57	1.0	23	B14.2	Ala	CA	−9.06	6.16	1.72	1.0	7
B4.2	Gln	C	−12.21	3.22	10.58	1.0	13	B14.2	Ala	C	−9.56	7.48	1.05	1.0	7
B4.2	Gln	O	−12.35	4.33	11.12	1.0	21	B14.2	Ala	O	−10.60	7.47	0.35	1.0	11
B5.2	His	N	−11.16	2.44	10.84	1.0	13	B15.2	Leu	N	−8.88	8.58	1.32	1.0	7
B5.2	His	CD2	−8.64	3.60	14.33	1.0	28	B15.2	Leu	CD2	−10.10	12.23	2.46	1.0	12
B5.2	His	NE2	−7.57	3.51	15.15	1.0	30	B15.2	Leu	CD1	−7.68	12.48	3.24	1.0	14
B5.2	His	CE1	−7.28	2.24	15.21	1.0	29	B15.2	Leu	CG	−8.76	11.58	2.65	1.0	9
B5.2	His	ND1	−8.07	1.53	14.44	1.0	35	B15.2	Leu	CB	−8.32	10.98	1.33	1.0	10
B5.2	His	CG	−8.93	2.40	13.86	1.0	26	B15.2	Leu	CA	−9.22	9.85	0.72	1.0	14
B5.2	His	CB	−9.98	1.96	12.88	1.0	22	B15.2	Leu	C	−9.11	9.81	−0.83	1.0	15
B5.2	His	CA	−10.08	3.00	11.78	1.0	17	B15.2	Leu	O	−9.96	10.39	−1.50	1.0	12
B5.2	His	C	−8.85	3.15	10.90	1.0	11	B16.2	Tyr	N	−8.05	9.15	−1.30	1.0	9
B5.2	His	O	−8.28	2.17	10.38	1.0	17	B16.2	Tyr	OH	−5.93	7.52	−8.59	1.0	25
B6.2	Leu	N	−8.53	4.40	10.60	1.0	11	B16.2	Tyr	CD2	−6.01	9.10	−5.34	1.0	13
B6.2	Leu	CD2	−10.02	5.60	6.67	1.0	15	B16.2	Tyr	CE2	−5.90	8.96	−6.74	1.0	14
B6.2	Leu	CD1	−8.62	3.52	6.77	1.0	18	B16.2	Tyr	CZ	−6.11	7.69	−7.26	1.0	17
B6.2	Leu	CG	−9.25	4.63	7.56	1.0	13	B16.2	Tyr	CE1	−6.35	6.58	−6.47	1.0	18
B6.2	Leu	CB	−8.23	5.40	8.41	1.0	14	B16.2	Tyr	CD1	−6.45	6.78	−5.05	1.0	17
B6.2	Leu	CA	−7.47	4.71	9.61	1.0	11	B16.2	Tyr	CG	−6.27	8.03	−4.52	1.0	11
B6.2	Leu	C	−6.40	5.60	10.16	1.0	11	B16.2	Tyr	CB	−6.49	8.25	−3.05	1.0	8
B6.2	Leu	O	−6.70	6.78	10.48	1.0	14	B16.2	Tyr	CA	−7.84	8.96	−2.69	1.0	9
B7.2	Cys	N	−5.18	5.07	10.11	1.0	10	B16.2	Tyr	C	−9.00	8.18	−3.28	1.0	11
B7.2	Cys	SG	−4.52	5.10	13.25	1.0	16	B16.2	Tyr	O	−9.51	8.50	−4.37	1.0	14
B7.2	Cys	CB	−3.43	5.11	11.76	1.0	16	B17.2	Leu	N	−9.43	7.11	−2.66	1.0	9
B7.2	Cys	CA	−4.06	5.84	10.57	1.0	11	B17.2	Leu	CD2	−9.14	3.07	−1.93	1.0	31
B7.2	Cys	C	−3.03	5.98	9.48	1.0	13	B17.2	Leu	CD1	−10.28	3.54	−4.16	1.0	35
B7.2	Cys	O	−2.95	5.20	8.57	1.0	19	B17.2	Leu	CG	−9.59	4.10	−2.92	1.0	29
B8.2	Gly	N	−2.18	6.99	9.54	1.0	12	B17.2	Leu	CB	−10.66	5.05	−2.33	1.0	11
B8.2	Gly	CA	−1.07	7.26	8.63	1.0	13	B17.2	Leu	CA	−10.57	6.29	−3.17	1.0	9
B8.2	Gly	C	−1.47	7.32	7.20	1.0	13	B17.2	Leu	C	−11.86	7.09	−3.14	1.0	11
B8.2	Gly	O	−2.47	7.88	6.74	1.0	12	B17.2	Leu	O	−12.65	7.05	−4.07	1.0	16
B9.2	Ser	N	−0.61	6.58	6.43	1.0	12	B18.2	Val	N	−12.12	7.79	−2.04	1.0	9
B9.2	Ser	OG	0.53	4.63	4.74	1.0	17	B18.2	Val	CG2	−13.86	7.76	0.39	1.0	13
B9.2	Ser	CB	0.31	5.92	4.21	1.0	18	B18.2	Val	CG1	−14.78	9.90	−0.28	1.0	12
B9.2	Ser	CA	−0.86	6.54	4.98	1.0	16	B18.2	Val	CB	−13.60	8.97	−0.46	1.0	14
B9.2	Ser	C	−2.18	5.87	4.58	1.0	10	B18.2	Val	CA	−13.35	8.55	−1.93	1.0	9
B9.2	Ser	O	−2.65	6.11	3.53	1.0	10	B18.2	Val	C	−13.43	9.71	−2.87	1.0	10
B10.2	His	N	−2.72	5.10	5.45	1.0	10	B18.2	Val	O	−14.47	9.94	−3.46	1.0	17
B10.2	His	CD2	−2.29	1.99	7.31	1.0	8	B19.2	Cys	N	−12.42	10.52	−2.96	1.0	9
B10.2	His	NE2	−1.59	0.94	6.88	1.0	9	B19.2	Cys	SG	−12.04	13.13	−1.34	1.0	10
B10.2	His	CE1	−2.08	0.57	5.77	1.0	11	B19.2	Cys	CB	−11.56	12.72	−3.01	1.0	11

TABLE 1.1. (*cont.*)

residue number	amino acid	atom type	x/Å	y/Å	z/Å	occ.	B/Å²
B19.2	Cys	CA	−12.43	11.76	−3.76	1.0	9
B19.2	Cys	C	−11.99	11.56	−5.21	1.0	15
B19.2	Cys	O	−12.41	12.30	−6.13	1.0	16
B20.2	Gly	N	−11.15	10.61	−5.46	1.0	17
B20.2	Gly	CA	−10.69	10.36	−6.85	1.0	22
B20.2	Gly	C	−10.27	11.65	−7.52	1.0	21
B20.2	Gly	O	−9.49	12.48	−7.05	1.0	20
B21.2	Glu	N	−10.78	11.84	−8.71	1.0	30
B21.2D1	Glu	OE2	−13.23	11.77	−12.96	0.5	50
B21.2D1	Glu	OE1	−12.52	10.19	−11.56	0.5	48
B21.2D1	Glu	CD	−12.71	11.35	−11.90	0.5	62
B21.2D1	Glu	CG	−12.31	12.52	−11.05	0.5	56
B21.2D2	Glu	OE2	−8.51	13.54	−13.78	0.5	27
B21.2D2	Glu	OE1	−10.54	12.77	−14.16	0.5	50
B21.2D2	Glu	CD	−9.69	13.29	−13.47	0.5	53
B21.2D2	Glu	CG	−9.80	13.41	−11.97	0.5	75
B21.2	Glu	CB	−10.78	12.72	−10.97	1.0	29
B21.2	Glu	CA	−10.40	13.04	−9.50	1.0	24
B21.2	Glu	C	−10.90	14.36	−9.06	1.0	19
B21.2	Glu	O	−10.43	15.33	−9.66	1.0	19
B22.2	Arg	N	−11.70	14.49	−8.03	1.0	15
B22.2D1	Arg	NH1	−16.74	17.67	−3.42	0.5	20
B22.2D1	Arg	NH2	−17.41	15.55	−3.87	0.5	35
B22.2D1	Arg	CZ	−16.61	16.58	−4.14	0.5	40
B22.2D1	Arg	NE	−15.74	16.40	−5.12	0.5	16
B22.2D2	Arg	NH2	−16.70	11.94	−4.18	0.5	20
B22.2D2	Arg	NH1	−16.99	14.05	−3.39	0.5	52
B22.2D2	Arg	CZ	−16.38	13.23	−4.28	0.5	33
B22.2D2	Arg	NE	−15.63	13.81	−5.27	0.5	18
B22.2	Arg	CD	−15.68	15.17	−5.96	1.0	31
B22.2	Arg	CG	−14.65	15.43	−7.04	1.0	58
B22.2	Arg	CB	−13.23	15.68	−6.46	1.0	22
B22.2	Arg	CA	−12.09	15.73	−7.46	1.0	16
B22.2	Arg	C	−10.80	16.22	−6.74	1.0	18
B22.2	Arg	O	−10.64	17.46	−6.61	1.0	21
B23.2	Gly	N	−10.01	15.25	−6.29	1.0	19
B23.2	Gly	CA	−8.84	15.67	−5.49	1.0	12
B23.2	Gly	C	−9.34	15.93	−4.07	1.0	13
B23.2	Gly	O	−10.52	15.92	−3.63	1.0	14
B24.2	Phe	N	−8.34	16.17	−3.19	1.0	13
B24.2	Phe	CD2	−6.67	14.20	−0.04	1.0	9
B24.2	Phe	CE2	−5.57	13.39	0.03	1.0	12
B24.2	Phe	CZ	−5.46	12.47	−1.01	1.0	10
B24.2	Phe	CE1	−6.38	12.34	−2.11	1.0	15
B24.2	Phe	CD1	−7.49	13.20	−2.07	1.0	6
B24.2	Phe	CG	−7.64	14.11	−1.03	1.0	6
B24.2	Phe	CB	−8.75	15.11	−1.03	1.0	4
B24.2	Phe	CA	−8.58	16.43	−1.76	1.0	10
B24.2	Phe	C	−7.49	17.22	−1.12	1.0	11
B24.2	Phe	O	−6.41	17.41	−1.66	1.0	11
B25.2	Phe	N	−7.72	17.61	0.12	1.0	14
B25.2	Phe	CD2	−6.00	21.54	0.90	1.0	57
B25.2	Phe	CE2	−5.65	22.56	−0.03	1.0	40
B25.2	Phe	CZ	−6.38	22.68	−1.24	1.0	30
B25.2	Phe	CE1	−7.45	21.79	−1.46	1.0	30
B25.2	Phe	CD1	−7.84	20.80	−0.54	1.0	61
B25.2	Phe	CG	−7.11	20.69	0.64	1.0	30
B25.2	Phe	CB	−7.39	19.63	1.68	1.0	17
B25.2	Phe	CA	−6.81	18.30	1.05	1.0	12
B25.2	Phe	C	−6.57	17.36	2.22	1.0	13
B25.2	Phe	O	−7.49	16.79	2.76	1.0	15
B26.2	Tyr	N	−5.34	17.20	2.58	1.0	11
B26.2	Tyr	OH	−1.85	12.45	7.73	1.0	16
B26.2	Tyr	CD2	−2.11	14.68	4.86	1.0	14
B26.2	Tyr	CE2	−1.58	14.01	5.94	1.0	11
B26.2	Tyr	CZ	−2.35	13.15	6.64	1.0	12
B26.2	Tyr	CE1	−3.65	12.97	6.25	1.0	16
B26.2	Tyr	CD1	−4.24	13.69	5.18	1.0	21
B26.2	Tyr	CG	−3.40	14.53	4.47	1.0	13
B26.2	Tyr	CB	−3.95	15.29	3.32	1.0	10
B26.2	Tyr	CA	−5.00	16.33	3.72	1.0	10
B26.2	Tyr	C	−4.45	17.35	4.71	1.0	15
B26.2	Tyr	O	−3.35	17.91	4.52	1.0	15
B27.2	Thr	N	−5.29	17.54	5.75	1.0	15
B27.2D1	Thr	CG2	−5.51	20.78	5.61	0.5	35
B27.2D1	Thr	OG1	−7.23	19.33	6.56	0.5	40
B27.2D2	Thr	CG2	−5.41	20.85	7.86	0.5	34
B27.2D2	Thr	OG1	−5.69	20.48	5.51	0.5	29
B27.2	Thr	CB	−5.86	19.83	6.75	1.0	26
B27.2	Thr	CA	−4.90	18.56	6.77	1.0	20
B27.2	Thr	C	−4.83	18.05	8.25	1.0	15
B27.2	Thr	O	−5.82	18.28	8.94	1.0	23
B28.2	Pro	N	−3.70	17.55	8.60	1.0	18
B28.2	Pro	CG	−1.54	16.62	8.60	1.0	21
B28.2	Pro	CD	−2.53	17.32	7.78	1.0	14
B28.2	Pro	CB	−2.25	16.21	9.81	1.0	22
B28.2	Pro	CA	−3.49	17.06	9.95	1.0	22
B28.2	Pro	C	−3.31	18.22	10.89	1.0	23
B28.2	Pro	O	−3.07	19.33	10.48	1.0	22
B29.2	Lys	N	−3.45	17.98	12.18	1.0	26
B29.2	Lys	NZ	−4.21	19.42	18.84	1.0	61
B29.2	Lys	CE	−5.07	18.49	18.09	1.0	53
B29.2	Lys	CD	−5.15	18.81	16.62	1.0	37
B29.2	Lys	CG	−3.99	19.39	15.80	1.0	48
B29.2	Lys	CB	−3.76	18.42	14.61	1.0	22
B29.2	Lys	CA	−3.23	18.94	13.31	1.0	23
B29.2	Lys	C	−1.71	19.00	13.46	1.0	53
B29.2	Lys	O	−1.03	17.92	13.41	1.0	39
B30.2	Ala	N	−1.17	20.05	13.78	1.0	53
B30.2	Ala	CB	0.03	22.11	13.69	1.0	53
B30.2	Ala	CA	0.15	20.54	13.90	1.0	45
B30.2	Ala	C	0.99	20.47	15.17	1.0	50
B30.2	Ala	O	0.43	20.71	16.27	1.0	63
B30.2	Ala	OE	2.23	20.20	15.00	1.0	76

residue number	zinc	x/Å	y/Å	z/Å	occ.[a]	B/Å²
M0.1	ZN1	0.00	0.00	−8.04	1.0	11
M0.2	ZN2	0.00	0.00	7.89	1.0	10

residue number	water		x/Å	y/Å	z/Å	occ.	B/Å²
W0.1	Wat	OW1	1.21	0.92	−0.24	1.0	44
W0.2	Wat	OW2	26.67	0.03	0.12	1.0	64
W1.1	Wat	OW1	13.44	19.18	0.63	1.0	37
W1.2D1	Wat	OW2	8.90	22.66	0.83	0.5	73
W1.3D1	Wat	OW3	5.43	4.63	0.35	0.5	68
W1.4	Wat	OW4	−9.60	22.80	0.34	1.0	86
W2.1	Wat	OW1	−11.31	20.40	0.77	1.0	76
W2.2D1	Wat	OW2	10.78	23.85	1.40	0.5	45
W1.5	Wat	OW5	14.91	22.15	0.11	1.0	64
W4.1	Wat	OW1	3.21	6.72	2.19	1.0	44
W4.2	Wat	OW2	15.23	18.99	2.77	1.0	48
W5.1	Wat	OW1	1.95	1.41	2.00	1.0	83
W5.2D1	Wat	OW2	−12.20	23.33	2.56	0.5	53
W6.1	Wat	OW1	3.29	3.44	2.90	1.0	29
W6.2	Wat	OW2	−0.46	3.09	2.99	1.0	35
W6.3D1	Wat	OW3	10.31	22.83	3.12	0.5	54
W6.4D1	Wat	OW4	10.18	24.20	3.00	0.5	48
W6.5	Wat	OW5	23.65	0.90	3.00	1.0	86
W7.1D1	Wat	OW1	8.26	18.78	3.33	0.5	14
W7.2D1	Wat	OW2	0.00	0.00	2.72	0.3	78
W7.3	Wat	OW3	−5.50	23.30	3.20	1.0	76
W7.4	Wat	OW4	−10.41	20.89	3.18	1.0	46
W7.5	Wat	OW5	5.79	4.69	3.30	1.0	79
W7.6D1	Wat	OW6	12.04	23.10	3.48	0.5	31
W10.1D1	Wat	OW1	−9.40	22.60	4.90	0.5	52
W9.1	Wat	OW1	9.78	14.00	4.35	1.0	38
W9.2D1	Wat	OW2	14.49	19.05	4.96	0.5	47
W9.4	Wat	OW4	24.80	2.00	4.68	1.0	51
W9.5	Wat	OW5	1.70	14.43	4.49	1.0	61
W11.1	Wat	OW1	−7.62	16.10	5.53	1.0	26
W11.3	Wat	OW3	4.07	5.89	5.14	1.0	61
W11.2D1	Wat	OW2	12.74	23.43	5.60	0.5	40
W11.5D1	Wat	OW5	10.35	21.08	4.83	0.5	34
W11.6D1	Wat	OW6	12.37	21.48	5.16	0.5	61

TABLE 1.1. (*cont.*)

residue number	water		x/Å	y/Å	z/Å	occ.	B/Å²
W12.2D1	Wat	OW2	−9.10	20.60	6.10	0.5	60
W12.3D1	Wat	OW3	−8.77	19.97	5.68	0.5	51
W13.1	Wat	OW1	13.10	13.72	6.03	1.0	21
W13.2D1	Wat	OW2	−7.99	22.59	6.58	0.5	57
W13.3	Wat	OW3	−12.27	21.57	6.10	1.0	55
W14.1	Wat	OW1	−2.65	21.39	6.53	1.0	65
W14.2	Wat	OW2	25.21	3.03	6.89	1.0	50
W14.3D1	Wat	OW3	10.60	10.10	6.80	0.5	50
W14.6D1	Wat	OW6	11.19	23.51	6.62	0.5	37
W15.1	Wat	OW1	14.53	16.19	7.07	1.0	51
W15.2D1	Wat	OW2	1.49	23.23	6.89	0.5	27
W15.5D1	Wat	OW5	14.50	18.88	6.98	0.5	60
W15.6D1	Wat	OW6	−9.48	23.46	7.07	0.5	41
W16.1	Wat	OW1	1.23	4.48	7.60	1.0	34
W16.2D1	Wat	OW2	0.17	22.96	7.61	0.5	29
W16.4D1	Wat	OW4	10.08	20.63	8.07	0.5	33
W14.4D1	Wat	OW4	8.28	22.35	6.63	0.5	71
W14.5D1	Wat	OW5	11.20	20.20	6.48	0.5	43
W14.7D1	Wat	OW7	9.47	21.71	6.55	0.5	61
W16.5D1	Wat	OW5	12.91	19.86	7.88	0.5	42
W16.6D1	Wat	OW6	5.79	10.58	7.55	0.5	19
W17.1	Wat	OW1	−3.19	10.01	8.36	1.0	20
W17.2	Wat	OW2	3.54	5.19	7.51	1.0	51
W17.3D1	Wat	OW3	−2.60	23.10	8.10	0.5	38
W17.4	Wat	OW4	24.91	0.40	8.40	1.0	60
W17.5	Wat	OW5	22.96	4.31	7.97	1.0	75
W18.1	Wat	OW1	3.16	12.35	8.53	1.0	60
W18.2D1	Wat	OW2	6.99	23.53	8.61	0.5	43
W18.3D1	Wat	OW3	7.75	22.51	8.74	0.5	51
W18.4	Wat	OW4	11.35	23.30	9.17	1.0	51
W18.5	Wat	OW5	13.64	22.50	8.40	1.0	61
W18.6	Wat	OW6	5.40	7.04	8.02	1.0	76
W19.1D1	Wat	OW1	−9.44	20.24	8.42	0.5	47
W19.3	Wat	OW3	21.20	5.80	9.00	1.0	47
W19.4D1	Wat	OW4	−0.56	20.34	9.23	0.5	31
W20.1	Wat	OW1	−0.07	1.53	9.45	1.0	26
W20.2	Wat	OW2	18.62	13.60	9.21	1.0	29
W20.3	Wat	OW3	0.47	13.40	9.00	1.0	61
W20.4	Wat	OW4	2.96	7.36	9.26	1.0	37
W20.5D1	Wat	OW5	11.91	19.70	9.07	0.5	36
W20.6	Wat	OW6	16.09	16.07	9.93	1.0	48
W20.7D1	Wat	OW7	−8.47	18.40	9.20	0.5	34
W20.8	Wat	OW8	13.48	17.84	9.05	1.0	66
W20.9D1	Wat	OW9	4.83	22.59	10.70	0.5	41
W21.1	Wat	OW1	4.50	3.27	9.80	1.0	21
W21.2	Wat	OW2	9.90	19.40	10.18	1.0	57
W21.3D1	Wat	OW3	−9.81	18.86	10.17	0.5	36
W21.4D1	Wat	OW4	8.30	21.90	9.77	0.5	63
W21.5	Wat	OW5	5.53	14.05	9.25	1.0	53
W21.6D1	Wat	OW6	0.50	21.50	9.74	0.5	41
W21.7D1	Wat	OW7	24.50	1.84	10.26	0.5	44
W21.8D1	Wat	OW8	14.40	20.41	9.88	0.3	63
W21.9D1	Wat	OW9	−9.50	22.80	10.02	0.5	41
W22.1	Wat	OW1	−5.37	9.30	10.39	1.0	29
W22.2	Wat	OW2	5.10	10.01	10.63	1.0	40
W22.3D1	Wat	OW3	3.40	22.24	10.23	0.5	39
W22.4D1	Wat	OW4	0.79	10.46	10.42	0.5	27
W22.5D1	Wat	OW5	7.10	20.64	10.12	0.5	46
W22.6D1	Wat	OW6	1.33	5.58	11.05	0.5	56
W22.7D1	Wat	OW7	−11.77	21.61	10.65	0.5	57
W22.8D1	Wat	OW8	−9.99	20.28	10.19	0.5	56
W23.1	Wat	OW1	−7.08	17.02	11.11	1.0	25
W23.2	Wat	OW2	10.11	17.14	11.05	1.0	38
W23.3D1	Wat	OW3	13.12	21.92	10.72	0.5	30
W23.4D1	Wat	OW4	−8.63	22.49	11.84	0.5	44
W23.5	Wat	OW5	24.00	4.40	11.20	1.0	64
W24.1D1	Wat	OW1	0.40	4.49	11.76	0.5	23
W24.2	Wat	OW2	20.20	6.30	11.10	1.0	62
W24.3	Wat	OW3	−6.06	22.69	11.49	1.0	62
W24.4	Wat	OW4	2.40	0.70	11.14	1.0	63
W24.5	Wat	OW5	1.43	8.00	10.90	1.0	62
W24.6D1	Wat	OW6	0.59	13.73	11.30	0.5	60
W24.7D1	Wat	OW7	−0.22	11.64	11.24	0.5	52
W24.9D1	Wat	OW9	3.58	13.15	11.26	0.5	53
W25.1D1	Wat	OW1	−0.02	−0.03	11.21	0.3	21
W25.2	Wat	OW2	3.72	4.94	12.00	1.0	60
W25.3	Wat	OW3	9.25	23.29	11.50	1.0	46
W25.4	Wat	OW4	−5.43	20.51	11.63	1.0	52
W25.5	Wat	OW5	13.71	17.82	11.96	1.0	66
W25.6	Wat	OW6	16.13	12.14	12.24	1.0	50
W25.7D1	Wat	OW7	12.90	20.42	11.49	0.5	31
W25.8	Wat	OW8	22.34	2.18	11.30	1.0	54
W26.1	Wat	OW1	−1.82	8.68	12.04	1.0	36
W26.2	Wat	OW2	2.94	2.62	12.78	1.0	46
W26.3	Wat	OW3	3.20	8.80	12.20	1.0	59
W27.1	Wat	OW1	6.25	11.75	12.89	1.0	50
W27.2D1	Wat	OW2	12.62	23.25	12.21	0.5	36
W27.3D1	Wat	OW3	6.08	21.44	12.95	0.5	46
W27.4	Wat	OW4	15.63	16.55	13.70	1.0	54
W27.5D1	Wat	OW5	4.21	21.40	12.87	0.5	43
W27.6	Wat	OW6	1.40	10.00	13.20	1.0	62
W28.1	Wat	OW1	−1.80	14.59	13.56	1.0	37
W28.2	Wat	OW2	10.68	15.95	13.46	1.0	38
W28.3D1	Wat	OW3	3.53	17.51	13.39	0.5	59
W28.4	Wat	OW4	4.99	14.66	13.14	1.0	52
W28.5D1	Wat	OW5	−3.06	10.93	13.00	0.5	55
W28.6	Wat	OW6	−8.65	24.05	2.00	1.0	56
W28.7D1	Wat	OW7	17.59	14.89	12.99	0.5	47
W28.8	Wat	OW8	−2.70	22.20	12.30	1.0	70
W28.9	Wat	OW9	−7.40	20.90	13.30	1.0	82
W29.1	Wat	OW1	7.84	13.59	13.24	1.0	37
W27.7D1	Wat	OW7	8.44	16.10	13.12	0.5	66
W29.2D1	Wat	OW2	20.77	2.76	14.13	0.5	50
W27.8D1	Wat	OW8	19.46	3.17	13.24	0.5	43
W29.3	Wat	OW3	10.60	21.20	12.80	1.0	65
W29.4	Wat	OW4	17.09	12.04	14.42	1.0	66
W29.5D1	Wat	OW5	26.23	1.15	12.98	0.5	43
W29.6	Wat	OW6	1.00	16.70	13.20	1.0	45
W29.7	Wat	OW7	−1.84	2.63	14.10	1.0	47
W29.8	Wat	OW8	8.20	19.89	13.34	1.0	60
W29.9D1	Wat	OW9	3.95	18.70	13.93	0.5	87
W30.1	Wat	OW1	3.00	8.25	14.50	1.0	39
W30.6	Wat	OW6	23.67	1.16	13.22	1.0	50
W30.7	Wat	OW7	6.71	17.66	14.20	1.0	62
W30.8	Wat	OW8	−10.44	21.80	14.70	1.0	83
W30.2	Wat	OW2	21.66	8.10	14.40	1.0	79
W30.3D1	Wat	OW3	16.00	14.16	14.20	0.5	36
W31.1	Wat	OW1	18.62	9.97	14.47	1.0	68
W31.2	Wat	OW2	12.35	18.36	14.32	1.0	62
W31.4D1	Wat	OW4	15.38	13.33	15.23	0.5	39
W31.5D1	Wat	OW5	8.90	16.07	15.11	0.5	56
W31.6	Wat	OW6	−8.81	19.80	14.80	1.0	84
W31.9D1	Wat	OW9	9.80	20.50	15.00	0.5	38
W32.1	Wat	OW1	−0.63	5.21	14.50	1.0	53
W30.5D1	Wat	OW5	0.06	7.37	13.48	0.5	44
W32.2	Wat	OW2	4.33	21.35	15.20	1.0	48
W32.4D1	Wat	OW4	20.47	6.31	15.33	0.5	42
W32.5	Wat	OW5	15.96	10.35	15.58	1.0	73
W32.6	Wat	OW6	4.17	16.12	15.12	1.0	65
W33.1	Wat	OW1	−3.50	10.46	15.32	1.0	42
W33.3	Wat	OW3	22.18	1.59	16.32	1.0	60
W33.4	Wat	OW4	7.23	21.58	16.03	1.0	80
W33.5D1	Wat	OW5	24.99	3.02	15.00	0.5	78
W34.1D1	Wat	OW1	0.00	0.00	15.94	0.3	48
W34.2D1	Wat	OW2	2.06	1.12	15.89	0.5	57
W34.3	Wat	OW3	−1.88	13.66	16.15	1.0	61
W34.4	Wat	OW4	−3.12	22.86	15.78	1.0	59
W34.5	Wat	OW5	19.20	4.00	16.26	1.0	52
W34.6D1	Wat	OW6	14.04	21.77	15.73	0.5	35
W34.7	Wat	OW7	1.90	3.40	16.73	1.0	43
W34.9D1	Wat	OW9	14.40	17.80	15.89	0.5	53
W33.6	Wat	OW6	13.70	15.60	15.00	1.0	65
W35.1	Wat	OW1	2.10	6.20	15.80	1.0	74
W35.2D1	Wat	OW2	13.10	18.67	16.69	0.5	58
W35.5	Wat	OW5	5.38	1.95	16.33	1.0	53
W35.6	Wat	OW6	19.39	1.65	16.54	1.0	39

TABLE 1.1. (*cont.*)

residue number	water		x/Å	y/Å	z/Å	occ.	B/Å²
W35.7	Wat	OW7	6.57	18.43	16.50	1.0	51
W35.8D1	Wat	OW8	−11.76	21.10	16.70	0.5	40
W35.9	Wat	OW9	18.27	7.54	16.57	1.0	60
W35.0	Wat	OW0	16.96	12.60	17.17	1.0	81
W35.4	Wat	OW4	19.27	10.58	17.79	1.0	63
W36.1	Wat	OW1	−2.56	5.64	17.14	1.0	48
W36.2	Wat	OW2	−3.81	12.17	17.18	1.0	40
W36.3	Wat	OW3	21.34	8.54	16.85	1.0	46
W36.4	Wat	OW4	11.36	23.31	16.93	1.0	57
W36.5D1	Wat	OW5	11.48	20.11	16.18	0.5	36
W36.6D1	Wat	OW6	10.43	19.61	16.70	0.5	62
W36.7	Wat	OW7	−0.46	10.10	16.80	1.0	96
W36.8D1	Wat	OW8	−5.60	22.74	16.75	0.5	61
W36.9D1	Wat	OW9	22.70	4.88	16.53	0.5	59
W36.0	Wat	OW0	−0.56	7.13	16.71	1.0	55
W37.0D1	Wat	OW0	22.35	3.89	16.63	0.5	45
W37.9D1	Wat	OW9	24.25	2.53	17.22	0.5	65
W37.1D1	Wat	OW1	−8.27	20.62	17.54	0.5	29
W37.6D1	Wat	OW6	−7.61	21.77	17.40	0.5	40
W37.2	Wat	OW2	17.31	15.07	17.58	1.0	54
W37.4	Wat	OW4	−1.70	21.23	17.21	1.0	50
W37.5	Wat	OW5	16.51	9.56	17.81	1.0	54
W37.7	Wat	OW7	−7.80	16.30	17.67	1.0	76
W38.1	Wat	OW1	4.68	20.38	17.88	1.0	49
W38.2	Wat	OW2	17.53	3.75	17.79	1.0	52
W38.3	Wat	OW3	−2.27	8.87	17.97	1.0	49
W38.6	Wat	OW6	−8.89	18.34	17.94	1.0	52
W38.7D1	Wat	OW7	0.50	3.00	18.60	0.5	35
W38.8	Wat	OW8	6.00	4.10	17.80	1.0	50
W39.0D1	Wat	OW0	2.70	3.90	19.00	0.5	40
W39.1D1	Wat	OW1	2.99	5.26	18.40	0.5	40
W38.9	Wat	OW9	22.32	6.62	18.38	1.0	76
W37.8D1	Wat	OW8	14.30	15.15	17.40	0.5	41
W39.2	Wat	OW2	12.38	14.00	18.49	1.0	69
W39.4	Wat	OW4	26.58	1.03	18.30	1.0	72
W39.5D1	Wat	OW5	15.05	0.59	18.65	0.5	28
W39.6D1	Wat	OW6	24.66	3.16	18.64	0.5	56
W39.7	Wat	OW7	16.00	6.40	18.50	1.0	75
W40.1	Wat	OW1	−4.84	15.67	19.01	1.0	34
W40.2	Wat	OW2	−1.20	22.84	19.14	1.0	36
W40.3	Wat	OW3	1.69	21.19	18.89	1.0	37
W40.4	Wat	OW4	7.59	19.90	18.90	1.0	74
W40.5D1	Wat	OW5	23.64	0.40	18.43	0.5	39
W40.7	Wat	OW7	7.20	0.00	19.36	1.0	80
W40.8	Wat	OW8	12.32	20.19	18.58	1.0	76
W40.9	Wat	OW9	8.87	21.92	18.72	1.0	88
W41.1	Wat	OW1	−5.30	12.85	19.15	1.0	51
W41.2	Wat	OW2	4.00	0.89	19.16	1.0	56
W41.3	Wat	OW3	11.50	16.20	18.50	1.0	55
W41.4D1	Wat	OW4	15.03	14.60	19.14	0.5	41
W41.5D1	Wat	OW5	16.70	11.30	19.50	0.5	53
W42.1	Wat	OW1	4.30	23.91	8.30	1.0	56
W42.2	Wat	OW2	17.59	8.31	19.86	1.0	44
W42.3	Wat	OW3	8.46	17.05	19.90	1.0	59
W42.4	Wat	OW4	11.60	18.40	20.00	1.0	60
W42.5	Wat	OW5	−9.41	23.12	19.63	1.0	66
W42.8	Wat	OW8	7.20	3.00	19.67	1.0	79
W42.9	Wat	OW9	21.00	4.00	19.96	1.0	59
W42.0	Wat	OW0	−0.80	4.80	19.70	1.0	50
W43.1	Wat	OW1	13.59	4.02	20.05	1.0	43
W43.2D1	Wat	OW2	−0.40	2.60	20.00	0.5	40
W43.3	Wat	OW3	19.48	6.22	19.80	1.0	63
W43.5D1	Wat	OW5	20.88	1.57	20.12	0.5	42
W43.7	Wat	OW7	18.87	3.20	20.30	1.0	65
W43.8D1	Wat	OW8	22.04	8.45	20.60	0.5	38
W43.9	Wat	OW9	15.42	7.74	20.81	1.0	62
W44.1	Wat	OW1	14.40	19.90	20.80	1.0	46
W44.2	Wat	OW2	8.90	4.30	20.90	1.0	67
W44.3	Wat	OW3	12.90	7.70	20.60	1.0	51
W44.4D1	Wat	OW4	23.37	3.91	20.78	0.5	38
W44.5D1	Wat	OW5	22.60	5.20	21.00	0.5	50
W44.6D1	Wat	OW6	11.00	20.30	21.10	0.5	50
W44.7	Wat	OW7	3.85	22.05	20.44	1.0	74
W44.8D1	Wat	OW8	15.90	13.40	20.50	0.5	37
W44.9D1	Wat	OW9	15.00	15.90	20.41	0.5	52
W45.1	Wat	OW1	−5.65	12.11	21.37	1.0	37
W45.3	Wat	OW3	−3.50	9.53	21.14	1.0	58
W45.2D1	Wat	OW2	11.45	5.67	21.40	0.5	45
W45.5	Wat	OW5	19.30	11.20	21.25	1.0	94
W45.6	Wat	OW6	12.87	15.31	21.38	1.0	105
W45.0D1	Wat	OW0	16.20	18.50	20.52	0.5	57
W45.9D1	Wat	OW9	17.90	14.20	21.40	0.5	54
W46.1	Wat	OW1	−2.44	17.55	21.49	1.0	28
W46.2D1	Wat	OW2	2.43	2.79	21.67	0.5	37
W47.3	Wat	OW3	5.59	3.51	22.64	1.0	67
W47.4D1	Wat	OW4	21.49	9.78	21.79	0.5	37
W47.5D1	Wat	OW5	22.46	7.10	21.91	0.5	54
W47.6D1	Wat	OW6	23.36	5.25	22.12	0.5	53
W47.7D1	Wat	OW7	10.03	5.99	22.15	0.5	37
W48.1	Wat	OW1	16.23	16.76	22.90	1.0	52
W48.2	Wat	OW2	7.95	1.50	22.40	1.0	74
W48.4D1	Wat	OW4	0.00	0.00	22.20	0.3	27
W48.5	Wat	OW5	6.00	0.33	22.81	1.0	82
W48.6D1	Wat	OW6	2.40	1.90	22.95	0.5	54
W48.7	Wat	OW7	16.83	9.36	22.92	1.0	65
W48.8	Wat	OW8	10.03	0.60	22.67	1.0	51
W49.1	Wat	OW1	1.80	22.10	22.79	1.0	34
W49.2	Wat	OW2	−2.16	10.17	22.90	1.0	35
W49.3	Wat	OW3	8.57	22.21	23.27	1.0	53
W49.4	Wat	OW4	19.04	7.32	22.97	1.0	45
W49.6D1	Wat	OW6	13.15	17.90	23.12	0.5	59
W49.7	Wat	OW7	25.12	2.79	23.67	1.0	50
W49.8D1	Wat	OW8	0.85	2.63	23.30	0.5	52
W50.1	Wat	OW1	12.80	4.91	23.30	1.0	45
W50.2	Wat	OW2	14.10	1.51	23.00	1.0	55
W50.3D1	Wat	OW3	18.00	5.20	23.40	0.5	40
W50.4D1	Wat	OW4	20.46	5.25	23.19	0.5	43
W50.5D1	Wat	OW5	18.50	13.24	23.93	0.5	60
W50.6D1	Wat	OW6	9.47	4.01	23.26	0.5	65
W51.1	Wat	OW1	3.33	4.80	23.76	1.0	24
W51.2	Wat	OW2	−6.41	11.03	24.09	1.0	36
W51.4	Wat	OW4	−1.80	3.54	24.06	1.0	81
W51.3	Wat	OW3	1.47	0.22	24.18	1.0	35
W51.5	Wat	OW5	−12.01	22.14	24.30	1.0	67
W51.6D1	Wat	OW6	13.05	19.51	23.13	0.5	63
W51.7	Wat	OW7	−9.52	22.73	24.41	1.0	72
W52.1	Wat	OW1	9.90	22.89	25.26	1.0	65
W52.2	Wat	OW2	16.50	17.13	26.00	1.0	40
W52.4D1	Wat	OW4	21.70	5.40	24.51	0.5	50
W52.5	Wat	OW5	6.64	20.68	24.90	1.0	59
W52.6D1	Wat	OW6	20.69	8.31	24.46	0.5	48
W53.1	Wat	OW1	−1.76	20.08	25.01	1.0	21
W53.2	Wat	OW2	−8.67	18.37	25.26	1.0	27
W53.3	Wat	OW3	7.70	4.22	25.06	1.0	34
W53.5	Wat	OW5	19.58	11.07	25.09	1.0	65
W53.6	Wat	OW6	−2.72	10.30	25.15	1.0	73
W53.7D1	Wat	OW7	22.45	8.70	25.16	0.5	48
W54.1	Wat	OW1	−1.29	22.96	25.51	1.0	30
W54.2D1	Wat	OW2	12.21	21.70	25.85	0.5	44
W52.7D1	Wat	OW7	11.50	21.60	24.16	0.5	71
W54.4	Wat	OW4	25.80	0.09	26.60	1.0	57
W54.5	Wat	OW5	23.39	1.29	25.73	1.0	49
W54.6	Wat	OW6	12.19	3.11	25.00	1.0	76
W55.1	Wat	OW1	4.71	1.34	26.04	1.0	31
W55.2	Wat	OW2	14.39	19.74	25.64	1.0	65
W56.1	Wat	OW1	5.42	3.67	26.43	1.0	31
W56.2	Wat	OW2	23.51	4.50	26.63	1.0	66
W56.3	Wat	OW3	−10.10	20.40	26.04	1.0	73
W57.1	Wat	OW1	−9.60	23.17	26.89	1.0	63
W57.2	Wat	OW2	16.39	14.06	27.17	1.0	75
W57.4	Wat	OW4	18.70	14.00	26.53	1.0	79
W58.2	Wat	OW2	10.93	10.68	26.94	1.0	45
W58.3D1	Wat	OW3	12.50	21.80	27.25	0.5	48
W58.1	Wat	OW1	15.04	21.55	5.27	1.0	56
W59.1	Wat	OW1	−1.74	18.45	27.70	1.0	12
W59.2	Wat	OW2	27.00	1.00	28.30	1.0	71

TABLE 1.1. (*cont.*)

residue number	water		x/Å	y/Å	z/Å	occ.	B/Å²	residue number	water		x/Å	y/Å	z/Å	occ.	B/Å²
W59.3D1	Wat	OW3	20.88	8.93	27.40	0.5	49	W65.1D1	Wat	OW1	0.00	0.00	30.94	0.3	24
W60.1D1	Wat	OW1	5.30	3.96	28.70	0.6	32	W65.2	Wat	OW2	3.56	3.11	30.84	1.0	35
W60.2	Wat	OW2	13.79	13.32	28.06	1.0	74	W65.3	Wat	OW3	−1.32	2.46	30.83	1.0	28
W60.3D1	Wat	OW3	6.58	4.37	28.78	0.4	29	W67.1	Wat	OW1	6.90	3.47	31.64	1.0	39
W61.1	Wat	OW1	19.25	14.02	28.83	1.0	55	W67.2	Wat	OW2	1.48	2.38	31.64	1.0	81
W61.2	Wat	OW2	14.28	1.57	28.96	1.0	65	W67.3	Wat	OW3	−11.70	22.92	32.10	1.0	83
W62.2	Wat	OW2	14.08	9.85	29.48	1.0	28	W67.4	Wat	OW4	12.20	21.56	31.60	1.0	55
W62.1D1	Wat	OW1	23.14	6.64	29.29	0.5	56	W69.1	Wat	OW1	25.19	1.53	33.00	1.0	78
W64.1D1	Wat	OW1	18.59	7.95	30.40	0.5	18	W70.1	Wat	OW1	11.09	22.75	33.20	1.0	96
W63.1	Wat	OW1	10.48	22.33	30.03	1.0	41	W70.2	Wat	OW2	3.30	3.65	33.20	1.0	86
W63.2	Wat	OW2	14.20	22.30	30.30	1.0	67	W71.1	Wat	OW1	24.24	4.74	33.83	1.0	39
W64.2	Wat	OW2	4.56	5.54	30.13	1.0	37	W72.1	Wat	OW1	6.36	23.61	−0.02	1.0	95

[a] The occupancy of the zinc ions within the hexamer. In calculating structure factors the zinc-ion occupancy is set to 0.33.

The relation between the rhombohedral and hexagonal unit cells and the crystal symmetry elements controlling these atomic positions is illustrated in figure 1.1 *a*. Figure 1.1 *b* shows the mass centres of the two insulin molecules in each asymmetric unit, repeated according to the formal listing of equivalent positions in the international tables. Between the molecules can be traced lines of non-crystallographic approximate twofold axes, marked at *d* and *d'* in figure 1.1 *a*. Within the crystal each molecule is in direct contact with five molecules of the alternative variety, i.e. 1 is closely surrounded by molecules of type 2 and vice versa. Only near the threefold axes can each molecule make a sixth contact with a like molecule.

1.2. *X-ray data and the analysis*

The X-ray data were 13761 structure amplitudes measured with Cu Kα radiation at room temperature to 1.5 Å spacing on a Hilger Y290 four-circle diffractometer; *ca.* 6 crystals were used. The number of amplitudes greater than 2σ was 10119 (74%). The intensities were corrected for the Lorentz and polarization effects and for absorption by the method of North *et al.* (1968).

The positional and thermal parameters of the atoms reached in the latest structure factor calculations ($R = 15.3\%$) are given in table 1.1. This includes the positions of 2 zinc ions and 831 positions corresponding to the 808 carbon, nitrogen, oxygen and sulphur atoms of two insulin molecules. Forty-six of the sites of carbon, nitrogen and oxygen atoms are weighted 0.5 and correspond with alternative disordered positions of seven residues: arginine, B22 of molecules 1 and 2, lysine B29 of molecule 1, and glutamine B4, valine B12, glutamic acid B21 and threonine B27, all of molecule 2. The 750 hydrogen atoms included in the calculation of the later structure factor sets, were placed by calculation and are not listed in table 1.1. Those in methyl groups were assumed to be non rotating and placed in staggered positions. There was usually electron density near their positions but this was too imprecise in difference maps, calculated when they were omitted from phasing, to be used to place them. Table 1.1 records also the parameters of 349 sites for water molecules of which 217 were weighted 1.00, 126, 0.5, 5, 0.33 and 1, 0.25 making a total of *ca.* 282, close to the figure 283 derived from the crystal density.

The coordinates of the atoms of the insulin molecule were first obtained from the isomorphously phased map by using data extending to 1.9 Å, then refined and corrected, first

through cycles of difference Fouriers, then later by least-squares fitting of the parameters to both the intensity data, extended to 1.5 Å, and to the expected geometrical parameters of the ideal protein structure. Throughout the process we have used the fast Fourier process to calculate structure factors and gradients (Agarwal 1978). At first the least-squares fitting of the geometric restraints was done in a separate calculation after shifts had been applied to the intensity data (model-fit) (Dodson *et al.* 1976); later further cycles were done where the information from the X-ray observations is passed to the Konnert–Hendrickson program, which minimized the fit to all the observations simultaneously (Konnert & Hendrickson 1980). Both systems require the user to choose a relative weighting of the various contributions to the minimization matrix. Those employed in the latest structure factor calculations are shown in table 1.2; the value of the agreement factor $R(=\Sigma||F_o|-|F_c||/\Sigma|F_o|)$ in ranges of $4\sin^2\theta/\lambda^2$ is shown in figure 1.2. The complete set of calculated structure factors and observed amplitudes is deposited at Brookhaven (Bernstein *et al.* 1977).

The water molecule sites were selected from Fourier maps with coefficient F_o, F_o-F_c and $2F_o-F_c$ and from maps where $\frac{1}{8}$ of the structure was systematically excluded from the phasing (Dodson 1981). They were assigned occupancies based largely on their appearance in the maps and limited by the constraint that no fully occupied site must be less than 2.3 Å from another

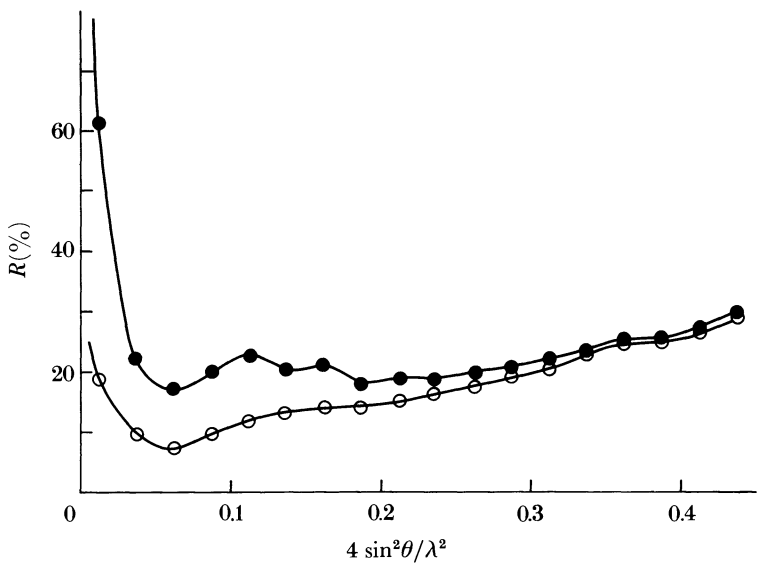


FIGURE 1.2. The agreement factor, R, evaluated for all terms in ranges of $4\sin^2\theta/\lambda^2$. The open circles are the values of R, calculated with protein atoms including hydrogens and the water molecules; the closed circles are the values of R calculated without the water molecules.

TABLE 1.2

molecular parameter	applied restraints (σ)/Å	r.m.s. deviation/Å
bond length	0.02	0.024
angle-related length	0.04	0.054
intraplanar distance	0.06	0.064
chiral centre	0.12	0.17
planarity	0.015	0.019
r.m.s. shift of coordinates in final cycle	—	0.1

site. *B*s were refined but not occupancies. In table 1.1 it is likely that occupancies and *B* values are only reliable for fully occupied sites with $B < 50$ Å^2 (compare p. 428). On the other hand, a considerable fall in *R* resulted for the inner terms from the inclusion of the less well defined water molecules sites. The set here listed is an extended version of that originally deposited at Brookhaven (Dodson *et al.* 1979).

The final reliability of a structure must be judged by the quality of the final Fourier maps. Figure 1.3 shows one section, at $z = 0$ in our latest electron-density map in which F_{000} is

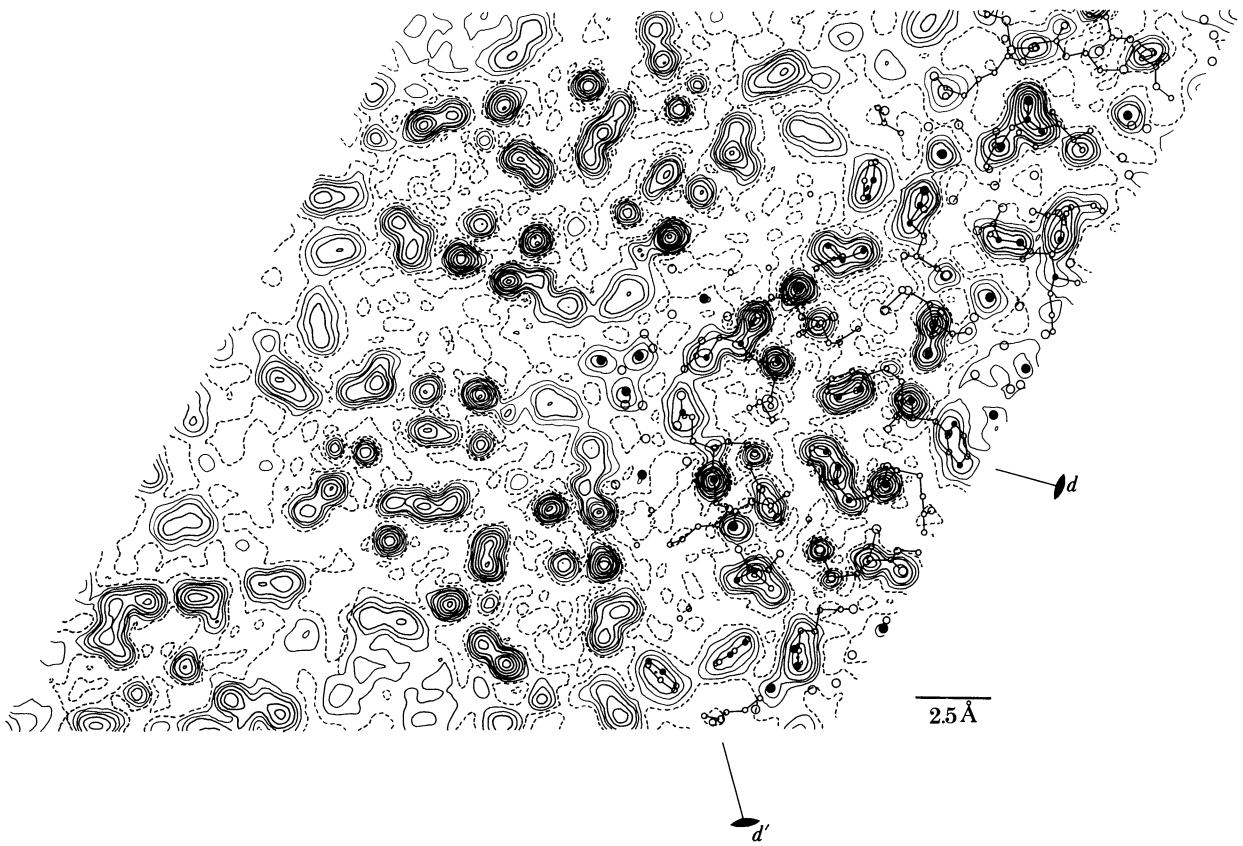


FIGURE 1.3. Section in the electron density map at $z = 0$, calculated with phases based on table 1.1 parameters. Contours: ---, 0, 0.2; ——, 0.4, 0.6; ——, 0.9; then at intervals of 0.3 e/Å^3. Over one third of the map, positions of atoms are shown, near section, open circles; within ± 0.5 Å of section, filled circles.

included in the summation for ρ_{xyz}. In this map the protein atoms appear as peaks generally between 1 and 2 e/Å^3 high, joined by ridges of up to 0.8 e/Å along the covalent bonds. Carbon and oxygen in carbonyl groups usually do not appear as separate peaks. Within the protein molecule, between the chains, the electron density falls sharply to zero or near it. By contrast, in the volumes occupied by water the contours are very diffuse, peaks are low, usually from 0.5 to 1 e/Å^3, and spreading; between them the electron density often falls as low as 0.2 e/Å^3, only very occasionally to zero. On the section at $z = 0$ the lines of the local twofold axes are clearly recognizable. The area round *d'* is characteristically empty, occupied by water a little away from the section except where penetrated by loosely interacting non-polar residues. The twofold symmetry round *d* is, on the other hand, broken at two points, by B12 valine and B25 phenylalanine, which make closer intermolecular contacts.

2. The organization of the crystal structure

Figure 2.1 *a*, plate 1 shows all the atomic positions as found in 42 insulin molecules projected along the trigonal ***c*** axis. In 2.1 *b* are shown, in the same projection, the water molecule positions over the same volume. The figure shows that within each rhombohedral unit cell six insulin molecules compose a hexamer round the two zinc ions, which lie 15.8 Å apart on the threefold axis; each zinc ion is coordinated octahedrally by three symmetry-related B10 histidine NE atoms and by three water molecules. Individually the hexamers are somewhat flattened spheroids, the packing of which in the crystal approximates to cubic body-centred close packing; each hexamer makes direct van der Waals and hydrogen-bonded contacts with eight other hexamers, two along the threefold axis at 34.0 Å and six along the rhombohedral ***a*** axis at 48.9 Å; six further contacts along the rhombohedral [10$\bar{1}$] axis 52.7 Å are looser; one or more water molecules intervene between the hexamer surfaces. Although the circumference of the hexamer appears rather smooth in projection, its upper and lower surfaces are in fact deeply grooved. As figure 2.2, plate 1, shows, this permits succeeding hexamers along the threefold axis to lock into one another; it also permits water to flow in channels between the threefold axis and water surrounding the hexamer circumference and along the threefold screw axes; as a consequence, the flow of water is continuous throughout the crystals as illustrated by figure 2.1 *b*.

Each hexamer (figure 2.3, plate 2) has the threefold symmetry axis of the crystal and also approximate twofold symmetry axes perpendicular to this, passing through the origin, interrelating the six molecules that compose it. The mass centres of these molecules, as shown in figure 1.1 *b*, form a rather flat irregular octahedron of edges 18.0–18.1 Å. The twist of the octahedron relative to the crystal axes permits an approximate close packing of the molecules throughout the crystal. Their exact relative positions appear to be controlled by the detailed contacts between residues on their surfaces (see p. 425). These are not very precisely defined; they adjust easily when the crystals dry. The hexamers then appear to turn a few degrees round the threefold axis into positions in which the close packing of the molecules is improved. The twist of the octahedron in the wet crystals carries the near twofold symmetry axes between molecules to positions *d* and *d'*, at 0 and $\frac{1}{2}$, 15° and 75° from the lines of the hexagonal ***a*** axis. The individual molecules, accordingly make different contacts in the crystal and, although similar, have not identical conformations.

The very specific nature of the contacts between the molecules around the dyad axis *d* at $z = 0$, d_0, which involve hydrogen bonds and many non polar interactions, make it *a priori* likely that this region is the site of aggregation of insulin molecules into dimers in solution. This conclusion is confirmed by the appearance of very similar insulin dimers in cubic pig insulin and in hagfish insulin crystals, where there is no further aggregation to hexamers (Dodson *et al.* 1979; Cutfield *et al.* 1979). The insulin dimer, as figure 2.4 shows, is an oblong of approximately 20 Å × 25 Å × 40 Å with a rather smooth surface.

In both the cubic and hagfish crystals the two molecules of the insulin dimer are identical, related by exact crystallographic twofold axes of symmetry. In the dimer found in the rhombohedral crystal one molecule is almost identical with these in conformation, the other differs in certain regions. For convenience in future we propose to call the molecule with the conserved conformation, molecule 1, and the second molecule, molecule 2, as in figures 1.1 *b* and 2.5 although this is the opposite to the convention we adopted earlier; it is the same as the convention used in the Chinese papers.

The two separate insulin molecules are shown projected along the axis d_0, around which molecule 1 has been rotated, in figure 2.5. Their construction is essentially similar. The two B chain residues between 9 and 19 form a helix from which the initial and terminal residues turn into generally extended conformations. The A chains are both compact and rest within the framework of the extended B chains and the B chain helices to which they are linked by disulphide bonds, A20–B19 and A7–B7. Viewed as individuals, the two insulin molecules are quite irregular in shape; the initial residues of the B chain extend somewhat loosely from the body of the molecule to make contacts with the neighbouring molecules in the hexamer. The relative positions and contacts of the amino acid residues in the two molecules may now be described in detail.

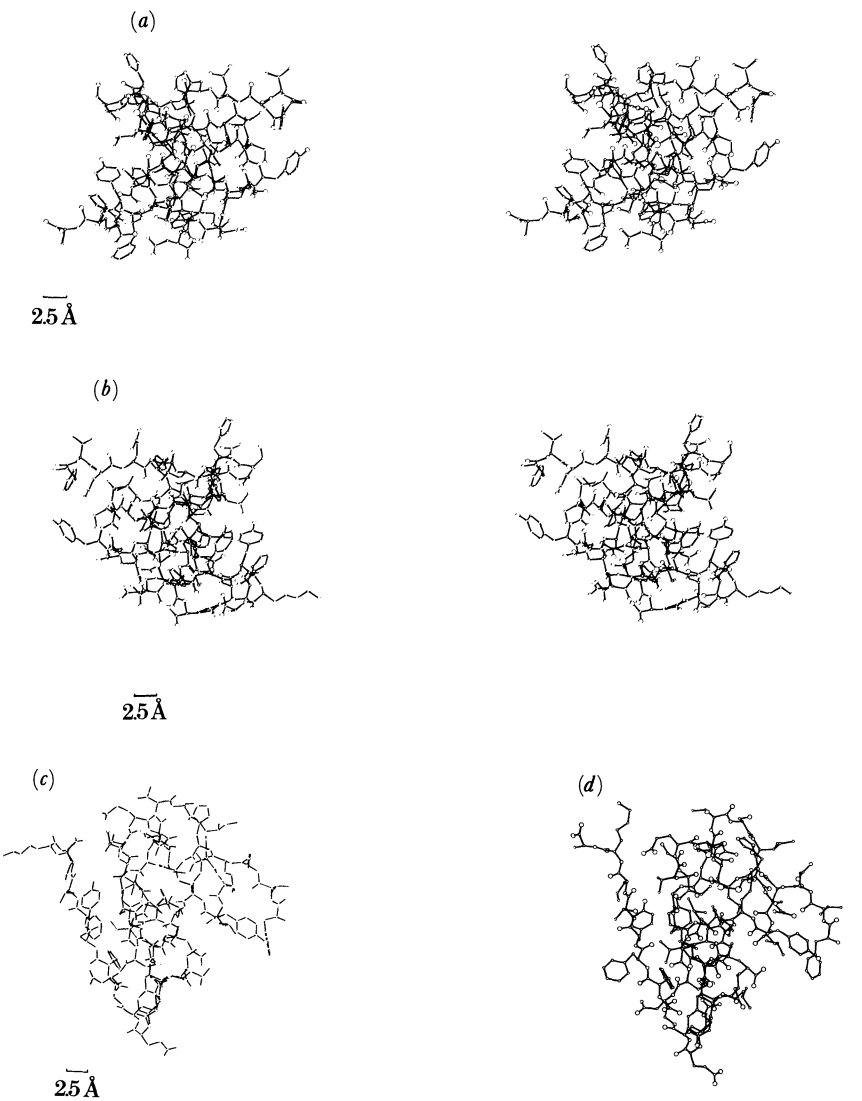


FIGURE 2.5. The insulin monomers viewed along the threefold axis (stereo views); (*a*) is molecule 1 (*b*) is molecule 2. The insulin molecules viewed perpendicular to the threefold axis. Molecule 1 (*c*) is rotated around the twofold axis d_0 giving an equivalent view to that for molecule 2 (*d*).

2.1. *Notation*

In this description the notation used for each residue is to give the chain letter (A or B) first, followed by its number in the sequence, followed by the number 1 or 2 representing molecule 1 or 2; e.g. Gly A1.1 represents A1 glycine of molecule 1.

Water molecules are described by a number (0–72) that indicates the level (in $\frac{1}{72}c$) on which the water molecule originally appeared, followed by W and a number 0–9; e.g. 1W2 sometimes shortened to 1,2.

This numbering is arbitrary but grew out of the system by which the waters were first identified. It has practical advantages over any rigorous and formal system.

Where symmetry relations are present in the molecular and atomic contacts these are given the conventions listed in figure 1.1*a*.

3. The A chain

3.1. *Conformations of the chains*

The projections of the A chains in figure 3.0 show that there is a marked difference between them that destroys twofold symmetry. This difference is largely because of a rotation of 32° about the bond CA–NH6 in molecule 2 relative to molecule 1, which carries with it the initial five residues. In both molecules these five residues follow somewhat distorted helical conformations (Dodson *et al.* 1980). Their relative rotation introduces different hydrogen-bonding patterns while leaving most of the residues in similar conformations, although necessarily making some different contacts. Following Höhne's suggested convention (Höhne & Kretschner 1982), we may describe the hydrogen bonding in the helix A1.1–A5.1, by the intervals, k, between bonded CO and NH groups as 4, 4, 4(5), 5W, W a mixture of α $(1\rightarrow5)$ and π $(1\rightarrow6)$ bonding, and that in molecule 2 as 4, 4, W, 4(3), 4 mainly α-helix, with some tendency to 3.10 helix. W indicates that a water molecule makes a closer contact than the NH, figures in brackets show that a second NH is also in contact.

Between the cystine residues, 6 and 11, both molecules form rings closed by disulphide links which are very similar to one another in conformation and are of the same size as the disulphide rings that occur in oxytocin and vasopressin. The recent solution of the crystal structure of desamido oxytocin (Wood *et al.* 1986) and pressinoic acid (Langs *et al.* 1986) show that the oxytocin and vasopressin rings in these crystals differ in conformation from the 6–11 insulin rings and from each other: as suggested from nuclear magnetic resonance (NMR) measurements (Walter 1969) there are direct hydrogen bonds across the oxytocin and vasopressin rings, absent in insulin, and the rings are less flat than the insulin rings. It is clear the various rings are flexible and able to react to different surroundings. In insulin A6, A7 and A8 are part of the initial helix; A9 is outside it, but near, whereas A10 and A11 have the extended β conformation. The individual residues have similar conformations, except threonine at A8. A7 is connected by a cystine link to SG7 of the B chain, which runs antiparallel with A7–A11 and is hydrogen bonded to it at intervals (figure 5.4). As a consequence both A7 NH and A10 NH make essentially helical contacts with A3 O and A5 O whereas both A7 CO groups contact B5 histidyl nitrogen atoms.

On Ramachandran plots, A12 still has the β structure parameters. However, the chain then turns round the bond A12 N–CA, and follows a helical course until the last two residues,

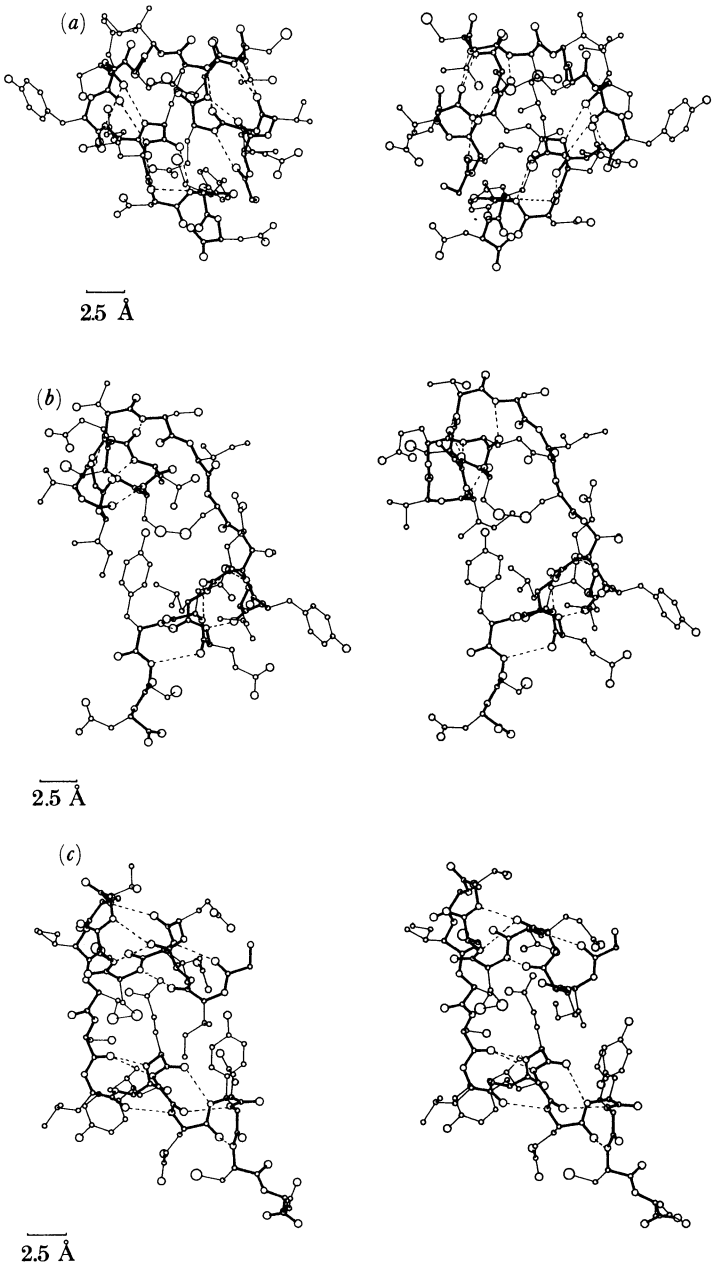


FIGURE 3.0. The isolated A chains, molecule 1 on the left, molecule 2 on the right. (*a*) View down the threefold axis. (*b*) View in the direction of the twofold axis, molecule 1 is given the same orientation as molecule 2. (*c*) View perpendicular to the twofold axis; molecule 1 is given the same orientation as molecule 2.

which are again extended. Both helices are very similar, well related by the twofold axis, but rather irregular. The carbonyl CO bonds are directed roughly parallel with the helix axis, making closer contacts towards the terminal residues with NH three residues away (3:10) than four residues (α), i.e. from 12 to 17, 3(4), 4, 3W, 3, 3, 3W. The chain here is on the outside of the hexamer and in contact with neighbouring hexamers in both molecules 1 and 2. There are more crystal contacts in molecule 2, however, leading to generally smaller values for its thermal parameters (see § 9.).

A tendency to helix formation in the A chain was expected from the sequence, A2 isoleucine and A3 valine, 3–4 residues from A6 and A7 cystine, and A13 leucine 3 residues from A16 leucine.

3.2. *The individual residues*

In the accounts of the individual residues of the A and B chains that follow, the headings show the natural variation of each residue including those present in mutant human insulin (italics), followed in brackets by some of the changes that have been carried out by synthesis. I denotes an invariant residue. Observations on the biological activity of the modified insulins are indicated briefly, as fractions, and apparent changes in immunological or structural integrity by letters, i = increased, n = normal, r = reduced, s = small.

The individual figures, (3.1–3.21 and 4.1–4.30) of the residues are viewed down the threefold axis. The twofold relation, or lack of it, of the corresponding residues of molecules 1 and 2 is illustrated by placing them in equivalent orientations to the nearest 'two'fold axis, placed vertically on the page. Where the two residues meet round the twofold axis this is shown. In the figures the described residue is drawn in thick lines, the attached peptides O and N from the adjacent residues are drawn in thin lines. Where additional protein structure or surrounding atoms make significant contacts they are drawn with thin lines. Hydrogen bonds are shown as broken lines; close van der Waals contacts by dotted lines. The chemical nature of the atoms C, N, O, S and Zn is represented by the radii of their circles (Zn > S > O > N > C). The water molecules, CA and occasional residues are labelled; ⓓ indicates disorder.

The tables list all the H-bonding contacts and angles made by the peptide O and NH and, to protein only, the H-bonding contacts and angles of the sidechain O and NH. The Ramachandran angles and the average B value of the main chain (m.c.) and side chain (s.c.) are given. The root-mean-square deviation from the local twofold axis (d) is given from the mainchain atoms (Δ_1 m.c.) and the whole residue (Δ_1 all). The root-mean-square deviation in the conformation of the two independent residues for the main-chain atoms and for the whole residue is given by Δ_2 m.c. and Δ_2 all, respectively.

The data for molecule 1 always precede those of molecule 2 in the tables. A short description of each residue and its environment follows the table.

4. The B chain

4.1. *Conformations of the chains*

The structure of the B chain is illustrated in figure 4.0a, b, c. It consists of three segments, an N terminal extended chain (B1–B8), a central helix (B9–B19) and a C terminal extended chain (B21–B30). In the figures the central position of the helix and the extended character of the N and C terminal segments are seen clearly. The two chains differ significantly only between B27 and B30, otherwise there is often near identity in their structures.

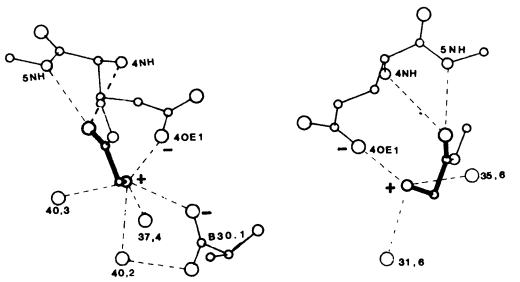
The N terminal segment is involved in hexamer formation, the B1 Phe side chain is actually buried in a non-polar crevice in the adjacent dimer. The peptide groups are solvated from B1–B3; there are H bonds between B4 and A11 and between B6 NH and A6 O.

At B8 Gly the chain twists into helix through a sharp 1→4 turn permitted by this glycine; thus the helix begins with 3:10 contacts made by B7, B8 and unfavourably, B9 peptide O. At B9 the helix H bonds 1→5 or α helical. This pattern of contact is mostly preserved for the

[*Text continues on page* 397.]

FIGURES 3.1–3.21. The individual amino acid residues of the A chains.

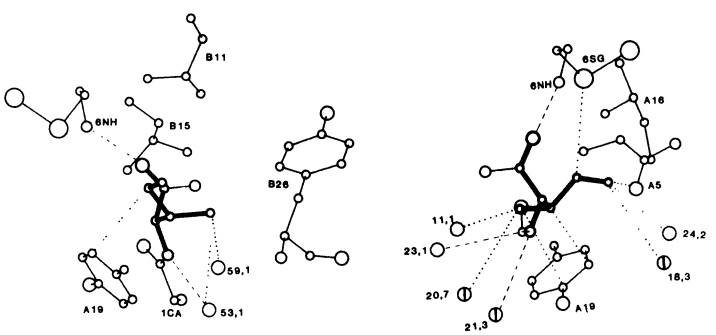
FIGURE 3.1. **A1 glycine**; I ((+)–Ala 0.9, (−)–Ala 0.05, CH_3CO 0.18, sarcosine 0.83 r, desamino 0.35)



NH_3^+	–	OE1 A4		2.4	2.5				Ĥ	123	112	ψ −150.5 −145.2	ϕ — —
		O B30.1		2.9					Ô	128		*B* m.c. 28.1 28.5	*B* s.c.
		40W3	35W6	2.8	3.3				N̂	101	90	Δ_1 m.c. 1.45	Δ_1 all 1.45
		40W2	31W6	3.0	2.9					104	95	Δ_2 m.c. 0.06	Δ_2 all 0.06
		37W4		3.2						151			
O	–	NH A5		3.1	2.9	H...O	2.1	1.9	Ô	154	157		
		NH A4		3.2	3.3		2.6	2.7	110	115			

A1 glycine has the same conformation in both molecules and, though differently situated to the surrounding atoms, is on the surface of each molecule at the beginning of a helix. Both NH_3^+ groups make an internal salt bridge with A4 carboxyl oxygen OE^- and are hydrogen bonded to water molecules. In addition, NH_3^+ A1.1 is beautifully situated to be in ionic contact with the terminal B30.1 carboxyl O^-. A1 O in both molecules is hydrogen bonded to NH A5 and also makes rather long contacts to NH A4.

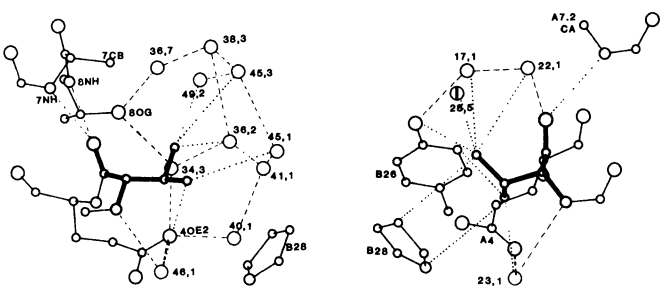
FIGURE 3.2. **A2 isoleucine**; valine (Pro 0.02 n, (+)-*Allo* Ile 0.0002, Leu 0.04)



NH	–	53W1	21W3	3.2	3.4	H...O	2.3	2.6	Ĥ	175	154	ψ −37.9 −45.5	ϕ −53.6 −55.2
												B m.c. 17.5 19.8	*B* s.c. 17.3 31.6
O	–	NH A6		2.9	2.9		1.9	1.9	Ô	150	155	Δ_1 m.c. 1.15	Δ_1 all 2.48
												Δ_2 m.c. 0.06	Δ_2 all 0.72

In both molecules, the residues belong to the inner core of the molecule but also make rather long contacts, 3.5–4 Å, with water molecules. They have different conformations and rather different contacts, mainly with non-polar residues. In 1, CB–CG1 runs nearly parallel with the tyrosine ring, A19, and touches this at 3.7 Å. CG1–CD is parallel to CA–CB and surrounded by B11 and B15 leucines; CG2 touches B26 tyrosine. In 2, it is the bond CB–CG2 that lies parallel with and touches A19 tyrosine; CG1–CD1 extends parallel with the 6–11 disulphide bond; CG1–A6 S is 3.8 Å. The residue contacts A16 leucine and A5 glutamine. The peptide CO is in the α helix; the NH groups are also attached to similarly placed water molecules, though the contact is long for a hydrogen bond in molecule 2.

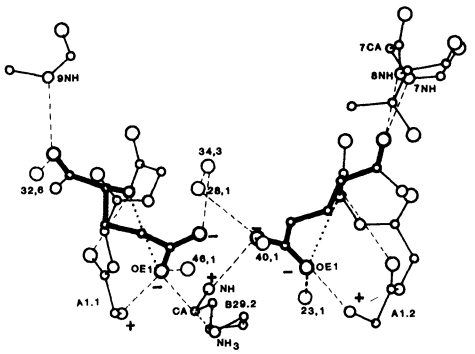
FIGURE 3.3. **A3 valine**; leucine, histidine, *leucine* (r)



NH	–	46W1	23W1	3.0	3.5	H...O	2.1	2.6	Ĥ	150	147	ψ −39.5 −50.3	ϕ −64.7 −55.8
												B m.c. 17.9 19.7	*B* s.c. 24.7 26.7
O	–	HN A7		2.9	3.5		2.0	2.7	Ô	127	147	Δ_1 m.c. 1.69	Δ_1 all 2.14
		HN A8		3.0			2.1			155		Δ_2 m.c. 0.04	Δ_2 all 0.10
			22W1		2.6						133		

The residues have similar conformations, but different contacts owing to the helix twist. Both are on the surface, loosely in touch with non-polar groups at 4.0 Å upwards, rather closer to networks of water molecules at 3.6–3.8 Å. Around molecule 1, the water molecules are linked in irregular five-membered rings (compare crambin (Teeter 1984)). Around molecule 2, they form chains and CG1 also contacts tyrosine B26 OH at 3.7 Å. The peptide NH touches water only; the peptide CO 2 is drawn closer to a water molecule than to NH within the helix; CO is in a bifurcated position making possible hydrogen bonds with NH A7 and NH A8.

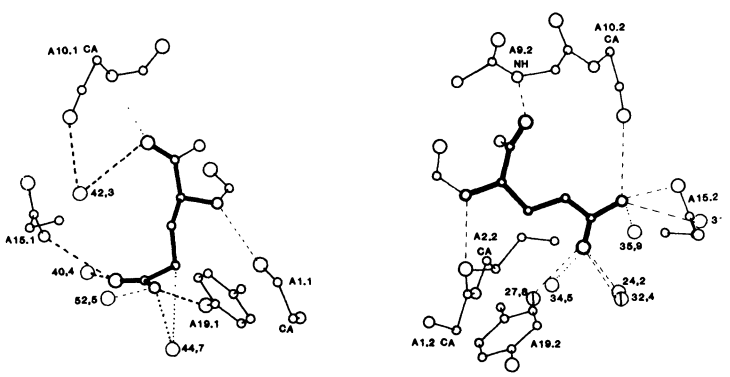
FIGURE 3.4. **A4 glutamic acid**; aspartic acid (Gln 0.75)



NH	–	O A1	3.2	3.3	H...O	2.6	2.7	Ĥ	119	115	ψ −39.0 −49.5 ϕ −70.0 −68.4 *B* m.c. 24.1 18.0 *B* s.c. 27.5 21.8
O	–	HN A7		3.2			2.5	Ô		101	Δ_1 m.c. 2.60 Δ_1 all 2.57
		HN A8		2.7			1.7			171	Δ_2 m.c. 0.06 Δ_2 all 0.15
		HN A9	3.0			2.2			135		
		OG A8		3.2						122	
		32W6	3.0						120		
OE2	–	HN B29.2		2.8			1.9			130	
OE1	–	NH_3^+A1	2.4	2.5		2.1			137	142	
		NH_3^+E B29.2	3.1						110		

In both molecules, glutamic acid A4 is in the gauche configuration with one carboxyl oxygen, OE1, in contact with NH and salt bridged to NH_3^+ A1. The same interaction can be made by aspartic acid, the only natural sequence variation of A4 so far observed. In the isolated dimer, the two A4 residues are about 30 Å apart but in the crystal, the groups in succeeding dimers along the *c* axis are within 5 Å of one another; OE2, 1 and 2, are linked by a chain of two water molecules, 28W1 and 34W3, but otherwise make very different contacts; in molecule 1, OE1 interacts with NH_3^+ of lysine 29.2 of the succeeding dimer; in molecule 2, OE2 interacts with lysine NH B29 within the same molecule. There are other contacts with water molecules. The peptide NH makes rather long helically directed contacts with A1 O; the peptide oxygen makes different helical H bonds in two molecules.

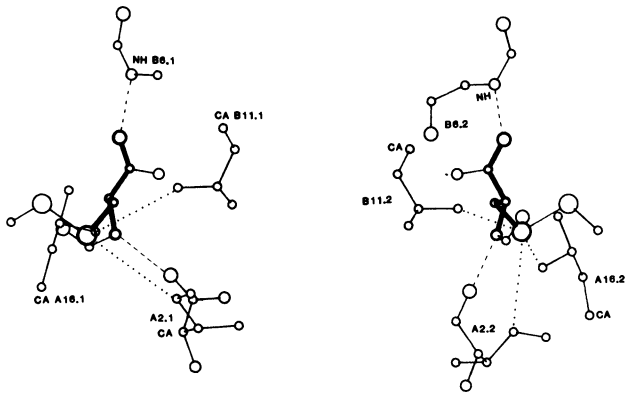
FIGURE 3.5. **A5 glutamine**; histidine (Ala 0.38 n, Thr 0.29, Leu 0.30)



NH	–	O A1	3.1	2.9	H...O	2.1	1.9	Ĥ	160	168	ψ −43.4 −44.3 ϕ −70.5 −54.8 *B* m.c. 21.4 16.2 *B* s.c. 44.4 33.5
O	–	42W3	3.2					Ô	113		Δ_1 m.c. 2.34 Δ_1 all 3.09
		HN A9		3.0			2.2			127	Δ_2 m.c. 0.02 Δ_2 all 4.13
$NE1H_2$	–	OH A19	2.7			1.9		Ĥ	132		
		CO A10		3.2			2.3			150	
		OE A15		3.5			2.8			132	
OE1	–	H_2NE1 A15	3.4			3.0			124		

The peptide NH groups are here both hydrogen bonded in the α helix and so is CO 5.2. The CO of 5.1, however, makes a good hydrogen bond with water and is directed to a long contact with A10 NH. Both the residues are extended on the surface, making loose connections with water and other surface residues. Their high *B* values reflect disorder in the atomic positions. In our interpretation, NEH_2 5.1 is placed in hydrogen-bonded contact with tyrosine 19 OH 1, whereas OE1 contacts NEH_2 glutamine 15.1. In molecule 2, NEH_2 contacts the carbonyl group CO 10.2 at 3.0 Å, and more loosely, at 3.5 Å, OE glutamine 15.2. OE1 5.2 touches water molecules and CD2 isoleucine; it has moved over 3.0 Å from the contact position with A19 OH observed in molecule 1.

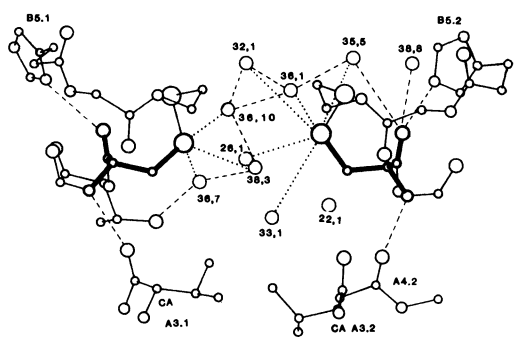
FIGURE 3.6. **A6 half cystine**; I ((+)–Cys 0.002 n, (+)–Cys 6:11 0.0003 vs, 6:11 Ala 0.08)



NE	–	O A2	2.9	2.9	H...O	1.9	1.9	Ĥ	157	165	ψ −20.5 −31.4 *B* m.c. 19.5 12.5	ϕ −107.8 −64.5 *B* s.c. 18.4 12.8
O	–	HN B6	2.8	2.9		1.9	2.0		142	121	Δ_1 m.c. 1.30 Δ_2 m.c. 0.07	Δ_1 all 1.36 Δ_2 all 0.10

Both residues are well defined and essentially buried. The sulphur atoms are in contact (although not always in the same way) with atoms of A16 leucine, A2 isoleucine and B11 leucine at distances between 3.5 and 4.5 Å. The peptide NH groups are in the α helix; the CO groups project from it to form hydrogen bonds with B6 NH groups.

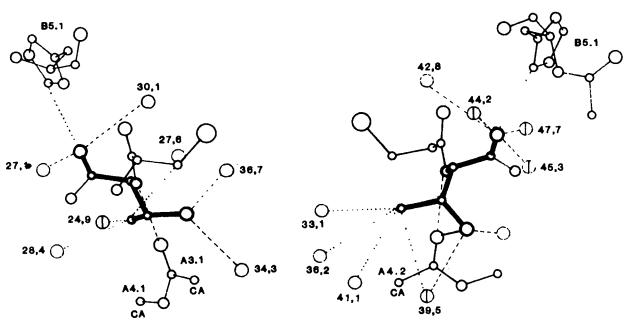
FIGURE 3.7. **A7 half cystine**; I ((+)–Cys 0.002 r, A7(& B7)-acetamidomethyl 0.003, A7(& A20) Homocys. 0 r, A7(& B7)-CH_2COOH 0.4)



NH	–	O A3 O A4	2.9	3.2	H...O	2.0	2.5	Ĥ	145	124	ψ −51.2 −42.6 *B* m.c. 19.6 15.2	ϕ −101.6 −87.3 *B* s.c. 18.5 17.0
O	–	HND B5.1	2.9			2.0			130		Δ_1 m.c. 1.34	Δ_1 all 1.33
		HNE B5.2		2.4			1.5			153	Δ_2 m.c. 0.05	Δ_2 all 0.09
		35W5		3.5				Ô	137			
		38W8		3.4					162			

Here the peptide NH groups are still in helices but the CO groups make contacts with the nitrogen atoms of B5 histidine which differ with the changed histidine conformations. The sulphur atoms are confined between A3 valine and the B chain. They are surrounded on other sides by water molecules at 3.6–3.9 Å, on the surface of the molecule and are indeed linked by short water-molecule chains between hexamers in the *c* direction.

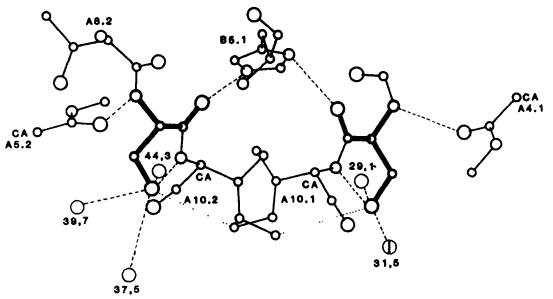
FIGURE 3.8. **A8 threonine**; alanine, histidine, glutamic acid (His 2.5 i, Lys 0.64, Phe 0.80 r)



NH	–	O A3	O A4	3.0	2.7	H...O	2.1	1.7	Ĥ	163	168	ψ −49.3 −32.4 *B* m.c. 24.3 21.1	ϕ −69.3 −78.0 *B* s.c. 36.6 24.7
O	–		42W8		3.4				Ô		116	Δ_1 m.c. 1.40	Δ_1 all 1.81
		30W1	44W2	3.4	2.4					110	143	Δ_2 m.c. 0.10	Δ_2 all 2.95
		27W1	45W2	3.2	3.1					119	116		
OG1	–	O A4			3.2						92		

Here the peptide NH still makes irregular α helical hydrogen bonded contacts and the CO groups interact with water molecules, whereas each touching B5.1 histidine CH groups at possibly hydrogen-bonded distances; 0.8.1–B5.1 CE is 3.2 Å, within molecule 1; 08.2–CD B5.1 is between hexamers along the *c* axis. The conformations of the two threonine residues are different and result in different contacts. Both hydroxyl groups hydrogen bond two water molecules; OG.2 in molecule 2 turns in to make contact with CO A4.2 and NH A9.2. Both methyl groups make several loose water molecule contacts.

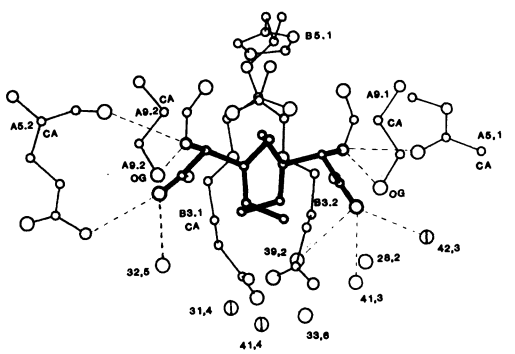
FIGURE 3.9. **A9 serine**; glycine, asparagine, arginine, histidine, lysine



NH	–	O A4	O A5	3.0	3.0	H...O	2.2	2.2	Ĥ	146	142	ψ −153.1 −123.2 ϕ −95.3 −113.0
												B m.c. 24.5 16.8 *B* s.c. 27.0 25.9
O	–	HND B5.1		2.8			2.5		Ô	153		Δ_1 m.c. 1.75 Δ_1 all 1.72
			HNE B5.1		2.6			1.6			132	Δ_2 m.c. 0.15 Δ_2 all 0.16

Both serine residues lie on the surface of the molecule near the contact between hexamers around the twofold axis, *d'*, at $z = \frac{1}{2}$. The peptide NH groups make different irregularly helical hydrogen bonds. Both CO groups make contact with histidine B5.1, which bridges the gap between hexamers. For O A9.1, the contact is intramolecular; for O A9.2, intermolecular. In spite of these differences, the residues are very similar. Both turn to bring the hydroxyl oxygen, OG, in contact with NH10, both touch water molecules and also the A10 CD atoms of neighbouring molecules around the *d* axis at 4.2 and 3.7 Å respectively.

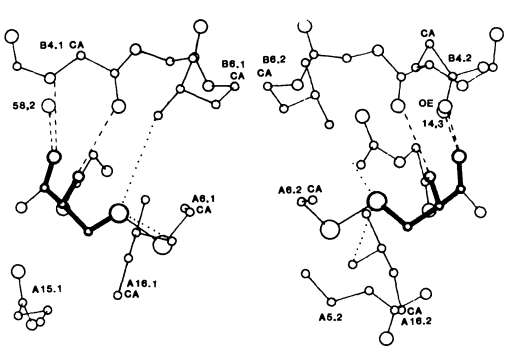
FIGURE 3.10. **A10 isoleucine**; valine, threonine, proline, arginine



NH	–	OG A9		2.9	3.0	H...O	2.0	2.3	Ĥ	140	127	ψ 148.3 159.1 ϕ −130.0 −147.8
		O A5		3.3	3.3							*B* m.c. 19.0 17.4 *B* s.c. 20.8 26.7
												Δ_1 m.c. 0.96 Δ_1 all 0.85
O	–	HNE A5.2			3.2			2.3			157	Δ_2 m.c. 0.06 Δ_2 all 0.15
		41W3	32W5	3.3	3.3				Ô	149	133	
		39W2		3.5						108		
		42W3		2.9						143		

The peptide NH makes contact with both O A5 and the serine hydroxyl OG; in molecule 2, the geometry is poor for hydrogen bonding. The peptide CO is in hydrogen-bonded contact with water and, in molecule 2, with glutamine NE A5. The residues are very similar, projecting into the solvent where the A chain turns around N–CA to adopt the β conformation. They are in loose contact (4.3 Å) with one another between the molecules related by the *d'* axis at $z = \frac{1}{2}$. They also touch histidine B5 and a semicircle of surrounding water molecules at 3.6–4.5 Å. Experimentally, there was difficulty in establishing their conformation. The latest difference map suggests a small occurrence of residues with an alternate conformation at A10.1 but not at A10.2.

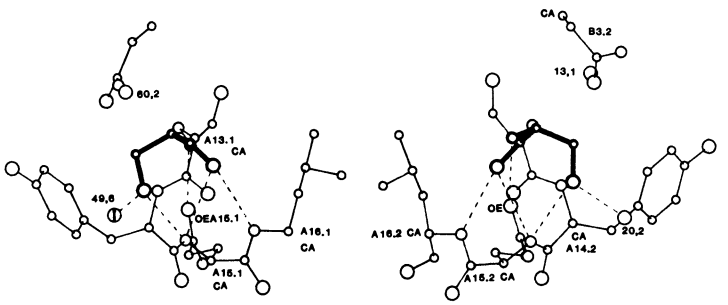
FIGURE 3.11. **A11 cystine**; I (11 (−)–6 (+)–Cys 0.007 s, (+)–Cys 6:11 0.0003 s, Ala 6:11 0.08)



NH	–	O B4		3.2	3.0	H...O	2.2	2.1	Ĥ	166	154	ψ 163.3 159.8 ϕ −143.9 −149.5
												B m.c. 20.8 14.2 *B* s.c. 23.5 13.3
O	–	HN B4		3.1	2.7		2.1	1.7		147	158	Δ_1 m.c. 0.62 Δ_1 all 0.86
		OE B4.2			3.1				Ô		139	Δ_2 m.c. 0.04 Δ_2 all 0.10
		58W2	14W3	3.2	3.3					136	133	

Here both peptide NH and CO groups are hydrogen bonded in a slightly twisted β sheet, to CO and NH, N4. The CO groups also each contact a water molecule, or in the case of A 11.2 CO, alternatively, the disordered B4 OE. Both cystine residues are surrounded, but rather loosely, by the B chain itself from B3–B4, by A16 and B6 leucines and a little differently, by parts of A15 and A5.

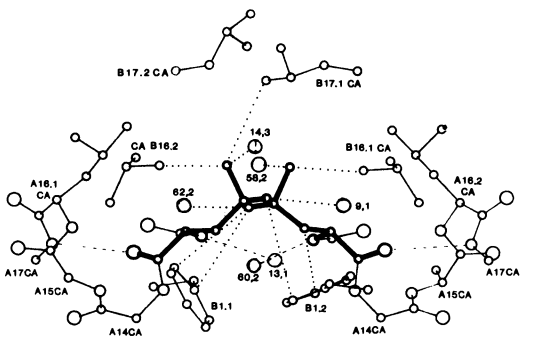
FIGURE 3.12. **A12 serine**; asparagine, aspartic acid, threonine



NH	–	OE A15	3.0	3.0	H...O	2.1	2.0	Ĥ	154	156	ψ 167.1 159.2	ϕ −93.1 −110.7
											B m.c. 21.2 9.6	*B* s.c. 23.8 16.9
O	–	HN A15	3.1	3.1		2.4	2.5	Ô	98	98	Δ_1 m.c. 0.46	Δ_1 all 0.47
		HN A16	3.2	3.2		2.3	2.2		140	143	Δ_2 m.c. 0.04	Δ_2 all 0.11
OG	–	NH A15	3.0	3.2					123	132		

The residues are on the surface, surrounded by leucine A16, glutamine A15, the main chain between 13 and 14, and solvent water molecules. The serine hydroxyl group is in contact with 15NH and also is hydrogen bonded to a water molecule. It also turns back to contact NH12 which itself is hydrogen bonded to the terminal OE of glutamine A15. The peptide CO group makes contact with NH15 and 16 both within the limits of hydrogen bond geometry.

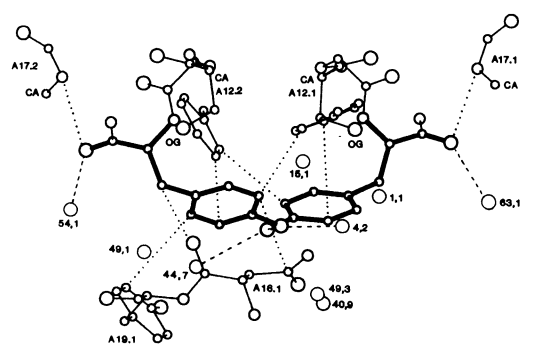
FIGURE 3.13. **A13 leucine**; isoleucine, arginine, lysine



NH	–	60W2 13W1	2.4	2.9	H...O	1.5	1.9	Ĥ	162	168	ψ −27.1 −30.9	ϕ −67.6 −62.2
											B m.c. 18.7 11.8	*B* s.c. 34.6 16.5
O	–	O HN A17	3.3	3.1		2.5	2.3	Ô	140	137	Δ_1 m.c. 0.23	Δ_1 all 0.33
											Δ_2 m.c. 0.03	Δ_2 all 0.14

The residues are on the surface of the molecules; they approach one another in the hexamer forming interface around *d* O at about 5 Å. They are otherwise surrounded by A16 leucine, B1 phenylalanine, B18 valine, the main chain A11–14 and water molecules in loose contact rather closer for molecule 2 than 1. The peptide NH is also on the surface and hydrogen bonded to water; the CO is directed helically towards NH17.

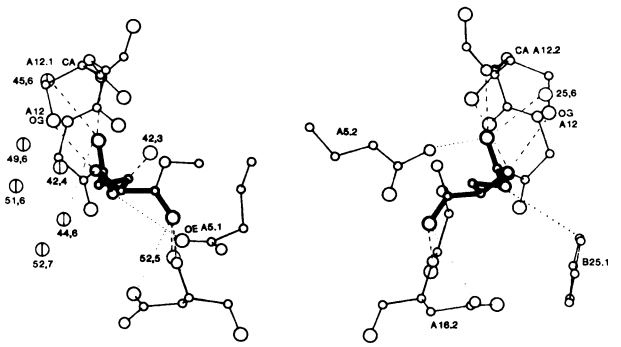
FIGURE 3.14. **A14 tyrosine**; phenylalanine, histidine, asparagine (Iodotyr, 1.0, n)



NH	–	OG A12	3.4	3.3	H...O	2.9	2.7	Ĥ	115	117	ψ −34 −32.3	ϕ −68.2 −67.4
		O A12	3.3	3.3		3.3	3.3		80	79	*B* m.c. 19.0 12.4	*B* s.c. 35.7 17.3
O	–	NH A17	3.3	3.3		2.6	2.7	Ô	89	90	Δ_1 m.c. 0.33	Δ_1 all 0.50
		63W1 54W1	2.7	2.8					114	132	Δ_2 m.c. 0.06	Δ_2 all 0.10
OH	–	OH A14	3.2	3.2					118	118		

The two residues are on the surface and are very similar. The tyrosine in molecule 2 is better defined than that in molecule 1, probably because it makes close contacts with tyrosine A19.1 and A18.1 asparagine in the neighbouring hexamer, whereas tyrosine A14.1 is only in external contact with water molecules. The two residues meet in the hexamer-forming interface around the axis, *d* 0, loosely bound, OH...OH, at 3.2 Å; the hydroxyl groups each bond with a second water molecule; both aromatic rings are in contact with both B1 phenylalanines. The peptide NH is shielded from hydrogen bonding by close contacts with serine 12 OG hydroxyl and O A12; the A14 CO is directed towards A17 NH, again with poor geometry for hydrogen bonding; it bonds to water.

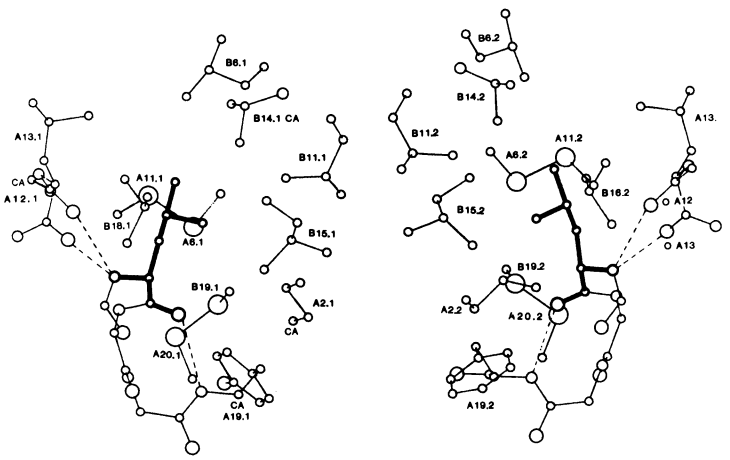
FIGURE 3.15. **A15 glutamine**; asparagine, aspartic acid, glutamic acid



NH	–	OG A12	3.0	3.3	H...O	2.1	2.3	Ĥ	150	153	ψ −39.8 −43.8	ϕ −65.0 −65.8
		O A12	3.1	3.1		2.4	2.5		123	124	*B* m.c. 16.1 11.0	*B* s.c. 34.5 22.0
											Δ_1 m.c. 0.31	Δ_1 all 0.42
O	–	NH A18	3.0	3.0		2.1	2.1		116	112	Δ_2 m.c. 0.05	Δ_2 all 0.28
		52W5	3.4					Ô	109			
OE1		HN A12	3.0	3.0		2.1	2.0		123	135		

The main chains are well defined and on the outside of the molecules. Both the peptide NH and CO are involved in hydrogen-bonded contacts approximating to 1→4 helices; NH also contacts the serine 12 OG. The side chains are extended and lie between A5 and the A chain 11–14. A15.2 is better defined than A15.1 packed against B25 phenylalanine in a neighbouring hexamer. The terminal OE atoms are bonded in each case to serine A12 NH, but the amide groups are differently tilted. In molecule 1, OE1 also contacts a water molecule whereas NE touches the A5 glutamine and a water molecule; in molecule 2, the OE1 touches both A12.2 serine NH and glutamine A5.2, and the NE only contacts water. The arrangement in molecule 1 is looser, perhaps disordered.

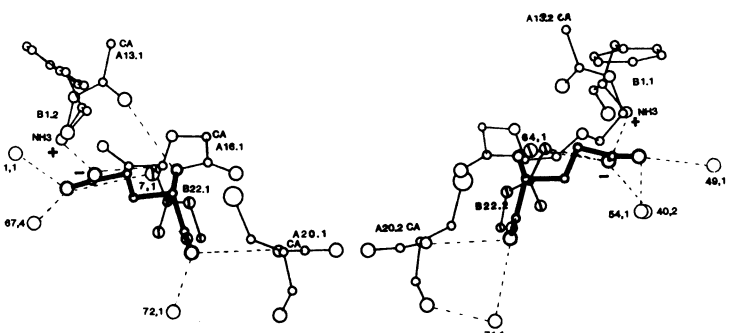
FIGURE 3.16. **A16 leucine**; I



NH	–	O A12	3.2	3.2	H...O	2.3	2.2	Ĥ	145	154	ψ −24.7 −32.4	ϕ −65.5 −63.6
		O A13	3.0	3.3		2.5	2.8		117	112	*B* m.c. 18.6 13.9	*B* s.c. 19.0 14.7
											Δ_1 m.c. 0.26	Δ_1 all 1.22
O	–	NH A19	3.1	2.9		2.1	2.0	Ô	116	124	Δ_2 m.c. 0.07	Δ_2 all 0.17

The A16 leucine residues are near the molecular centres and are completely buried. They touch the A6–A11 and, through CO, the B19–A20 disulphide bond. They also make contacts from 3.5 to 4.8 Å with atoms belonging to B6 leucine, B11 leucine, B15 leucine, B14 alanine, B18 valine, A2 isoleucine and A13 leucine. The peptide NH and CO groups make somewhat irregular helical contacts.

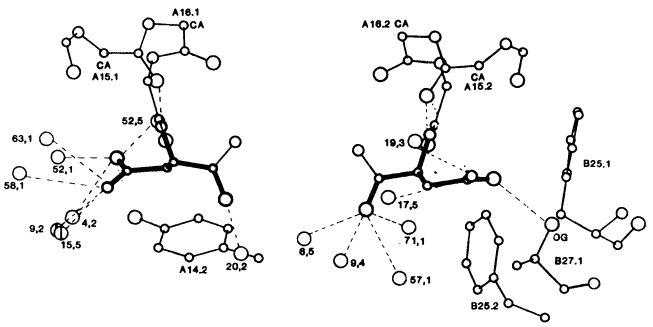
FIGURE 3.17. **A17 glutamic acid**; glutamine, leucine, methionine



NH	–	O A13		3.3	3.1	H...O	2.5	2.3	Ĥ	141	143	ψ −18.4 −17.4	ϕ −68.1 −56.7
												B m.c. 13.1 18.2	*B* s.c. 24.6 27.3
O	–	HN A20		3.3	3.3		2.4	2.3	Ô	106	108	Δ_1 m.c. 0.19	Δ_1 all 0.41
		72W1	71W1	2.7	3.3					148	144	Δ_2 m.c. 0.04	Δ_2 all 0.22
OE2	–	NH_3^+ B1		2.8	3.0	H...O				143	142		
		NH_2^+ B22		2.9	3.5		2.0	2.6		126	138		

The residues run in an extended conformation along the outside of the molecule, bounded on one side by the A chain, on the other by the B chain of its own neighbouring molecules. OE2 makes intermolecular salt bridges with phenylalanine B1 NH_3^+ at 2.8 Å, 3.0 Å and also intramolecular bridges with the disordered arginine B22 NH_2^+ or the water molecule replacing it, at 2.9 Å, 3.5 Å; it also touches one further water molecule. OE1 is in contact with two water molecules. The peptide NH and CO make helical contacts.

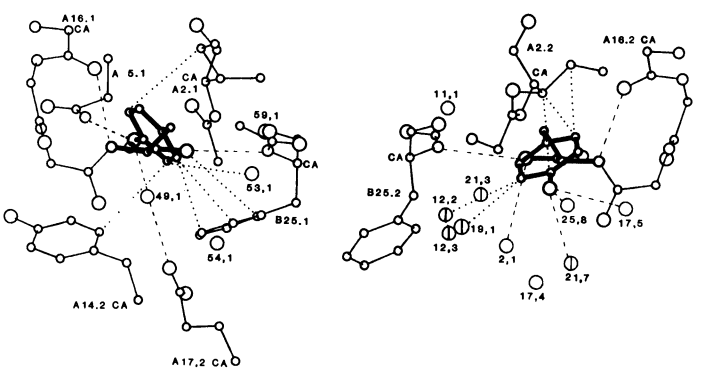
FIGURE 3.18. **A18 asparagine**; serine, threonine, glycine



NH	–	O A15		3.0	3.0	H...O	2.1	2.1	Ĥ	157	156	ψ −18.5 −13.4 *B* m.c. 14.8 18.0	ϕ −68.1 −72.7 *B* s.c. 23.1 30.2
O	–	20W2	9W4	2.6	2.8				Ô		134	Δ_1 m.c. 0.26	Δ_1 all 1.14
			6W5		2.6						137	Δ_2 m.c. 0.03	Δ_2 all 0.24
			71W1		3.5						95		
			57W1		3.2								
OD1	–		OG1 B	27.1	2.8						128		

The residues turn at CB to lie packed against the main chain on the outside of the molecules. They make contacts both with water molecules and with atoms in the neighbouring hexamer. OD A18.1 touches tyrosine A14.2 hydroxyl group at 3.6 Å. OD A18.2 is hydrogen bonded to B27.1 threonine hydroxyl, whereas CG and ND1 touch the B25.1 and 2 phenylalanines. The peptide NH still makes a helical contact, with OA15, which in molecule 2 also touches ND2. The CO group turns out to connect with water molecules, more in molecule 2 than molecule 1.

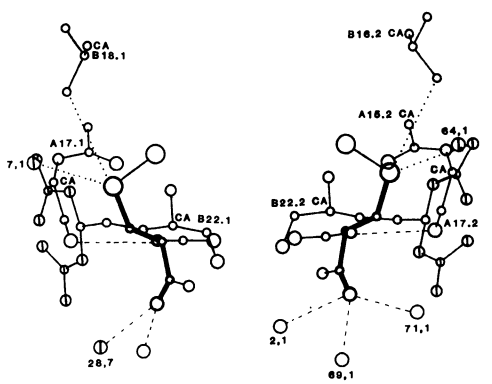
FIGURE 3.19. **A19 tyrosine**; I (Iodo-tyr 0.5, (+)–Tyr 0.02 s, Phe 0.1–0.03 r)



NH	–	O A16		3.1	2.9	H...O	2.1	2.0	Ĥ	165	165	ψ 9.8 −2.4 *B* m.c. 13.9 14.8	ϕ −101.7 −95.0 *B* s.c. 18.0 25.7
O	–	NH B25		3.2	3.4	H...O	2.2	2.4	Ô	169	160	Δ_1 m.c. 0.33	Δ_1 all 0.74
		2W1			3.4				Ô		99	Δ_2 m.c. 0.00	Δ_2 all 0.29

The two residues have essentially the same conformation. They contribute to the stability of the contacts both within the A chain, particularly between the initial and final residues, and between the A and B chains. In both, certain aromatic ring atoms are in contact with A2 isoleucine; in both, CO A19 is hydrogen bonded to A14.2 (8) in the neighbouring hexamer whereas tyrosine A19.2 is open on two sides to water molecules: as a consequence, the ring of A19.2 is free to swing, the B values are larger on one side than the other. The hydroxyl group OH 19.1 makes close contacts with A5 NE and one water molecule, while OH 19.2 touches four water molecules but these are less well defined. The peptide NH is in the 3.10 helix.

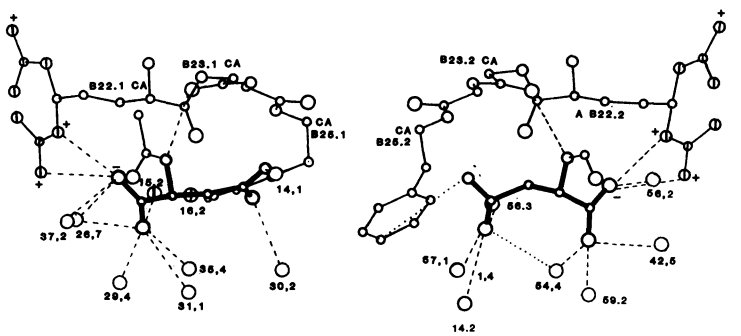
FIGURE 3.20. **A20 cystine**; I (7:20 homocystine 0)



NH	–	O A17		3.3	3.3	H...O	2.4	2.3	Ĥ	165	167	ψ 151.5 157.0 *B* m.c. 18.4 17.5	ϕ −79.0 −77.3 *B* s.c. 12.9 14.5
O	–	20W2	69W1	3.2	2.5				Ô	156	136	Δ_1 m.c. 0.25	Δ_1 all 0.24
		28W7	71W1	3.0	2.8					120	132	Δ_2 m.c. 0.06	Δ_2 all 0.09
			2W1		3.5						104		

The CH_2–S residues are confined by the peptide chain A16–18, the B22 side chain and valine B18. They each make a short contact to water, SG1–7W1, disordered B22 NH_2^+, 3.9 A, and SG2–64W1 and the disordered B22 NH_2^+, 3.4 A. They can therefore be reached by solvent. The peptide NH contacts O A17 in the 3.10 helix; the CO group only touches water molecules.

FIGURE 3.21. **A21 asparagine**; I (Iso asn 0.87 n, (+)–Asn 0.33 r, Arg 0.4 r, asnamide 0.7 r, Des asn 0.2 s)



NH	–	O B23		3.0	2.9	H...O	2.0	2.0	Ĥ	163	161	ψ −170.0 155.1 *B* m.c. 28.9 27.2	ϕ −91.5 −80.6 *B* s.c. 34.0 41.9
O	–	NH2 B22		3.1	3.2	H...O	2.5	2.4	Ô	117	101	Δ_1 m.c. 0.52	Δ_1 all 1.18
		NHE B22		3.0	2.8		2.2	1.9		154	166	Δ_2 m.c. 0.23	Δ_2 all 0.93
		28W7	56W2	3.2	2.9					84	136		
OE	–	15W2	54W4	2.8	3.3					99	116		
		29W4	59W2	2.6	2.3					109	159		
		35W4	56W3	3.1	3.0					145			
		35W10	57W1	2.7	2.8					146	158		
		31W1		3.2						114			
		37W2		3.4						103			

Both residues are external, tethered to the body of the molecule by the hydrogen bond NH–CO B23 and salt bridges from $-CO^-$ to the B chain arginine NH^+ and NE^+ atoms. The arginine sites appear sometimes replaced by a water molecule; the second O^- of the carboxyl group only contacts water, as do the atoms of the amide group. ND2 in molecule 2 is hemmed in by B25 Phe and makes only long contacts, whether with water or other atoms.

remainder of the helix. At each end of the helix (at B7 and B19) there is a cystine linking it to the A chain. At the B20 peptide the chain continues straight on and the helix converts into a $1 \rightarrow 4$ (β) bend in which B19 O is H bonded to B22 NH and B20 O is H bonded to B23 NH. The peptide oxygens at B17 and B18, not involved in the helix, are solvated or make H-bond contact to the disordered B22 arginyl sidechain. The peptide oxygen at B21 makes no H-bond contacts owing to the steric blocking by the nearby protein structure.

As seen in figure 4.0*c* the hydrophobic residues are generally on the inside of the helix facing towards the body of the molecule. Figure 4.0*b* shows that the helix is not quite straight, curving towards the inside of the molecule. From their low-temperature coordinates, Sakabe *et al.* (1981) have estimated the radius of curvature as approximately 86 Å for molecule 1 and 60 Å for molecule 2.

From B21 the 1–4 turn leads to the C terminal chain extending back in the reverse direction rather more than the length of the helix. Between B24 and B26 the chain is involved in an antiparallel H bond with its twofold-related equivalents in the dimer. The other main chain protein contacts made by this segment are from B23 O to A21 NH, from B25 NH to A19 O, B27 O to B30 N and in molecule 1 only, B30 O to A1 NH_3^+. All the other peptide groups are solvated.

The angle of 45° made by the B chain C terminal segments of extended chain against the helix is found in other proteins. (Chothia 1984). This favours the efficient packing of their non-polar side chains.

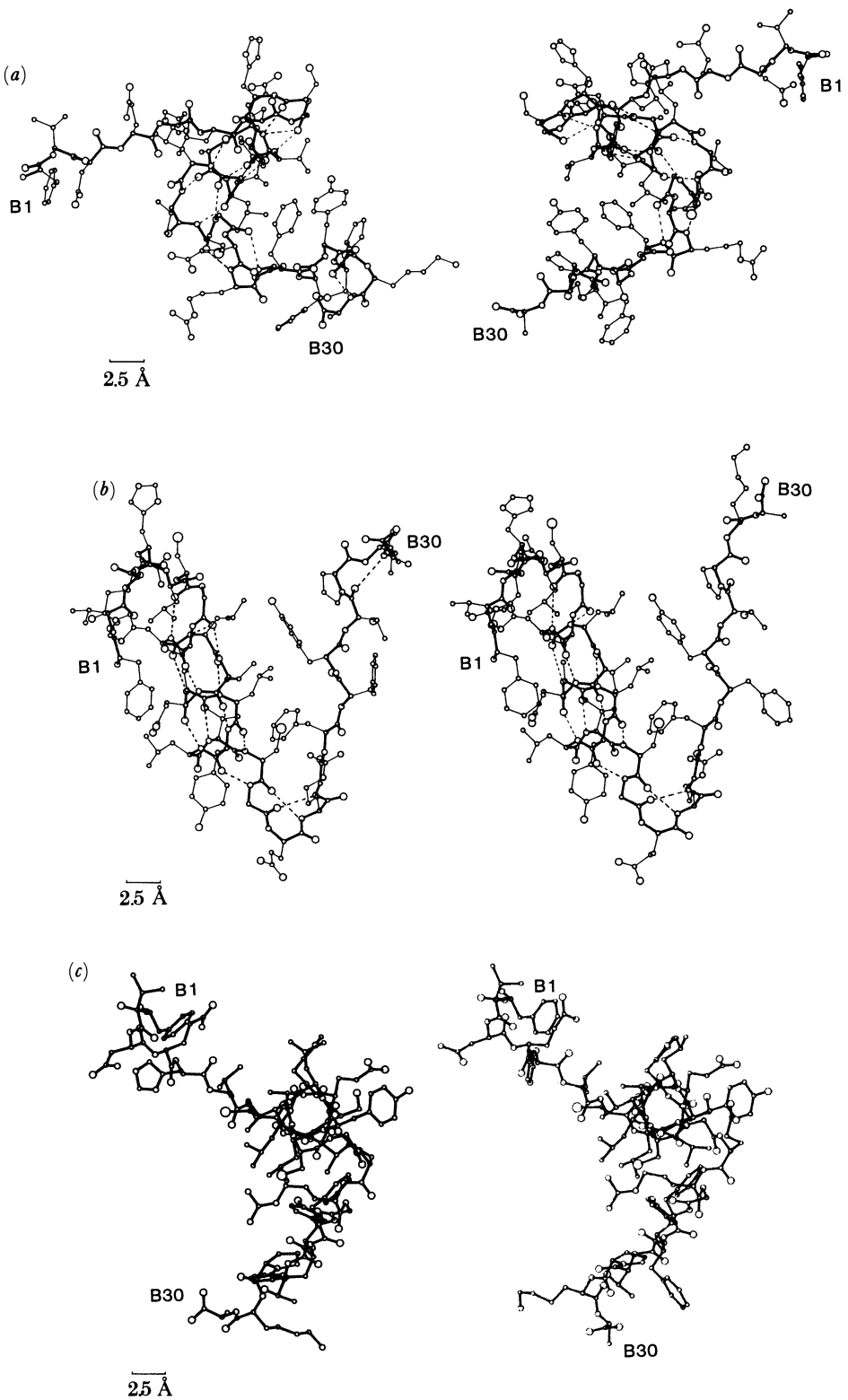
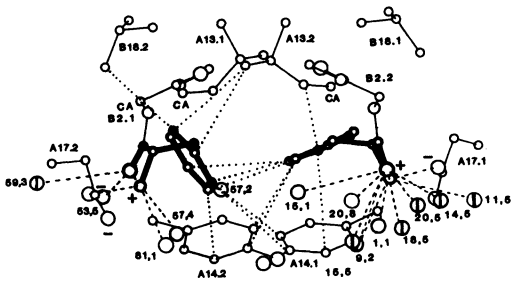


FIGURE 4.0. The isolated B chain, molecule 1 on the left, molecule 2 on the right. (*a*) View down the threefold axis. (*b*) Molecule 1 and 2 viewed with their dimer forming surfaces facing the reader. Molecule 1 is given the orientation of molecule 2 (cf. figure 2.5). (*c*) View down the helix B9–B19. Molecule 1 has the orientation of molecule 2.

FIGURES 4.1–4.30. The individual amino acid residues of the B chains.

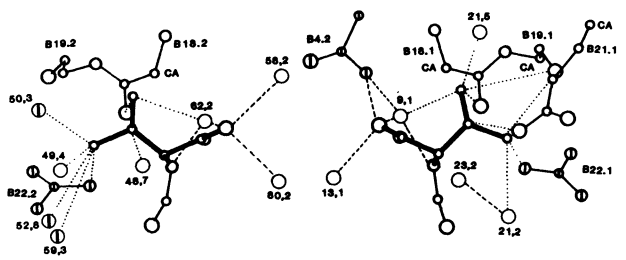
FIGURE 4.1. **B1 phenylalanine**; Tyr, Ala, Leu, Arg, Gly (des B1, 1.0, n; phenylthiocarbamoyl, 0.6 n)



NH	–	OE2	A17.2, 17.1	3.0	2.8	H...O			N̂	123	113	ψ 168.7 159.4	ϕ — —
		61W1	9W2	2.7	3.1					116	111	*B* m.c. 46.5 32.0	*B* s.c. 36.5 34.6
			1W1		3.5						131	Δ_1 m.c. 0.43	Δ_1 all 0.81
			11W5		3.4						118	Δ_2 m.c. 0.07	Δ_2 all 0.38
CO	–	53W5	14W5	2.8	2.3				Ô	153	135		
		57W4	15W1	3.1	3.5					120	95		
		59W3	15W5	3.5	3.1					120	139		
		61W5	9W2	3.4	3.2					102	117		
			16W5		2.8						171		
			20W5		3.4						146		
			20W8		3.4						127		

The B1 residue begins an extended stretch of chain on the outside of the dimer. To form the hexamer, the chains cross in the region of B4 and hook into one another, burying the B1 phenyl groups in the adjacent dimer surface. Here they are surrounded by the B1 Phe, A13 Leu and A1 Tyr from the adjacent dimer and within the monomer, by B18 Val and A14 Tyr. The peptide N and O are on the hexamer surface and H bond to solvent. There is a salt bridge between B1 NH_3^+ and the A17 Glu (COO^-) of the adjacent dimer helping to link the hexamer structure together.

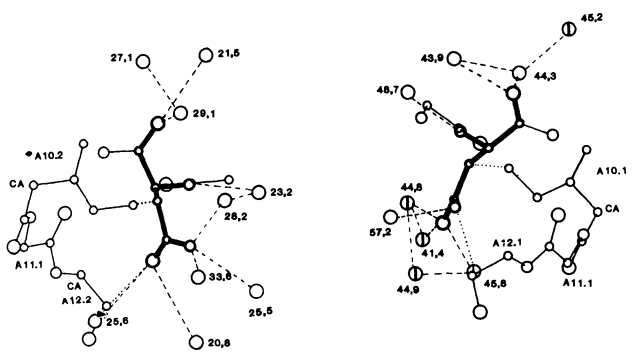
FIGURE 4.2. **B2 valine**; Ala, Pro, Thr (des B2, 0.88, n)



NH	–	62W2	9W1	2.7	2.7	H...O	1.7	1.8	Ĥ	172	163	ψ 156.6 158.4	ϕ −123.8 −119.3
											159	*B* m.c. 28.5 25.2	*B* s.c. 50.0 26.2
O	–	HNE2 B4.2			2.3			1.7				Δ_1 m.c. 0.41	Δ_1 all 0.41
		58W2		2.8					Ô	146		Δ_2 m.c. 0.03	Δ_2 all 0.12
		60W2	13W1	3.1	2.8					123	115		
		62W2	9W1	3.1	3.0					115	117		

The residue is on the hexamer surface and well related by the local twofold symmetry; one face of the side chain lies against the peptide chain B17–B19 and the side chain of B22 Arg from the other dimer. Water molecules at distances of 3–4 Å cover the other face. Both the peptides are H bonded to solvent; one molecule (62W2, 9W1) bridges the peptide O and N. In molecule 2, one of the disordered B4 Gln side chains makes a close H-bonding contact, displacing a water molecule.

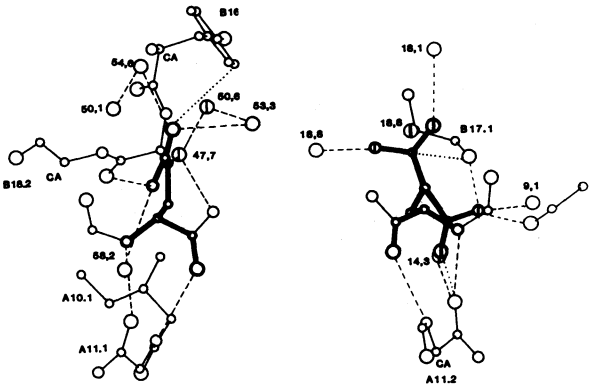
FIGURE 4.3. **B3 asparagine**; Lys, Gly, Ser, Pro, Ala, Thr (des B1–B3, 0.75)



NH	–	48W7	23W2	2.9	3.2	H...O	1.9	2.2	Ĥ	169	166	ψ 136.3 145.8	ϕ −84.9 −76.1
												B m.c. 29.0 16.8	*B* s.c. 51.6 35.9
C	–	43W9	21W5	3.4	3.2				Ô	135	166	Δ_1 m.c. 0.82	Δ_1 all 1.41
		44W3	29.1	2.5	3.5					127	106	Δ_2 m.c. 0.10	Δ_2 all 0.39
										135			

The residue is on the hexamer surface, approached on one side by A10 Ile sidechain. The side chains differ slightly in the position of their amide groups. These and the peptide groups are H bonded to the network of water surrounding the hexamer. There is a markedly larger thermal parameter in molecule 1 for this residue that appears to make less favourable H-bond contacts through its amide.

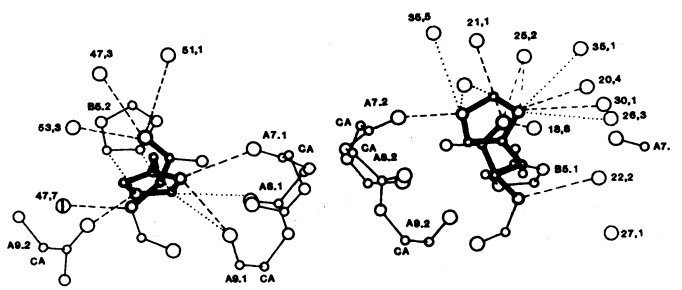
FIGURE 4.4. **B4 glutamine**; Arg, Lys, Pro, Gly (des B1–B4, 0.45, r)



NH	–	O	A11	3.1	2.7	H...O	2.1	1.7	Ĥ	171	167	ψ 158.5 153.2	ϕ −147.9 −129.6
												B m.c. 27.1 17.3	*B* s.c. 37.7 21.9
O	–	HN	A11	3.2	3.0		2.2	1.7	Ô	139	141	Δ_1 m.c. 0.84	Δ_1 all 1.43
												Δ_2 m.c. 0.06	Δ_2 all 0.98
NEH2	–	O	B17	3.4	3.1		2.5	2.4		148	167		
		O	B2.2		2.3			1.7			159		

This residue is on the hexamer surface between the dimers. Its side chain is directed downward from the extended peptide where it is in contact with water. The side-chain conformation is different for the two molecules; in molecule 2, the side chain appears to be disordered with one structure turning back towards the A11 peptide. There is a pair of β-type H bonds between B4 and A11 peptide that run antiparallel. The slightly longer peptide H bonds between B4 and A11 in molecule 1 are associated with it having a distinctly larger thermal parameter.

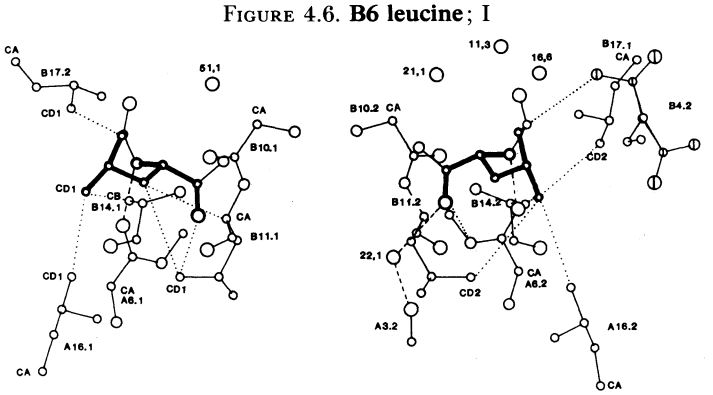
FIGURE 4.5. **B5 histidine**; Arg (Ala 1.0, n; r; des 1–5 0.21, s)



NH	–	47W7	22W2	2.5	3.0	H...O	1.6	2.1	Ĥ	153	150	ψ 116.9 108.2	ϕ −84.7 −114.4
												B m.c. 17.1 14.5	*B* s.c. 28.0 28.4
O	–	47W3	18W6	2.9	2.6				Ô	146	126	Δ_1 m.c. 0.97	Δ_1 all 2.42
		53W3	25W2	3.3	3.0					148	118	Δ_2 m.c. 0.03	Δ_2 all 4.29
		51W1	21W1	3.2	3.3					117	120		
			20W4		3.5						158		
HNE2	–	O A9.2		2.6			1.6			159			
		O A7.2			2.4			1.5			144		
HND1	–	O A7.1		2.9			2.0			158			
		O A9.1		2.8			2.5			100			

The residues lie on the hexamer surface where they H bond (in conformations different by about 90° in the two molecules) through the imidazole ring to the adjacent A7 O (via B5 ND1 in molecule 1 and B5 NE2 in molecule 2). There is one less-favourable contact in molecule 1 to A9 O (via ND1) and also through crystal packing to A9 O (via NE2) in the twofold related hexamer. The H bonding to carbonyl O at ND1 and NE2 in molecule 1 indicates this imidazole is protonated; that of molecule 2 need not be. Alternative H-bonding interactions of this kind are seen in other structures (Bhat & Vijayan 1981). In 2Zn insulin, the structural requirements of packing dictate that one histidine (that of molecule 2) is displaced by the hexamer assembly along the threefold axis (see §8). The two sidechains are roughly parallel and make some close contacts. Both peptide O and N are directed outwards and are solvated.

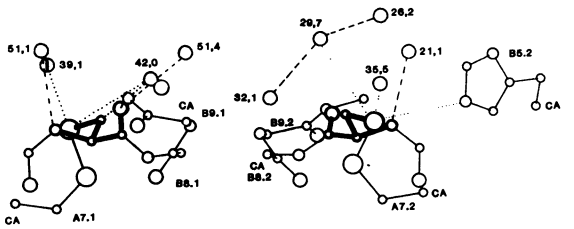
FIGURE 4.6. **B6 leucine**; I



NH	–	O A6		2.8	2.9	H...O	1.9	2.0	Ĥ	167	152	ψ 110.9 121.1	ϕ −123.0 −125.4
												B m.c. 15.0 11.8	*B* s.c. 17.9 15.1
O	–		22W1		2.9				Ô		135	Δ_1 m.c. 0.80	Δ_1 all 0.80
												Δ_2 m.c. 0.03	Δ_2 all 0.13

The residue packs against the central helical structure. The leucyl side chain extends down against B11 leu in the non-polar core. There are four leucines packed together across the B chain helix, only B6 is not completely buried. This concentration of leucines within the monomer is remarkable; moreover they appear nowhere else. The peptide N is directed towards the molecule (H bonding to A6 O) whereas the peptide CO is directed outwards. In molecule 2, it is H bonded to water 22W1, which is linked to A3 O. In molecule 1, the A chain N terminus (positioned differently in molecule 2) prevents approach of water and B6 O is not H bonded.

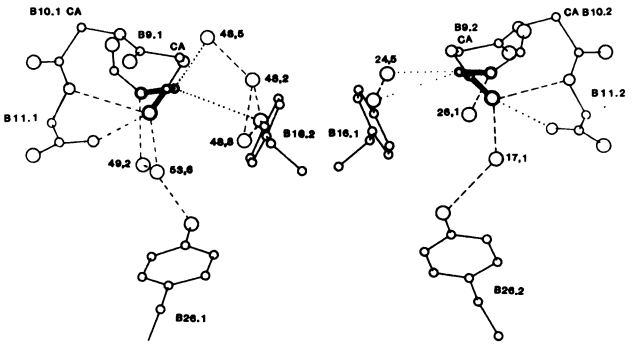
FIGURE 4.7. **B7 cystine**; I ((+)-Cys 0.01, r, B7 (and A7) acetamidomethyl 0.003, B7 (and A7) $-CH_2$, COOH 0.4)



NH	– 51W1	21W1	3.0	2.8	H...O	2.0	1.9	Ĥ	152	157	ψ 150.3 157.2	ϕ −122.1 −122.2
											B m.c. 17.2 13.3	*B* s.c. 25.7 16.1
O	– HN	B10	3.1	3.1		2.4	2.3	Ô	139	143	Δ_1 m.c. 0.67	Δ_1 all 0.79
	51W4		3.4						105		Δ_2 m.c. 0.06	Δ_2 all 0.08

In both molecules, this residue is on the surface, in contact with water to which the peptide NH is H bonded. The peptide makes the initial 1 → 4 H-bond contact to B10 N; in molecule 1, it is also H bonded to water whose approach is permitted here by the side chain turning away slightly more from the peptide.

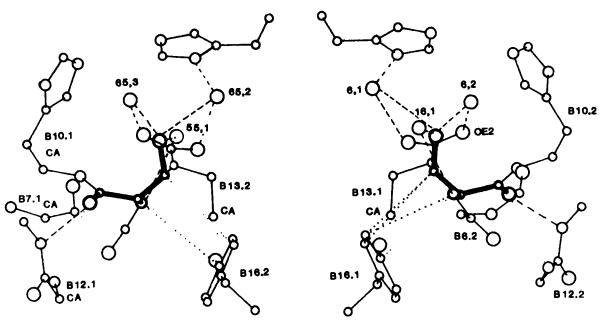
FIGURE 4.8. **B8 glycine**; I



NH	– 49W2	26W1	3.0	3.0	H...O	2.1	2.1	Ĥ	162	156	ψ −132.9 −133.2	ϕ 58.4 54.2
											B m.c. 15.1 12.6	*B* s.c.
O	– HN	B11	3.1	3.1		2.4	2.2	Ô	100	116	Δ_1 m.c. 0.53	
	HN	B12	3.4	(3.8)		2.4	(2.5)	Ô	148	(110)	Δ_2 m.c. 0.04	

The B8 Gly is on the molecule's surface at the N terminal end of the α helix. In the dimer it is approached by B16 Tyr of the twofold related molecule. The folding of the chain into helix is permitted by B8 Gly, which has D amino acid ϕ, ψ angles and contacts through peptide O, B11 NH, in both molecules. There are longer contacts to B12 NH. Both NH contact, via water, B26 OH though with different geometry, presumably caused by the different conformations at the adjacent A chain N terminus.

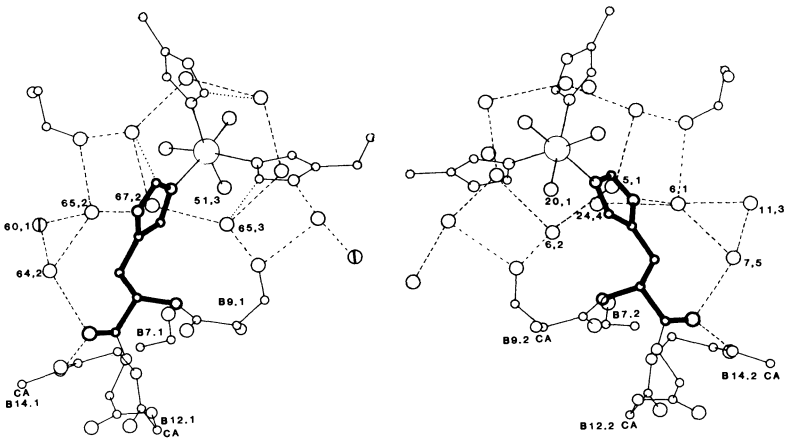
FIGURE 4.9. **B9 serine**; Lys, Pro (Leu 0.6, n)



NH	– 55W1	16W1	3.1	3.0	H...O	2.1	2.1	Ĥ	165	158	ψ −29.0 −19.1	ϕ −62.3 −64.5
											B m.c. 10.7 12.0	*B* s.c. 18.0 17.6
O	– HN B13		3.1	3.1		2.2	2.2	Ô	144	149	Δ_1 m.c. 0.14	Δ_1 all 0.14
OG	55W1	16W1	3.0	3.0					107	116	Δ_2 m.c. 0.06	Δ_2 all 0.07

The serine side chain and peptide NH is turned out from the molecule and is part of the polar B chain helical N terminal surface, which forms the hydrophilic central core of the hexamer. In the hexamer it lies between each pair of crystal-related histidyl side chains. The OG makes H-bonding contact to 3 waters making a roughly tetrahedral structure about the O. All of these waters H bond back to the protein; (65W2 and 6W1) to ND2 of the threefold related B10 imidazole; (65W3 and 6W2) to B13 COO^- of the other molecule in the dimer; (55W1 and 16W1) to the B9 NH of the same molecule. The peptide O is in the helix, H bonded to B13 NH.

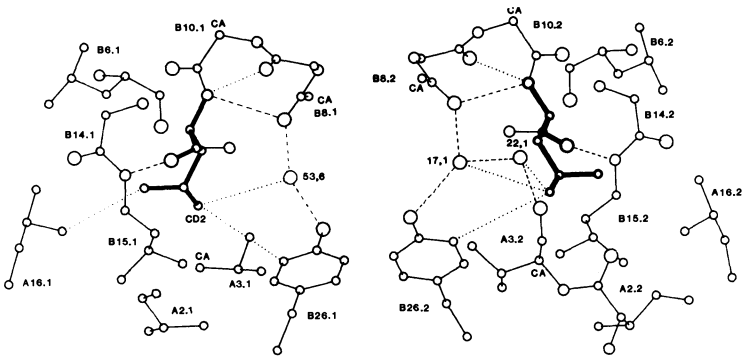
FIGURE 4.10. **B10 histidine**; Asn, Glu, Asp (Asp 2–4 n; Leu 0.41 r)



NH	– O	B7	3.1	3.1	H...O	2.4	2.3	Ĥ	128	136	ψ −30.4 −33.6	ϕ −66.5 −67.9
											B m.c. 9.3 9.7	*B* s.c. 10.6 10.6
O	– HN	B14	3.2	3.2		2.2	2.2		137	144	Δ_1 m.c. 0.17	Δ_1 all 0.24
	64W2	7W5	2.8	2.8				Ô	121	131	Δ_2 m.c. 0.05	Δ_2 all 0.05
NE2	– Zn		2.1	2.1								

The histidyl side chain extends from the helix axis alongside the B9 OG, which it contacts at about 3.3 Å. In the hexamer, the imidazole group and its threefold related equivalents coordinate through NE2 to the Zn ions. There are only very small differences in the Zn imidazole geometry and the bond lengths are essentially identical (2.05 Å and 2.06 Å respectively). The water molecules completing the zinc ion's octahedral coordination are close (3.0 Å) to the imidazole NE2 atoms. The threefold related histidines alternate with the B9 serine side chains around the threefold axis. There is one water (65W2, 6W1) strongly H bonded to B10 ND1; it links to B9 OG of the threefolded related serine.

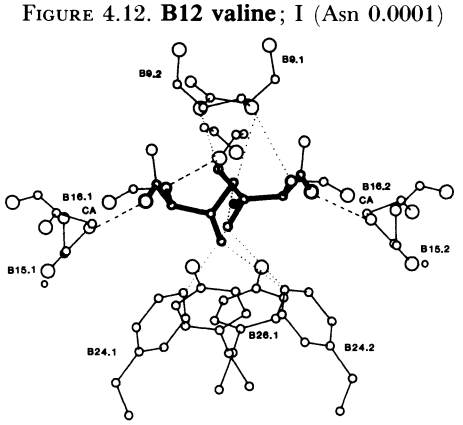
FIGURE 4.11. **B11 leucine**; I



NH	– O	B8	3.1	3.1	H...O	2.4	2.2	Ĥ	131	138	ψ −47.1 −38.4	ϕ −67.2 −68.0
											B m.c. 8.5 9.1	*B* s.c. 13.0 15.1
O	– HN	B15	3.1	3.0		2.1	2.1	Ô	149	153	Δ_1 m.c. 0.20	Δ_1 all 0.28
											Δ_2 m.c. 0.07	Δ_2 all 0.09

The NH makes a 1→4 H bond and the CO a 1→5 H bond, part of the B chain central α helix. The residue is buried in the molecule near its centre of mass; the side chain is closely surrounded by three other leucines, A6, B15 and A16 and more distantly it is also adjacent to A2 in molecule 1. This cluster of leucines constitutes a major portion of the monomer's non-polar core. In molecule 2, the arrangement of the A chain N terminal residues also brings A3 Val alongside B11 leucine side chain. A water molecule (53W6 and 17W1) attached to B8 O is at van der Waal's distance to the side chain; in molecule 2, A30 stabilizes and traps water (22W1) H bonded to 17.1 near the leucyl atoms.

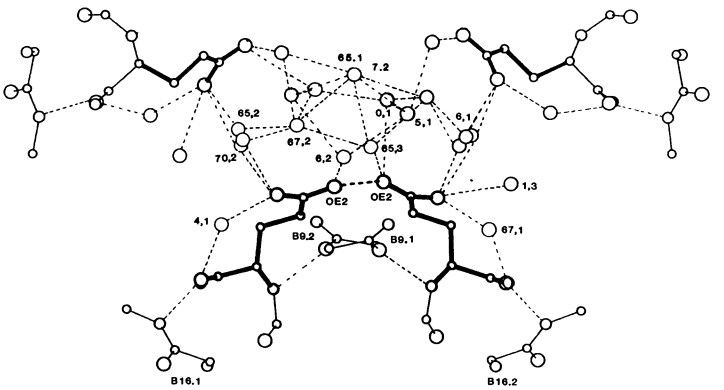
FIGURE 4.12. **B12 valine**; I (Asn 0.0001)



NH	– O B9		3.3	3.2	H...O	2.9	2.6	Ĥ	109	114	ψ −24.2 −33.8	ϕ −71.9 −65.6
	O B8		3.4	3.8		2.4	2.8	Ĥ	159	155	*B* m.c. 6.9 9.5	*B* s.c. 14.9, 11.3
											Δ_1 m.c. 0.26	Δ_1 all 0.80
O	– HN B16		2.9	3.0		2.0	2.1	Ô	146	143	Δ_2 m.c. 0.06	Δ_2 all 0.57

The valine side chain is directed out from the B chain helix and fills space in the non-polar and mainly aromatic surface buried by dimer formation. The side chain in both molecules makes internal and external contacts; those in molecule 1 are tighter (to B9 particularly) and fix the valine in one conformation. There are two conformations in molecule 2 ($x \sim 120°$). Both B12 valines make loose but defining contacts within the monomer to B8 Ser, B24 Phe and B26 Tyr. The two side chains also make contacts within the dimer, the closest approach being made by one of the disordered CH_3 groups in molecule 2 to the B12 Val across the diad axis. There are also longer non-polar contacts to B16 Tyr within the monomer and dimer, and B24 Phe and B26 Tyr, in the opposite molecule ranging between 3.6 and 4.2 Å. The peptide NH in molecule 1 only forms one good 1→5 H bond (to B8 O); the peptide O forms a well-defined 1→5 bond within the α helix.

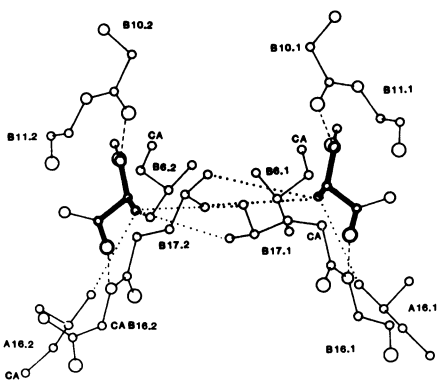
FIGURE 4.13. **B13 glutamic acid**; Asp, Asn (Gln 18, s)



NH	– O	B9	3.1	3.1	H...O	2.2	2.2	Ĥ	151	149	ψ −39.3 −47.5	ϕ −66.7 −63.1
											B m.c. 8.7 9.6	*B* s.c. 32.6 32.6
O	– HN B17		3.1	3.0		2.1	2.0	Ô	147	151	Δ_1 m.c. 0.16	Δ_1 all 0.21
	4W1	67W1	2.9	2.7				Ô	111	110	Δ_2 m.c. 0.03	Δ_2 all 0.10
OE2	OE2B13.1	2.6	2.6					Ô	122	111		

The two side chains are in close contact (*ca.* 2.6 Å) across diad d_0 axis at the hexamer centre where they form three pairs about the threefold crystal axis. There is an extensive network of water about the carboxylate groups; there is low density corresponding approximately to one water molecule linking carboxylate ions across d_0 (6, 1; 1, 3; 70, 2; 65, 2). Each carboxylate group is also H bonded to a water molecule (4W1, 67W1) which is H bonded in turn to B13 peptide O. The peptide O (which makes two H bonds, one to water) and N are also involved in 1 → 5 helical H bond contacts. The dimer packing brings B9 Ser close to B13 of the diad-related molecule.

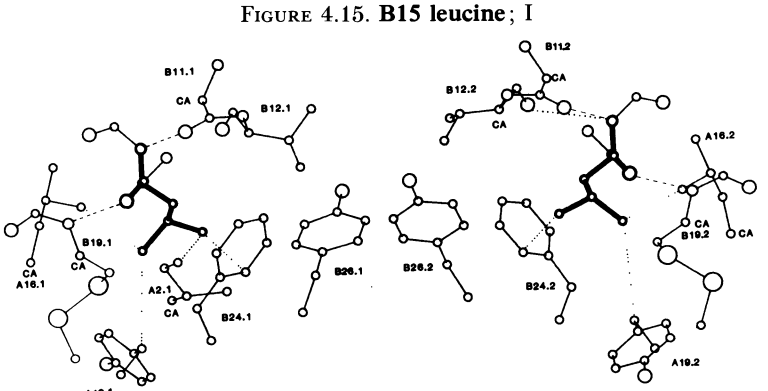
FIGURE 4.14. **B14 alanine**; Thr



NH	– O B10		3.2	3.2	H...O	2.2	2.2	Ĥ	152	169	ψ −39.9 −44.9	ϕ −65.1 −56.1
											B m.c. 10.5 8.1	*B* s.c. 9.2 12.1
O	– HN B18		2.9	2.9		2.0	1.9	Ô	151	151	Δ_1 m.c. 0.12	Δ_1 all 0.12
											Δ_2 m.c. 0.04	Δ_2 all 0.10

The peptide NH and O are both in the α helix. The residue is within 6 Å of the centre of the molecule but on the surface of the dimer, part of the non-polar region buried in hexamer formation. As a consequence each methyl group makes rather long contacts with the B17 leucine residue of both molecules 1 and 2. The electron density is very well defined, showing evidence of hydrogen atoms in staggered positions.

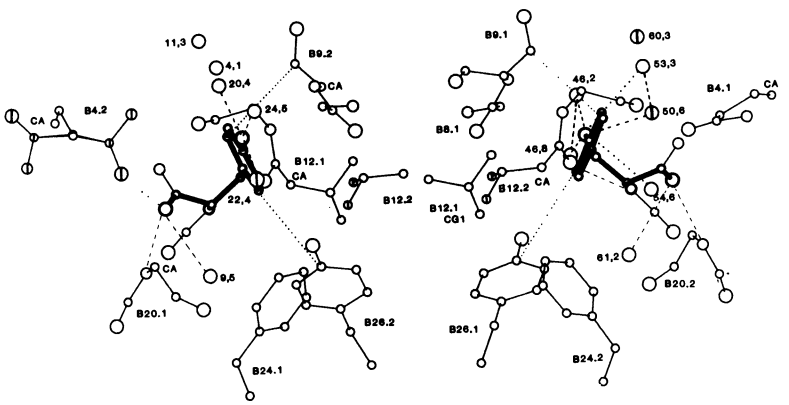
FIGURE 4.15. **B15 leucine**; I



NH	– O B11		3.1	3.0	H...O	2.1	2.1	Ĥ	154	153	ψ −45.5 −40.5	ϕ −62.9 −61.3
											B m.c. 11.7 11.8	*B* s.c. 12.2 11.3
O	– HN B19		2.8	2.9		1.8	1.9	Ô	148	148	Δ_1 m.c. 0.08	Δ_1 all 0.12
											Δ_2 m.c. 0.04	Δ_2 all 0.07

The side chain is well buried in the monomer making no solvent contacts. It is surrounded by B11 Leu, B12 Val and B19 Cys (all one turn of the helix away), A16 Leu and B24 Phe. The residue obeys the local axis very well; this symmetry extends to its immediate environment. More distant, A19 Tyr and A2 Ile (closer in molecule 1), complete B15 Leu's non-polar environment. The peptide O and N are in the centre of the B chain helix; both make 1 → 5 H bonds.

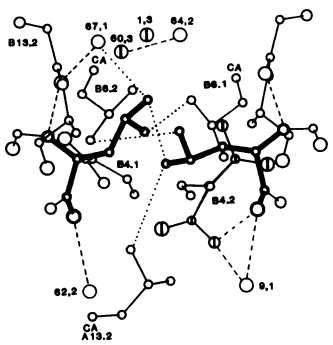
FIGURE 4.16. **B16 tyrosine**; Phe ((+)-Tyr, 0.17, n; Glu, 0.3, r)



NH	–	O B12		2.9	3.0		2.0	2.1	Ĥ	146	143	ψ −46.0 −46.0	ϕ −57.2 −61.0
												B m.c. 10.6 10.7	*B* s.c. 12.8 15.3
O	–	HN B20		2.7	2.9		1.9	2.0	Ô	132	127	Δ_1 m.c. 0.13	Δ_1 all 0.38
		9W5	61W2	3.1	3.3				Ô	109	107	Δ_2 m.c. 0.03	Δ_2 all 0.12

The side chain is directed away from the helix and contributes to the non-polar dimer-forming surface. The twofold symmetry in the residue conformation and much of the environment is obeyed closely. Within the monomer the only significant contact is to B24 Phe. On dimer formation the side chain makes contacts through one edge to B8 Gly, B9 Ser, B12 Val and B26 Tyr of the other molecule. The other edge is exposed to solvent, which is part of a network of water molecules leading from the tyrosine hydroxyl group and stretching across the protein surface. The peptide NH makes a good 1→5 H bond; the peptide O is in the last turn of helix and its interaction with B20 NH is shared with a water molecule (9W5 and 61W2).

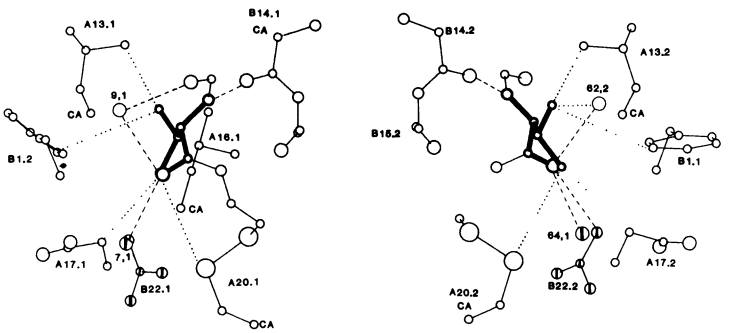
FIGURE 4.17. **B17 leucine**; Ile, Phe, Ser



NH	–	O B13		3.1	3.0	H...O	2.1	2.0	Ĥ	160	150	ψ −45.9 −45.7	ϕ −60.3 −65.0
												B m.c. 11.2 11.0	*B* s.c. 25.7 26.3
O	–	9W1	62W2	3.0	3.0				Ô	137	139	Δ_1 m.c. 0.13	Δ_1 all 1.21
		HNE2 B4		3.1			2.4		Ô	167		Δ_2 m.c. 0.03	Δ_2 all 0.73

The residue is in α helix with its leucyl side chains on the dimer surface that is buried by hexamer formation. The peptide NH forms a 1→5 H bond; the peptide O is at the helix C terminus; it interacts with solvent. The side chains are disposed opposite each other across the local diad-relating dimers; their close approach (*ca.* 4 Å) prevents them assuming the same conformation. Thus the molecule 2 side chain, which makes water contacts, is turned away from the simpler extended conformation in molecule 1. Within the monomer, B17 Leu is surrounded by B14 Ala and B18 all at distances of 4–5 Å. Within the hexamer, B4 Gln, B6 Leu, A14 Ala and B17 Leu from the adjacent dimer all make approaches of about 4 Å. Molecule 1 B17 is distinctly closer to Leu A13.2 (2).

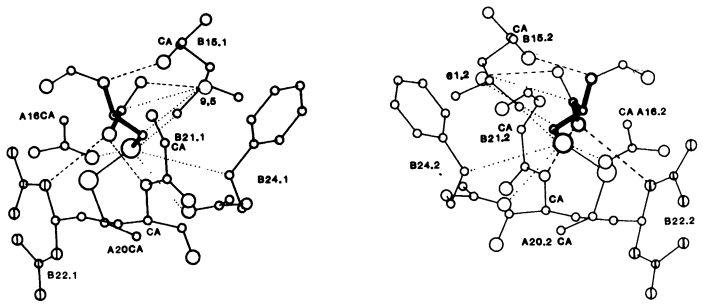
FIGURE 4.18. **B18 valine**; Ala



NH	–	O B14		2.9	2.9	H...O	2.0	1.9	Ĥ	161	151	ψ −41.8 −47.6	ϕ −73.6 −7.2
												B m.c. 10.2 11.1	*B* s.c. 14.2 12.9
O	–	HNH B22		2.8	3.1		2.1	2.6	Ô	156	159	Δ_1 m.c. 0.12	Δ_1 all 0.15
		7W1	64W1	3.0	2.8					145	134	Δ_2 m.c. 0.04	Δ_2 all 0.10
		9W1	62W2	3.2	3.1					102	108		

The residue is situated on the final turn of helix B9–B20. The peptide NH is in a 1→5 H bond, the peptide O is H bonded to the disordered B22 side chain (NH_2^+) and to a water molecule (7W1 and 64W1) also disordered, which replaces it. There is a second less favourable H bond from the B18 peptide O to 9W1 and 62W2, which in turn H bonds to B2 peptide NH and O. There are contacts within the monomer and dimer to the aliphatic groups in A17 Glu (*ca.* 3.1 A) and to A13 and at a longer distance to A16 Leu. This surface in the hexamer packs against B1 Phe and, more distantly, B2 Val. The local symmetry is rather well obeyed.

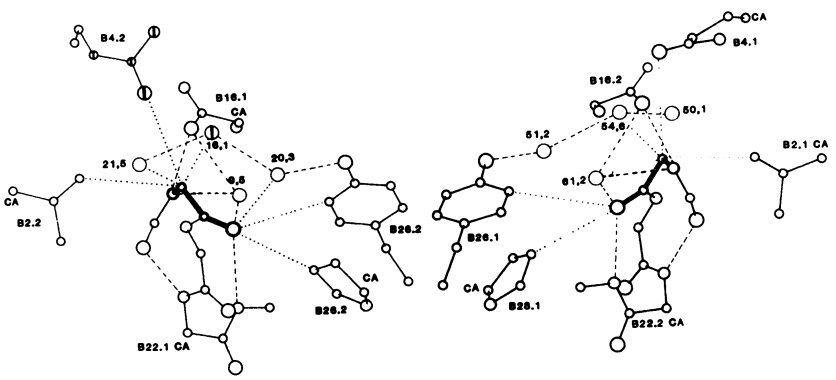
FIGURE 4.19. **B19 cystine**; I



NH	–	O B15	2.8	2.9	H...O	1.8	1.9	Ĥ	156	163	ψ −43.2 −29.9	ϕ −86.2 −88.7
											B m.c. 14.3 12.3	*B* s.c. 11.5 10.7
O	–	HN B22	3.1	3.0		2.2	2.1	Ô	135	151	Δ_1 m.c. 0.22	Δ_1 all 0.18
		HNE2 B22	3.2	(3.7)		2.4	(2.7)	Ô	126	(112)	Δ_2 m.c. 0.09	Δ_2 all 0.10

The residue B19–A20 Cys forms a disulphide bond with very similar conformations in the two molecules. The B19 sulphur is completely buried, surrounded by the non-polar side chains of B15 Leu, B18 Val and B24 Phe; there are also close approaches to the peptides B20–B23 in a 1→4 turn. The peptide NH is in a 1→5 H bond, the peptide O (no longer in helix) H bonds to the disordered B22 side chain.

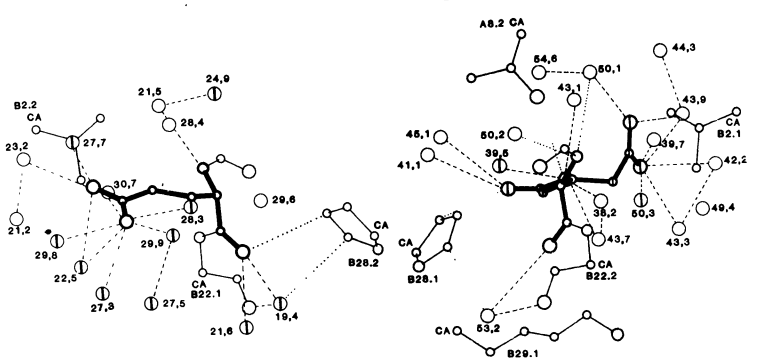
FIGURE 4.20. **B20 glycine**; Gln, Lys, Arg



NH	–	O B16		2.7	2.9	H...O	1.9	2.0	Ĥ	132	141	ψ −125.7 −130.7 ϕ 57.5 47.6
		9W5	(61W2)	2.5	(2.9)		2.3	(2.9)	Ĥ	92	83	*B* m.c. 17.8 20.1 *B* s.c.
												Δ_1 m.c. 0.14
O	–	HN B23		3.1	2.9		2.2	2.0	Ô	111	119	Δ_2 m.c. 0.06
		9W5	61W2	2.6	2.4				Ô	100	109	

The peptide NH is in helix (1→4); the peptide O is H bonded to B23 NH in a 1→4 turn to a water molecule (9W5 and 61W2) which is H bonded to B16.O. The conformation of the glycine is that of a D-amino acid, allowing the chain to change direction sharply. Being largely exposed, the CA is surrounded by solvent molecules; the residue makes a few internal contacts along the helix. There are non-polar contacts from the peptide O to B16 and B26 tyrosine belonging to the twofold-related molecule in the dimer. The dimer packing in the hexamer brings B2 Val and B4 Gln side chains into non-bonded contact to B20 Gly.

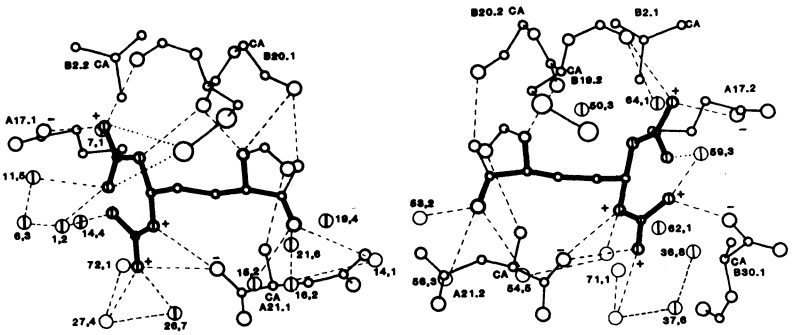
FIGURE 4.21. **B21 glutamic acid**; Asp, Asn, His, Pro, Val



NH	–	21W5	(50W1)	3.0	(3.8)	H...O	2.2	(3.1)	Ĥ	138	(125)	ψ −24.0 −2.29 ϕ −63.9 −71.5
												B m.c. 21.7 23.1 *B* s.c. 60.3 49.1
O	–	19W4	(53W2)	2.4	(3.6)	H...O			Ô	158	(135)	Δ_1 m.c. 0.27 Δ_1 all 1.40
		21W6		2.9		H...O			Ô	137		Δ_2 m.c. 0.11 Δ_2 all 0.54

The residue is situated at the centre of the 1→4 turn between B20 and B23 glycines and projects out from the molecule. The main-chain conformation of the peptides is significantly different. Both peptide N and O are directed into solvent. In molecule 1, the N makes one good H bond whereas the O makes H bonds to half-occupied and fairly mobile water molecules. Molecule 2 makes longer (3.5 Å) and less favourable contacts owing to the nearby B29 Lys (molecule 1) packed along the threefold axis. The side chain in molecule 1 is poorly defined, with many contacts to mostly partially occupied water sites. It appears to have only one conformation. In molecule 2 the side-chain density is confused and suggests a number of conformations exist. Two of these are shown; they make complex contacts with mostly fully occupied solvent molecule sites. In dimer formation, the residue is approached by B28 and in hexamer formation, B2 Val. The A8.2 Thr contacts one of the side-chain conformations of B21.2 along the threefold axis.

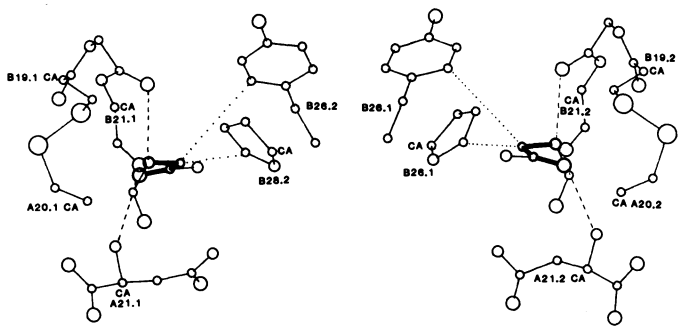
FIGURE 4.22. **B22 arginine**; Asp, Lys (Lys 0.36, r; Glu in desheptapeptide, active)



NH	–	O B19		3.1	3.0	H...O	2.2	2.1	Ĥ	135	156	ψ −21.2 −31.0 ϕ −72.7 −71.5 *B* m.c. 22.4 17.8 *B* s.c. 25.3 31.8
O	–	16W2	54W5	2.8	3.5				Ô	125	119	Δ_1 m.c. 0.18 Δ_1 all 0.33
		15W2	56W3	2.8	3.3					116	149	Δ_2 m.c. 0.10 Δ_2 all 0.16
		14W1	53W2	3.3	3.0					131	107	
NEH	–	O A21		3.0	2.8	H...O	2.2	1.9	Ĥ	154	155	
		O B19		3.2	3.7		2.4	2.7		136	160	
NH1	–	O A21		3.1	3.2		2.5	2.4		117	131	
NH2	–	OE B30.1			2.9							
		OE2 A17		2.9	3.5		2.0	2.6		156	158	
		O B18		2.8	3.1		2.1	2.6		129	108	

The residue NH is in a 1 → 4 turn formed at the end of the B9–B19 helix, with similar geometry in both molecules; the peptide O is directed into solvent and H bonded to 3 waters. The side chain lies on the surface and is disordered with two fairly well defined principal conformations related by 180° rotation about CG–CD. Both conformations make salt bridges: one to the A21 carboxylic acid, the other to A17 Glu carboxylic acid. The second arrangement also brings NE into H-bonding interaction with B18 O perhaps cancelling the negative charge produced there by the helix dipole (Hol 1978). In molecule 2, there is a salt bridge to B30 OE from hexamer (6) related by the threefold screw axis. Some well defined complementary water molecules, displaced from H bond interactions by the disordered guanadinium group, have been identified (7W1 and 64W1). In spite of the disorder, the two residues obey the local twofold axis quite well.

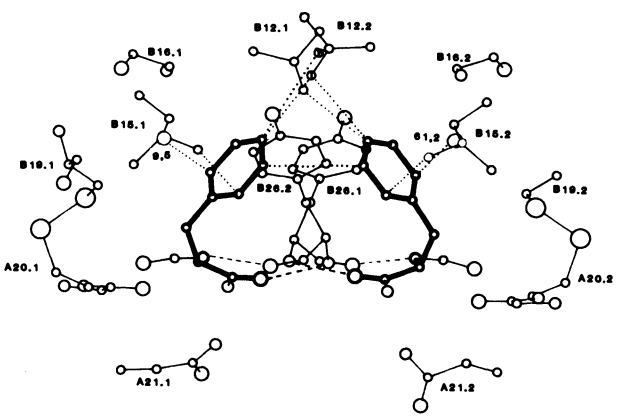
FIGURE 4.23. **B23 glycine**; I ((+)–Ala, 1.0; (−)–Ala, r, desheptapeptide insulin, r)



NH	–	O B20		3.1	2.9	H...O	2.2	2.0	Ĥ	150	148	ψ 174.5 171.9 ϕ 79.9 81.4 *B* m.c. 13.6 14.6 *B* s.c.
0	–	HN A21		3.0	2.9		2.0	2.0	Ô	119	133	Δ_1 m.c. 0.14 Δ_2 m.c. 0.06

The B23 glycine is the first residue in the final stretch of largely extended B chain. Its peptide N is in a 1 → 4 bonding turn with B20 O, its peptide O makes a structurally important H bond to A21 NH. The residue lies on the monomer surface adjacent to A21 Asn side chain. But it is buried by dimer formation, approaching in that structure, B26 Tyr and B28 Pro. The ψ, ϕ angles for this peptide are only possible for a D-amino acid. The absence of side chain favours both the turn between B20–B23 and the close approach of the monomer in the dimer.

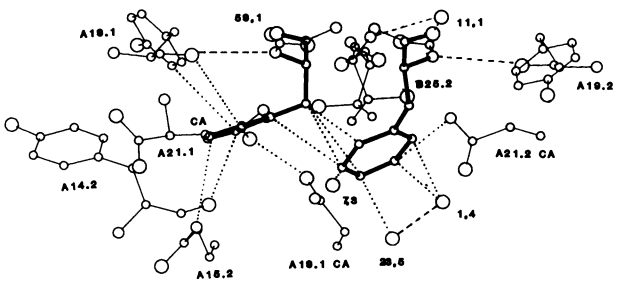
FIGURE 4.24. **B24 phenylalanine**; Ser (Leu, 0.2, s; (+)–Phe, 2.0, r; Ser 0.01, r)



NH	–	O B26 (2)		3.1	2.9	H...O	2.2	2.1	Ĥ	139	140	ψ 171.3 173.0 ϕ −153.1 −154.8 *B* m.c. 10.9 11.1 *B* s.c. 12.8 8.6
O	–	HN B26 (2)		2.9	2.9		1.9	1.9	Ô	166	172	Δ_1 m.c. 0.12 Δ_1 all 0.16 Δ_2 m.c. 0.03 Δ_2 all 0.05

The peptide NH and O are in the antiparallel B sheet with B26 tyrosine around the dimer twofold axis. The phenyl groups pack against the body of the molecule, making contacts with valine B12, leucine B15 and B19 SG, cystine. They lie near the centre of the dimer-forming surface, in contact with each other and with tyrosine B26 in the twofold related molecule. The single water molecule in contact with each phenyl group is essentially buried, hydrogen bonded to peptide groups but touching no other water molecules.

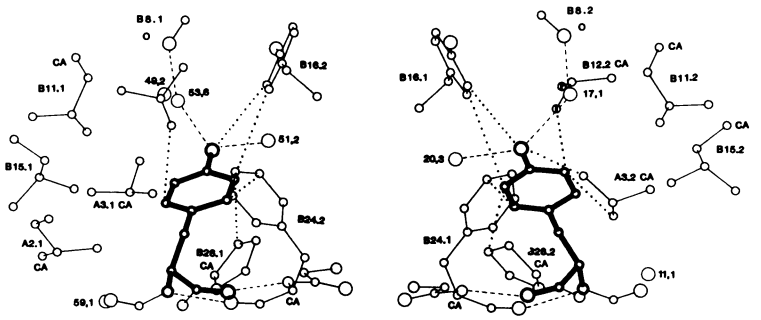
FIGURE 4.25. **B25 phenylalanine**; Tyr, *Leu* (Leu, 0.01, n; Ser, 0.01)



NH	–	OA19		3.2	3.4	H...O	2.2	2.4	Ĥ	160	159	ψ 138.1 132.3 *B* m.c. 10.3 13.5	ϕ −127.1 −115.7 *B* s.c. 15.4 38.0
O	–	59W1	11W1	3.1	2.9				Ô	117	126	Δ_1 m.c. 0.19 Δ_2 m.c. 0.06	Δ_1 all 3.66 Δ_2 all 1.41

The peptide NH group is hydrogen bonded to the A chain carbonyl O at A19 tyrosine in a continuation of the sheet structure between the twofold related B24 and B26 peptides. The carbonyl oxygen is hydrogen bonded to a very well-defined buried water molecule, which is also bonded to B27 NH. The side chains are on the hexamer surface and have different conformations. B25.1 phenylalanine turns in towards the A chain, touching tyrosine A19.1. It is also in contact with residues in the neighbouring hexamer (8), particularly tyrosine A14.1. B25.2 phenylalanine turns out from molecule 2 to make contact with B25.1 phenylalanine across the twofold axis. In this position it has more freedom of movement, its other contacts are mostly with water molecules. A persistent low streak of density in F_o, $2F_o-F_c$ and ΔF maps suggest that a small proportion of the residues have swung over to an alternative conformation, similar to that of B25.1 (see §8).

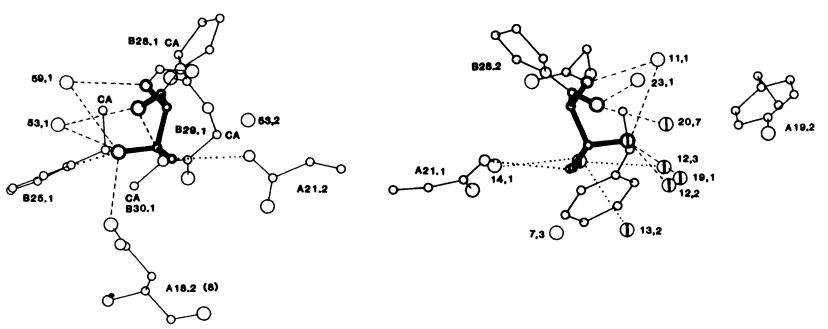
FIGURE 4.26. **B26 tyrosine**; Arg, (des B26–B30, 0.4, n)



NH	–	O B24		2.9	2.9	H...O	1.9	1.9	Ĥ	178	175	ψ 105.0 108.6 *B* m.c. 10.8 12.8	ϕ −119.2 −114.4 *B* s.c. 15.3 14.0
O	–	HN B24		2.9	3.1		2.1	2.2	Ô	139	139	Δ_1 m.c. 0.26 Δ_2 m.c. 0.08	Δ_1 all 0.34 Δ_2 all 0.04

The residue lies on the dimer forming surface. Its peptide O and N are both H bonded (to B24) in the antiparallel B sheet structure between monomers. The extended main-chain structure directs the side chain internally, where it packs against B28 Pro on one side and B24 Phe on the other. Each tyrosyl OH is bound to two water molecules, through one of which it is connected back to the main-chain peptide B8 O.

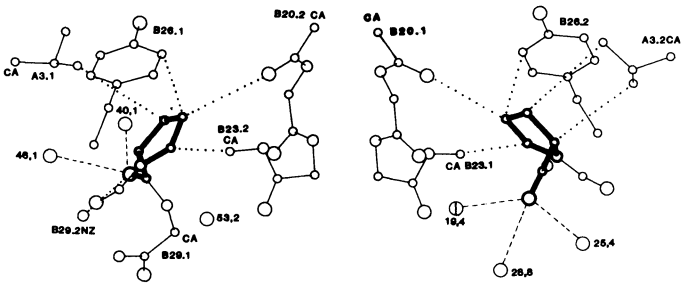
FIGURE 4.27. **B27 threonine**; Ile, Arg, Ser, Glu, Asn, Leu, Asp (des B27–B30, 0.71, n)



NH	–	59W1	11W1	3.0	2.8	H...O	2.0	1.8	Ĥ	163	165	ψ 86.1 87.2 *B* m.c. 14.9 18.3	ϕ −126.4 −124.7 *B* s.c. 17.7 33.6
O	–	HN B30		3.3			2.3			154		Δ_1 m.c. 0.71	Δ_1 all 0.87
			23W1		2.8				Ô		135	Δ_2 m.c. 0.07	Δ_2 all 0.15
		53W1	20W7	3.1	2.7					133	150		
OG	–	HN B18.2		2.7									

The residue is on the hexamer surface and makes all its H-bonding contacts in molecule 2 to solvent and in molecule 1 all but one. In each residue there are two water molecules H bonded to the OG; one of these is in turn H bonded to the peptide N and the other to the peptide O. The water at the peptide N is further H bonded to B25 peptide O. In molecule 1, the C terminal structure brings B27 O into a 1→4 turn with B30 NH as well. Dimerization brings the side chain to van der Waals separation from A21 Asn side chain. In the crystal, B27 OG of molecule 1 is stabilized by an H bond to A18 OD Asn in the adjacent hexamer, (8). This interaction does not occur in molecule 2 where the side chain takes up two conformations in both of which OG is H bonded to a network of surrounding water.

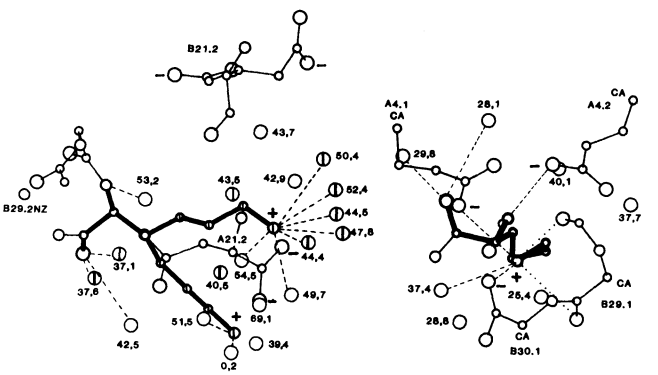
FIGURE 4.28. **B28 proline**; Ser, Arg, Asn (des B28–B30, 1.0, n)



O	–	40W1	19W4	3.0	3.0			Ô	141	128	ψ −27.8 162.9	ϕ −62.3 −76.1
											B m.c. 17.9 21.1	*B* s.c. 17.3 19.1
		46W1	25W4	3.0	2.9				120	95	Δ_1 m.c. 1.06	Δ_1 all 0.83
			28W8		3.4					128	Δ_2 m.c. 2.44	Δ_2 all 0.67

This residue is on the monomer surface and is largely buried by dimer formation. The extended chain structure packs the prolyl sidechain against B26 tyrosyl sidechain. The rearrangement of molecule 2 in the crystal produces rotation of about 180° at the B28–B29 peptide (see §8). In both molecules, the peptide O is H bonded to solvent; in molecule 1, there is a van der Waals contact to the NZ of B29.2 Lys of the hexamer packed below in the *c* direction. Dimer formation packs the proline against the residues B20–B23, which are in a 1→4 turn.

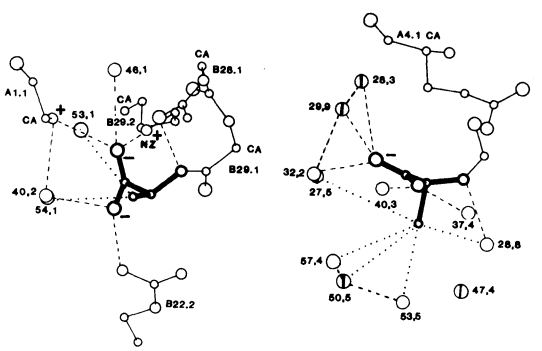
FIGURE 4.29. **B29 lysine**; Met, Ser, Asn, Arg, Thr (succinyl 0.4, acetyl; 1.0 n)



NH	–	OE2 A4			2.8	H...O		1.9	Ĥ		166	ψ −16.7 145.1	ϕ −73.4 −77.6
		53W2		3.1			2.1		Ĥ	170		*B* m.c. 23.7 35.4	*B* s.c. 44.4 17.3
O	–	37W1	28W1	3.1	3.4				Ô	137	134	Δ_1 m.c. 3.62	Δ_1 all 7.02
		37W6	29W6	3.0	2.4				Ô	161	153	Δ_2 m.c. 2.49	Δ_2 all 5.03
		42W5		3.3					Ô	127			
NZ	–	OE1 A4.1			3.1				N̂		102		
		O B30.1			2.5						153		

The residue is on the hexamer surface. In molecule 1, the residue is part of a 1→4 turn between B27 and B30; both peptide O and N are H bonded to water. Although the packing of the hexamers along the *c* direction brings B21.2 (glutamic acid) nearby, the side chain in molecule 1 makes only loose contact with water; it has at least two conformations and is poorly defined. Crystal packing has imposed a quite different conformation on B29 in molecule 2. The rotation around the B28–B29 peptide allows B29.2 NH to H bond to A4.2 Glu OE. At the same time B29 turns up and the NZ is able to salt bridge across the gap between the two stacked hexamers to B30.1 O and to A4.1 Glu OE (see §8).

FIGURE 4.30. **B30 alanine**; Thr, Ser, Asp, Glu, Asx, Lys (des B30 1.0, n)



NH	–	O B27.1		3.3		H...O	2.3		Ĥ	153		ψ 137.8 141.0	ϕ 76.0 −90.0
			28W8		3.3			2.4			120	*B* m.c. 26.4 52.9	*B* s.c. 34.4 64.8
O	–	NH_3^+ A1.1		2.9					Ô	128			
		NZ B29.2		2.5						106		Δ_1 m.c. 8.13	Δ_1 all 8.12
		46W1	37W4	3.1	2.4					161	142	Δ_2 m.c. 0.19	Δ_2 all 0.13
			40W3		3.0						128		
OE	–	NH_2^+ B22.2'	(8)	2.9			1.9		Ĥ	154			
		40W2	27W5	3.0	3.2				Ô	113	128		
			28W5		3.4						127		
			29W9		2.5						147		
			32W2		2.4						137		

The two residues are both on the hexamer surface but have very different conformations and contacts. In molecule 1, the peptide NH interacts in a 1→4 turn with B27 peptide O. The terminal carboxyl group is in contact with surrounding positive groups, in the same and different molecules; O^- with A1 NH_3^+ at the beginning of the A chain and B29.2 lysine NH_3^+ of the hexamer at Z, OE with B22.2 NH_2^+ arginine, in the adjacent hexamer (8). Both also form hydrogen bonds with water molecules. The network appears to stabilize the crystal packing. By contrast, molecule 2 residue 30 is essentially floating, loosely packed against the A4 glutamic acid of molecule 1 across the twofold axis between hexamers. All the close contacts of the NH and carboxyl group are with water molecules.

5. The connections between the A and B chains in each molecule

The A and B chains are drawn together through non-polar contacts between the hydrocarbon residues distributed along their sequence. The chains are covalently linked to each other by the disulphide bonds A7–B7 and A20–B19 and their precise relative positions further secured by interchain hydrogen bonds.

5.1. *Non-polar contacts*

The mass centre of each of the two molecules is empty, but surrounded, within a sphere of 6 Å radius, by nine predominantly and mostly wholly buried non-polar residues. These residues include all the invariant leucine residues, A16, B6, B11 and B15, the invariant cystine A6–A11, and the A19 tyrosine ring, also A2 isoleucine, which varies in the green monkey, and alanine, B14, which only varies in the guinea pig. The atoms of this inner core make contact with other non-polar residues, A3 valine, A13 leucine, B12 and B18 valine, B24 phenylalanine and B19 cystine, which, in turn, contact most of the remaining non-polar residues. Because members of the last two groups are all partly on the outside of the molecule and make contact with other non-polar residues in the dimer and hexamer interfaces, there is, in fact, an irregular belt of continuous non-polar character within the hexamer around the threefold axis (see figure 7.6). Only the non-polar residues, A10 isoleucine, A7.2–B7.2 cystine, and B30 alanine do not take part in this system.

Figure 5.1 illustrates the actual arrangement of the atoms in the inner core, which at the isoleucine A2 particularly, is different in the two molecules. Although contact distances are rather long, 4–5 Å, all the atoms are well defined except A2 isoleucine, molecule 2. This residue has been moved rather far from the close-packed position of molecule 1 by the A chain conformational change.

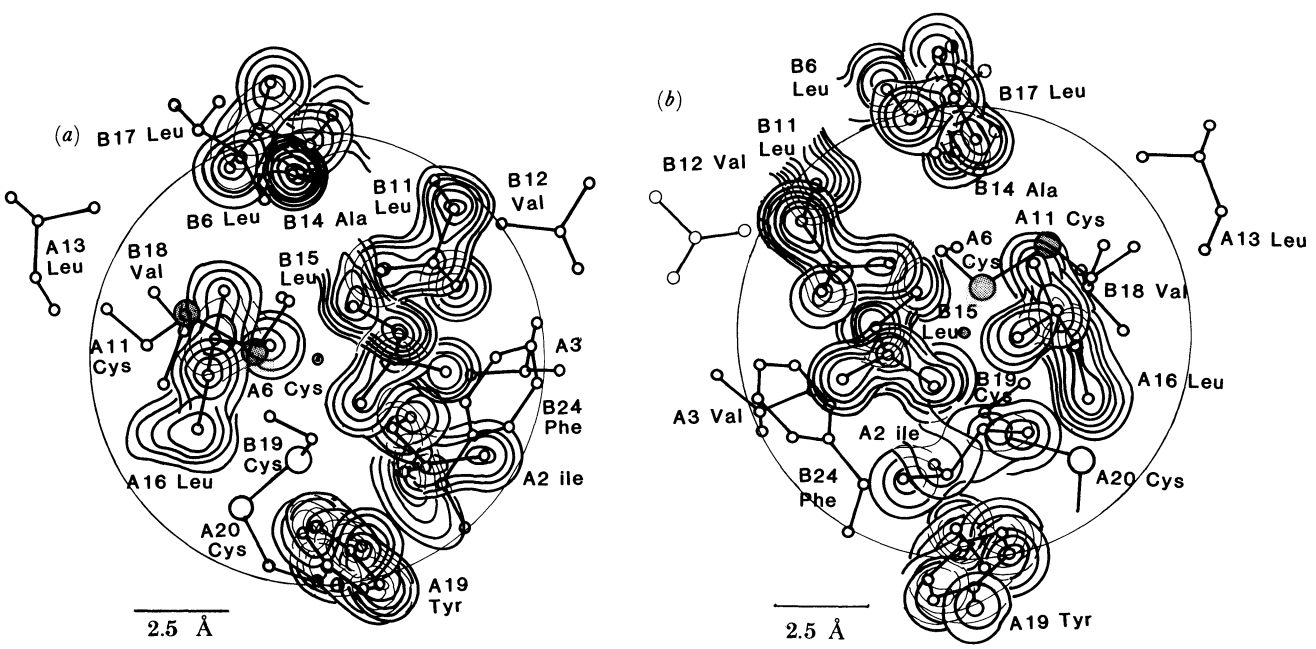


Figure 5.1. The atoms and electron density in the inner core of (*a*) molecule 1 and (*b*) molecule 2, viewed down the *c* axis. Contours are drawn at 0.3 e/Å^3, starting at 0.7 e/Å^3.

The non-polar residues shown in figure 5.2 are evidently brought together by the chain folding; the compact structure formed by the A chain rests on the B chain helix between the extended N and C terminal residues of the B chain. Of the core residues, A6, A11, A16 and B11, B15 and B19 are completely buried. It is notable that none of these buried residues is aromatic, in marked contrast with the situation in the aggregation surfaces.

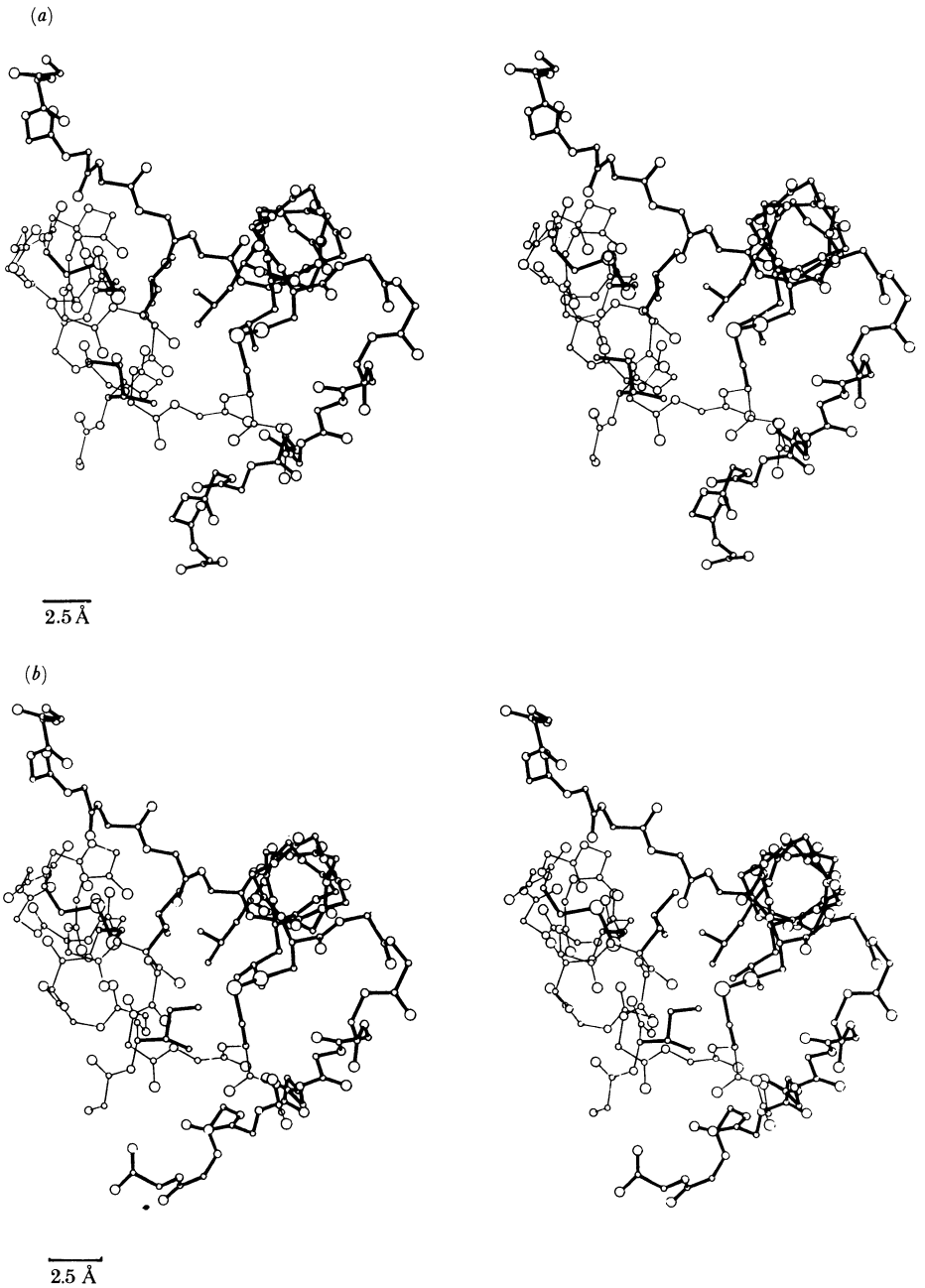


FIGURE 5.2. Stereo views of the buried sidechains of (*a*) molecule 1 and (*b*) molecule 2 viewed down the B chain helix B9–B20. The B main chain and buried side chains are drawn with thick lines and the A main chain with thin lines.

5.2. *The disulphide bonds, including* A6–A11

Figure 5.3 illustrates the geometry of the three disulphide bonds in each molecule, A6–A11 within the A chains, A7–B7 and A20–B19 between the chains. Some stereochemical details are listed in tables 5.1 and 5.2; two of the cystines are right handed in each molecule and one left handed; a similar arrangement is found in the three disulphide bonds of the protein crambin (Hendrickson & Teeter 1981).

In general, the disulphide bonds in the two molecules are well related by the twofold axes. Of the three disulphide bonds in each molecule, A6–A11 is completely buried, A20–B19 is largely buried but the A20 SG can be reached by the water molecule that replaces the

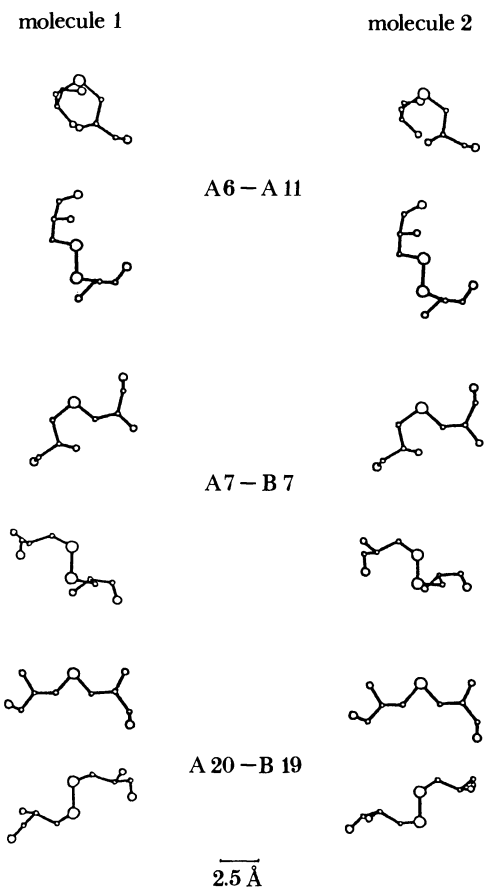


FIGURE 5.3. The individual disulphide bonds viewed along the bond and perpendicular to the bond.

TABLE 5.1

cystine		hand	dihedral angle	bond length (restrained)/Å
A6–A11	(1)	R	108° ± 1.5°	2.06 Å
	(2)	R	110° ± 1°	2.05 Å
A7–B7	(1)	R	102° ± 1°	2.01 Å
	(2)	R	102° ± 1°	1.97 Å
A20–B19	(1)	L	82° ± 1°	2.02 Å
	(2)	L	81° ± 1°	2.00 Å

disordered arginine B22 residue, whereas A7–B7 is exposed; both sulphur atoms are in contact, rather differently, with water molecules at distances between 3.4 and 3.9 Å.

The positions of all of the cystine residues close to the ends of α helices probably contribute to the rigidity and stability of the core of the molecule. They may also help in the chemical synthesis of insulin, encouraging the correct combination of the separate chains, particularly when A6–A11 is preformed in the A chains.

5.3. *Hydrogen-bond and salt-bridged contacts*

The precise relative positions of the A and B chains are strengthened at a number of points by strategic hydrogen bonds and ionized contacts.

(*a*) *Between the peptide chains*

The hydrogen bonds occur in two regions and are shown below (table 5.2) and illustrated in figure 5.4.

In the first group, the hydrogen bonds secure the A chain to the B chain between A6..A11, the region of the disulphide ring. The pair between A11 and B4 can be seen as part of an

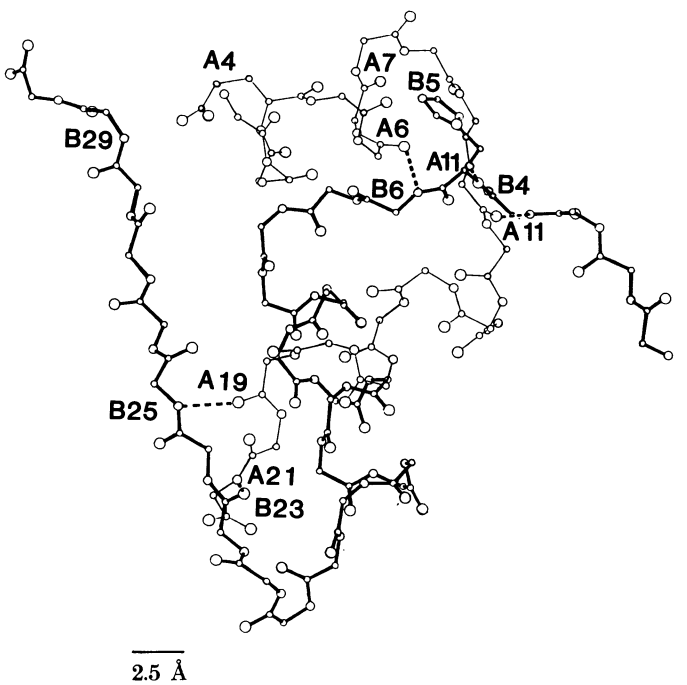


Figure 5.4. The insulin A and B chains, with the H bonds between them.

TABLE 5.2

		distances		angles at H/deg	
		molecule 1	molecule 2	molecule 1	molecule 2
(*a*)	B4 NH...OC A11	3.1	2.7	171	167
	B4 CO...HN A11	3.2	3.0	139	141
	B6 NH...OC A6	2.8	2.9	167	152
(*b*)	B23 CO...HN A21	3.0	2.9	163	160
	B25 NH...OC A19	3.2	3.4	160	159

antiparallel pleated sheet, twisted by about 20° formed at the point where the A chain turns to run antiparallel with the B chain.

The second group secures residues towards the end of the A chain to the B chain at the point where the B chain turns to adopt the extended configuration.

(*b*) *Between side chains and the peptide chains*

Side-chain to main-chain bonds reinforce the arrangement in these two regions. In the first, histidine B5.1 ND1 bonds to CO A7, and histidine B5.2 NE2, bonds to A7.2 CO. The difference is because of the difference in the histidine ring conformations, which also introduce other intra and interhexamer contacts. In the second region, near the A chain termini, B22 arginine is bonded in one of its two positions to the A21 carboxyl oxygen, in the other to A17 glutamic acid carboxyl oxygen.

A third important connection, different in the two molecules, links together the N terminal A chains and C terminal B chain structures. In molecule 1, A1 NH_3^+ is in ionic contact with B30 COO^- at 2.8 Å. In molecule 2 the conformation is very different; glutamic acid A4.2 COO^- here hydrogen bonds with the NH of B29.2 lysine, which permits the lysine NH_3^+ itself to interact with the terminal carboxyl in another hexamer. It is interesting that only the initial B chain residues are free of interchain links, and can move away from the body of the molecule. They adopt different conformations in different insulin crystal structures (Cutfield *et al.* 1981). In 4Zn insulin (Cutfield *et al.* 1981) and in the monomeric despentapeptide insulin where the molecules undergo considerable rearrangement, the H bonds between B4 and A11 and B25 and A19 are not formed (Bi *et al.* 1984).

6. The formation of the dimer

Seen in isolation, the insulin molecule has one obviously flat surface studded with aromatic and aliphatic residues. In the dimer as found in this crystal, molecules 1 and 2 are rotated relative to the twofold axis d_0 parallel with this surface and packed face to face. The forces leading to aggregation are clearly predominantly non-polar, reinforced and given direction by hydrogen bonds. Under these influences the B chain helices pack together at about $-40°$ and the extended chain at B24–B26 forms a β sheet.

6.1. *Non-polar contacts*

Figure 6.1 shows different views of the non polar surfaces in molecules 1 and 2. Van der Waals radii, drawn in at 3.5 Å, demonstrate the close packed arrangement of the atoms. In figure 6.2*a*, the dimer is shown as a whole with the residues in contact projected onto planes in each molecule defined by their mean positions and viewed normal to the twofold axis. All but one of the contacts involve B chain residues, B12 valine and B16 tyrosine in the α helix and B23 glycine, B24 and B25 phenylalanine, B26 tyrosine and B28 proline in the extended chain. B27 threonine loosely makes contact with the amide of A21 asparagine. B8 glycine and B9 serine (Cβ) fit in between B13 glutamic acid and B16 tyrosine; a view of the residues only down the threefold axis is shown in figure 6.2*b*.

The packing of the aromatic rings in the interface is particularly interesting, reminiscent of that found in some aromatic molecule crystals though necessarily more irregular. The six residues, illustrated in figure 6.2, two B16 tyrosines, two B24 phenylalanines and two B26

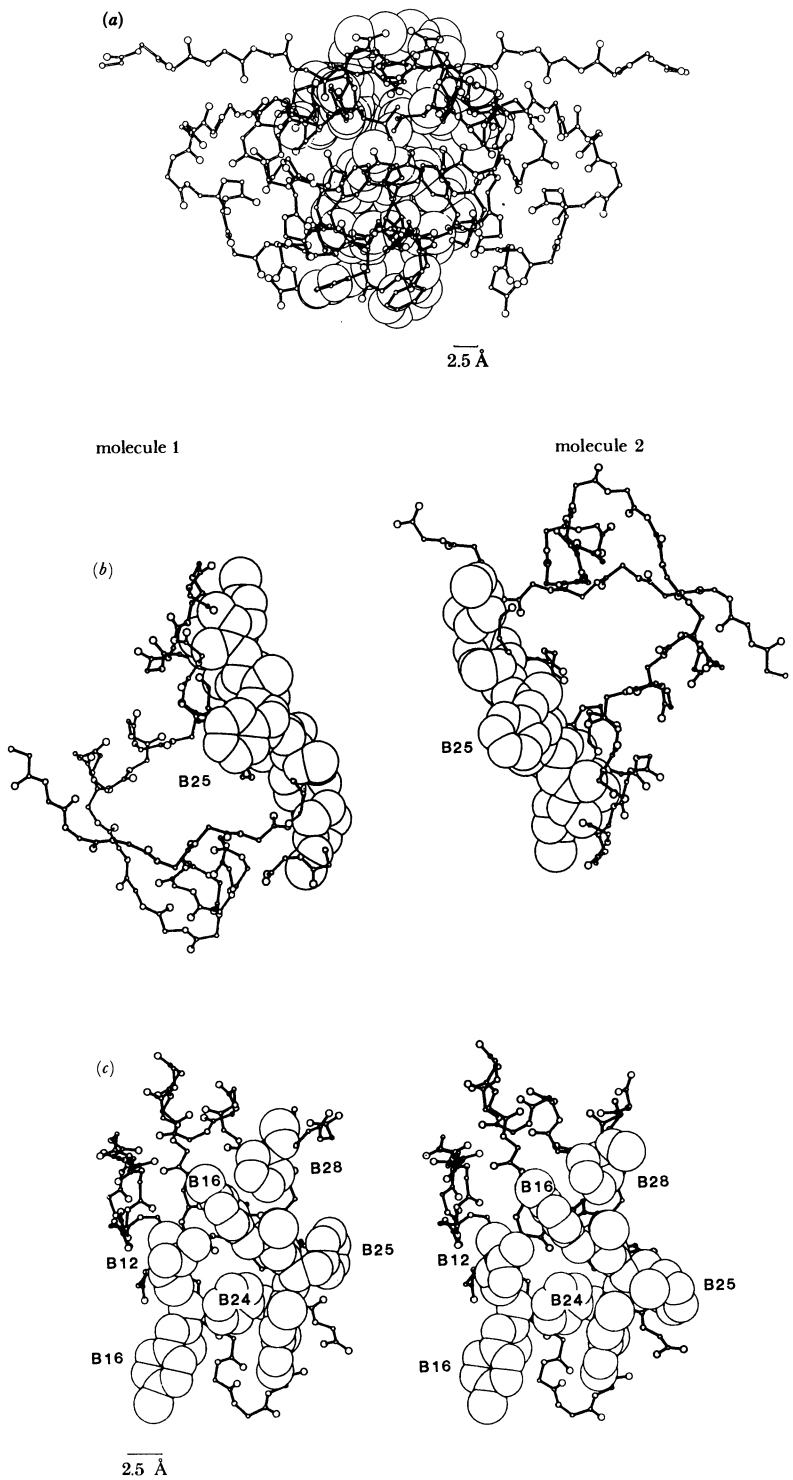


FIGURE 6.1. The dimer-forming surfaces drawn with van der Waals radii; only the structure of the main chains is shown for the remainder of the molecule. In (*a*) the view is in the direction of the threefold axis, (*b*) along the twofold axis, and (*c*) perpendicular to the twofold axis.

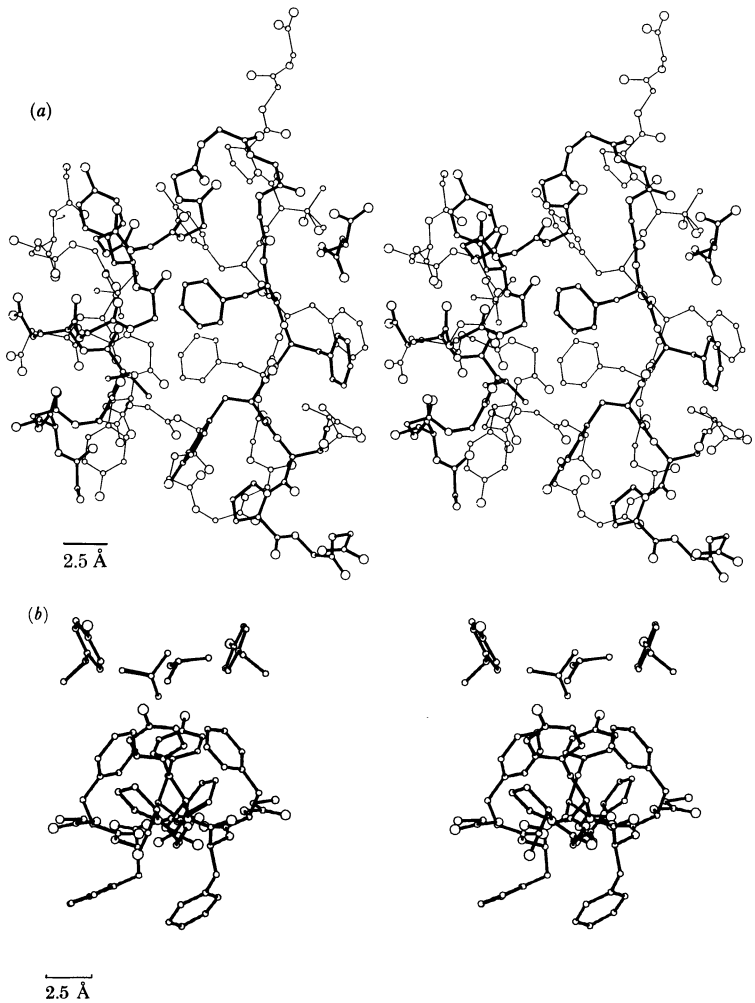


FIGURE 6.2. Stereo views of the dimer contacts seen (*a*) perpendicular to the twofold axis and (*b*) down the threefold axis; only side chain shown.

tyrosines are linked together by a network of loose contacts between 4 and 4.5 Å long. A second smaller group of three aromatic residues is formed by the two B25 phenylalanines of which B25.1 is in contact with A19.1.

On the whole, the positions of the atoms in the dimer surface are very well defined with low *B* values and well related by the twofold axis.

The main-chain structures agree within a r.m.s. discrepancy of only 0.14 Å. Separate comparison of the α helix and β sheet structure with their equivalents indicated that there is a small relative shift of 0.3 Å between them (Chothia *et al.* 1983). The twofold symmetry is broken in two places by two residues, B25.2 and B12.2. The B25.2 phenylalanyl side chain is rotated around the bond Cα–Cβ to a position lying across the twofold axis itself where it is in contact with B25.1. There is, however, evidence of disorder in the electron-density maps (p. 427) that suggests that the phenyl group sometimes occupies the symmetrical position, with B25.2 in contact with A19.2 tyrosine (p. 424). B12.2 valine is also disordered with two equally populated conformers. One conformation is symmetrically related to B12.1; the other, formed by *ca.* 100° rotation about the Cα–Cβ bond, brings CG into contact with B12.1. The residues fill in the cavity formed at the centre of the aromatic cluster.

6.2. *Hydrogen bonds*

The relative positions of the extended chains in the dimer surface are secured by the small antiparallel β sheet of hydrogen bonds involving the B24 and B26 residues. This sheet is extended less regularly to the A chain by B23 CO and B25 NH. The projections in figure 6.3*a*, *b* show the extent and twist, very small, of the sheet between the antiparallel B chain residues B23–B28.

There is also a close contact between carboxyl oxygen ions in the two B13 residues that suggests they are hydrogen bonded. This is discussed in the section on the hexamer.

6.3. *The stability of the dimer*

The insulin dimer is stable in aqueous solutions between pH 2–8. Its stability constant is about 2×10^5 (Jeffrey & Coates 1966; Goldman & Carpenter 1971). The area of the monomer

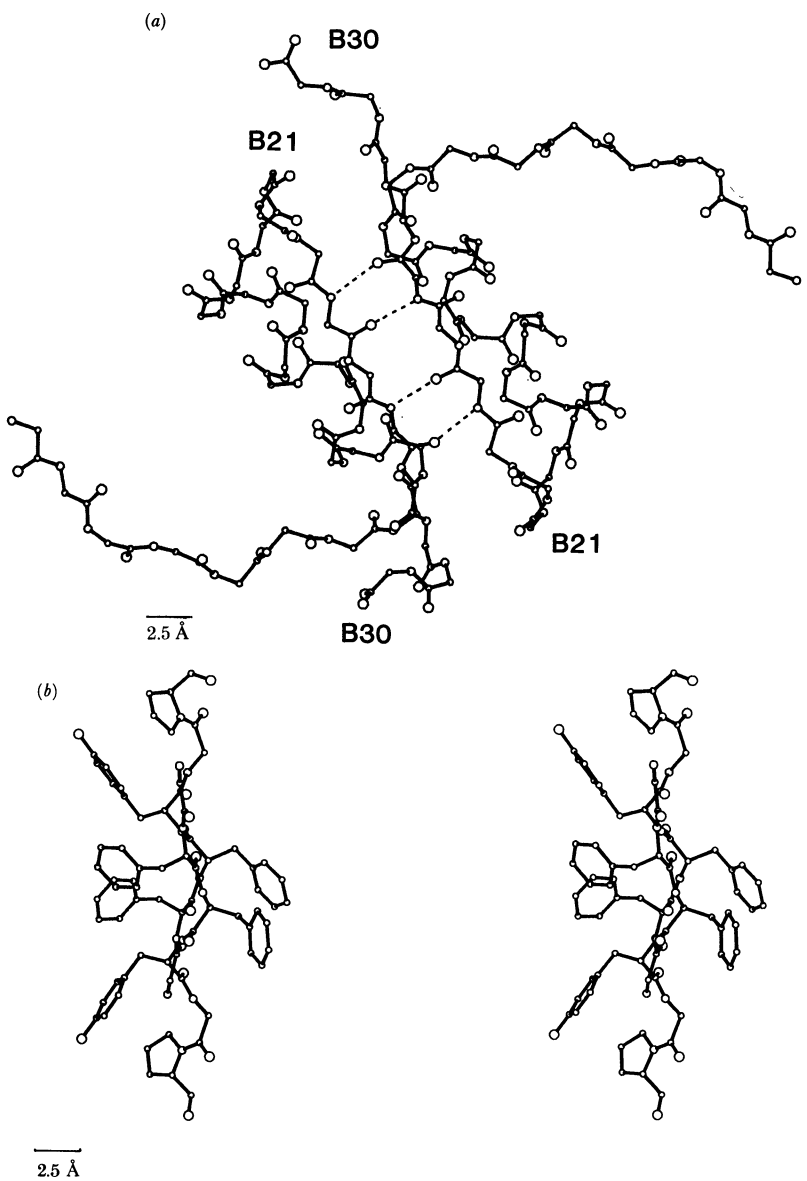


FIGURE 6.3. The antiparallel β sheet H-bonding structure formed within the dimer. (*a*) The view down the local twofold axis with the H bonds shown. (*b*) A stereo view perpendicular to the twofold axis illustrating the packing of the aromatic side chains.

surface excluded from contact with water by dimer formation has been calculated by Chothia & Janin (1975) to be 1100 Å^2, sufficient to provide the favourable entropic increase necessary to drive the assembly process. The buried surface is itself approximately square with sides about 23 Å.

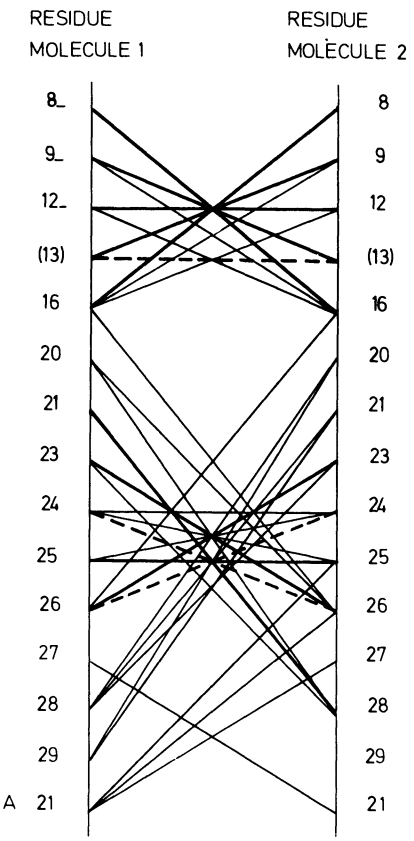


FIGURE 6.4. A schematic view of the contacts betwen the residues within the dimer. Broken lines indicate H bonds, thick lines non-bonded contact distances of 3.5 Å, or less and thin lines non-bonded contact distances 4.0 Å or less. The brackets about residue B13 glutamic acid indicate that this possible H-bond contact is generated within the dimer by hexamer packing.

The removal of B26–B30 in despentapeptide insulin does not significantly alter the rest of the molecule (Bi *et al.* 1983). This modification abolishes the dimerization of the molecules by removing the residue B26 essential for hydrogen bonding, together with approximately a third of the dimerizing surface. Both H bonding and the non-polar contacts are obviously important in stabilizing the dimer.

Contacts between atoms within the dimer are represented schematically in figure 6.4.

7. THE FORMATION OF THE HEXAMER

7.1. *The zinc coordination*

In aqueous solution between pH 5 and pH 8 insulin aggregates in the presence of zinc ions to form hexamers with a zinc binding constant of *ca.* 10^6 M^{-1} (Goldman & Carpenter 1976). In the 2zinc crystals, three dimers are assembled around two zinc ions, 15.82 Å apart on the

threefold axis as illustrated in figure 7.1. Each zinc (figure 7.2) is coordinated to three symmetry-related NE atoms of histidine B10, at 2.05 and 2.05 Å, and to three water molecules, at 2.36 and 2.21 Å, in molecules 1 and 2 respectively. The coordination is in each case a rather irregular octahedron, more so in 1 than 2.

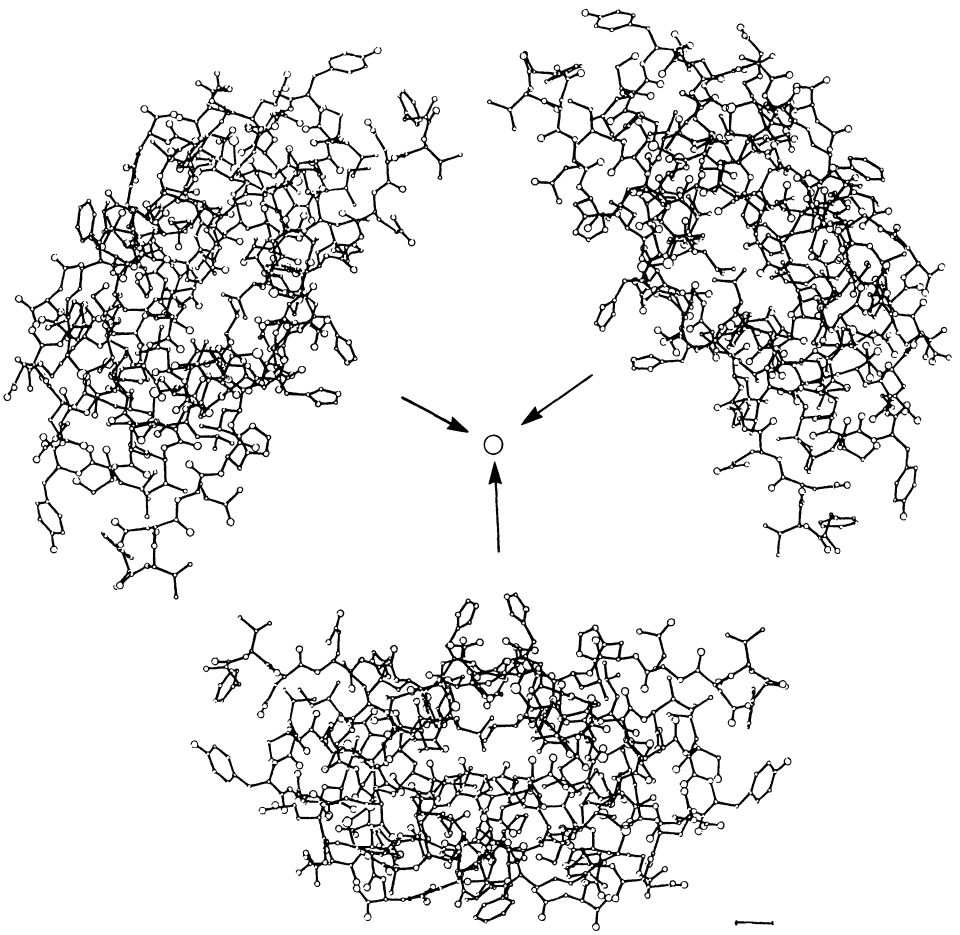


FIGURE 7.1. The construction of the hexamer from three dimers.

The coordination geometry and bond lengths are similar to those generally found in zinc complexes, e.g. in zinc *p*-toluene sulphonate and zinc benzene sulphonate hexahydrates, octahedral Zn–O, 2.05–2.14 Å (Hargreaves 1957; Broomhead & Nicol 1948). In the tetrahedral complexes of di(histidino) pentahydrate (Harding & Cole 1962) and di(L-histidino)zinc(II) dihydrate Zn–N varies from 2.00 to 2.05 Å. The double positive charge on the zinc in these complexes is balanced by separate negative ions in the hexahydrates and by the negatively charged carboxyl groups within the histidine complexes Zn...O, 2.8–2.9 Å. We have failed to find groups of peaks in the electron density maps, that could constitute any of the ions most likely to be present in our crystals although there is some evidence for citrate ions (p. 438). It seems most probable that here the zinc positive charges are balanced by the negatively charged glutamic acid B13 carboxyl groups that lie in a layer midway between them. The histidine B10 ND of one molecule is linked by a hydrogen-bonded water molecule,

65.2 or 6.1, to OE1 glutamic acid B13 of the other molecule of the dimer as shown in figure 7.3, suggesting lines for charge transfer.

If all the zinc glutamate B13 residues are negatively charged, the proposed arrangement would give two net negative charges. These could be balanced by a positive ion, such, for example, as the sodium ion found in the divalent complex ferroverdin, where sodium rests on

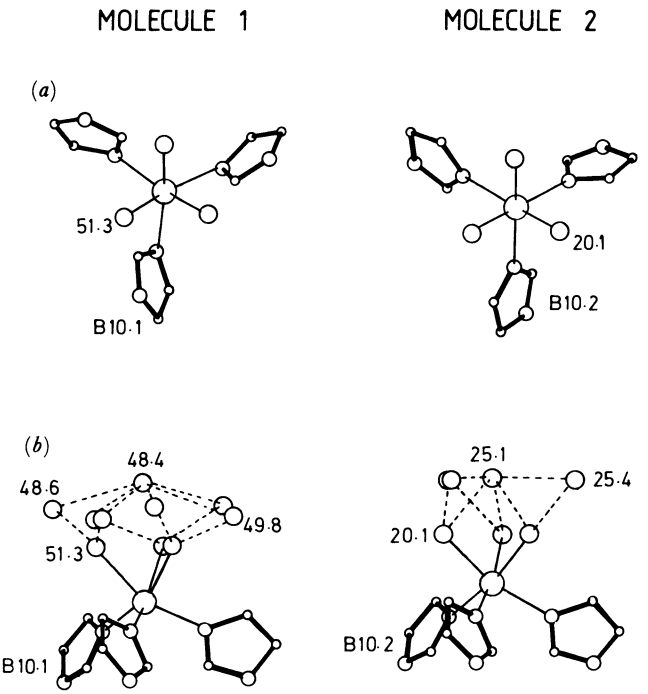


FIGURE 7.2. The Zn^{II} coordinating groups (*a*) down the threefold axis and (*b*) perpendicular to the threefold axis. The water molecules in contact with those coordinated to the Zn^{II} ions are also shown.

a triangle of oxygen atoms attached to iron (Candeloro *et al.* 1973). There are two peaks in similar positions, 2.1 and 2.3 Å from the water molecules attached to zinc in our crystals. But neither has the octahedral or tetrahedral surroundings expected for a sodium ion and it seems they are most likely water molecules. Here a different possibility exists suggested by the actual arrangement of the glutamate residues. These form a circle in which pairs of adjacent oxygen atoms are at 2.6 Å from one another, suggesting the presence of a hydrogen bond (figure 7.3), i.e. that only three of the six residues are negatively charged. In this case we should seek a single additional negative ion, to balance the zinc, perhaps an OH^- ion, which could appear as one of the peaks on the threefold axis.

There are differences connected with the two zinc positions that may result from the history of the crystals. The small distortions from ideal geometry in the water molecules attached to zinc and their surroundings are most probably a consequence of the packing of the hexamers in the crystals and changes in the protein conformation in their neighbourhood; B7 CO, for example, is more accessible to water in 2 than 1 and affects the solvation pattern. Between the two zinc ions the distribution of water is not symmetrical above and below the zero plane. The asymmetry here may result from the sequential introduction of zinc ions in the formation of

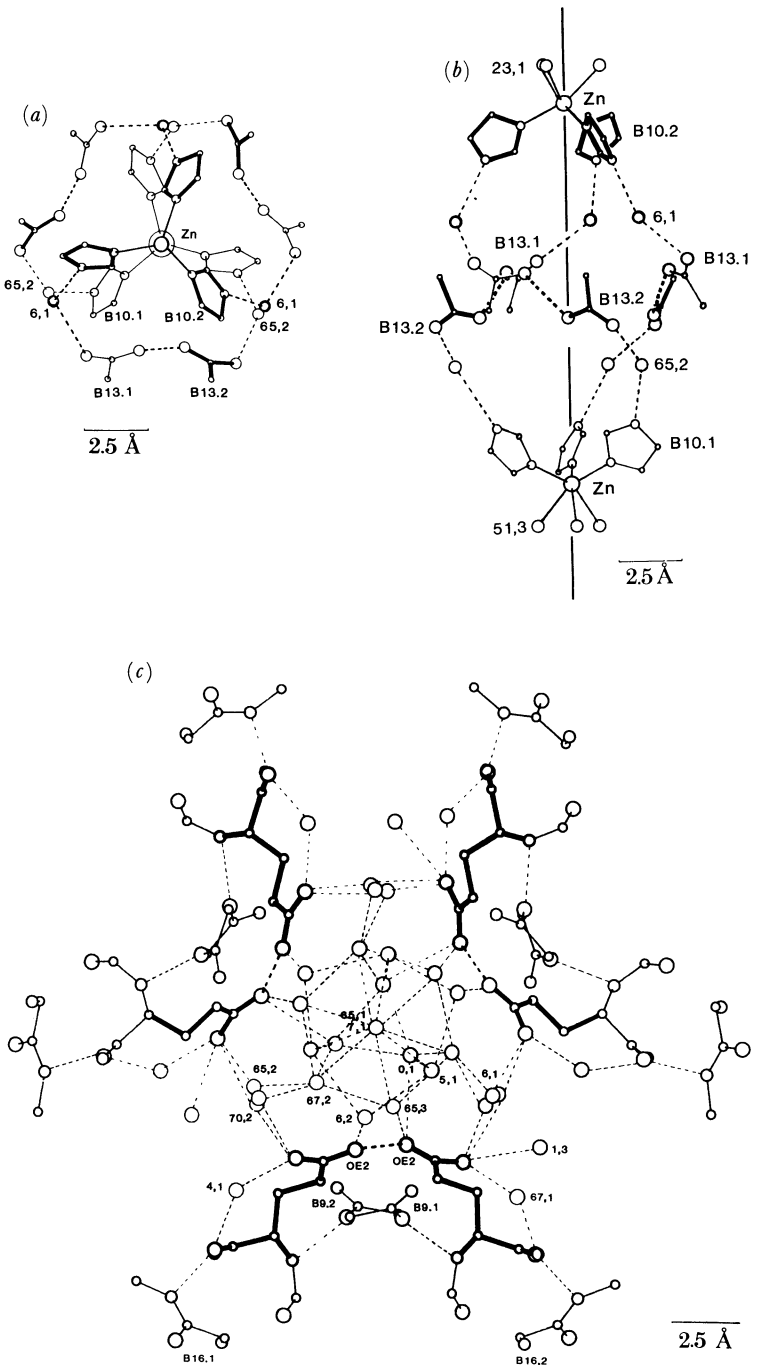


FIGURE 7.3. The water molecules H bonded between B10 ND1 and B13 Glu OE1 are shown (*a*) in a view down the threefold axis and (*b*) in a view perpendicular to the threefold axis, (*c*) the central region of the 2Zn insulin hexamer bounded by the 6 B13 glutamic acid groups. The H-bond contacts of the water molecules are indicated by dotted lines. Note the contacts (2.6 Å) between the OE1 atoms of adjacent B13 glutamic acids. The water molecules 7, 1 and 65, 1 lie superimposed on the threefold axis and are not labelled.

the hexamer in solution, first suggested by the observations of Brill & Venables (1967) on differences in the magnetic orientation of the two cupric ions in cupric insulin. Recent enthalpy measurements (J. Hansen, personal communication) indicate that the two zincs bind to the hexamer with very different enthalpies and entropies. The overall free energy of binding for the two metal ions, however, is similar (*ca.* 34 kJ mol^{-1}).

The partly ionic character of the zinc complex when formed is emphasized by the fact that zinc can easily be removed from the crystals when they are placed in solutions of EDTA. The zinc passes out of the water channels leaving the hexamers behind with the crystals preserved although a little disorganized, cloudy and cracked but still giving good diffraction effects. It can then be replaced by cadmium and lead, by soaking in cadmium chloride and lead acetate solutions respectively.

7.2. *The hexamer centre*

As shown in figure 7.3 *c* and discussed above, the centre of the hexamer is surrounded by six glutamic acid B13 residues brought together in a circle. They are arranged in pairs, the two within each dimer being in direct contact at 2.6 Å. The arrangement suggests a H bond between them, corresponding with the appearance of low positive electron density at the bond centres. The situation might be similar to that found between aspartic acid carboxyl groups in the aspartic proteinases (Blundell & Pearl 1984). However, at the pH of the insulin crystals, all the carboxyl groups would be expected to be fully ionized. The observed contact, though short between negative ions, is not impossible, however, given the large standard deviations of their positions; these and their relatively low density may be partly a product of imperfect resolution. The second pair of carboxyl oxygens, between the dimers, are over 5 Å apart. There is weak spreading water density between them.

Both pairs of carboxyl oxygen ions are, in addition, hydrogen bonded to well-defined water molecules in the central cavity, two of which are linked in each case to ND histidine B10 and serine B9 OH. The network is illustrated in figure 7.3*b* and discussed in detail in §9. Less well defined water molecules fill gaps between those linked to protein. When the crystals are placed in suitable solutions the cavity can be penetrated by divalent ions calcium, lead, cadmium, zinc, samarium, ytterbium and uranyl, which take up definite positions in relation to the surrounding carboxyl ions, displacing water molecules or perhaps other ions such as sodium if present.

7.3. *The protein contacts in hexamer formation*

When the hexamer is assembled through the zinc coordination, both polar and non-polar residues are buried between the dimers. The packing is correspondingly much looser than in the molecular interface within each dimer; it is also very intricate (figure 7.4, plate 3).

Between A11 and B4, the A and B chains of each molecule are hydrogen bonded together. From B4 to B1 the B chains swing away from their attached chains, crossing at B5 when viewed down ***c*** and making close contacts with the A chains in the neighbouring molecules. The initial residue of each B chain, B1 phenylalanine, is fitted into a pocket between the main A chain and A14 tyrosine residue of its neighbour. Because the two A14 residues are in contact, another close aggregate, here of four aromatic residues, is formed. In addition, the connection is strengthened by ionic attraction between the terminal NH_3^+ B1 and A17 glutamic acid carboxyl

ion COO^-. The B2 valine on each chain is also turned in towards the neighbouring A chain, making rather loose contacts, mainly with part of the peptide chain.

The remaining non-polar residues of the A chain and B chain helix approach one another. The two A13 leucines are in contact at 4.3 Å, the B17 leucines touch one another at 3.5 Å and each touches alanine B14 of the opposite molecule at 4–4.4 Å. These are part of the circle of non-polar aliphatic residues running through the hexamer illustrated in figure 7.5. There are water molecules penetrating the interface that make contact with the loosely extended polar residues, asparagine B3 and serine A12.

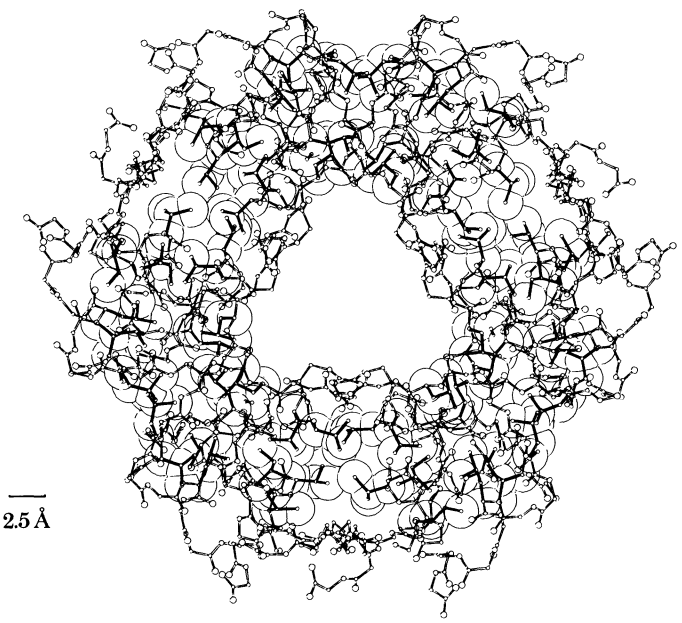


Figure 7.5. The ring of aliphatic contacts made within the hexamer.

Near the threefold axis, hexamer formation is reinforced by direct contacts of 3.3 Å, possibly H bonded, between histidine B10 ND in each molecule and serine B9 OH belonging to the threefold symmetry related molecule, i.e. (100) to (200) and (300) (p. 402).

Throughout this region the twofold symmetry is obeyed only approximately. Although the main chains are well related, similar residues often differ markedly from one another in side-chain conformation. B values also are higher than in the interior of the molecules.

7.4. *The framework of the hexamer*

As a whole the hexamer is similar in dimension to many enzymes and has an overall packing of secondary structure reminiscent of some of them; around a central channel helices are packed at angles of $-40°$ monomer to monomer and $-80°$ dimer to dimer to one another (Chothia *et al.* 1981). They are connected with extended β chains running antiparallel in pairs on the outside of the complex (figure 7.6). The comparable enzyme structures usually have chains running in antiparallel twisted β sheets surrounding the central channel with helices packed on the outside. However, the plant protein trichosantin has the helices around the centre and β chains outside, as in insulin (Pan Kezheu *et al.* 1985).

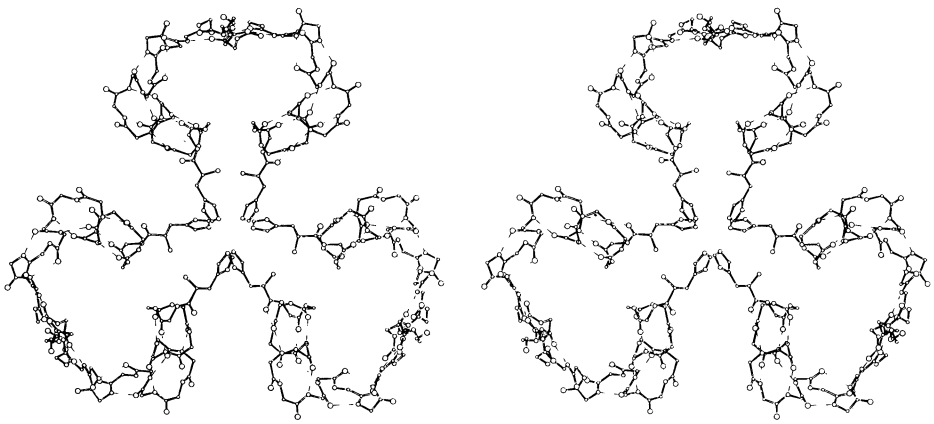


FIGURE 7.6. A stereo view down the threefold axis showing the packing of the secondary structure in the B chains of the insulin hexamer.

8. THE PACKING OF HEXAMERS IN THE CRYSTAL: THE DEPARTURE FROM TWOFOLD SYMMETRY

8.1. *Hexamer packing*

When models of hexamers derived from low-resolution electron-density maps are examined, the way they should pack together appears obvious. The three A chains from A8–A10 project above the hexamer surface leaving gaps into which the projecting A chains of the lower surface will clearly fit. Stacked chains of hexamers may then be packed in essentially hexagonal rod closed packing. The actual crystal structure conforms approximately to this picture (figures 2.1 and 2.2).

There seems to be nothing in the interactions through which the dimer and hexamer are formed that requires more than minor departures from 32 symmetry. The observed departures appear to be introduced by the fact that the simple packing of projecting A chains, as described, brings the B5 histidine groups of succeeding molecules into direct collision along the NE–CE bonds near the $d'_{\frac{1}{2}}$ axis. To avoid this, B5.2 histidine is seen to rotate away to a new, conformationally stable, position (figure 8.1). In this it preserves contact with A7 CO through NE rather than ND. The change is correlated with a considerable movement of the A chain from its original position into contact through A9.2 CO with B5.1 NE of the colliding histidine. Histidine B5.1 as a consequence adopts a bridging position between two peptide chains in neighbouring molecules closely similar with that found recently in the crystal structure of a tetrapeptide derived from ACTH (Admiraal & Vos 1983). The movement of the A chain of molecule 2 required for these contacts is followed by further movements to make new hydrogen bonds; the symmetry-breaking rotation of the initial A chain helix described earlier. Nearby there is a loose but fairly symmetrical contact at 4.3 Å between the two A10 isoleucines in the adjacent hexamer surfaces (figure 8.1).

A further area of contact between the succeeding hexamers is marked by a network of ionic contacts formed initially on the surface of molecule 1 where the B terminal B30 alanine COO^- meets the initial A1.1 glycine NH_3^+, which is also in contact with A4.1 glutamic acid COO^- (figures 8.1 and 8.2). This region in addition is penetrated by the B29 lysine of molecule 2.

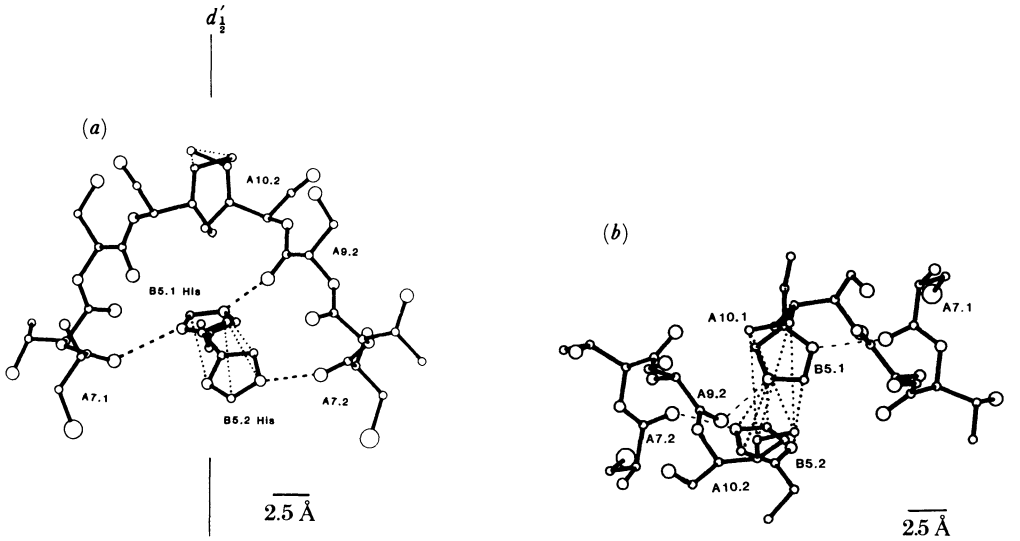


FIGURE 8.1. The B5 imidazole and A10 isoleucyl contacts between the hexamers as found in the crystal. Selected contact distances between 3.5 and 4.2 Å are shown by dotted lines. The view in (*a*) is down the threefold axis and in (*b*) perpendicular to the threefold axis and along the twofold axis.

Here the terminal residues B29, B30 have turned away by rotation about C28–C29 from their position in molecule 1, permitting the lysine chain to extend through the water filled gap between the hexamers. There is thus a continuous chain of close interactions, atom to atom, throughout the crystal in the direction of the trigonal axis.

As figure 2.1 shows, the hexamers project down the ***c*** axis as cylinders in a hexagonal array, one occupying each rhombohedral cell. These cylinders are in direct contact along the rhombohedral axes. The exact positions in the circumference at which they touch are illustrated in figure 8.2*a–c*. They appear to be determined by the existence of still unburied aromatic residues on the surface, A19 tyrosine, B25 phenylalanine and the A14.2 tyrosine (801) of the adjacent hexamer that are again grouped together; A14.2 tyrosine in turn links the system to the adjacent A14.1 tyrosine and the B1 phenylalanines, themselves in contact. There are further consequent interactions between the residues and the peptide chains, a hydrogen bond between B27.1 threonine OH and A18.2 asparagine amide NH2, and through arginine B22.2 (801) an ionized contact with B30.1 alanine COO^- extending the region described above. The surface over which contacts of 4.5 Å or less occur between hexamers is approximately circular, of area about 120 Å around the ***a*** axis.

8.2. *The departure from twofold symmetry*

The positions that the atoms adopt in the crystal structure appear throughout to be governed by the tendency for like atoms to pack together, aromatic residues to aromatic, aliphatic hydrocarbons to aliphatic hydrocarbons, ions to ions, modified by specific lines of attraction to maximize hydrogen bonding. There seems nothing to prevent essentially identical molecules associating to dimers and hexamers with only small modifications of residue conformation where they actually meet (e.g. at B12 valine or B17 leucine) in an interface between two molecules. The marked changes in the conformation of molecule 2 compared with molecule 1 appear to arise from the packing together of the B5 histidine residues when forming crystals.

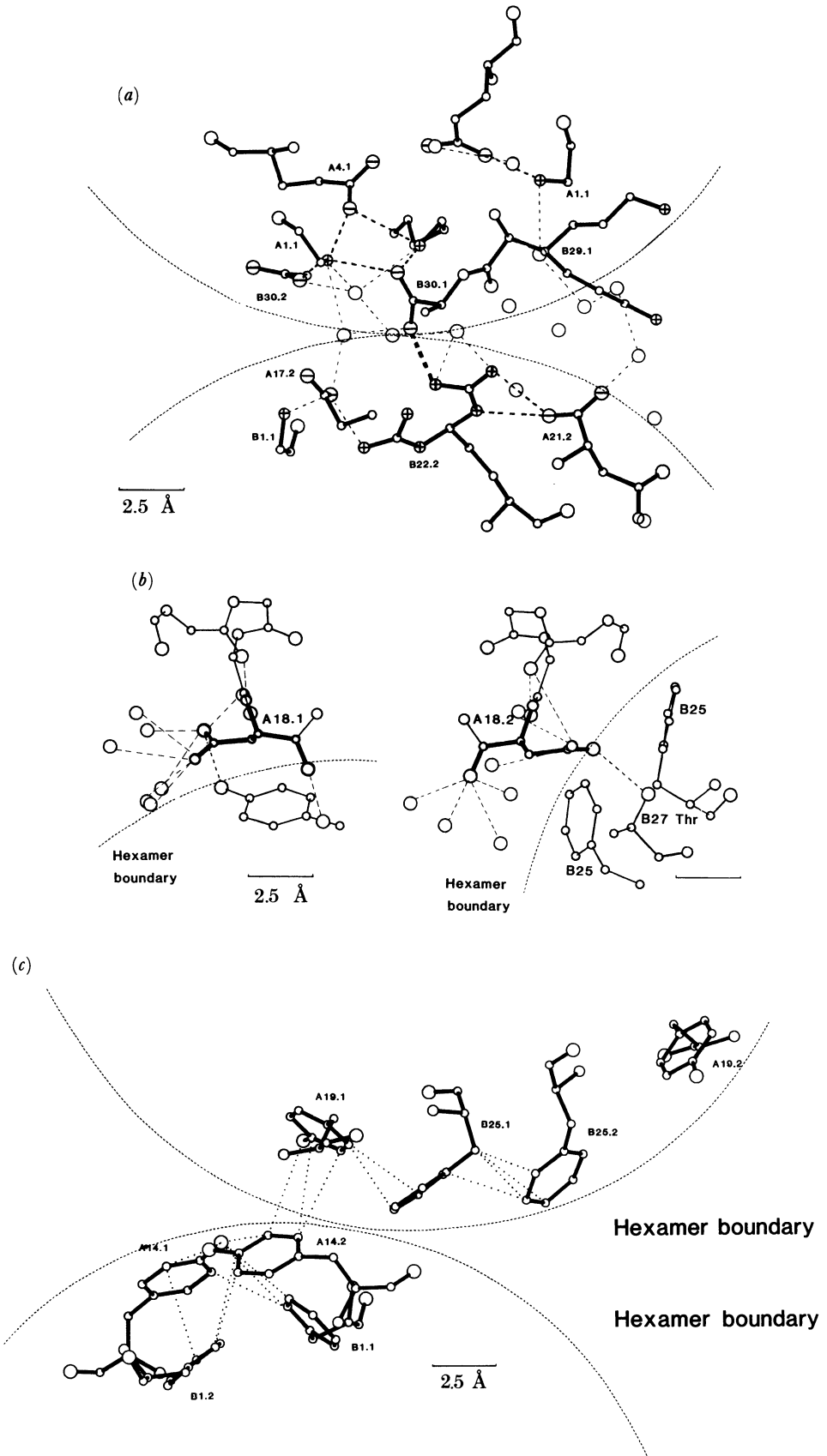


FIGURE 8.2. Three kinds of interhexamer contacts seen in the crystals. (*a*) Ionic interactions, (*b*) H bonding and (*c*) aromatic. Broken lines represent H bonds, dotted lines correspond to distances of less than 4.0 Å.

The adoption of such different conformations in contact is found in many histidine containing crystals (see, for example, Bhat & Vijayan 1978). In 2Zn insulin the change in conformation of histidine B5.2 breaks the original hydrogen-bonding pattern; the new one is established by a movement of 3 Å. The change is strengthened by the entry of histidine B5.1 into a hydrogen-bonded relation with A9.2 CO but at the same time other bonds are broken, e.g. A4.2 CO–A10.2 NH. A radical swing around A6.2 CA–N permits a new hydrogen-bond pattern to develop: 5.2 CO...9.2 NH among others. The initial A chain helix A1–A5 and its attached side chains is correspondingly carried as a whole into new positions where actually a more regular α helix is established from A5–A9. It is noticeable that the strongly hydrogen-bonded region A11...B4 remains unaltered.

The initial A chain movements appear to have consequences in other regions of molecule 2 (Chothia *et al.* 1983). A2.2 isoleucine moves more than 4 Å from its initial position. A5.2 glutamine moves 2 Å out from hydrogen-bonded contact with A19 OH into a new contact with A10 CO. A19.2 is accordingly free to move a little, packing against the differently positioned A2.2 isoleucine. This may be the reason for the change in the predominant conformation of B25.2 phenylalanine, which turns from contact with A19.2 to contact with B25.1. Again in the neighbourhood of A19 and B24, B26, strong hydrogen bonds hold most of the atoms of molecule 2 in their original and closely twofold related positions.

A further change occurs in the region occupied by B28.2 proline, B29.2 lysine and B30.2 alanine. In the new position of A4.2 glutamic acid COO^-, still firmly held by its salt bridge to NH_3^+ A1.2 glycine, one O^- would be in collision with B28.2 CO. Rotation of the group around B28.2 CA–C removes B28.2 O from close contact, *ca.* 1.5 Å, as illustrated in figure 8.3, and

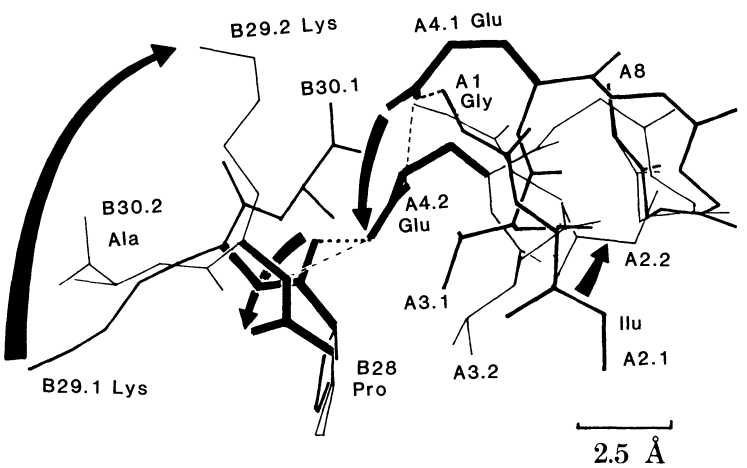


FIGURE 8.3. The residues A1–A8 and B28–B30 are shown overlapped after optimizing on residues A7–A9. Molecule 1 is drawn with thick lines and molecule 2 with thin lines; the sidechains A4 and the peptide B28–B30 are emphasized with very thick bonds. H bonds are represented as dashes and the impossible contact (*ca.* 1.5 Å) generated by movement of A4.2 towards B28–B29 peptide is shown by dots. Arrows indicate the movement of molecule 1 to molecule 2.

brings B29.2 NH instead into H-bonded contact with A4.2 glutamate OE2. This rotation breaks the contact of B30.2 COO^- with NH_3^+ A1 and sends the lysine chains out parallel with the trigonal axis to make the direct inter molecular contacts mentioned above between the lysine NH_3^+ and the carboxyl groups of B30.1 and A4.1 both at (101). These interactions

appear to stabilize the change in the crystal structure, although they leave the terminal B30.2 alanine residue essentially floating.

Some of the changes appear to depend on rather small precipitating movements; it is difficult to balance all the factors leading to the pattern we see. A19.2, for example, looks as if it could stay closer to the position found in molecule 1, when B25 would not need to move. Judging by the electron density this happens occasionally; see figure 8.4. Both A19.2 and B25.2 side chains have rather high B values as seen in these crystals.

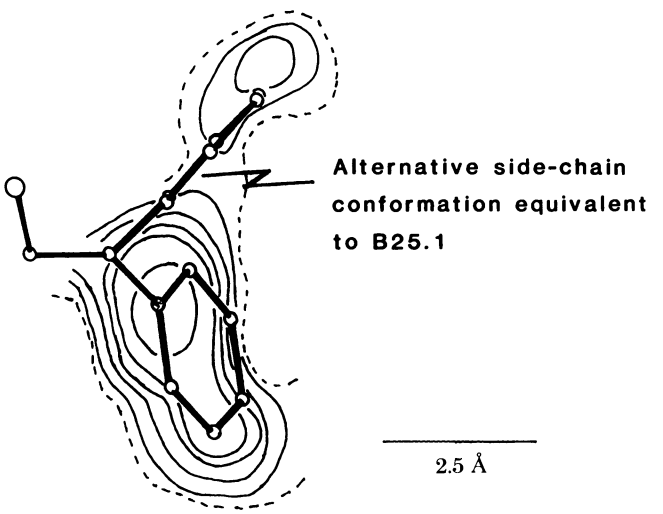


FIGURE 8.4. The electron density at B25.2 Phe side chain from the F_o Fourier synthesis on section $z = 0$. The broken contours represent 0.3 e/Å^3, the continuous contours are at 0.5, 0.6, 0.7 and 1.0, 1.3 e/Å^3 all on the absolute scale. An alternative conformation, roughly equivalent to that in molecule 1, can be seen to lie in low, but persistent, electron density.

9. THE DISTRIBUTION OF WATER MOLECULES IN THE CRYSTAL

9.1. *A general description*

The 2zinc insulin crystals contain 30.6% (by mass) of solvent, mainly water, a rather lower proportion than that found in most protein crystals but similar to that present in crambin and rubredoxin (Teeter 1984; Watenpaugh *et al.* 1978). In insulin the water is essentially continuous in different directions from one side of the crystal to the other but confined within irregular channels leading in and out of cavities, formed by the protein molecules. At their maxima the spaces filled with water are 14–16 Å across so that no water molecule is further than *ca.* 8 Å from protein atoms, which accordingly affect both its position and movement in varying degrees.

From the first electron-density maps, it was easy to see a number of well-defined peaks, corresponding to water molecules hydrogen bonded to the protein. These were refined and added to as the analysis proceeded. At a stage when almost all the protein structure had been essentially established, a map was calculated in the phasing for which all water-molecule contributions were excluded. The resulting difference map showed in a striking way the characteristics of the water-molecule distribution as a whole. A fringe of peaks marked the outline of the protein framework, many very strong where the channels were narrow, others

much weaker where long hydrophilic side chains extended into stretches of water. In contact with these, further peaks could be traced in an increasingly diffuse distribution. A number of these peaks were small, less than 2 Å apart, suggesting one water molecule occupying alternative sites, which could be weighted 0.5. Similar peaks in other situations were similarly weighted. In the following refinement calculations an attempt has been made to introduce water-molecule density up to the number of water molecules, 283, expected in the crystal asymmetric unit. This density is divided at present in table 1 between 217 whole molecule sites and 132 sites weighted 0.6 or less. As mentioned earlier, the occupancies and thermal parameters are probably only reliable for molecules weighted 1.00 with $B = 50$ Å^2 or less. Higher B values will often indicate lower occupancy than recorded. Vice versa, low B values of half-weighted water molecules suggest higher occupancy. The electron-density distribution that appears following the refinement illustrates the more complicated, character of the real situation. It has been tested in a number of calculations using a variety of coefficients: $|F_o|$, $|2F_o - F_c|$, $|F_c|$, $|F_o - F_c|$ with and without contributions from certain of the 'water molecules'. The calculations will be described elsewhere.

In the latest electron-density distribution calculated on the parameters of table 1.1 the water molecules appear as peaks varying in height from 2.5 to 0.4 e/Å^3. They are often linked by continuous threads of density *ca.* 0.4 e/Å in height. Between them the electron density falls off slowly, occasionally reaching 0.1 e/Å^3, more commonly 0.2 or 0.3 e/Å^3. The peaks can be seen to form chains, sometimes zig-zag, sometimes nearly straight, with interpeak distances from 2.5 to 3.0 Å, and also rings of varying sizes: three, four, five, six, seven or more membered. The individual peaks are often diffuse or oddly shaped, suggesting they do not represent single water-molecule sites, but a number of close alternative positions, the stopping places of water molecules moving through the crystal. The situation they suggest seems close to that given in water molecule simulation calculations by, for example, Stillinger & David (1978) and Hermans (1984).

Owing to the complex character of many of the peaks listed as individuals, statistics of water-molecule contacts have to be treated with caution. But some observations can be made. There are, first, a very few water molecules entirely surrounded by protein atoms, making no contacts less than 3.8 Å with other water molecules: 9.5, 11.1, 14.3, 46.1, 58.2, 61.2 and 64.1. Of these, 14.3 and 64.1 are 'half weight' substitutes for disordered residues, glutamine B4.2 and arginine B22.2; 58.2, near which glutamine B4.1 is not disordered, and 14.3 are twofold equivalents, 9.5 and 61.2 are also twofold equivalents, in pockets between glycine B20 and tyrosine B16; 11.1 and 46.1 are particularly well defined. All make good hydrogen-bonded contacts with active groups on the protein; 46.1 makes 4, one of the highest observed. There are, on the other hand, 91 water molecule sites entirely surrounded by other water molecules, making no protein contacts less than 3.8 Å. Some of these in the central cavity of the hexamer are well defined. The large majority of water molecules are in contact both with protein atoms and other water molecules, as shown in table 9.1. This is calculated only for contacts under 3.8 Å and over 2.3 Å, half-molecule contacts counted as $\frac{1}{2}$, totals rounded up.

Figure 9.1 *a*, *b* shows histograms of the number of atoms surrounding whole water molecule positions at distances (i) less than 3.25 Å and (ii) less than 3.8 Å, half-molecule contacts being counted as $\frac{1}{2}$. In 9.1 *a* (ii) the maximum of the distribution is at 5–6 water molecules or protein atoms within 3.8 Å of each water molecule; there are a few water molecules that make many more contacts, often near protein side chains. In 9.1 *a* (i) only potentially hydrogen-bonded

TABLE 9.1. WATER CONTACTS

contacts to proteins	contacts to other waters: 0	1	2	3	4	5	6	7	8	9	10	total
0	—	—	2	1	15	25	19	17	7	4	1	91
1	—	1	2	5	9	14	13	12	—	4	—	60
2	—	1	6	5	8	13	8	2	1	1	—	45
3	—	—	1	10	11	8	7	4	1	2	1	45
4	—	—	—	7	6	5	3	3	—	—	—	24
5	—	—	5	13	3	6	—	—	—	—	—	27
6	2	3	2	4	4	3	1	—	—	—	—	19
7	1	2	2	2	3	—	—	—	—	—	—	11
8	1	2	—	2	1	—	—	—	—	—	—	6
9	—	3	2	2	—	—	—	—	—	—	—	7
10	—	—	1	—	—	—	—	—	—	—	—	1
11	—	—	—	—	—	—	—	—	—	—	—	—
12	2	1	—	—	—	—	—	—	—	—	—	3
13	—	1	—	—	—	—	—	—	—	—	—	1
14	—	—	—	—	—	—	—	—	—	—	—	—
15	1	—	—	—	—	—	—	—	—	—	—	1
16	—	1	—	—	—	—	—	—	—	—	—	1
total	7	15	23	51	60	74	51	38	9	11	2	342

This table lists the number of waters making contacts ($d \leqslant 3.8$ Å). Across, to other waters (0–10) and down, the number of contacts by water to protein.

atoms were included; the maximum here is much sharper, at 3 contacts per water molecule. The distribution of angles at water molecules in the two sets illustrates further differences between them. Figure 9.1*b* (ii) shows two main maxima, one around 60–80° and the other, as expected, around 110°. The first maximum probably indicates a tendency to close packing among molecules not specifically hydrogen bonded to one another which gives an appearance of three-membered rings. In figure 9.1*b* (i), of potentially hydrogen-bonded molecules, the maximum at 110° is stronger and that around 60° relatively less significant, although there are a few positions, e.g. around 000, where water molecules do appear to be hydrogen bonded in three-membered rings.

We can distinguish a number of regions in the crystal where the water structure is worth separate discussion. Around the centre of the hexamer, 000, there is (1) a small confined group of water molecules in a cavity. These are linked (2) by two single chains of water molecules running past the histidine B10 residues above and below the hexamer centres to join an extensive pool of water (3) around $00\frac{1}{2}$ between hexamers succeeding one another along ***c***. The pool is linked by two streams (4) two or more molecules across, and centred on levels $\frac{1}{3}$ and $\frac{2}{3}$ *c*, to the large volume of water (5) separating neighbouring hexamers. This water connects with water channels along the two threefold screw axes (6) which are continuous throughout the crystal parallel with ***c*** (compare figures 9.7 and 2.1*a*, *b*).

9.2. *The central cavity*

The central cavity is approximately spherical, about 10 Å across, bounded on its circumference (viewed along ***c***) by the carboxyl groups of glutamic acid B13, 1 and 2, and above and below, $z = 0$, by the two B10 histidine NDs and B9 serine OGs. It is illustrated in figure 9.2, which shows 24 water molecule sites, and in figure 9.7 (plate 4), which contains a

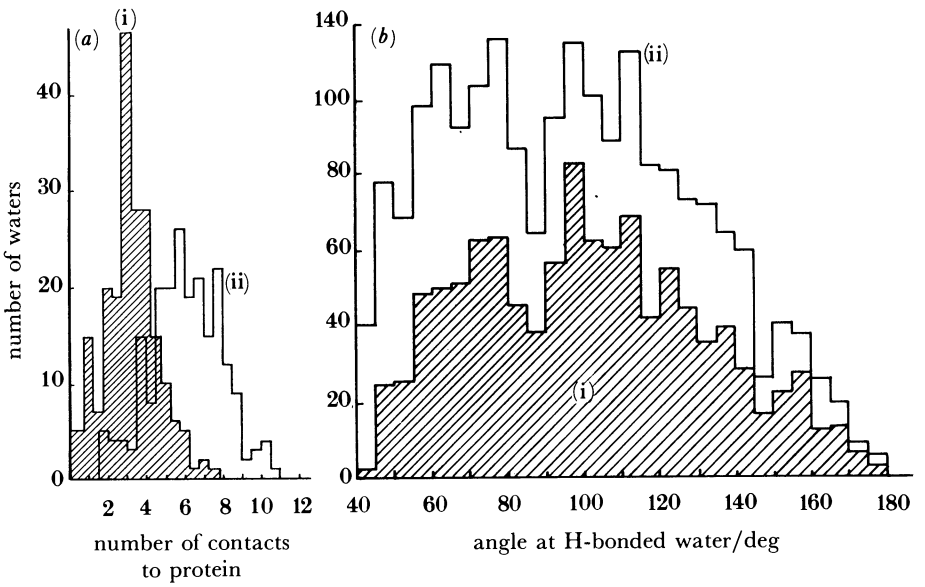


FIGURE 9.1. Histograms of the distribution of water-molecule contacts. The shaded regions (i) correspond to contact distances less than 3.25 Å and the outlines (ii) to distances less than 3.8 Å. In (*a*) the number of contacts to protein and water atoms are shown. In (*b*) the distribution of angles at the water molecules ($d < 3.25$ Å) is illustrated. Maxima at *ca.* 60° and *ca.* 100° can be seen; that at 60° is accentuated when the contact distances are extended to $d < 3.8$ Å.

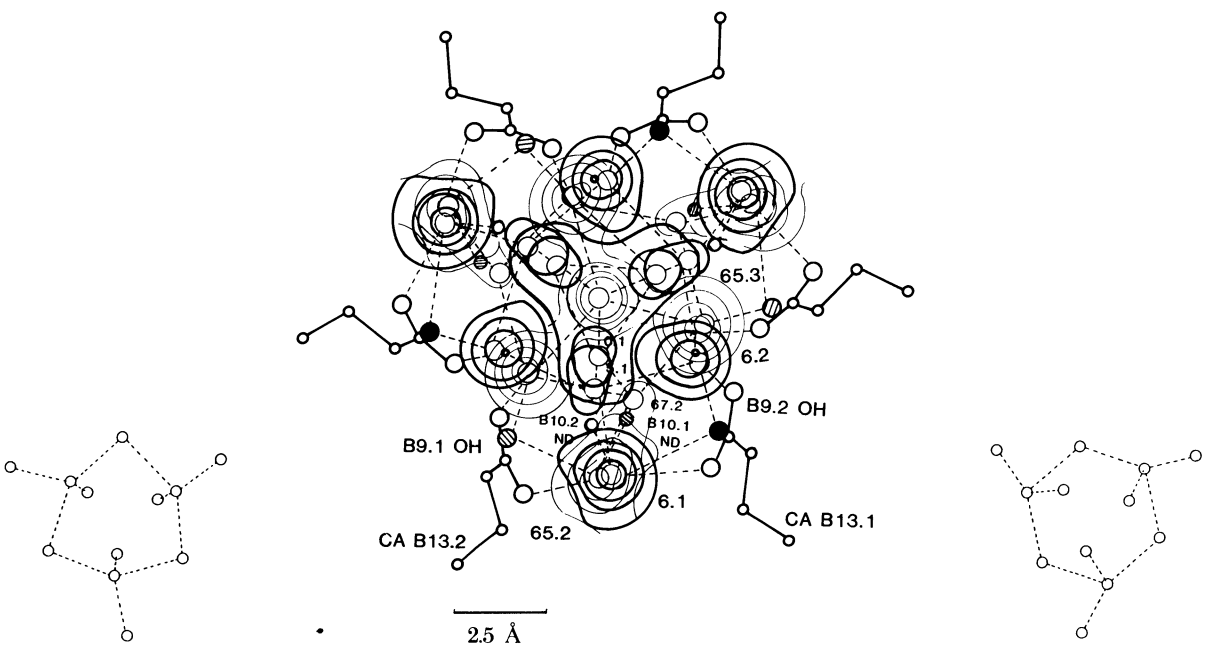


FIGURE 9.2. Distribution of water molecules in the central cavity around 000. The contours over the centre of each molecule are drawn at intervals of 0.3 e/Å^3, starting at 0.7 ē/Å^3. Strong contours over molecules above $z = 0$, fine contours, below $z = 0$. Filled circles mark histidine ND and serine OG above, and hatched circles, below $z = 0$; small diagrams show partial water-molecule arrangements above (left) and below (right) $z = 0$, illustrating the similarity to ice (I) about $z = 0$.

view of the cavity perpendicular to the threefold axis. This latter figure also shows the networks of water molecules up the threefold axis connecting the cavity to other solvent volumes. Thirteen very well defined water molecules (B value 24–35 Å^2) (figure 9.2) lie on the central cavity's periphery; twelve of these, 6.1, 6.2, 65.2, 65.3 each make two hydrogen bonds, one to carboxyl oxygen, one to histidine N or serine OH. The thirteenth lies on the three fold axis, hydrogen bonded to 65.3. The twofold-related position above $z = 0$, 7.2, is comparatively low in electron density, ($B = 76$ Å^2). Above $z = 0$ with the omission of 7.1, (right), the arrangement appears to be ice like; below (left), the distortions produce a different pattern. There is further low density near $z = 0$ suggesting water between the two carboxyl oxygens, related by d'.

Divalent ions, particularly lead, uranyl, cadmium, zinc, samarium, ytterbium, and calcium readily enter the cavity when insulin crystals are soaked in appropriate solutions. They occupy positions at approximately $x = 0.35$, $y = 0.31$, $z = 0$ displacing the water 0.1 (very near Zn and Cd sites), 5.1 and 67.2, and making contact with the glutamic acid oxygen ions. They enter presumably via the water-molecule chains attached to 65.2 and 6.1.

9.3. *Chains of water molecules linking those of the central cavity to those of the pool at* $z = \frac{1}{2}$

Very well defined chains of water molecules can be traced as in figure 9.3*a*, *b* running approximately from the water molecules 65.2 and 6.1 of the central cavity to those on the edge of the pool, 51.5, 53.3 and 21.1, 20.4. The chains fill the gap between B chain atoms of neighbouring molecules related by the three fold rotation, particularly the atoms between B9 and B10.

In the centre of the group are two molecules 60.1–60.3 and 11.3 which are linked by continuous electron density parallel with the **c** axis to 56.1 and 17.2 respectively. Each is hydrogen bonded to three other very well defined water molecules, attached to protein oxygen

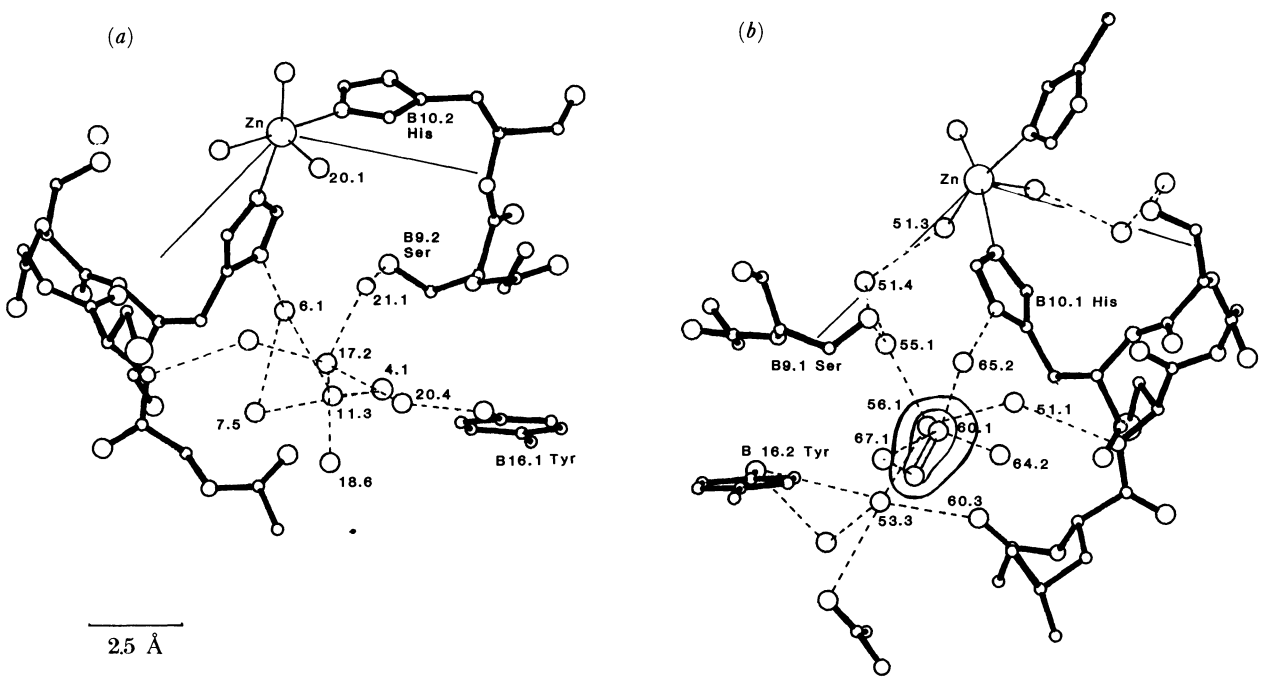


FIGURE 9.3. Short chains of water molecules linking the central cavity around $z = 0$ to the pool around $z = \frac{1}{2}$. Contours of electron density at 0.7 and 1.0 e/Å^3 over 60.1–60.3 (*a*) above $z = 0$, (*b*) below $z = 0$.

or nitrogen. The central peaks are distorted, suggesting they represent close alternative positions, which have been assigned for 60.1 and 60.3 at different weights, 0.6 and 0.4, contributing one molecule, whereas 11.3, though elongated, is recorded as a single site. The distortions are probably determined by the attached water molecules, 65.2, 64.2, 67.1 around 60.1–60.3, and 4.1, 7.5 and 6.1 around 11.3. Around molecules 56.1 and 17.2 the packing is easier. Very good contacts are made with 55.1, 51.1, 53.3 and 16.1, 21.1 and 20.4 respectively. Lines of marked density *ca.* 0.5 e/Å^3 can be traced along some of the hydrogen bonded links involved here. There are less well marked connections with the water molecules attached to zinc on one side and to the tyrosine B16 OH on the other.

In this region substitution of water by *o*-chlormercury benzaldehyde occurred to a very small extent, no doubt owing to the restricted volume involved and difficulty of access.

9.4. *The pool of water around* $00\frac{1}{2}$

Between the two hexamers succeeding one another along the trigonal axis there is a large, roughly spherical, volume of water of diameter 16–18 Å bounded by the two zinc ions and residues A7–10 and B5–7 of the insulin molecules. On its surface are a number of well-defined water molecules attached to B5.2 histidine (30.1), the peptide nitrogen and oxygen atoms and the three water molecules attached to each zinc ion. Only slightly less well marked is a group of water molecules, two in contact with each of the four sulphur atoms of B7 and A7 cystine 1 and 2, at distances between 3.4 and 3.7 Å suggesting possible hydrogen bonds. Several others make slightly longer contacts of 3.9 Å. Figure 9.4*a*, *b*, *c* shows the peak density of the water molecules in three layers. Near the centre of the pool (9.4*b*) is a rather diffuse peak on the threefold axis from which irregular chains lead to the periphery; above and below it, the water molecule distribution is similar though not identical. Chains and rings can be traced in figure 9.4*a*, *b* between the water molecules attached to zinc, 20.1 and 51.3 and the very strong peaks attached to the peptide chains, 20.4, 21.1, 51.1 and 53.3.

9.5. *The streams at* $z \approx \frac{1}{3}$ *and* $z \approx \frac{2}{3}$

'Streams' of water molecules join the pool of water at $z = \frac{1}{2}$ to the main body of water around the hexamer circumference. These 'streams' are 2–5 water molecules across and occur in a similar pattern on the equivalent hexamer surface, as shown in Figure 9.5*a*, *b*. Along one side, the first line of water molecules is hydrogen bonded to the peptide NH and CO residues of the B chain, from the asparagine B3, to B7 NH. Usually only one well-defined water molecule is attached to NH, two or three, sometimes rather weak, to the carbonyl oxygen. Parallel with the B chain, there is increased electron density between them, giving the effects of streaming. Away from the B chains the water molecules are bounded by a variety of groups, on one side by tyrosine B26, valine A3 and glutamic acid B21, above (or below) by A10 isoleucine and below (or above) by B1 phenylalanine, which close gaps around the *d'* axis. There are small volumes of considerable disorder around B21.1 glutamic acid and B16.1 tyrosine; the density is spreading and low peaks suggest several ill-defined water molecules. But there is also considerable order; witness the long zig-zag chains of water molecules passing from B8 CO through tyrosine B26 OH across to B4 CO, making somewhat different links in the interhexamer contacts of molecules 1 and 2.

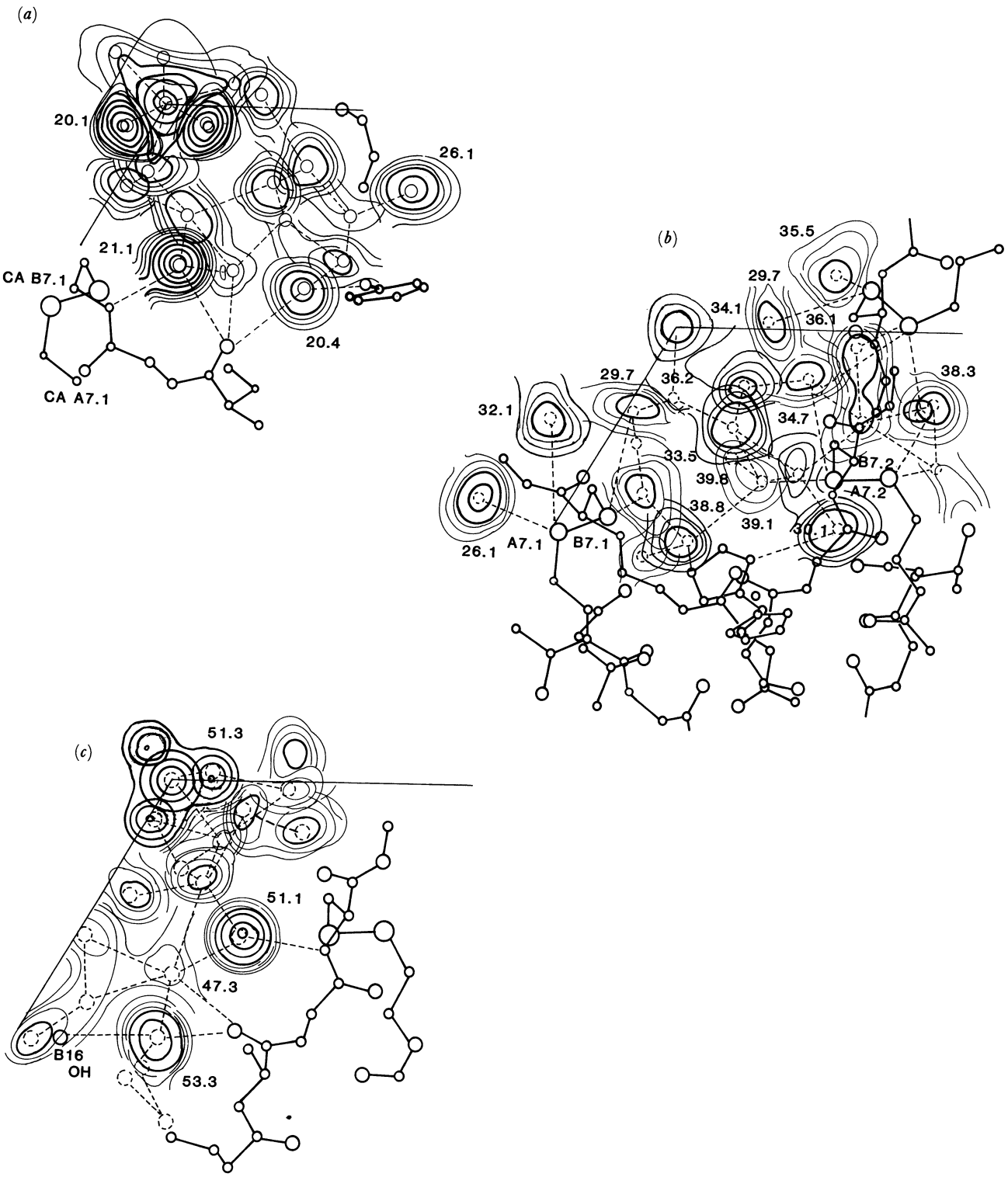


FIGURE 9.4. The pool around $z = \frac{1}{2}$. The water molecule distribution is shown in layers approximately between $\frac{20}{72}$–$\frac{54}{72}$ in z (*a*) 20–26, (*b*) 27–41, (*c*) 42–54, with some overlapping. The electron-density contours are at 0.4, 0.5, 0.6 e/Å^3 fine lines, 0.7, 1.0 e/Å^3 strong lines, here and in later drawings. They are taken from the nearest section to the water molecule positions. Broken lines mark contacts less than 3.5 Å; they are not all necessarily hydrogen bonds. The similarities in water structure in (*a*) and (*c*) can be seen.

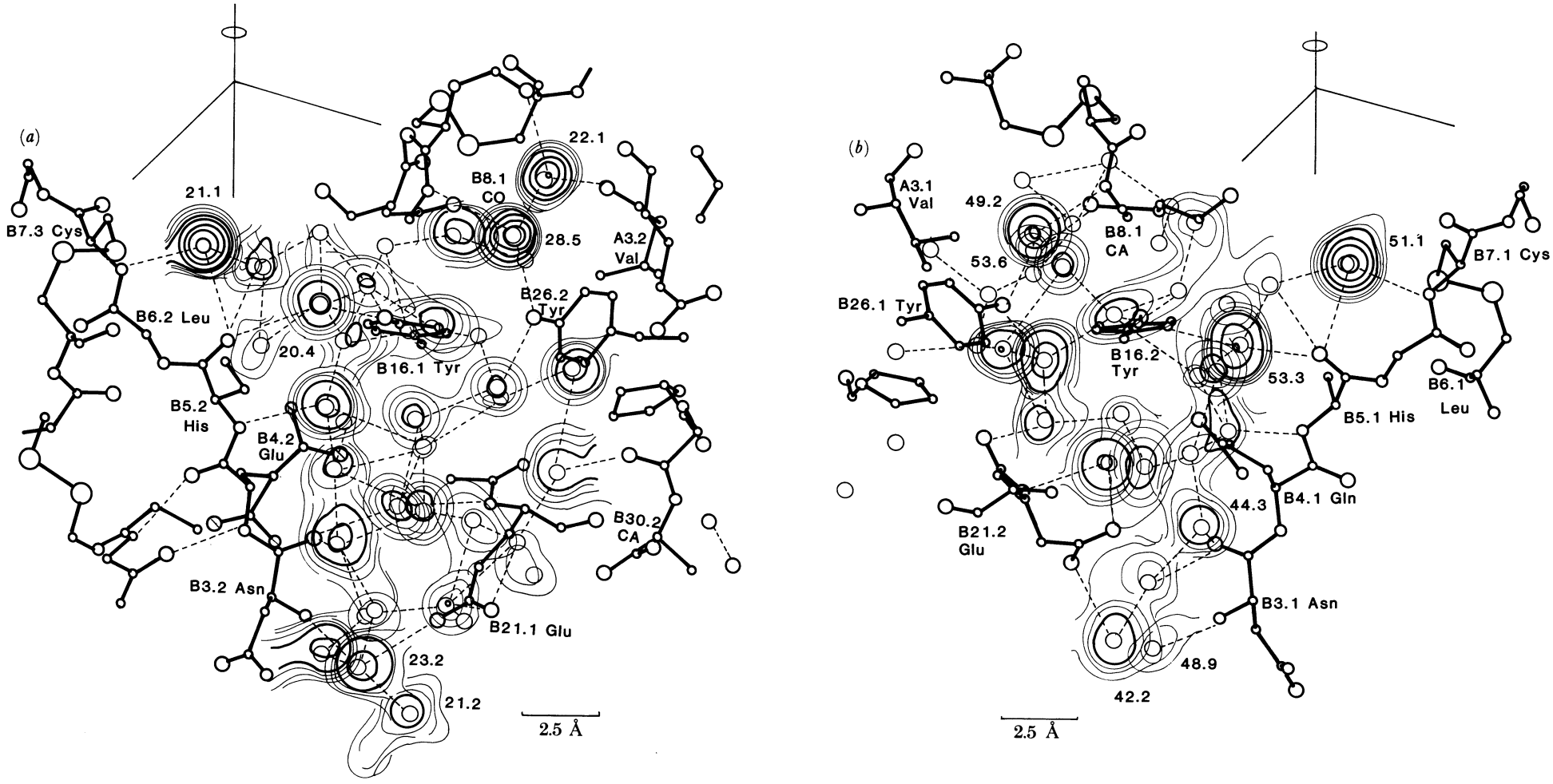


FIGURE 9.5. The water molecule streams at $z \sim \frac{1}{3}$ (*a*) and $z \sim \frac{2}{3}$(*b*). The electron density for the water molecules linking the central channel of the hexamer and the threefold screw axes. The lines of water molecules along the peptide bonds between B3 glutamine and B7 cystine are seen to be similar in both their positioning and definition for molecule 1, (*a*), and molecule 2, (*b*). The marked contrast in the peak height of the water molecules is apparent; notice particularly 51.1 and 21.1, which are very well defined but in contact with poorly ordered water molecules. There are distinct differences in the water electron density in several regions e.g. near B16 tyrosine OH; these arise from the small changes in protein conformation affecting H-bonding geometry and the space available for water.

9.6. *Volume of water separating neighbouring hexamers*

In the direction of the rhombohedral ***a*** axes the hexamers are in direct contact with one another over an area of *ca.* 14×12 Å^2. The large and irregular region between them is filled by water, which spreads out as shown in figures 2.1 *b* and 9.6 *a*, *b*. Around any one hexamer, water is continuous, penetrated by the causeways of the intermolecular contacts and connecting the water channels along the threefold screw axis. Figure 9.6 *a* shows a section in the electron-density distribution at $z = (40, 16, 64)/72$ around the threefold screw axis. In the top right-hand corner, the hexamer surface of dimer (1) meets that of dimer (8) in the neighbouring hexamer. Water molecules in the crevice near their contact are very well defined, 40.2 and 40.3 hydrogen bonded to glycine A1.1 and alanine B30.1 carboxyl group. Further away the peaks are lower and spreading and form parts of different patterns, strands including peaks about 3 Å apart,

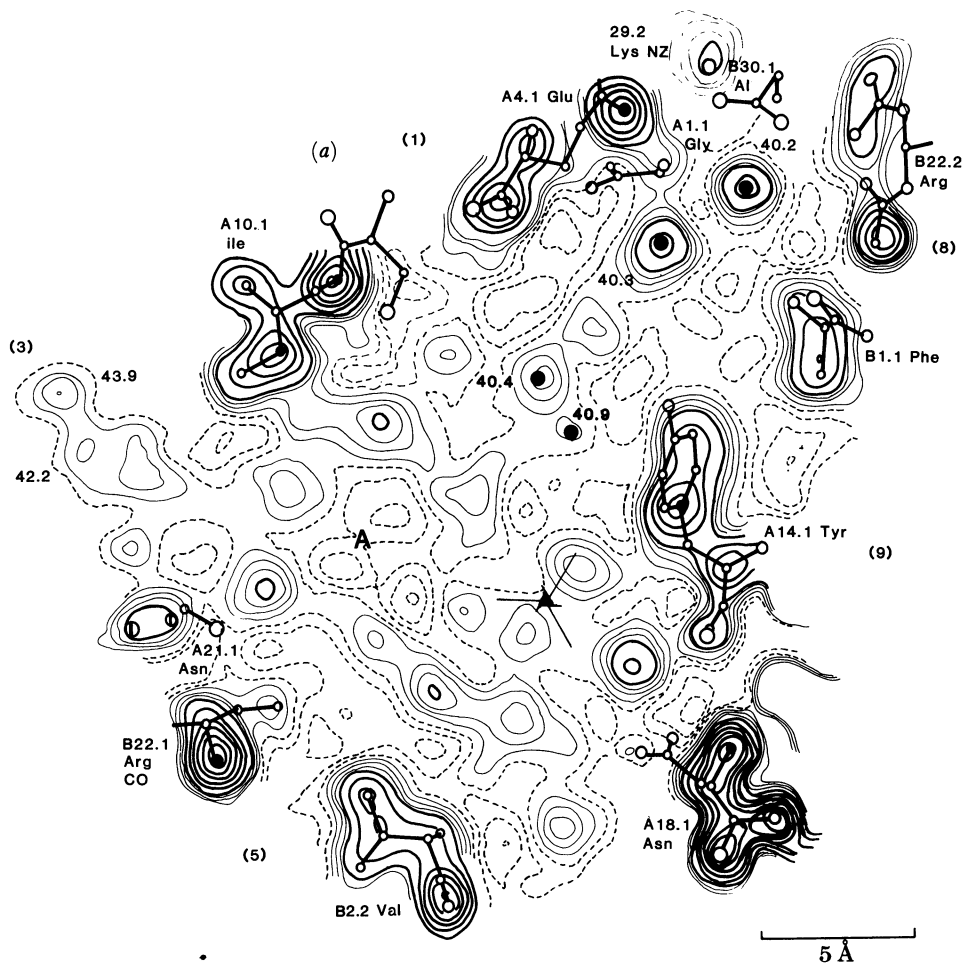


FIGURE 9.6. Volumes of water around neighbouring hexamers. (*a*) Section at $z/72$, 40, 16, 64 surrounding the 3.1 axis. The dotted contours are at about 0, 0.1, 0.2, 0.3 e/Å^3. Other conventions as before. (*b*) Part of the distribution of water molecules between $z/72 = 39$–49, extended round 3.1 axis. This shows extended loosely arranged water peaks; at A there is a cavity closed above the sections by 48.1 and 49.6 (contours not shown) and below by those surrounding A in (*c*). (*c*) Illustrates some marked rings and chains of water molecule positions near asparagine A21.1 and arginine B22.1. It covers part of the electron density distribution between $z/72 = 27$–37 beneath the cavity A and extending round the 3_1 screw axis.

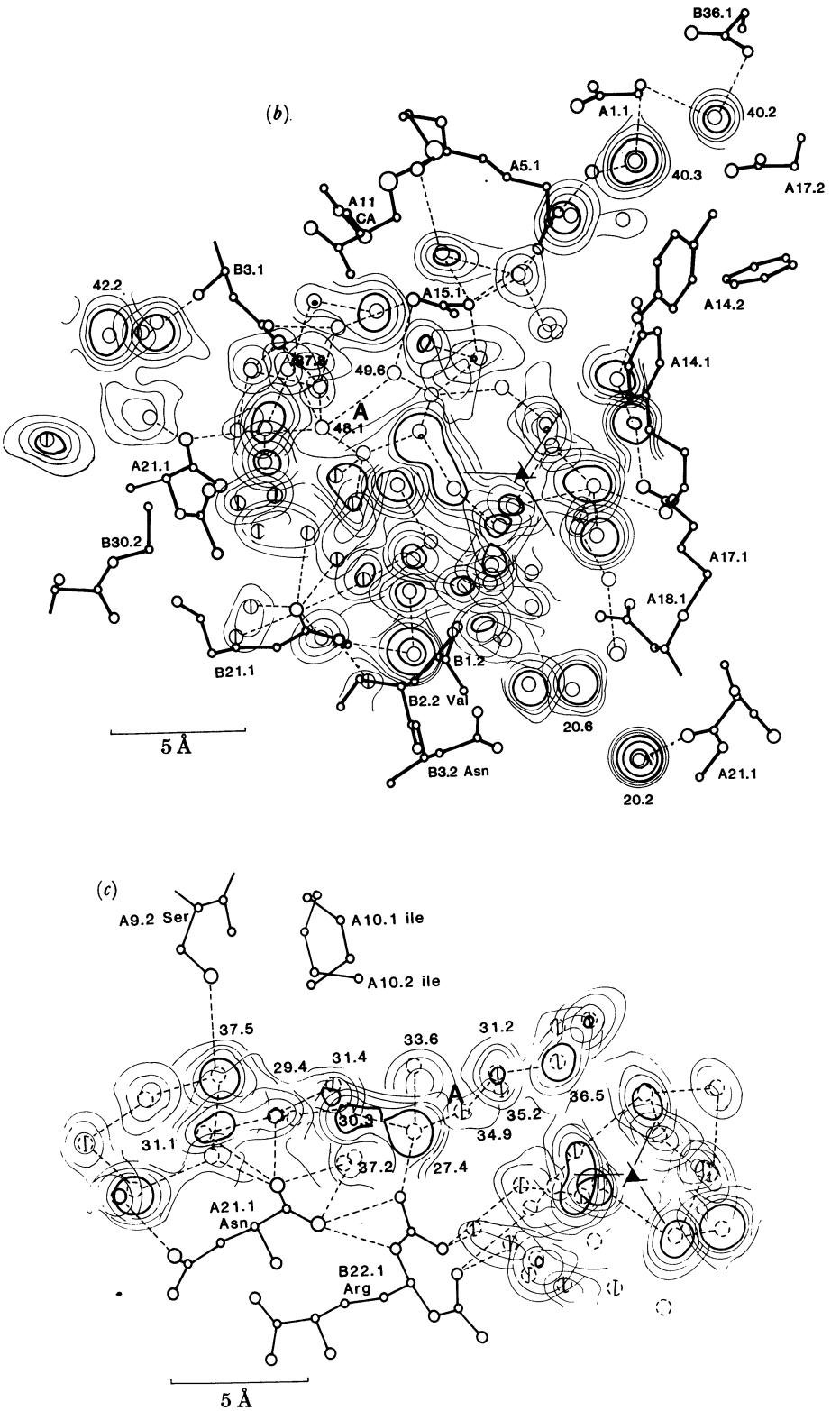


FIGURE 9.6 *b*, *c*. For description see opposite.

rings of various sizes and frameworks enclosing empty spaces. Some of the complexities are further illustrated by 9.6*b*, *c*, which lie one above the other, including section *a* which is contained in figure 9.6*b*. Here overlapping peaks of density are shown in a total depth of about 10 Å. Near the centre of the field a confused framework of peaks surround a hole A where the electron density falls nearly to zero. This cavity is about 6 Å across and could enclose a water molecule. That it does not is probably because of the fact that an enclosed molecule would make too many close contacts. It is closed above by the uncontoured water molecules, 48.1 and 49.6 and below by 34.9 seen in 9.6*c*. There are other examples of relatively large holes in the water distribution.

The hexamer surface includes both non-polar and polar groups and the latter vary considerably in mobility; the surface water molecules correspondingly vary in their definition. Sometimes, near carboxyl groups such as B21.1 glutamic acid, a number of small, clearly partly occupied alternative sites can be traced. Though a number of contacts occur, no definite characteristic arrangement of water molecules near the non-polar groups has been traced, with the exception of the five-membered rings observed near valine A3.1 (p. 390).

Heavy ions such as lead penetrate this region in soaking experiments to settle near A17.1 glutamic acid and B30.2 alanine (terminal CO_2^-).

9.7. *Water channels along the threefold screw axes*

Water is continuous from one side of the crystal to the other along the screw axes as shown by the empty regions over 3_1 and 3_2 in figure 2.1, and the chains in figure 9.7, plate 5.

The channel around the 3_1 screw axis is actually rather smaller than that round the 3_2 axis (generally 10–12 Å across) because certain residues, particularly asparagine A21.2 and the disordered lysine B29.1, approach close to the screw-axis line. The distribution is confused in the region of disorder but otherwise quite well-defined water molecules often follow irregular chains, held through contacts to hydrophilic residues bordering the chains. Around the 3_2 axis the channel is wider, 12–16 Å across. The chains branch, often rings, three, four or five membered can be observed as in figures 9.4, 9.5, 9.6 and 9.7.

9.8. *Possible ions in the crystals*

(*a*) *Sodium ions*

Early chemical analysis suggested there should be 1–2 sodium ions in the crystal. These would have very much the same electron density as water molecules but might be expected to have geometrically different surrounding, 4–6 peaks in tetrahedral or octahedral coordination with Na–H_2O distances of *ca.* 2.3 Å.

The most likely positions among those listed for the water molecules to be actually sodium ions are 25.1 and 48.4, both of which are on the trigonal axis making contacts of 2.1 and 2.5 Å with the three water molecules surrounding the axis and attached to the zinc. Their positions here would be similar to those found in ferroverdin (Candeloro de Sanctis *et al.* 1983). However, the expected coordination is not completed in this crystal by other water molecules to form either regular octahedra or tetrahedra (figure 7.3). Alternative positions for sodium ions would be that of 5.1, possibly also 67.2 in the central cavity (figure 9.2); these both have four near neighbours and are occupied by other metal ions in soaking experiments. 1.1 has a similar claim: it makes four close contacts, two with OE_1 and OE_2 of glutamic acid A17.1, and is

displaced by lead. There are other water molecules that make short contacts with carboxyl groups but nothing that will make their identification as a sodium certain.

(*b*) *The citrate ion*

The marked ridge of peaks in the centre of the central pool near $z = \frac{1}{2}$ suggested at one stage of refinement the backbone of the citrate ion. However, it proved impossible to find acceptable positions for the carboxyl groups attached to it and refinement improved with the substitution of water-molecule chains.

In the map obtained by Sakabe *et al.* (1985) at Nagoya, an alternative site for citrate has been found near the 3_2 axis and $z = \frac{1}{2}$. The peak distribution that we have here, although parts are similar, does not fit well with citrate. It may, if present, appear better from data obtained at low temperatures.

10. Hydrogen bonding in the crystals

10.1. *The identification of hydrogen bonds*

The term hydrogen bond has been used in proteins to cover a variety of situations where positively charged hydrogen atoms attached to oxygen or nitrogen, and occasionally carbon, or sulphur, make contact with another negatively charged atom. Here we have identified 'hydrogen bonds' by the angles and distances between the polar groups. In the commonest situation NH...O a calculated position of the amide hydrogen atom has been assigned and the following limits have been allowed.

this paper	Höhne & Kretschner (1982)
1.3 Å $\leqslant$ O...H $\leqslant$ 2.5 Å	1.45 Å $\leqslant$ O...H $\leqslant$ 2.45 Å
90° $\leqslant$ O...H—N $\leqslant$ 180°	105° $\leqslant$ O...H—N $\leqslant$ 180°
90° $\leqslant$ C=O...H $\leqslant$ 180°	85° $\leqslant$ C=O...H $\leqslant$ 160°

These limits are similar to those proposed recently by Höhne & Kretschner (1982), which are shown for comparison. Höhne & Kretschner derived their figures from a number of accurately determined crystal structures combined with a theoretical treatment of hydrogen bonding that included the influence of the C=O electrons. Where the hydrogen atoms cannot be placed accurately by calculations, as in the bond between carbonyl oxygen and water, we have allowed the distance range 2.3–3.5 Å between the two negative atoms. The limits we have set are clearly arbitrary and the natural variations in distances are aggravated in our analyses by further inaccuracies in the atomic positions. The forces involved in hydrogen bonding are obviously most marked by short interatomic distances but the actual atomic positions often suggest that effects continue to larger atomic separations.

To represent the hydrogen bonding systematically, parameters for all possible hydrogen bonds from the atoms in the insulin molecule were calculated in the ranges given above. These are shown in the following figures and discussed in the text. Figures 3.1–3.21 and 4.1–4.30 also record a relatively small number of examples where apparently favourable interactions occur outside the limits set. These are not included in the tables and figures in this section.

10.2. *The distribution of hydrogen bonds*

(*a*) *The main chains*

The H-bonding contacts from the amide hydrogen atoms and the carbonyl oxygen atoms of the A and B main chains are shown in figure 10.1. It illustrates nicely the character of the

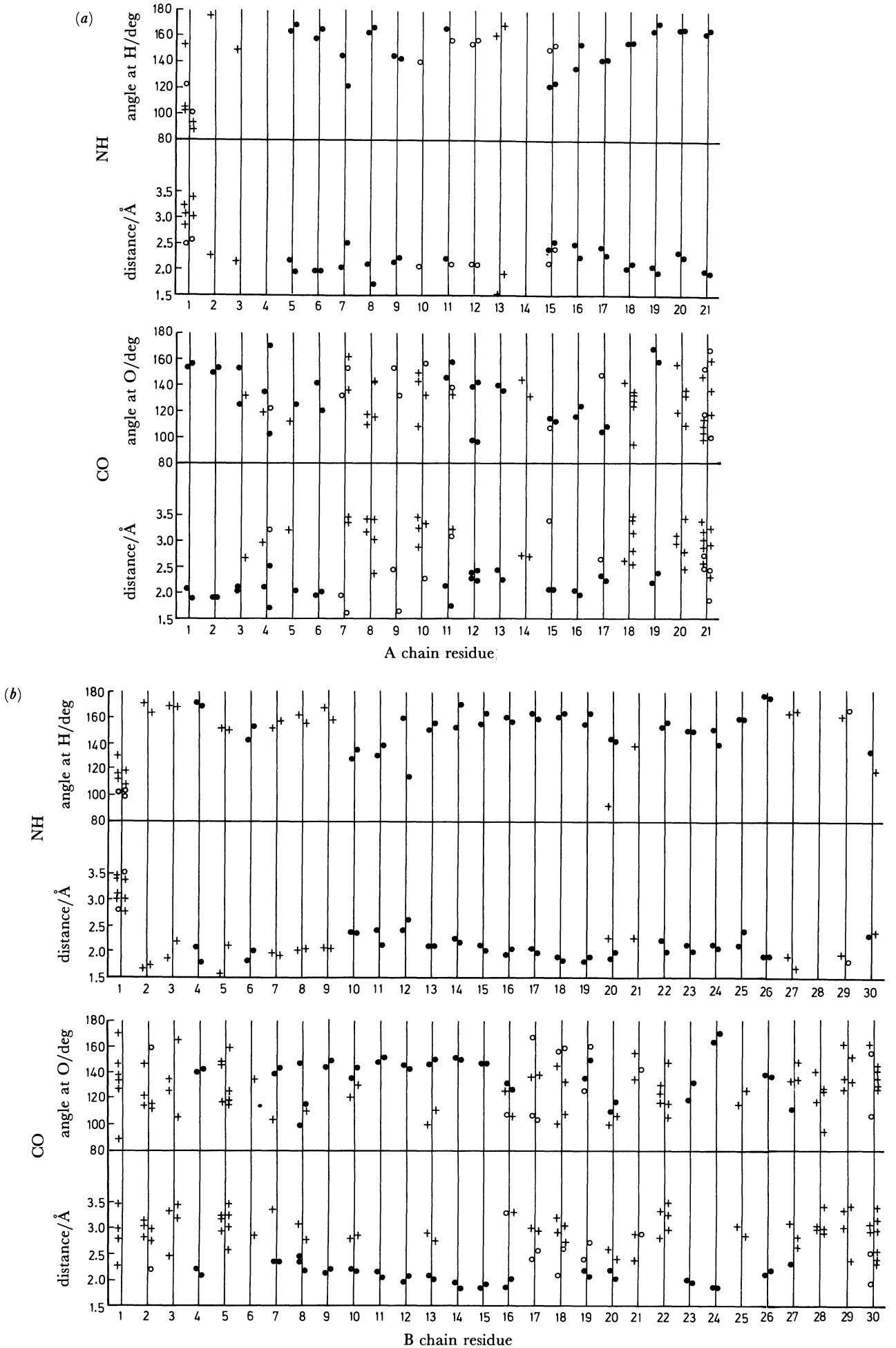


FIG. 10.1 For description see opposite.

hydrogen bonding in proteins involving other peptide and side-chain groups and water molecules. At each hydrogen-bonding centre there are often 2–3 geometrically acceptable contacts but with one interaction usually preferable both by the criteria of bond angle and distance. Where one H...O distance is shorter, it almost always correlates with the large N–$\hat{H}$...O angle. Atoms in regions of secondary structure, α helices as illustrated in figure 10.2, or β sheets, have greater steric restrictions and the H bond capacity of the carbonyl O is more often reduced from two to one bond. Where carbonyl O can make more than one, e.g. at A14 CO, B10 CO and B13 CO where there are additional H-bond contacts to water molecules, the hydrogen bonds are longer within the helices, (Baker & Hubbard 1984).

As Höhne & Kretschner pointed out, within generally helical geometry, the closest contacts vary in particular helices from 3:10–α–π (intervals $1 \rightarrow 4$, $1 \rightarrow 5$, $1 \rightarrow 6$) with at times NH atoms occupying 'bifurcated' positions between the two carbonyl groups (compare Pauling & Corey 1951). The N terminal A chain helix is rather irregular but in the C terminal helix A13–A20 there is a smooth transition from $1 \rightarrow 5$ to $1 \rightarrow 4$ contacts. The B chain helices are more regular than those in the A chains. There are also $1 \rightarrow 4$ interactions at the initial (B7 and B8) and terminal turns (B19 and B20) in both chains where they reverse direction, and also at B27, B30 in molecule 1. These features are readily seen in figure 10.2. There is occasional H bonding to water from peptide oxygen atoms within the helix; never, however, from NH (see figure 10.1). This reflects the tilting of the peptide planes away from the helix axis with the peptide CO out and the peptide NH in (Bolin *et al.* 1982; Baker & Hubbard 1984).

It is notable that at the A chain N terminus the A2, A3 and A4 NH groups in molecule 2 are blocked by A19 Tyr and the A4 Glu sidechains and can only make contacts greater than 3.5 Å to solvent. In molecule 1 water is able to bond to A2 and A3 NH but not to A4 NH, which, however, is in contact with the A4 COO^- group. In both molecules A14 NH, at the N terminus of helix, is completely covered by the approach of A12 serine sidechain, preventing any H bonding to water. The initial turns in both the A chain helixes contain H bonding contacts involving the NH groups and side chain three residues away. These contacts A1NH_3^+----A4COO^-----A4NH and A12NH----A15OG, A15NH----A12OG appear to stabilize the helixes, or perhaps help to initiate them, by giving the initial NH, H-bond interactions.

To summarize: of the 204 main chain polar groups, 53 are in $1 \rightarrow 5$ helices, 29 in $1 \rightarrow 4$, 2 in $1 \rightarrow 6$ helices and 10 in β sheet structures. The remaining atoms mostly interact with side chains or water molecules, there are 6 NH and 2 CO groups apparently unbonded by the above criteria. Of these, only 3 NH groups, A14.1, A14.2, A10.2 and one C=O group, B6.1, are in situations too crowded for any hydrogen bonded contacts to occur with protein or water. In all other cases there are water molecules or H-bonding groups a little outside our limits; these presumably are interacting favourably even if weakly.

There are 35 main-chain carbonyl oxygens (out of 98), which make only 1 H bond, the second contact being prevented by the steric packing and geometry of the folded polypeptide chain when buried and excluded from solvent.

The initial NH_3^+ groups and terminal COO^- all make salt bridges except the terminal COO^- of molecule 2, which is directed into solvent.

FIGURE 10.1. The H-bonding contacts with main chain (●), side chain (○) and water molecules (+) made by main chain atoms in (*a*) the A chain and (*b*) the B chain. Each line represents a residue with molecule 1 contacts to the left and molecule 2 contacts to the right.

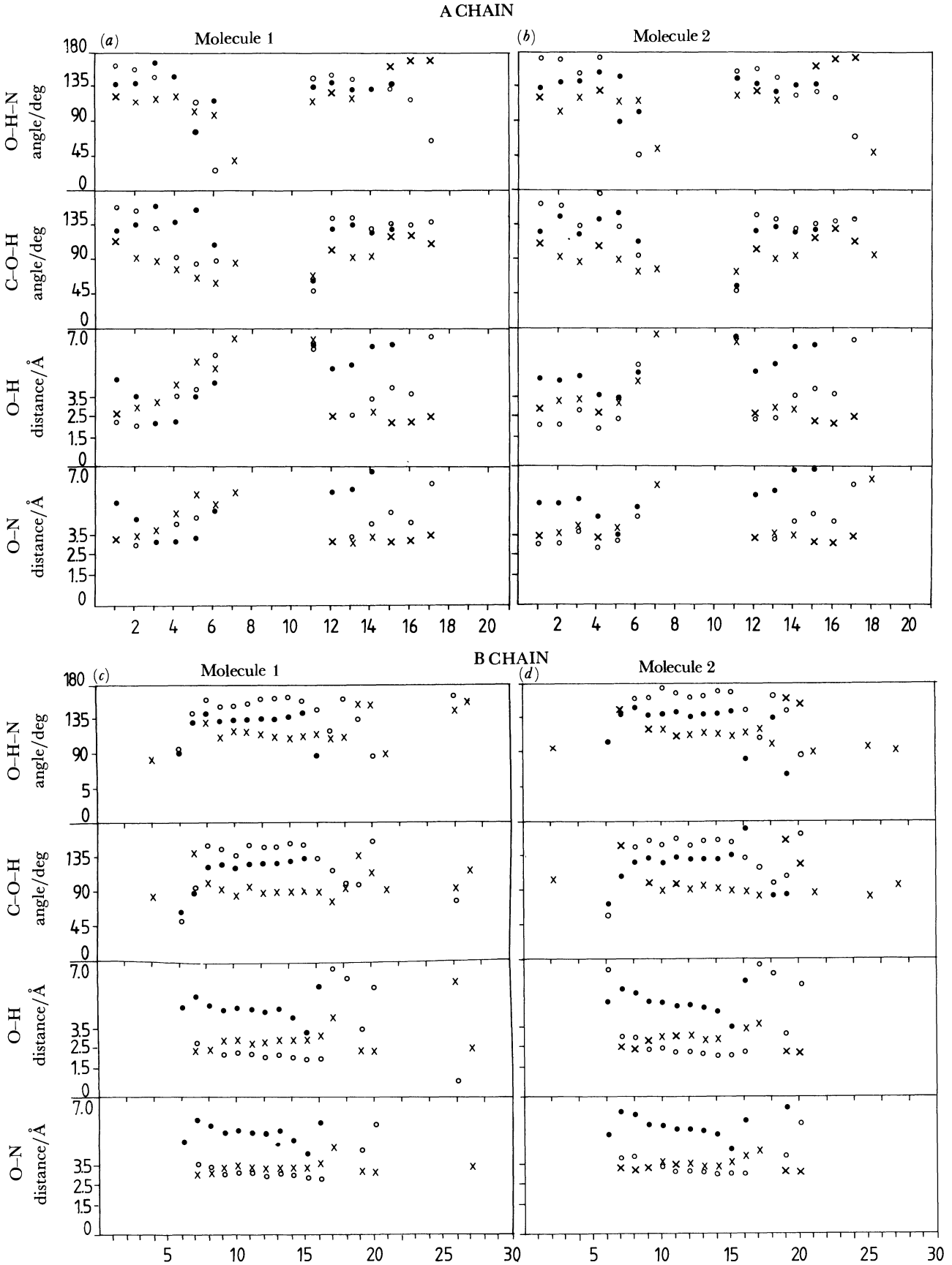


FIGURE 10.2. The pattern of helical contacts produced by $1 \rightarrow 4$ (x), $1 \rightarrow 5$ (○) and $1 \rightarrow 6$ (●) interactions; (*a*) A chain molecule 1, (*b*) A chain molecule 2, (*c*) B chain molecule 1 and (*d*) B chain molecule 2. The contacts that satisfy H-bond geometry almost entirely belong to either $1 \rightarrow 4$ or $1 \rightarrow 5$ helices. There are just two $1 \rightarrow 6$ (or π) helix contacts in molecule 2 between A3 and A8 and A4 and A9. The tendency for $1 \rightarrow 4$ contacts to be most favourable at the beginning and end of helical segments is evident.

(*b*) *The side chains*

Of the 102 amino acids in the two molecules, 46 have hydrophilic side chains. Of these, 20, including disordered side chains, make H-bond contacts with main-chain atoms. Seven side chains interact with each other; two (A14 Tyr and B13 Glu) to their equivalents by bonding across a local axis and one (A5.1 Gln to A15.1 glutamine) bonds within the A chain. The four side-chain H bonds remaining are made between the A and B chains; two of these are produced by crystal contacts. All the other hydrogen-bonded contacts appear to be with water molecules, usually one or two. Every hydrophilic side chain makes at least one hydrogen-bonded contact. There appear to be two examples, both asparagine, A18.2 and A21.2 where the asparaginyl amide ND2 side-chain atoms are not hydrated at all; they are prevented in part from H-bond contact, curiously enough, by the same residue B25.2 Phe. For A21.2 this is through an intramolecular approach, for A18.2 through a hexamer to hexamer approach.

The H-bonding capacity of the individual side-chain atoms can be considered in more detail. The amide NH_2 can make 2 H bonds, the carbonyl O 2 H bonds and the hydroxyl O 3 H bonds. The steric structures in the protein prevent many of these additional bonds being formed to either protein atoms or water molecules. In all there are 18 bonds in 15 side chains whose full H-bonding potential is not achieved, excluding B4.2 where disorder complicates the interpretation of the electron density.

In summary, for main-chain and side-chain atoms in the asymmetric unit there are 156 NH bonding sites of which 91 H bond to protein; 57 to water; 8 are unbonded; and 314 O H bonding sites; 114 bond to protein; 153 to water; 47 are unbonded. However, only 4 main-chain groups are quite excluded by steric factors from water and all side groups make at least one H bond.

10.3. *The protein* H *bonds to water*

The extent of the contacts by the water molecules to the protein is broken down in table 10.1. There are 251 water molecule sites making 869 contacts of less than or equal to 3.8 Å to the protein; the waters make a further 1052 contacts to the other water molecules in the cell. Of these contacts, 298 satisfy the H-bonding criteria involving 184 water molecules from 135 fully and 49 partly occupied sites. The main chain H bonds to 149 sites and also, coincidentally the side chains H bond to 149 sites; there are 48 sites H bonding to both main-chain and side-chain atoms.

The effect of the protein H-bond contacts on the waters' freedom of motion, as measured by their *B* values, only becomes obvious when there are three or more (see table 10.1). Single H

TABLE 10.1. PROTEIN AND WATER H-BOND CONTACTS

no. of H bonds • to protein	no. of water (full and fractional)	B/Å^2 (avg.) water	no. of waters (full weight)	B/Å^2 (avg.) water
0	160	56.5	75	60.9
1	109	55.3	75	60.2
2	56	50.4	44	50.5
3	20	37.8	18	38.6
4	3	29.8	2	32.8
5	1	38.0	1	38.0

bonds to the protein will allow considerable movement of the water molecules, only reduced by more H bonds or a restricted environment. The main-chain contacts, especially those by NH with their more restricted access, are the most stabilizing. The charged environments of the chain termini affect 4 of the best defined waters and steric hindrance affects 7 other well-defined waters.

The distribution of the distances less than or equal to 3.5 Å from the protein H-bonding centres (main chain and side chain) to the water molecules at fully occupied sites is shown by the histogram in figure 10.3. The most frequently occurring distances in a rather smooth population appears to be between 2.7–3.1 Å, although there are many shorter and longer contacts than these. There is not a sharp fall off of numbers as the protein water distance increases to 3.5 Å, this is a consequence of water molecules packing around the H-bonded waters at distances of 3–4 Å from the protein.

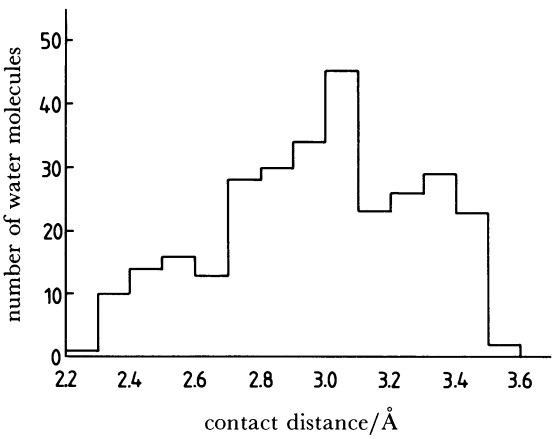


FIGURE 10.3. The distribution of the bond distances of the water molecules H bonded to the insulin molecules.

10.4. *Hydrogen-bond geometry*

Analysis of the angles at the oxygen and hydrogen centres and their separation within the protein structure and in the side chains shows rather broad distributions like those found with other proteins (Baker & Hubbard 1984).

The most favoured angles at the main-chain oxygen and nitrogens are between 130–160° and 140–170°, respectively. The distortion from 120° at oxygen to higher values and at the hydrogen to values less than 180° reflects the accommodation required, particularly in the peptide secondary-structure H bonding. The distance O...H is mostly distributed between 1.9–2.3 Å.

The side-chain H-bonding atoms are found to have more scattered values of angle and distance; this in part reflects the poorer accuracy of the atomic positions. However, the angles at the side-chain oxygens fall, the most populated values now lying between 100–140°. Concomitantly there is an increase in the proportion of angles greater than 170° at the hydrogen.

The correlation between the angle at the nitrogen, oxygen and hydrogen centres and the X...H distances are illustrated in figure 10.4*a*, *b* and *c*. The more directional character of the peptide nitrogen H bond in 10.4*a* is clear; the absence of any detectable correlation between

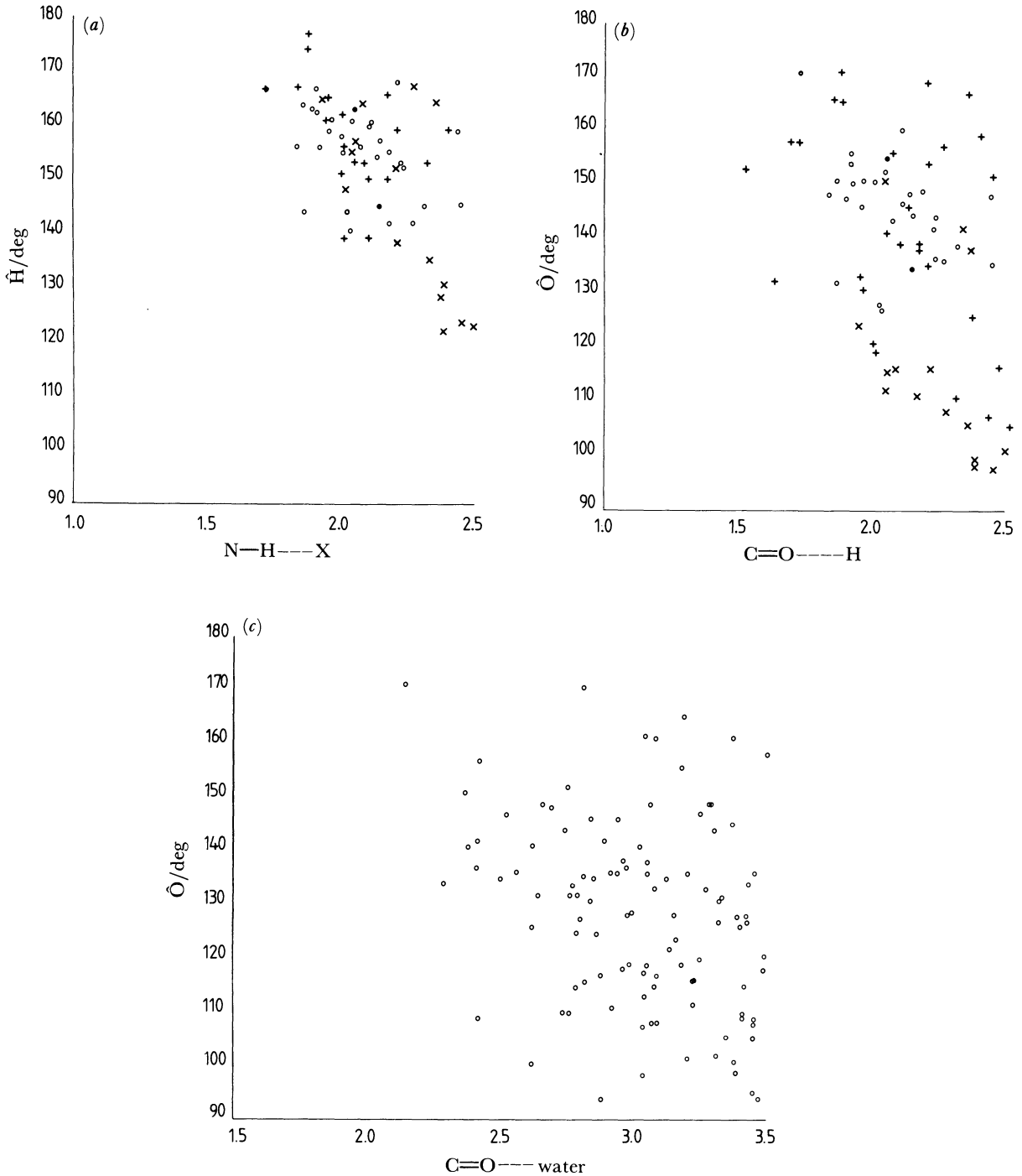


FIGURE 10.4. The correlation between the angles and the separations of the H-bonded peptide backbone atoms. (*a*) the peptide NH to peptide CO; (*b*) the peptide CO to peptide NH; (*c*) the peptide CO to H-bonded water molecules. Symbols: ● represents π helix (1→6) contacts; ○ represents α helix (1→5) contacts; × represents 3_{10} helix (1→4); + represents β sheet. In (*c*), the symbol ○ represents water.

carbonyl oxygen and attached water seen in 10.4*c* reflects the more tolerant geometry at the oxygen and the weak steric restrictions on water molecules. In 10.4*b* the bond parameters of the $1 \rightarrow 4$ and $1 \rightarrow 5$ helices are seen to be separated into two distinct regions, a reflection of how the differing internal geometry of the helices dictates the approach of the amide H to the more flexible carbonyl O.

11. The apparent atomic motion in the crystal

11.1. *The relation between atomic motion and the atomic thermal parameter (B)*

The magnitude of the atomic thermal parameter, $B = 8\pi^2\overline{U}^2$, is formally a reasonably good quantitative estimate of the mean square displacement of each atom from its mean position. It is observed that in organic crystal structures the B value is small (2–5 Å^2) for those atoms that make strong stabilizing H bonds and other close contacts and larger where there are few interactions.

The values obtained for the protein atoms in our crystal range between 4 and 76 Å^2 (equivalent to displacements of 0.2 and 1 Å, respectively). For water molecules, treated as oxygen atoms, the range is between 11.0 and 105 Å^2 (displacements of 0.37–1.1 Å). The higher thermal-parameter values in the protein are the sum of atomic vibrations and the movements of the peptide groups, usually the side chain, as a whole in the loosely packed spaces on the surface. When there is unrecognized disorder reducing the occupancy of the refined atoms, the B value will also be increased. The thermal parameters and occupancy of the poorly defined water positions are particularly uncertain in this respect, sometimes being given estimated values only. For these reasons thermal parameters greater than 50 Å^2 must not be regarded as an accurate measure of atomic motion.

11.2. *Individual atomic thermal parameters*

The figure 11.1 shows the plots of individual atomic B values for a number of residues, selected to illustrate the generally good correlation between the value of B and the residue's structural environment. Thus there is usually a steady increase in the B value (up to 50 Å^2) along the length of the side chains, much more pronounced for surface (e.g. A10 isoleucine), than for buried (e.g. A16 leucine), side chains. And the crystal contacts made by A14.2 tyrosine can be seen to have reduced its B values relative to the freer A14 tyrosine in molecule 1. In five cases only, B2.1 valine, B3.1 asparagine B21.1 and B21.2 glutamic acid and B30.2 alanine do side-chain terminal atoms have B values greater than 70 Å^2, indicating very mobile, disordered structures. There are a few examples of atoms whose B parameters refine to values markedly different to those of the three attached atoms. These anomalies can disappear if restraints are applied to the thermal parameter (H. Savage, personal communication). The most anomalous refined B value is that of A2.2 isoleucine CB, which is 70 Å^2, its surrounding atoms have B values of 20.8, 23.0 and 36.7 Å^2. In the final structure-factor calculation, B for A2.2 CB was set at 40 Å^2 because its high value was considered to result from deficiencies in the minimization approximations with limited data (Dodson 1981).

11.3. *The average thermal parameter at each residue*

Figure 11.2 shows the average value of the thermal parameter for the main-chain and side-chain atoms for the residues of the A and B chains for each molecule. The main-chain atoms are almost always less mobile than the side-chain atoms, which is to be expected. The

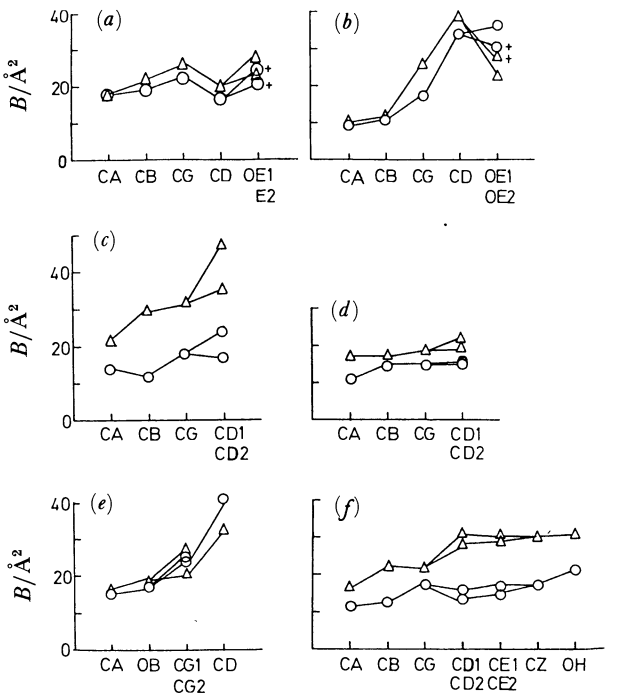


FIGURE 11.1. The atom thermal parameters (B) for selected amino-acid residues illustrating their different behaviour in different environments. (a) A4 glutamic acid: surface residue in which the terminal COO^- is salt bridged to the A_1 amino group. (b) B13 glutamic acid: in the hexamer centre, the terminal COO^- in 2.6 Å contact with local twofold related equivalent. (c) A13 leucine: loosely buried (in hexamer formation). Molecule 2 is better packed than molecule 1. (d) A16 leucine: buried within the monomer. (e) A10 isoleucine: surface residue making non-polar contacts in the crystal. (f) A14 tyrosine: surface residues; molecule 1 is unobstructed, molecule 2 makes crystal contacts with a symmetry related hexamer. Symbols: △, molecule 1; ○, molecule 2.

exceptions are the cystine bridge at A20–B19 and the buried phenylalanine B24. It is also noticeable that the A chain thermal parameters are generally larger than those of the B chain, a consequence of their larger surface distribution. Also that the A chain in molecule 1 is more mobile than in molecule 2, reflecting its fewer crystal contacts. There is a close correspondence in both molecules within the B chain central helix, within the β turn and, excepting B25 side chain, in the subsequent β sheet. In contrast, however, the helical segments in the A chain do not have significantly lower thermal parameters. The distinctly different B values at the N and C termini in the two molecules are a consequence of their different and looser contacts in the crystal.

Residues that are buried in the monomer or by dimer formation are illustrated in the figure 11.2. These residues, mostly from the B chains, have small values for their average thermal parameters and have about equal values for their side-chain and main-chain atoms. Two exceptions are A2.2 isoleucine, moved during hexamer packing in the crystal and B25.2 phenylalanine whose side chain contacts and buries the B25.1 Phe side chain across the d_0 axis, but is itself on the surface of the hexamer and relatively free to move.

The hexamer-forming residues, indicated as H, are sometimes well defined, e.g. B10 histidine, which coordinates to the central zinc ions; but more often these residues have relatively mobile side chains (B13, A13, A14 and A17) indicating looser interactions between the dimers.

There are close contacts made in the crystal by the side chains of B27.1 threonyl, A18.2

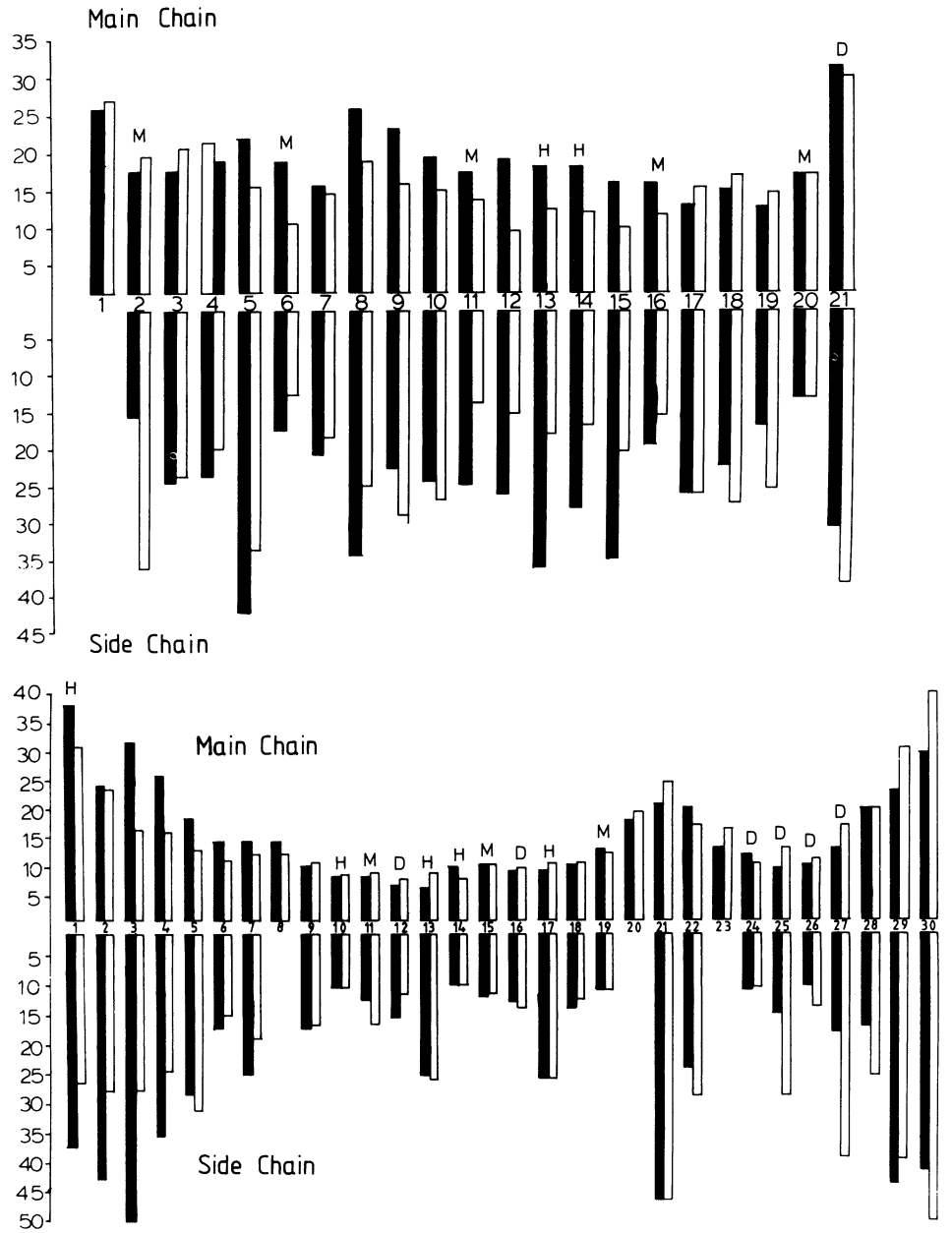


FIGURE 11.2. The average value of the thermal parameters B for the A and B chains. The average value of the main-chain atoms are shown above and that of the side-chain groups below. M, D, H refer to residues buried by monomer folding, dimer formation and hexamer formation respectively. Filled areas, molecule 1; open areas, molecule 2.

asparagine, B5.1 histidine and A14.2 tyrosine that affect the thermal parameters. The side chains of B29.2 lysine and B22.2 arginine form part of an ionic patch in which they make salt-bridge contacts to A4.1 glutamic acid and the carboxylate group of B30.1 Ala.

11.4. *The overall motions in the hexamer*

The relation between the B, or $\bar{U}$ value of individual atoms and the overall dynamic motion of the free molecule is of great interest and the sensible values obtained for the atomic thermal parameters encourages this kind of analysis (Artymiuk *et al.* 1979).

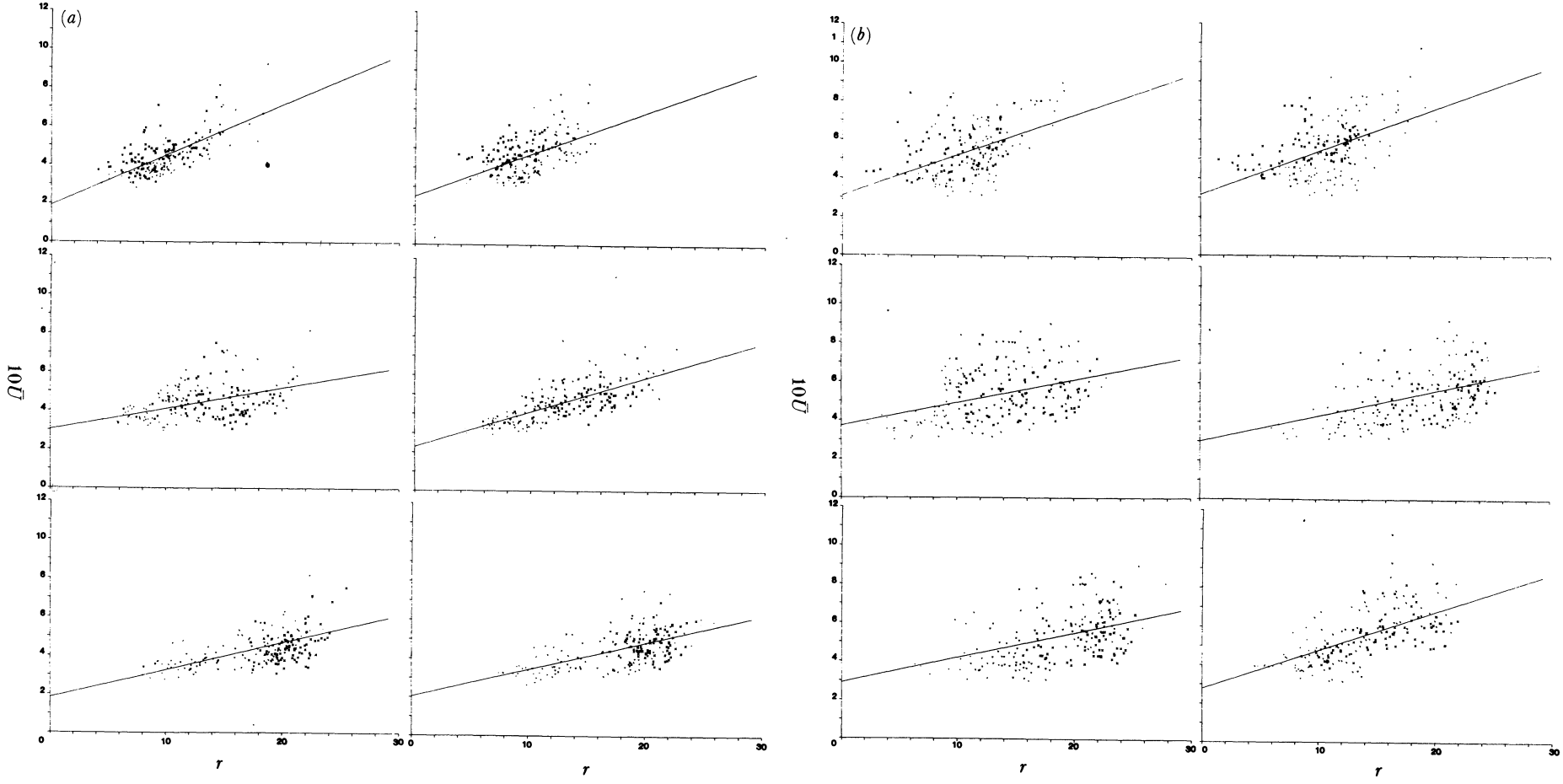
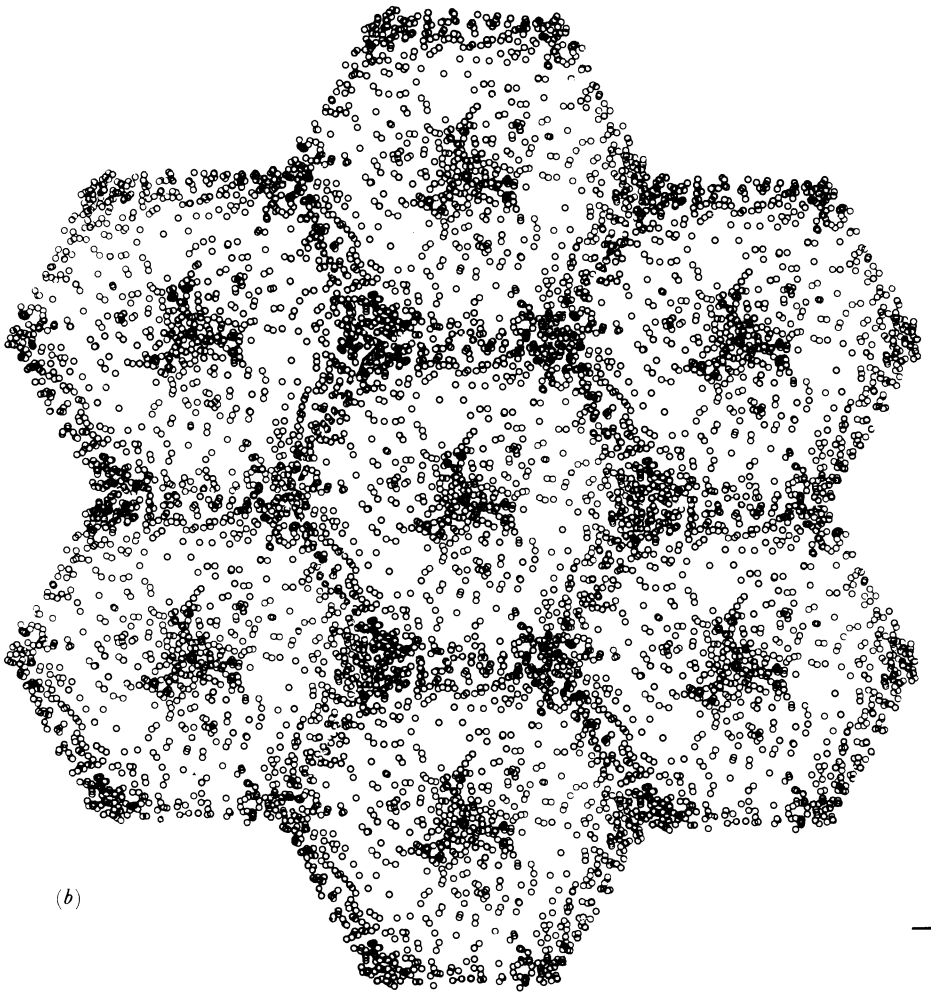


FIGURE 11.3. Plots of $\bar{U}$ against the radial distance r from the centre of mass of (*a*) the main-chain atoms for molecule 1, left, and molecule 2, right, as the monomer (top), the dimer (middle) and hexamer (bottom) and (*b*) of the side-chain atoms in the same order. The A chain atoms are represented by dots and the B chain by crosses.



(*b*)

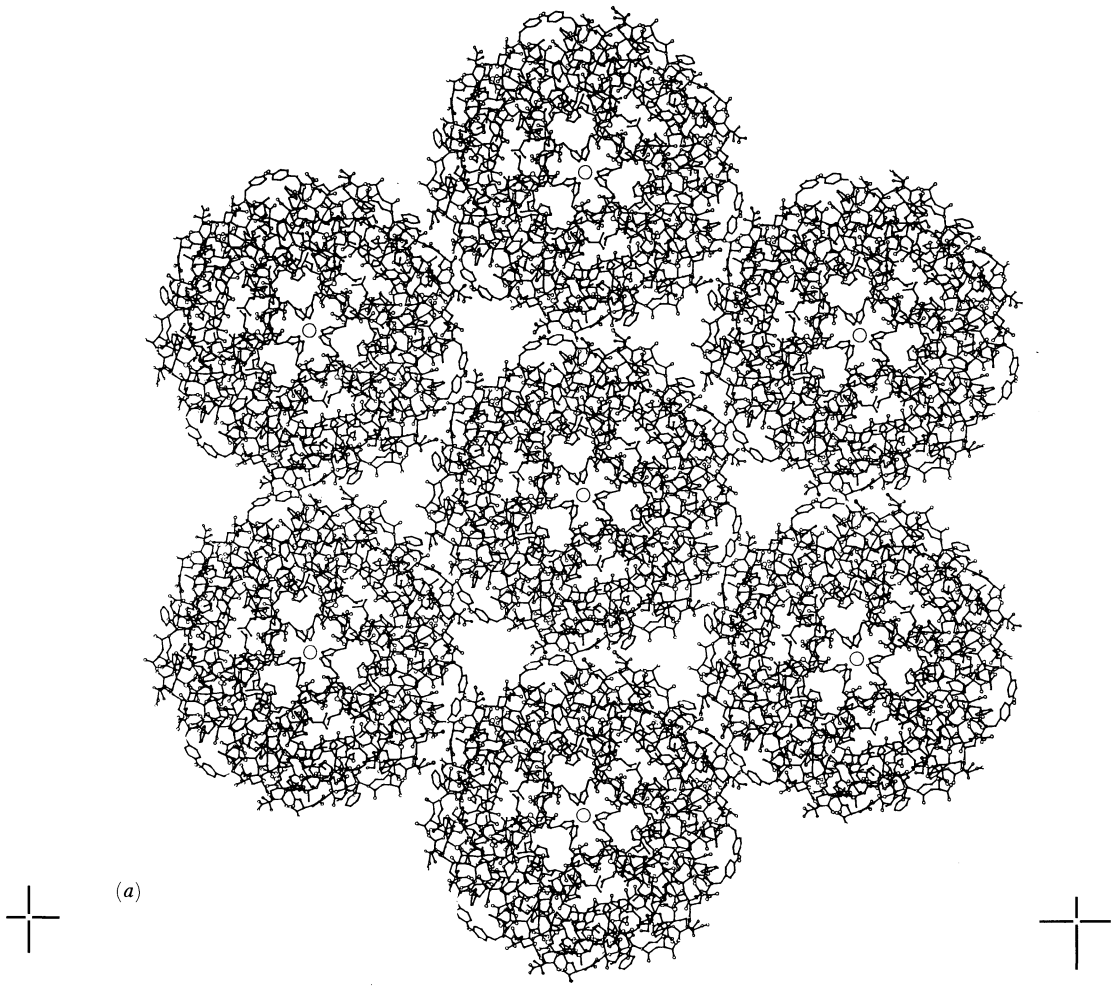


FIGURE 2.1. (*a*) The atoms in 42 2Zn insulin molecules (excluding hydrogen) projected along the crystal threefold axis. (*b*) Overlay showing water molecule positions found in the same volume. Not all positions are fully occupied.

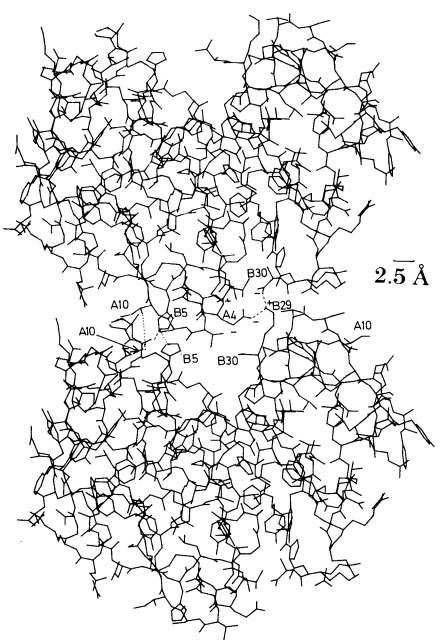


FIGURE 2.2. Projection of the atomic positions, excluding hydrogen, perpendicular to the crystal threefold axis. The positions of three insulin molecules from the hexamer are repeated by one unit cell translation along the ***c*** axis.

(*Facing p.* 448)

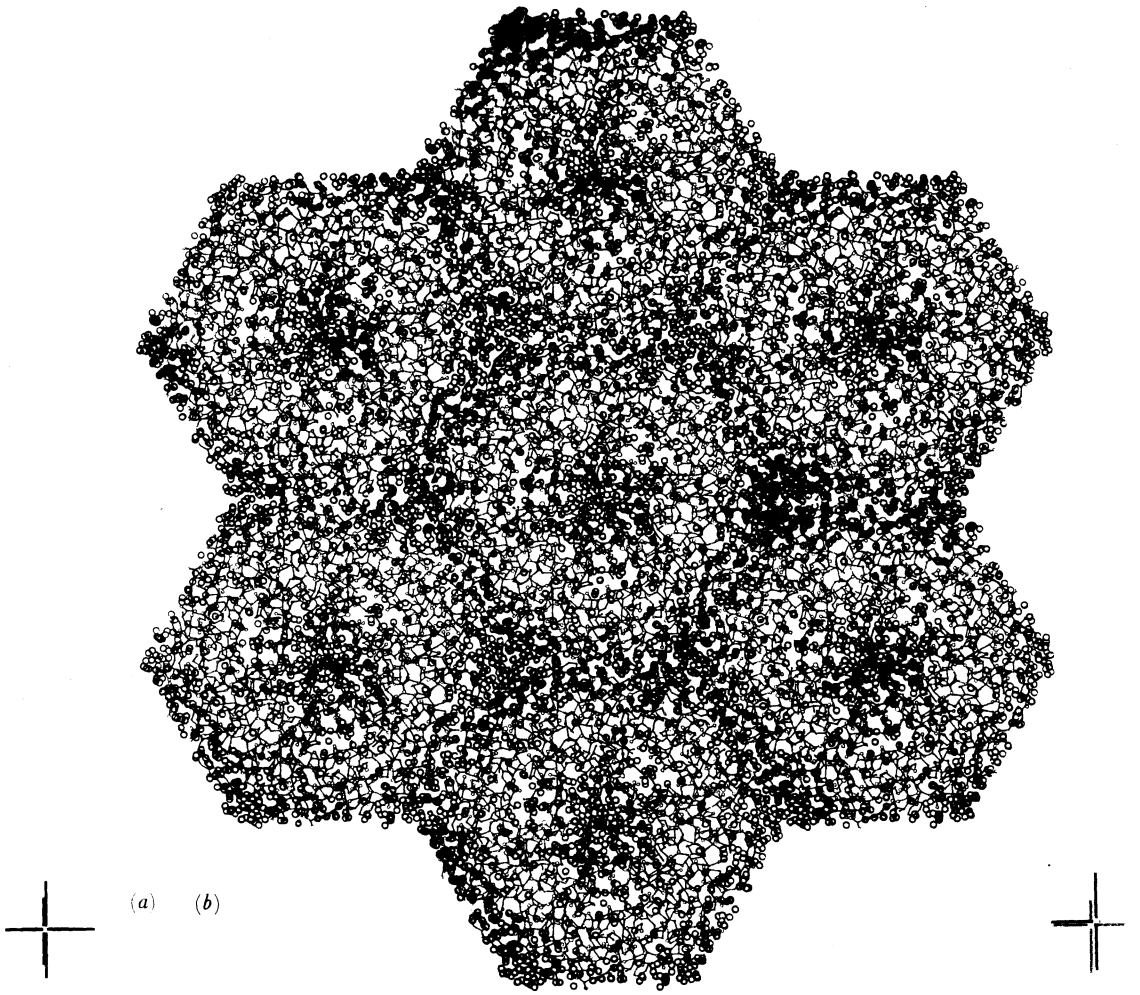


FIGURE 2.1. (*a*) The atoms in 42 2Zn insulin molecules (excluding hydrogen) projected along the crystal threefold axis. (*b*) Overlay showing water molecule positions found in the same volume. Not all positions are fully occupied.

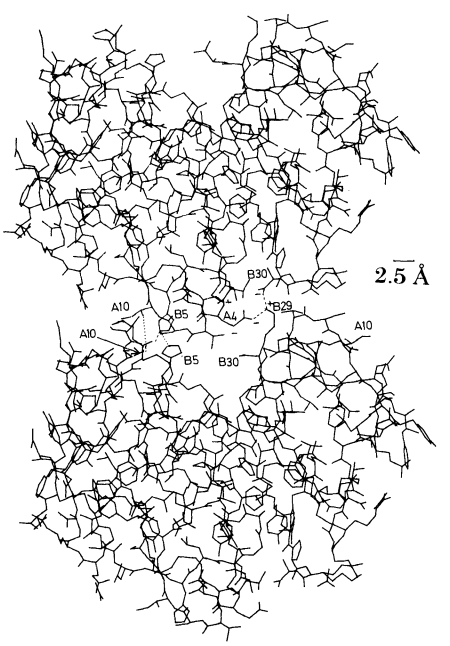


FIGURE 2.2. Projection of the atomic positions, excluding hydrogen, perpendicular to the crystal threefold axis. The positions of three insulin molecules from the hexamer are repeated by one unit cell translation along the ***c*** axis.

(*Facing p.* 448)

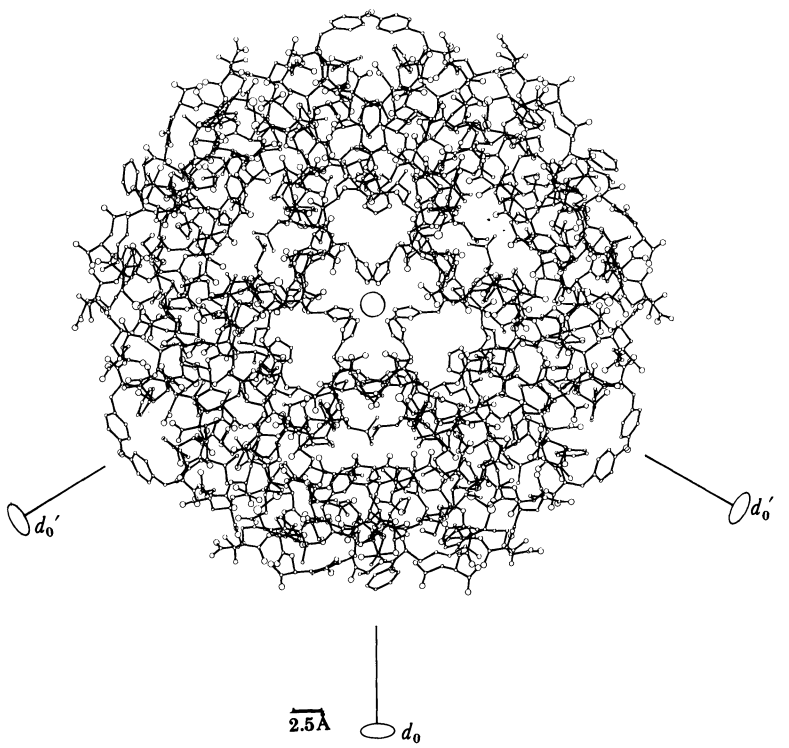


FIGURE 2.3. Atoms in the 2Zn insulin hexamer viewed along the crystal theefold axis. The molecules in red represent molecule 1 and those coloured black, molecule 2. The directions of d_0 and d_0' are shown.

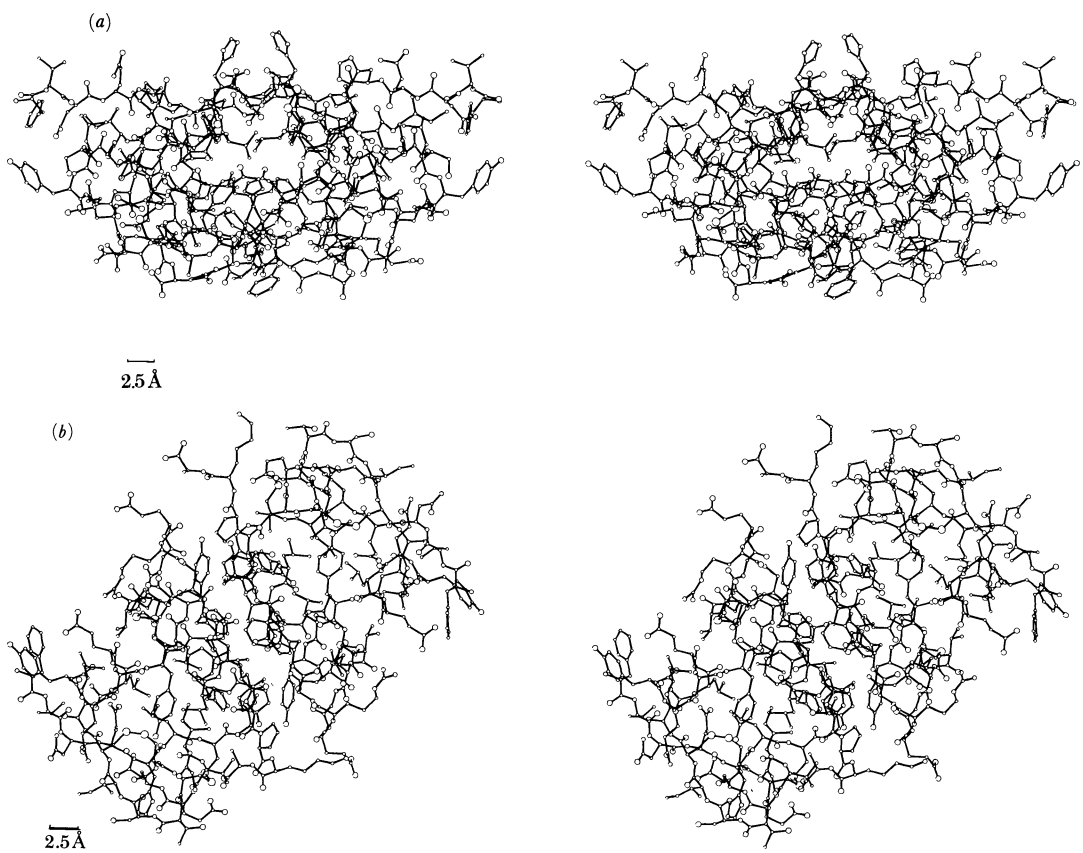


FIGURE 2.4. Stereo pictures of the atoms in the insulin dimer projected (*a*) along the threefold axis and (*b*) along the local dyad axis d_0.

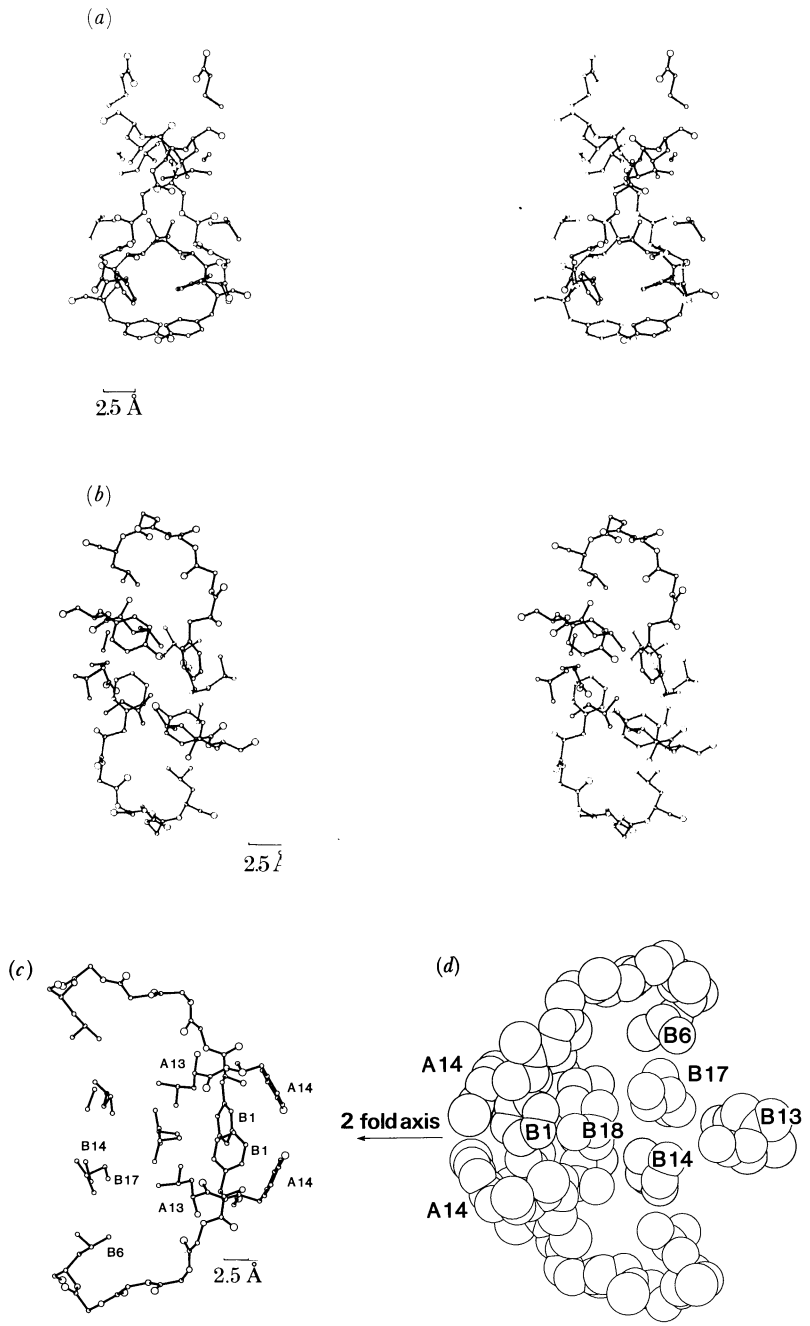


FIGURE 7.4. The hexamer-forming contacts between adjacent dimers. The stereo view (*a*) down the threefold axis and (*b*) down the local two-fold axis. The mono view perpendicular to the twofold axis (*c*) shows the relatively loose packing emphasized by the van der Waals surfaces shown in (*d*).

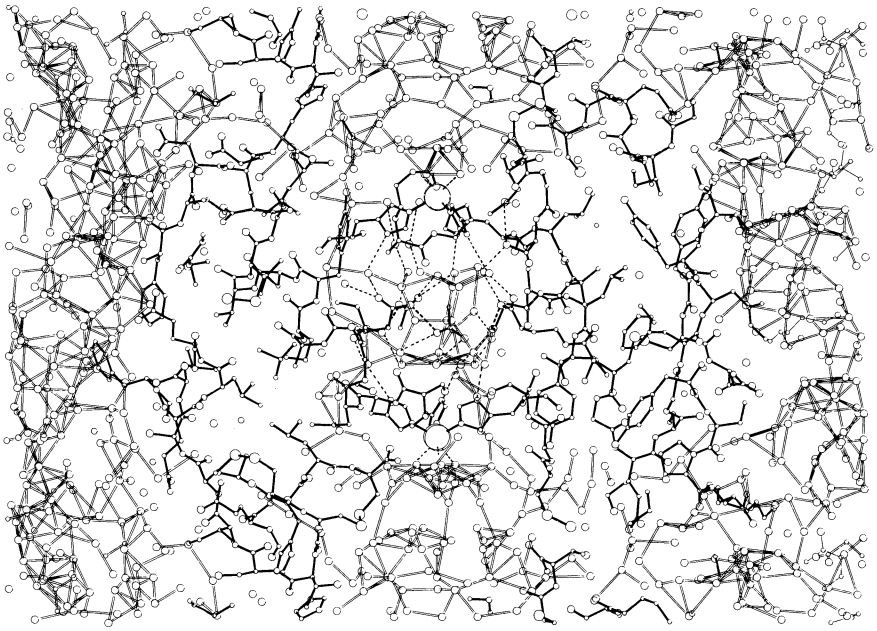


FIGURE 9.7. A view of the 2Zn insulin crystal perpendicular to the threefold axis including both the theefold axis and the 3_2 screw axis. The two axial zinc ions are in the centre of the figure. The dimensions of the box shown are 54 Å in x and 34 Å in z; the box is 6 Å in depth. Contacts between water molecules of less than 3.5 Å are drawn in red lines; not all of these are H bonds. The H bonds between some protein atoms and water molecules in the central cavity are indicated by broken lines.

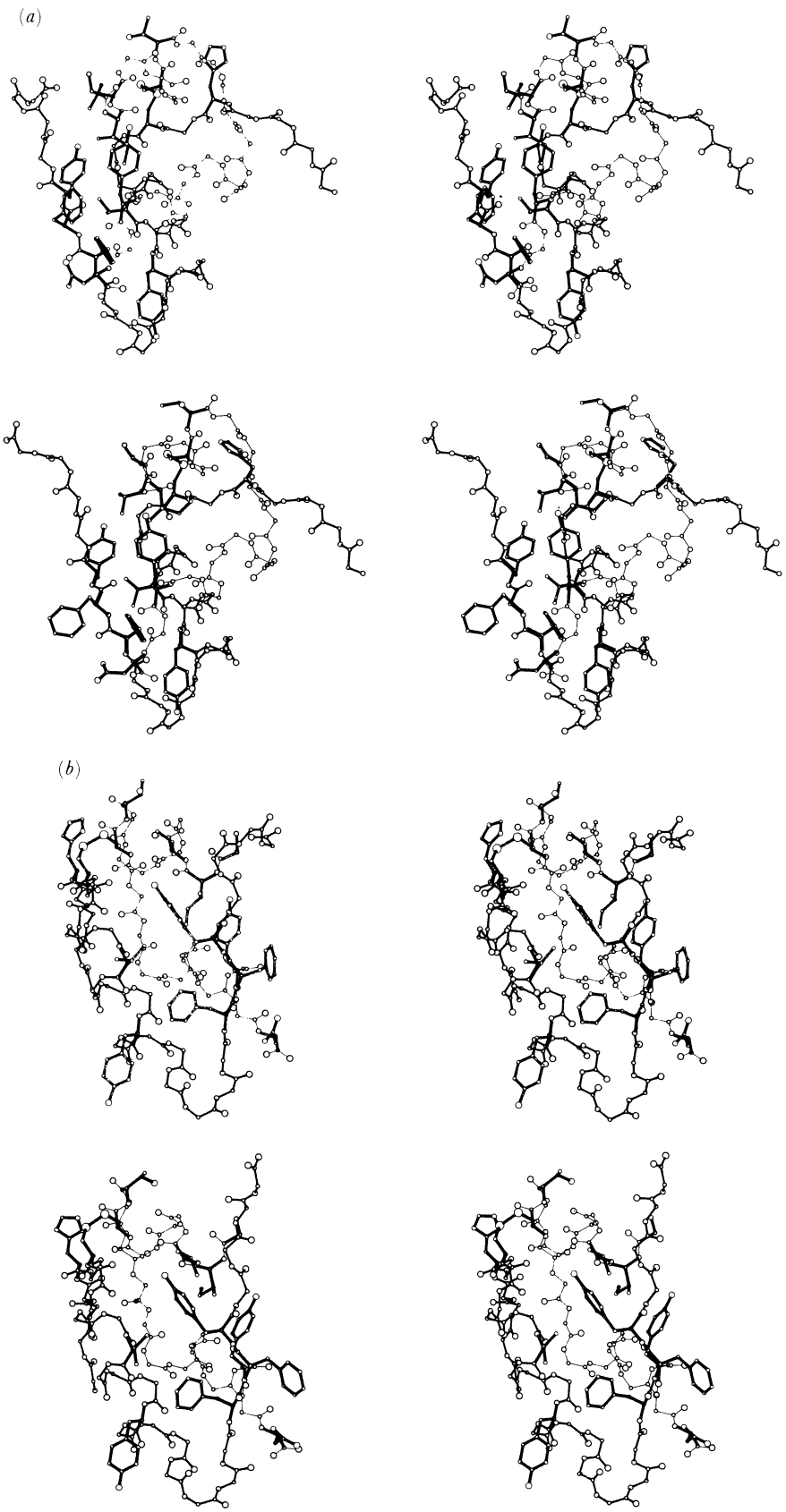


FIGURE 12.2. Stereo pictures of molecule 1 (upper) and molecule 2 (lower) showing the spatial distribution of residues whose structure appears important in insulin's binding and biological action. The residues that appear to be involved in activity are drawn in red, for the remainder of the molecule only the main chain is shown, the A chain as thin lines, the B chain as thicker. The six carbonyl oxygens possibly available for H bonding are filled in red. In (*a*) the view is perpendicular to the threefold axis and 30° away from the direction of the local twofold axis of the dimer. In (*b*) the view is perpendicular to the local twofold axis and faces the dimer-forming surface.

The distance, r, of each residue from the centre of mass when plotted as r against $\bar{U}$ will correlate sensibly if the structure as a whole is vibrating about its centre.

The plot of $\bar{U}$ against r (where r is the distance from the molecular centroid) is shown in figure 11.3 for each insulin molecule separately with the centroid of the monomers, the dimer and the hexamer. The plots show a much larger spread of values for the side chains, reflecting their generally higher thermal parameters and the marked effects molecular and crystal contacts can have in stabilizing them and reducing the atomic motion.

The correlation between $\bar{U}$ and r appears best, however, for molecule 1 with the dimer centroid. It is still good for molecule 2 and for both molecules with the hexamer centroid. This behaviour suggests that the molecules are oscillating more in motions about the dimer and hexamer centres than about the monomer centres. The existence of these two modes of motion is perhaps not surprising, the hexamer although a compact, well H-bonded and symmetrical structure, has contacts at the hexamer centre and between the dimers that are weaker than those within the dimers.

12. The crystal structure and the biological action of insulin

In any discussion of the reactivity of a molecule in relation to its structure as observed in a crystal we need to consider whether crystal-packing forces have modified the atomic arrangement from that required for biological action. In the case of the 2Zn insulin crystal it is certain that the atomic arrangement in at least one of the two molecules present is altered by crystal packing and that other changes, small and large, have been observed in other crystals containing insulin or insulin-related molecules. We do not know which, if any, of the observed conformations is that involved in biological reactions; it is clear they are easily interconvertable, and there is some indication that further changes from the crystalline structure may be necessary for insulin action.

The major part of the atomic arrangement we find is both relatively rigid and common to the molecules in 2Zn insulin, (a hexamer), 4Zn insulin (a hexamer) (Cutfield *et al.* 1981), hagfish insulin and zinc-free pig insulin (dimers) (Bentley *et al.* 1978), and despentapeptide insulin (monomer) (Bi *et al.* 1984), which as a biologically active molecule may represent the insulin monomer, the biologically active species. The major flexibility we observe in 2Zn insulin at the A chain N terminus (A1–A6) and the B chain C terminus (B25, B28–B30) may actually be important for the expression of insulin activity, combined with the rigid structure of the rest of the molecule. The changes observed do in any case provide a model for the way in which packing pressures in one region of a protein molecule can affect atomic movement at a distance and so produce the changes we must expect when the insulin molecule interacts with the receptor proteins through which most, if not all, of its biological actions are transmitted. Only receptor-mediated actions will be discussed here.

The insulin receptor has been found to be a large glyco-protein molecule of relative molecular mass about 300000, consisting of two α and two β chains organized in a generally similar way to that in immunoglobulin molecules. The very recent determination of the amino acid sequence in the receptor by Ullrich *et al.* (1985) and Ebina *et al.* (1985) has revealed the exact composition of the two chains, which include many suggestive details, a cysteine-rich region in the α chain, many isolated cysteine residues, a short stretch of hydrophobic residues in the β chain, a homology with the epidermal growth factor receptor. Figure 12.1 shows a

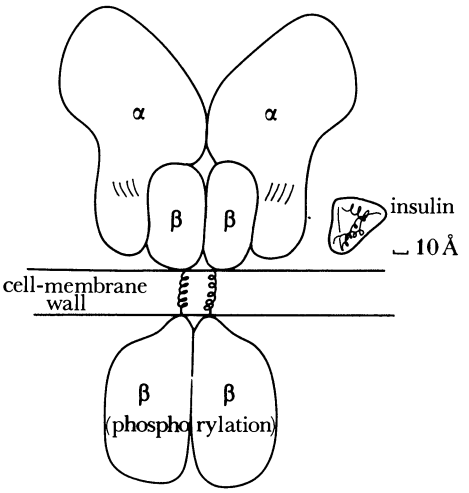


FIGURE 12.1. A schematic view of the insulin receptor and the insulin monomer drawn approximately to scale.

conceivable receptor structure, very diagrammatically, and its size relation to insulin, suggesting that it may partly or wholly envelop the insulin molecule, binding at several points leading to pressures that precipitate conformation changes both in insulin and in remote parts of the receptor. The observations that insulin binds to the receptor α chain, perhaps near the β chain (Massague *et al.* 1981), provide a framework by which it is possible to visualize the complexities of insulin interaction. The β chain itself passes through the membrane and includes within the cell a cytoplasmic region with kinase activity that can be initiated by the distant binding of insulin. Insulin binding has been shown to cause the phosphorylation of β subunit tyrosine residues; these may further affect the enzymes involved in phosphorylation and dephosphorylation in the cell and so may activate the processes of glycolysis, lipogenisis, protein synthesis, etc. (Kasuga *et al.* 1982; Obberghen *et al.* 1983).

From this picture of the course of insulin action, it would be expected that the size and shape of the insulin molecule and the distribution of its surface residues would affect both binding and the course of the various catalytic reactions that follow in possibly different ways. This is sometimes observed; binding and potency in the different biological activities of insulin do not always run parallel in different modifications of the insulin structure (Emdin *et al.* 1977; Schuttler *et al.* 1980). Most of the evidence we have is derived from glucose oxidation in the fat cell (*in vitro*), the mouse convulsion test (*in vivo*) and receptor-binding measurements (*in vitro*), all rather imprecise measurements that do, however, generally agree qualitatively. There are a few examples where potency for glycolysis is different from that for DNA synthesis (King & Kahn 1981).

Our original view of the structure–activity relation was based initially on the distribution of the insulin residues unchanged in Nature. Of those listed in 1969, five, A2 isoleucine, A5 glutamine, B18 valine, B16 tyrosine and B24 phenylalanine, have since been found to change, leaving sixteen still called invariant. These can be divided into two groups.

I	A6 Cys A7 Cys A11 Cys A16 Leu A20 Cys B6 Leu B7 Cys B8 Gly B11 Leu B12 Val B15 Leu B19 Cys B23 Gly
II	A1 Gly A19 Tyr A21 Asn

The first group can all be seen as essential to the construction of the molecule; the second contains surface residues that may be extended to include certain others conserved in all the more active insulins, A2 isoleucine, A3 valine, A5 glutamine, B22 arginine, B24, B25 and B26, all aromatic; they lie on two contiguous surfaces, the B chain non-polar dimer-forming region and a polar surface containing the A chain residues and perhaps B22 arginine. Some of these are in the flexible regions where changes of the molecular arrangement in different crystals affect their conformation.

Figure 12.2, plate 5, illustrates the surfaces containing the exposed residues that have been implicated in the hormone's binding and activity together with residues such as A2 isoleucine, which have an important structural role in supporting the active surface and which are mentioned in the text.

That certain of the group I and II residues are individually important for biological activity has received some confirmation from recently determined sequences of the two insulin-like growth factors, IGFI and II (Blundell & Humbel 1980) and of various relaxins (Hudson *et al.* 1983). The IGFs contain all the invariant insulin residues of type I, the relaxins, the cystines and glycines. Of those of type II the growth factors only lack A21 asparagine, replaced by alanine and a continuing chain. The relaxins, which have no insulin-like activity, on the other hand, lack A21 asparagine and as well B24 and B25 phenylalanine and B26 tyrosine: some also lack A1 glycine and A19 tyrosine. The importance of the B chain aromatic residues to insulin's activity is emphasized.

By now chemical experiments have been carried out introducing modifications to all but six of the invariant residues in group I (A16, B6, B8, B11, B15 and B19) with some reassuring and some puzzling effects (Brandenburg *et al.* 1983; Gammeltoft 1984; Katsoyannis 1979; Marki *et al.* 1979). Replacement of cystine S at 6 and 11 by methyl groups or enlargement of the ring by the introduction of homocystine, reduces but does not abolish activity. The introduction of homocystines at 7 and 20, however, largely extinguishes activity. More curious, garbling of the S–S links, joining A6–B7 and A7–A11 or A6–A7, A11–B7 produces molecules with different physical characteristics to insulin but potencies from 13 to 37 %. In spite of the reorganization of the A chain loop, enough of the rest of the molecule can maintain its normal conformation, including most of the dimer forming and A chain surfaces, and so presumably contribute to activity. Of the remaining structurally involved non-polar residues, A2 isoleucine and B12 valine appear critical. Change of A2 to glycine practically removes activity, to alanine or leucine reduces it. The residue packing against tyrosine A19 is important for the stability of the A chain helix, which changes in solution when A2 isoleucine is changed (Kitagawa *et al.* 1984*b*). The change of B12 valine to isoleucine (D. Brandenburg, personal communication) evidently does not affect the structure but reduces potency to about 10 %; the change of valine to asparagine (Kitagawa *et al.* 1984*a*, *b*) has a profound effect on the formation of the molecule and results in an inactive product, possibly unfolded. This residue is concerned both with the development of the B chain helix, the non-polar core of the molecule and the dimer-forming surface. It is possible that similar profound effects may appear with changes in the other of the non-polar core residues. Again, so far, only the advantage of the glycine conformation with right-handed dihedral angles at B23 on chain bending has been checked, replacement by (−)-alanine reduces, whereas (+)-alanine retains activity (Zhang 1981).

Modification of the invariant A chain surface residues, A1 gly, A19 tyr and A21 asn, generally reduces the activity, although in a variable way. A1 glycine may be replaced by

sarcosine, (+)-alanine and (−)-alanine with 80%, 95% and 5% activity, respectively. Tyrosine A19 is exchanged for phenylalanine with reduction to 22% activity; (+)-tyrosine and (−)-leucine A19 result in an insulin with very low activity. Substitution of A21 by arginine reduces activity, removal practically abolishes it.

The particular importance of B24 and B25 phenylalanine in the B chain surface has been confirmed by observations on the human mutant insulin at B25 → Leu (Chicago) and B24 → Ser (Los Angeles) (Shoelson *et al.* 1983). To check the identities of these, leucine and serine were substituted first at B24 and then at B25 by semisynthesis. It is notable that both B25 mutant insulins have very low activities, *ca.* 1% and that for B25 its structure is undisturbed. A number of other aromatic modifications of these two residues have been made, and their properties explored. In the B25 series potency and binding are retained but reduced to various degrees, becoming very low (*ca.* 1%) when the phenyl group is moved relative to the peptide Cα (Nakagawa *et al.* 1985; Kobayashi *et al.* 1982). This makes it all the more surprising that potency is enhanced in B24 (+)-phenylalanine. Models show, however, that it will fit into the dimer surface if small movements of the B chain occur similar to those found in despentapeptide insulin. Alternatively it can project rather further out of the surface than (−)-phenylalanine.

The wide variation in the potency of natural insulins shows that many other than the invariant residues influence its activity, which is not surprising, given the dimensions of the molecule. It also shows there are regions where the residue character has little effect. That these regions are not concerned with insulin action is confirmed by chain-shortening experiments. Whereas the removal of the first two residues (both invariant) of the A chain gives an inactive product, the removal of the first four residues of the B chain hardly changes the activity; only as B5 histidine is removed does the activity fall. The effects of C terminal B chain removal are more gradual; at first, there is little change; from despentapeptide onwards the activity falls to near zero at the desoctapeptide. It returns as soon as B24 phenylalanine is again built back into the chain. Although the activity of despentapeptide and desheptapeptide insulins is lower than normal, it seems that the position of the charge at the chain terminus is critical here; amidation restores full activity (D. Brandenburg, personal communication). This is particularly important because despentapeptide insulin is monomeric and is structurally very similar to undegraded insulins, confirming other evidence that the monomer is the active species (Shanghai Insulin Group 1976) It also points to a contrast: the power of insulin to dimerize is lost with the loss of B30–B26; only with the loss of B24 is activity largely removed.

There are other complexities to consider. First, there are regions in the molecule where modification of surface residues other than those so far discussed affects activity. For example, the removal of B5 histidine, mentioned above, markedly reduces activity. Near this in space, the change in human insulin of A8 threonine to histidine doubles the activity whereas the mutation A3 Val → Leu reduces activity significantly (Nanjo *et al.* 1986). Both histidine positions are close to the cystine group A7–B7, which, as a surface residue, may itself take part in binding or pressure, although clearly it is also important for structure, under which heading its activity sensitive character was classified. On the other hand there are analogues, for example, hag fish insulin and DAS insulin (where diaminosuberic acid cross links B29 lysine to A1 glycine) that have apparently normal insulin structures and all the residues usually associated with insulins of high activity, but that still have low activity. In hag fish insulin, binding is relatively higher than activity in the fat cell (Cutfield *et al.* 1979). This might be because of the replacement of a residue such as B21 glutamic acid by valine, which increases

binding but changes it in a way prejudicial to activity. In the crosslinked DAS insulin, the low activity favours one of the ideas with which we started, that freedom to move the initial A chain and B chain terminus is important for full activity, (Dodson *et al.* 1983; Peking Insulin Structure Group 1974). These observations and the high activity of despentapeptide insulin suggest the B chain C terminal residues are actually displaced on binding and it is possible that this brings an aliphatic surface including B12 and A3 valine into contact with the receptor surface.

One interesting general observation is that although removal or change of specific individual residues may almost abolish activity it is difficult wholly to extinguish it. Insulins with activities varying 2000-fold have been obtained. Those with very small activities still are able to produce maximum response if present in sufficient quantities. One can imagine a slower cooperative process of fitting to the receptor in the absence of the reactive residues; such a process may be assisted not only by remaining active residues but also by the specific character of the peptide-chain surface atoms. In the reactive regions there are two exposed clusters of carbonyl oxygens, A17, A18 and A20 in the 3.10 helix, and B21 and B22 in a β bend. These in the crystal are hydrated, not in contact with other protein atoms. They may well interact directly to form hydrogen bonds with the receptor, displacing water molecules and making a favourable contribution to the entropy of binding and strengthening its specific character (compare the interchain hydrogen bonds in the molecule and dimer).

In general, it seems clear that the interaction of insulin molecules with their receptor includes most of the factors we have observed in the packing of molecules within insulin crystals. We can identify the individual aromatic residues that probably contact similar groups within the receptor, the likely effects of ionic charge, the possibility of surface peptide interaction and pressure from conflicting contacts leading to conformational change. We cannot yet give any details of the actual course of the reactions that occur. But the future is very promising. The rapid development of genetic-engineering techniques should make it possible soon to obtain sufficient material for the structural analysis of the receptor and of the complexes it forms with insulins of different structure and potencies. This should provide the explanation for the biological behaviour of the insulin molecule.

This research was supported over many years by grants from the Rockefeller Foundation, the Science Research Council, the Medical Research Council, the British Diabetic Association, the Wellcome Foundation and the Wolfson Professorship fund of the Royal Society, which contributed to the cost of apparatus, research expenses and to individual salaries. Fellowships from N.Z. University Grants Committee supported E.N.B. and J.C., from IBM, N.W.I., and from the Royal Society exchange programme with Japan, K.S. and N.S.

Dr Claire Gudex, Linda Cornfield and Elizabeth Hubbard assisted with the computer drawings for the paper. Valmai Firth typed the manuscript, and John Olive prepared many of the very difficult figures.

At critical periods of our researches, we were lent working space in the Department of Pharmacology by Professor W. B. Paton, F.R.S., and later in the Department of Psychology by Professor Weiskrantz, F.R.S., at Oxford University. The crystalline insulin we used was given by Dr Schlichtkrull, Novo Terapeutisk Laboratories, Copenhagen. We acknowledge gratefully all this support and also many useful discussions with colleagues working on insulin, particularly D. Brandenburg, S. Gammeltoft, P. Katsoyannis, Liang Dong-cai, D. F. Steiner, D. Stuart, Wang Ying Lai and H. Zahn.

References

Abel, J. J. 1926 Crystalline insulin. *Proc. natn. Acad. Sci. U.S.A.* **12**, 132–135.

Adams, M. J. Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, R. & Sheet, S. 1969 Structure of rhombohedral 2 zinc insulin crystals. *Nature, Lond.* **224**, 491–496.

Admiraal, G. & Vos, A. 1983 Structure of the tetrapeptide L-methionyl L-alpha-glutamyl L-histidyl L-phenylalanine monohydrate. *Acta crystallogr.* C**39**, 82–87.

Agarwal, R. C. 1978 New least squares refinement techniques based on fast Fourier transform algorithm. *Acta crystallogr.* A**34**, 791–809.

Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C. & Sternberg, M. J. E. 1979 Crystallographic studies of the dynamic properties of lysozyme. *Nature, Lond.* **280**, 563–568.

Baker, E. N. & Hubbard, R. E. 1984 H-bonding in globular proteins. *Prog. Biophys. molec. Biol.* **44**, 97–179.

Bentley, G. A., Dodson, E. J., Dodson, G. G. & Levitova, A. 1978 Zinc-free cubic pig insulin: crystallization and structure determination. *J. molec. Biol.* **125**, 387–396.

Bernstein, F. C., Koetzle, T. F., Williams, G. B., Meyer, G. F., Price, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. 1977 The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. molec. Biol.* **122**, 535–542.

Bi, R. C., Dauter, Z., Dodson, E. J., Dodson, G. G., Giordano, F. & Reynolds, C. D. 1984 Insulin structure as a modified and monomeric molecule. *Biopolymers* **32**, 391–395.

Bi, R. C. *et al.* 1983 Structural changes in the monomeric despentapeptide (B30–B26) insulin crystal. *Proc. Indian Acad. Sci.* (*Chem. Sci.*) **92**, 478–483.

Blundell, T. L. & Humbel, R. E. 1980 Hormonal families – pancreatic hormones and homologous growth factors. *Nature, Lond.* **287**, 781–784.

Blundell, T. L. & Pearl, L. 1984 Structural studies of the aspartic proteinases. *FEBS Lett.* **174**, 96–100.

Blundell, T. L., Dodson, G. G., Hodgkin, D. C. & Mercola, D. A. 1972 Insulin: the structure in the crystal and its reflection in chemistry and biology. *Adv. Protein Chem.* **26**, 279–402.

Bolin, J. T., Filman, D. J., Mathews, D. A., Hamlin, R. C. & Kraut, J. 1982 Crystal structures of *E. coli* and *L. casei* dihydrofolate reductase refined at 1.7 Å resolution. *J. biol. Chem.* **257**, 3650–3662.

Brandenburg, D., Saunders, D. & Schuettler, A. 1983 Pancreatic hormones. In *Amino acids, peptides and proteins, specialist reports, London*, Vol. 14, pp. 461–476.

Brill, A. S. & Venable, J. H. 1967 Effects of site symmetry and sequential metal binding upon protein titration (zinc insulin). *J. Am. chem. Soc.* **89**, 3622–3627.

Brill, A. S. & Venable, J. H. 1968 The binding of transition metal ions in insulin crystals. *J. molec. Biol.* **36**, 343–353.

Broomhead, J. M. & Nichol, A. D. I. 1948 The crystal structures of zinc and magnesium benzene sulphonates. *Acta crystallogr.* **1**, 88–92.

Brown, H., Sanger, F. & Kitai, R. 1955 The structure of pig and sheep insulins. *Biochem. J.* **60**, 556–565.

Candeloro, S. de S., Grdenit, D., Taylor, N. & Hodgkin, D. C. 1973 The structure of ferroverdin. II. Rhombohedral ferroverdin crystals. *Proc. R. Soc. Lond.* B **184**, 137–148.

Chothai, C. & Janin, J. 1975 Principles of protein:protein recognition. *Nature, Lond.* **256**, 705–708.

Chothai, C., Dodson, G. G., Hodgkin, D. C. & Lesk, A. 1983 Transmission of conformational change in insulin. *Nature, Lond.* **302**, 500–505.

Chothia, C., Levitt, M. & Richardson, D. 1977 Structure of proteins; packing of α-helices and β-pleated sheets. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4130–4134.

Chothia, C., Levitt, M. & Richardson, D. 1981 Helix to helix packing in proteins. *J. molec. Biol.* **145**, 215–250.

Crowfoot, D. M. 1935 X-ray single-crystal photography on insulin. *Nature, Lond.* **135**, 591–592.

Crowfoot, D. M. 1937 Two crystalline modifications of insulin. *Nature, Lond.* **140**, 149–150.

Crowfoot, D. M. & Riley, D. 1939 X-ray measurements on wet insulin crystals. *Nature, Lond.* **144**, 1011–1012.

Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Emdin, S. O. & Reynolds, C. D. 1979 Structure and biological activity of hagfish insulin. *J. molec. Biol.* **132**, 85–100.

Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Reynolds, C. D. & Vallely, D. 1981 Similarities and differences in the crystal structures of insulin. In *Structural studies on molecules of biological interest* (ed. G. G. Dodson, J. Glusker &.D. Sayre), pp. 527–546. Oxford University Press.

Dodson, E. J. 1981 Block diagonal least squares refinement using fast Fourier techniques. In *Refinement of protein structures, Proceedings of the Daresbury Study Weekend* (ed. P. Machin, I. Campbell & M. Elder), pp. 29–39. SERC Daresbury.

Dodson, E. J., Dodson, G. G. & Hodgkin, D. C. 1980 The conformations observed in the N terminal A chain residues of insulin. In *Frontiers of bioinorganic chemistry and molecular biology* (ed. S. N. Ananchenko), pp. 145–150. Oxford: Pergamon Press.

Dodson, E. J., Isaacs, N. W. & Rollett, J. S. 1976 Block diagonal least square refinement using fast Fourier techniques. *Acta crystallogr.* A**32**, 311–320.

Dodson, E. J., Dodson, G. G., Hodgkin, D. C. & Reynolds, C. D. 1979 Structural relationships in the 2Zn insulin hexamer. *Can. J. Biochem.* (Best Memorial Volume) **57**, 469–479.

Dodson, G. G., Hubbard, R. E. & Reynolds, C. D. 1983 Insulin's structural variations and their relation to activity. *Biopolymers* **22**, 281–292.

Dodson, G. G. 1981 Some refinement experiences with 2Zn insulin. In *Refinement of Protein Structures, Proceedings of the Daresbury Study Weekend* (ed. P. Machin, J. Campbell & M. Elder), pp. 95–98. SERC Daresbury.

Ebina, Y., Leland, E., Jarnagin, K., Eden, M., Lanzlo, G., Clauser, E., Ou, J. H., Masiarz, F., Kan, Y. N., Goldfine, I. D., Roth, R. A. & Rutter, W. J. 1985 The human insulin receptor cDNA; the structural basis for hormone activated transmembrane signalling. *Cell* **40**, 747–758.

Emdin, S. O., Gammeltoft, S. & Gliemann, J. 1977 Degradation, receptor binding affinity and potency in insulin from Atlantic hagfish. *J. biol. Chem.* **252**, 602–608.

Gammeltoft, S. 1984 Insulin receptors binding kinetics and structure function relationship of insulin. *Physiol. Rev.* **64**, 1321–1378.

Goldman, J. & Carpenter, F. 1976 Zinc binding, circular dichroism and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives. *Biochemistry* **13**, 4566–4574.

Harding, M. M. & Cole, S. J. 1963 The crystal structure of di(histidino) zinc pentahydrate. *Acta crystallogr.* **16**, 643–650.

Hargreaves, A. 1957 Crystal structure of zinc *p*-toluene sulphonate hexahydrate. *Acta crystallogr.* **10**, 191–195.

Hendrickson, W. A. & Teeter, M. M. 1981 Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature, Lond.* **290**, 107–113.

Hermans, J. 1985 Potential functions for simulation of large molecules. In *Proceedings of a Workshop*; *Molecular Dynamics and Protein Structure* (ed. J. Hermans), pp. 31–34. Western Springs, Illinois: Polycrystal Book Service.

Höhne, E. & Kretschner, G. 1982 New interpretation of helical structures in polypeptides. *Studia Biophysica* **87**, 23–28.

Hol, W. 1978 The α helix dipole and the properties of proteins. *Nature, Lond.* **273**, 443–446.

Hudson, P., Huley, J., John, M., Cronk, M., Crawford, R., Haralambidis, J., Tregear, G., Shire, J. & Niall, H. 1983 Structure of a genomic clone encoding biologically active human relaxin. *Nature, Lond.* **301**, 628–630.

Jeffrey, P. D. & Coates, J. H. 1966 An equilibrium ultracentrifuge study of the self-association of bovine insulin. *Biochemistry* **5**, 489–498.

Kasuge, M., Zick, Y., Blith, D. L., Karkson, F. A., Harring, H. U. & Kahn, C. R. 1982 Insulin stimulation of phosphorylation of the β subunit of the insulin receptor: formation of both phosphoserine and phosphotyrosine. *J. biol. Chem.* **257**, 9891–9894.

Katsoyannis, P. G. 1979 New synthetic insulins. In *Treatment of early diabetes* (ed. R. A. Camerini-Davalos & B. Hanover), pp. 319–328. New York: Plenum Publishing Corporation.

King, G. L. & Kahn, R. C. 1981 Non parallel evolution of metabolic and growth promoting functions. *Nature, Lond.* **292**, 644–646.

Kitagawa, K., Ogawa, H., Burke, G. T., Chanley, J. D. & Katsoyannis, P. G. 1984*a* Interaction between the A2 and A19 amino acid residues is of critical importance for high biological activity in insulin. *Biochemistry* **19**, 4444–4448.

Kitagawa, K., Ogawa, H., Burke, G. T., Chanley, J. D. & Katsoyannis, P. G. 1984*b* The critical role of the A^2 amino acid residue in the biological activity of insulin: [2-glycine-A] insulin and [2-alanine-A] insulin. *Biochemistry* **23**, 1405–1413.

Kobayashi, M., Ohgaken, S., Iwasaki, M., Maegawa, H., Shigata, Y. & Inouye, K. 1982 Supernormal insulin (D-PHEB24) insulin with increased affinity for insulin receptors. *Biochem. biophys. Res. Commun.* **107**, 329–336.

Konnert, J. H. & Hendrickson, W. A. 1980 A restrained parameter thermal factor refinement procedure. *Acta crystallogr.* A**36**, 344–350.

Langs, D. A., Smith, D. G., Stezowski, J. J. & Hughes, R. E. 1986 Structure of pressinoic acid: the cyclic moiety of vasopressin. *Science, Wash.* **232**, 1240–1242.

Marki, F., de Gasparo, M., Eisler, K., Kanker, B., Riniker, B., Rottel, W. & Siever, P. 1979 Synthesis and biological activity of 17 analogs of human insulin. *Hoppe-Seyler's Z. physiol. Chem.* **360**, 1619–1632.

Massague, J., Pilch, P. F. & Czech, M. P. 1981 A unique proteolytic cleavage site on the β subunit of the insulin receptor. *J. biol. Chem.* **256**, 3182–3190.

Nakagawa, S. H. & Tager, H. S. 1988 Role of phenylalanine B25 side chain in directing insulin interaction with its receptor. *J. biol. Chem.* (In the press.)

Nakagawa, S. H., Sheet, M. J. & Tager, H. S. 1985 Insulin analogues bearing unnatural aromatic amino acid substitutions at B25. In *Peptide Structure and Function, Proceedings of the 9th American Peptide Symposium* (ed. C. M. Deber, V. J. Hruby & K. D. Kopple), pp. 695–698.

Nanjo, K., Snake, T., Miyano, M., Okai, K., Sowa, R., Kondo, M., Nishimura, S., Iua, K., Miyurama, I. C., Given, B. D., Chan, S. J., Tager, H., Steiner, D. & Rubinstein, A. 1986 Diabetes due to secretion of a structurally abnormal insulin (insulin Wakayama). *J. clin. Invest.* **77**, 514–519.

North, A. C. T., Phillips, D. C. & Matthews, F. S. 1968 A semi-empirical method of absorption correction. *Acta crystallogr.* A**24**, 351–359.

van Obberghen, E., Rossi, B., Kowakski, A., Euzzano, H. & Ponzio, G. 1983 Receptor mediated phosphorylation of the hepatic insulin receptor. Evidence that the M_r 95,000 receptor subunit is its own kinase. *Proc. natn. Acad. Sci. U.S.A.* **80**, 945–949.

Pauling, L., Corey, R. B. & Branson, H. R. 1951 Two hydrogen bonded helical configurations of the polypeptide chain. *Proc. natn. Acad. Sci. U.S.A.* **37**, 205–211.

Pauling, L. & Corey, R. B. 1951 Configuration of polypeptide chains with favoured orientations around single bonds: two new pleated sheets. *Proc. natn. Acad. Sci. U.S.A.* **37**, 729.

Peking Insulin Structure Group 1971 Insulin's crystal structure at 2.5 Å resolution. *Peking Rev.* **40**, 11–16.

Peking Insulin Structure Group 1974 Studies on the mechanisms of insulin action. *Sci. Sin.* **17**, 752–778.

Sakabe, N., Sakabe, K. & Sasaki, K. 1985 X-ray studies of water structure on 2Zn insulin crystals. In *Proc. Int. Symp. Biomol. Struct. Interactions. J. Biosci.* (Suppl.) **8**, 45–55.

Schlichtkrull, J. 1958 *Insulin crystals.* Copenhagen: Munksgaard.

Schutler, K., Peterson, K. G., Schuttler, A., Brandenburg, D. & Kerp, L. 1980 Biological activity and receptor binding of six different covalent dimers of insulin. In *Insulin chemistry, structure and function and insulin and related hormones* (ed. D. Brandenburg & A. Wollmer), pp. 433–438. Berlin: de Gruyter.

Shoelson, S. M., Hameda, P., Blix, A., Nanjo, T., Sanke, K., Inouyo, D., Steiner, D., Rubenstein, A. & Tager, H. 1983 Three mutant insulins in man. *Nature, Lond.* **302**, 540–543.

Sieber, P., Eisler, K., Kamber, P., Riniker, B., Rittel, W., Marki, R. & de Gasparo, M. 1978 Synthesis and biological activity of 2 disulfide bond isomers of human insulin. *Hoppe-Seyler's Z. physiol. Chem.* **359**, 113–123.

Stillinger, F. H. & David, C. W. 1978 Polarization model for water and its ionic dissociation products. *J. chem. Phys.* **69**, 1473–1484.

Tager, H., Given, D., Baldwin, D., Maks, M., Markese, J., Rubinstein, A., Olefsky, J., Kobayashi, M., Koltermann, O. & Poucher, R. 1979 A structurally abnormal insulin causing human diabetes. *Nature, Lond.* **281**, 122–175.

Teeter, M. M. 1984 Water structure of a hydrophobic protein at atomic resolution – pentagon rings of water molecules in crystals of crambin. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6014–6018.

Ullrich, A., Bell, J. R., Chen, E. Y., Herrera, R., Petrucelli, L. M., Dull, T. J., Gray, A., Coussens, L., Liao, Y. C., Tsubokawa, M., Mason, A., Seeburg, P. H., Grumfeld, C., Rosen, O. M. & Ramachandran, J. 1985 Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. *Nature, Lond.* **313**, 756–761.

Walter, R., Schwartz, I. L., Darrell, J. H. & Urry, D. W. 1971 Relation of conformation of oxytocin to biology of neurohypophyseal hormones. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1353–1359.

Watenpaugh, K. D., Margulis, T. N., Sieker, L. C. & Jensen, L. 1978 Water structure in a protein crystal: rubredoxin at 1.2 Å resolution. *J. molec. Biol.* **122**, 175–190.

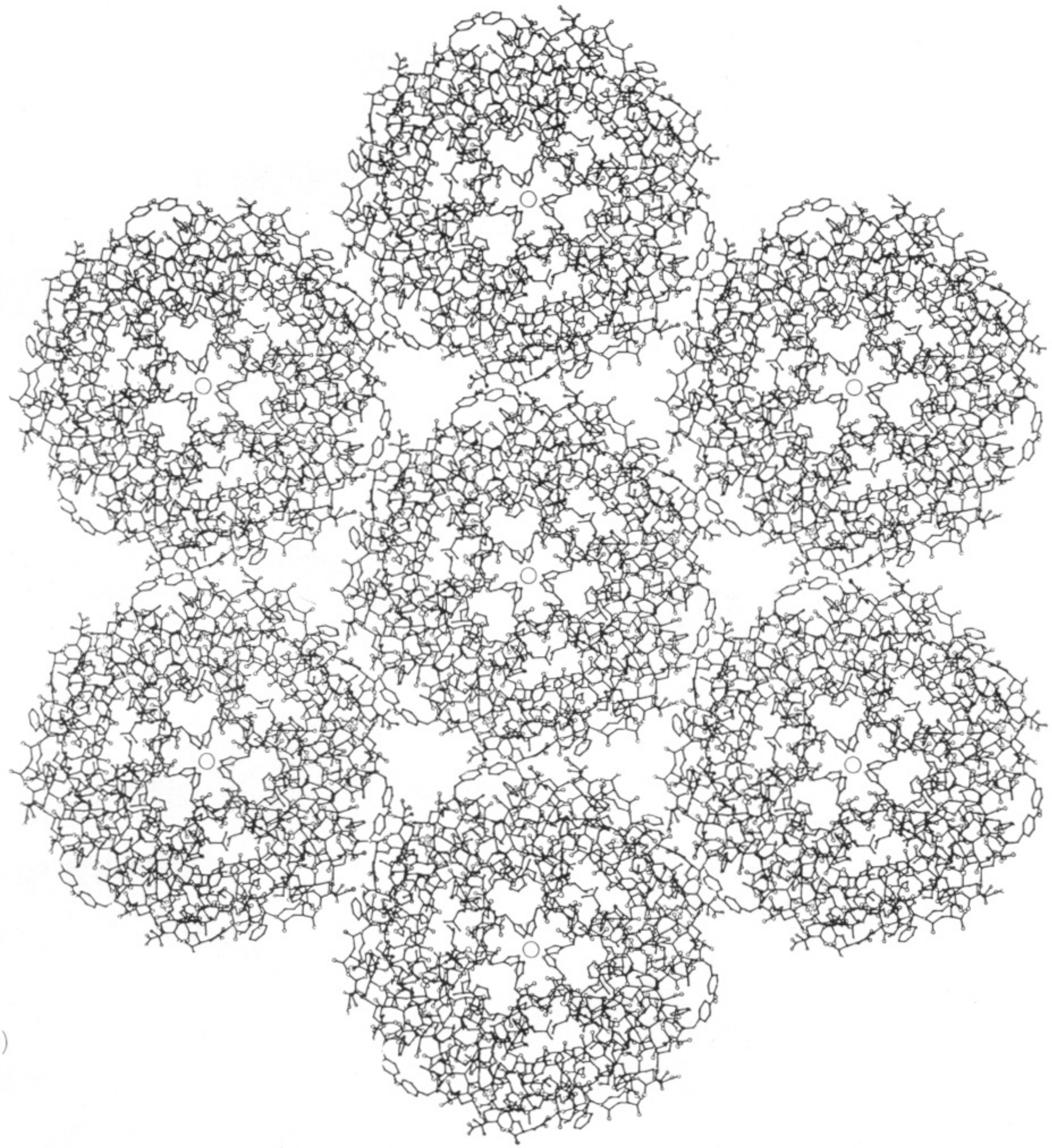
Wollmer, A., Strassburger, W., Glatter, U., Dodson, G. G., McCall, M., Danks, W., Brandenburg, D., Gattner, H. G. & Ruttel, W. 1981 2 mutant forms of human insulin structural consequences of the substitution of invariant B24-phenylalamine of B25-phenylalamine by leucine. *Hoppe-Seyler's Z. physiol. Chem.* **362**, 581–591.

Wood, S. P., Tickle, J. J., Treharne, A. C., Potts, J. E., Mascarenhas, Y., Li, Y.-J., Husain, J., Cooper, S., Blundell, T. L., Honby, V. J., Wyssbrod, H. R., Buka, A. & Frichmann, A. J. 1986 Crystal structure analysis of deamino-oxytocin: conformational flexibility and receptor binding. *Science, Wash.* **232**, 633–636.

Zhang, Y.-S. 1981 The study of the insulin molecule. In *Structural studies on molecules of biological interest* (ed. G. Dodson, J. Glusker & D. Sayre), pp. 492–500. Oxford University Press.

(*a*) (*b*)

FIGURE 2.1. (*a*) The atoms in 42 2Zn insulin molecules (excluding hydrogen) projected along the crystal threefold axis. (*b*) Overlay showing water molecule positions found in the same volume. Not all positions are fully occupied.



(*a*)

FIGURE 2.1. (*a*) The atoms in 42 2Zn insulin molecules (excluding hydrogen) projected along the crystal threefold axis. (*b*) Overlay showing water molecule positions found in the same volume. Not all positions are fully occupied.

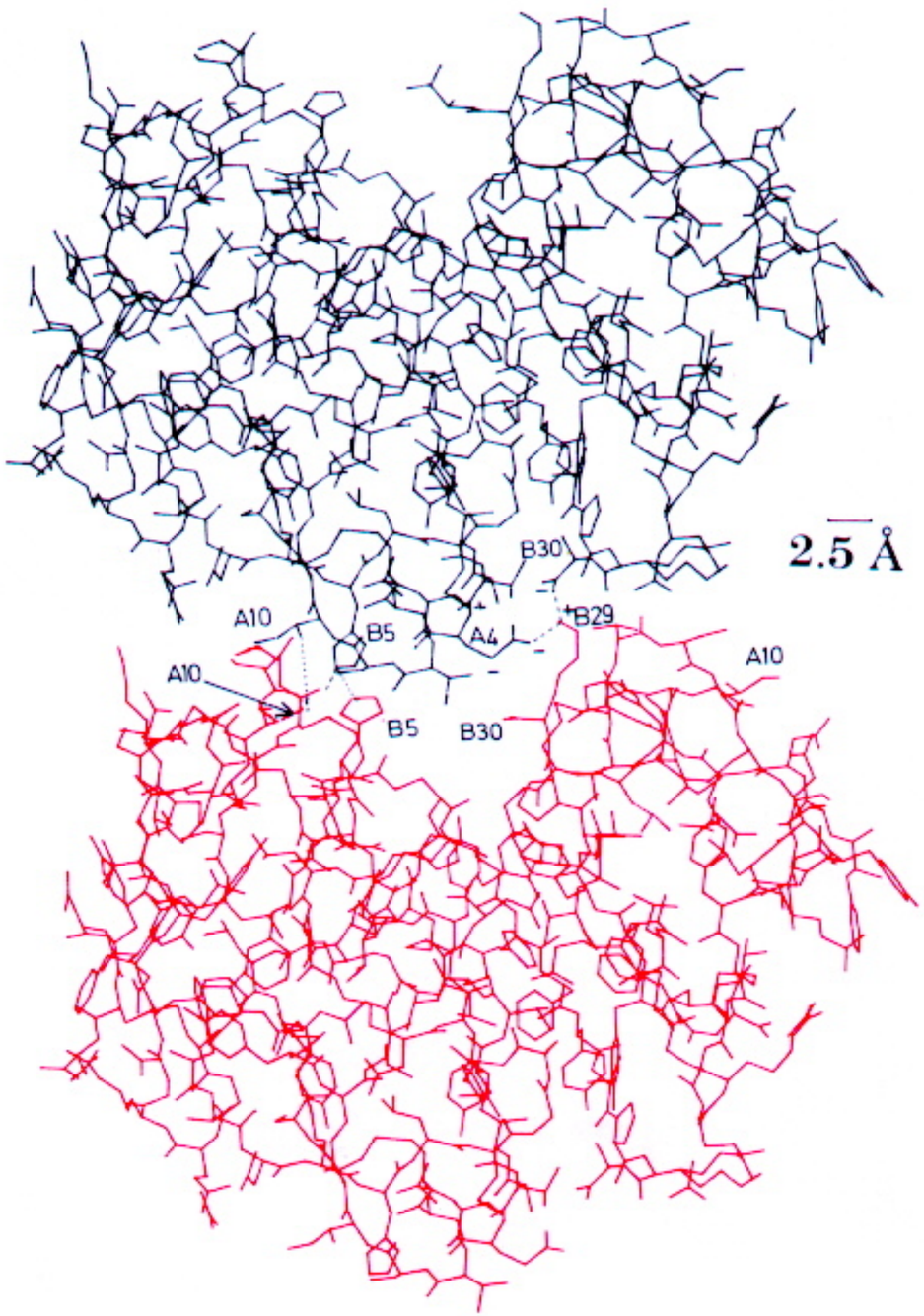


FIGURE 2.2. Projection of the atomic positions, excluding hydrogen, perpendicular to the crystal threefold axis. The positions of three insulin molecules from the hexamer are repeated by one unit cell translation along the ***c*** axis.

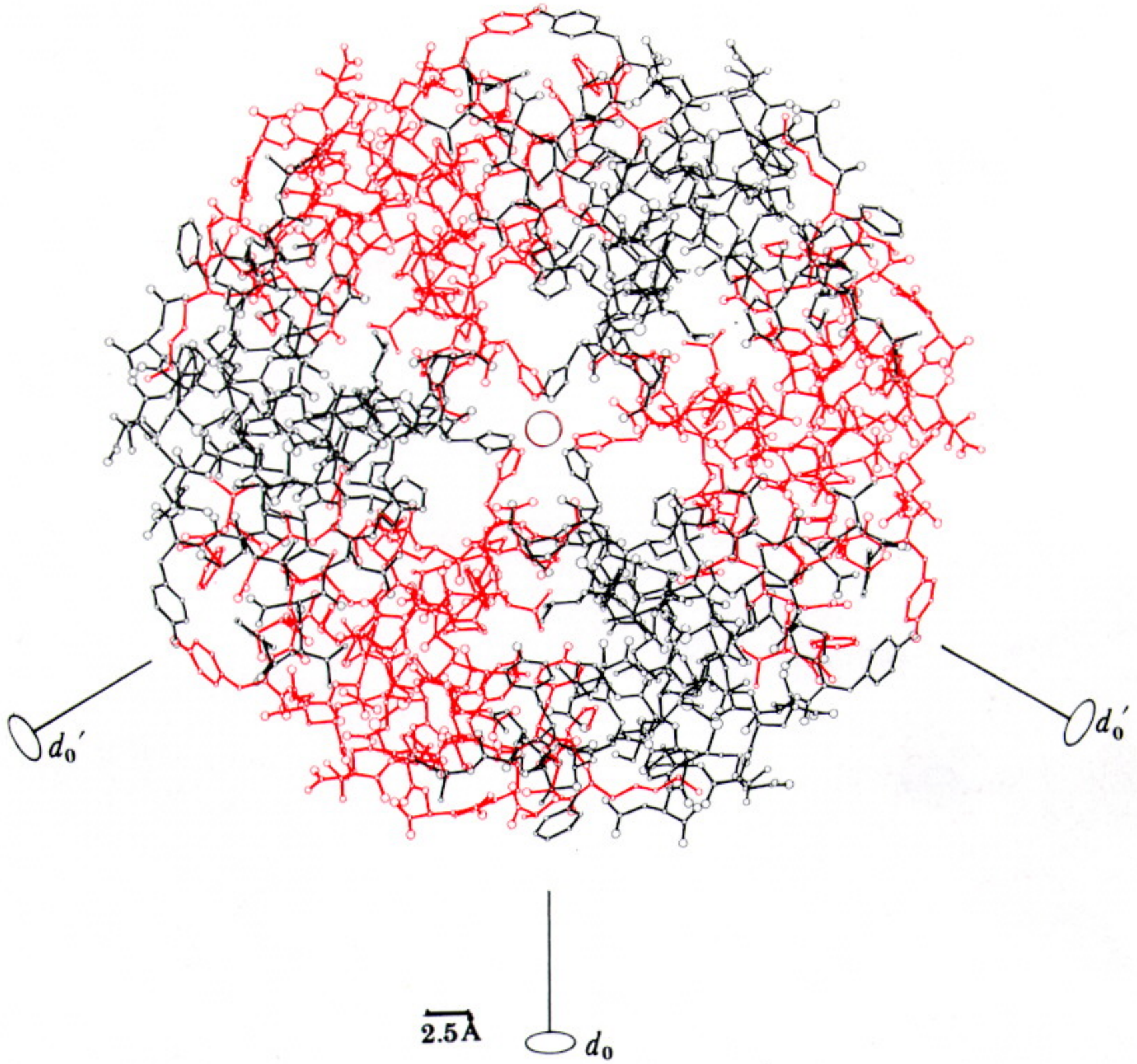


FIGURE 2.3. Atoms in the 2Zn insulin hexamer viewed along the crystal theefold axis. The molecules in red represent molecule 1 and those coloured black, molecule 2. The directions of d_0 and d_0' are shown.

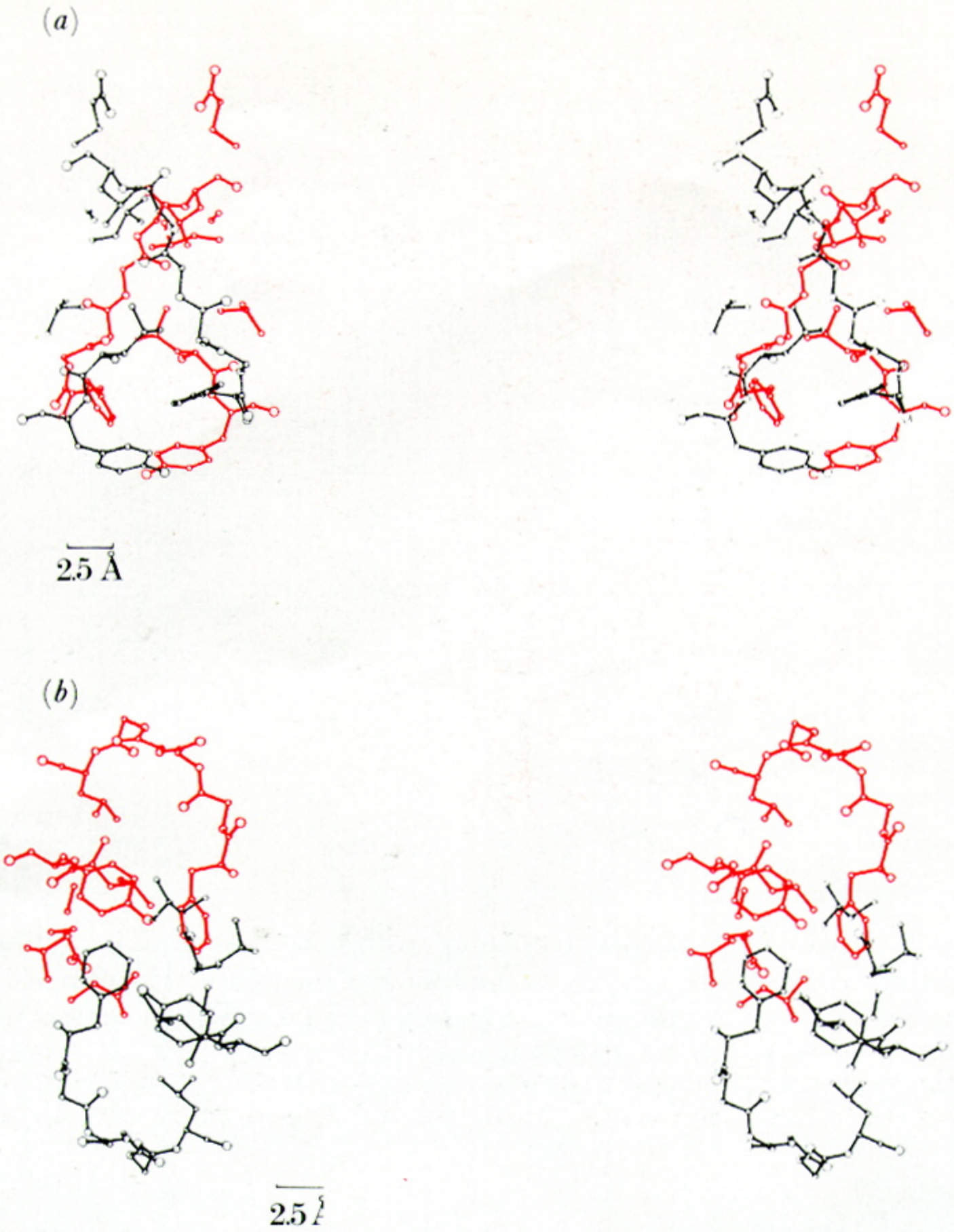


FIGURE 7.4. The hexamer-forming contacts between adjacent dimers. The stereo view (*a*) down the threefold axis and (*b*) down the local two-fold axis. The mono view perpendicular to the twofold axis (*c*) shows the relatively loose packing emphasized by the van der Waals surfaces shown in (*d*).

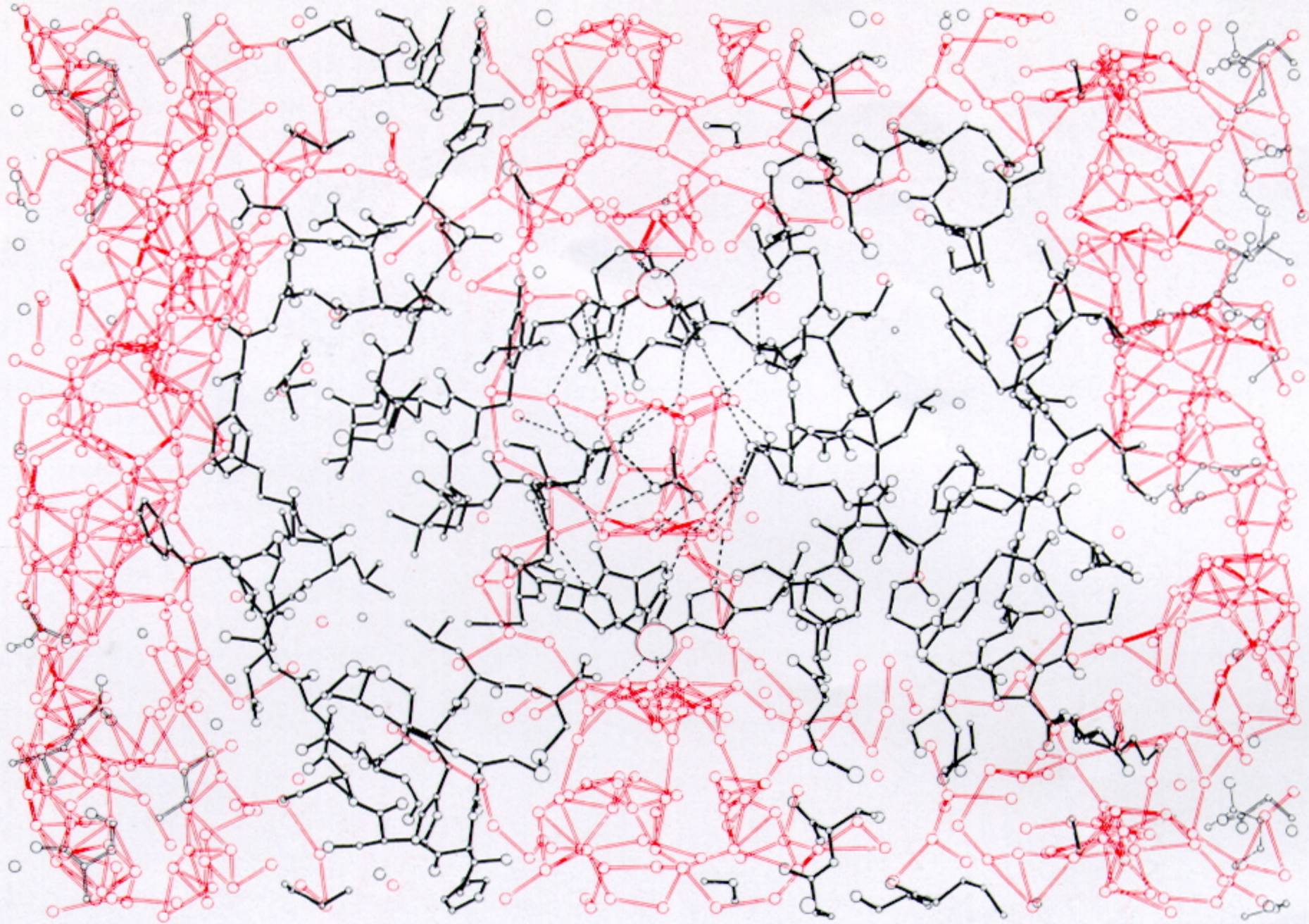


FIGURE 9.7. A view of the 2Zn insulin crystal perpendicular to the threefold axis including both the theefold axis and the 3_2 screw axis. The two axial zinc ions are in the centre of the figure. The dimensions of the box shown are 54 Å in x and 34 Å in z; the box is 6 Å in depth. Contacts between water molecules of less than 3.5 Å are drawn in red lines; not all of these are H bonds. The H bonds between some protein atoms and water molecules in the central cavity are indicated by broken lines.

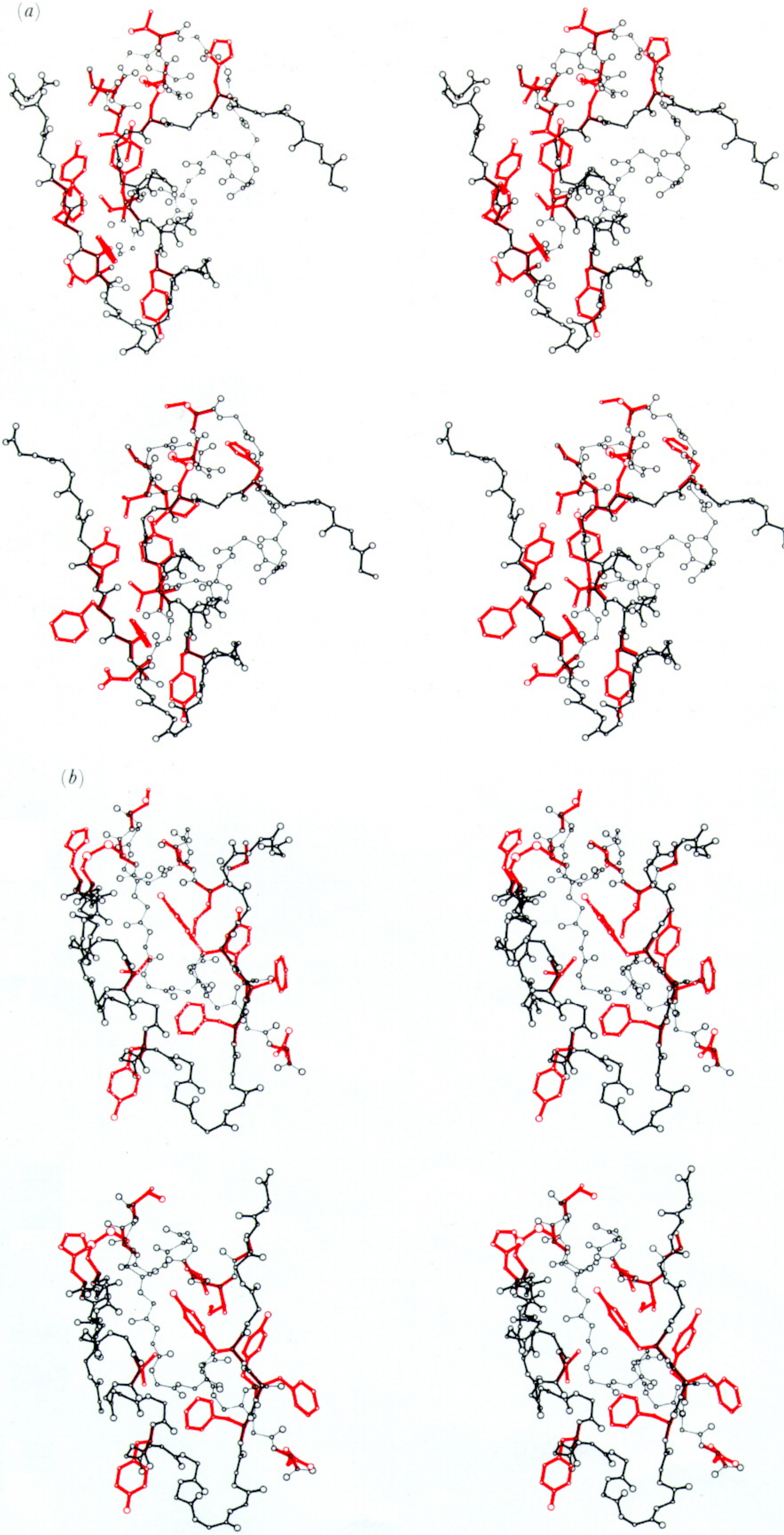


FIGURE 12.2. Stereo pictures of molecule 1 (upper) and molecule 2 (lower) showing the spatial distribution of residues whose structure appears important in insulin's binding and biological action. The residues that appear to be involved in activity are drawn in red, for the remainder of the molecule only the main chain is shown, the A chain as thin lines, the B chain as thicker. The six carbonyl oxygens possibly available for H bonding are filled in red. In (*a*) the view is perpendicular to the threefold axis and 30° away from the direction of the local twofold axis of the dimer. In (*b*) the view is perpendicular to the local twofold axis and faces the dimer-forming surface.