LIDE, D. R. (1960). *J. Chem. Phys*, **33**, 1514–1518.
McMULLAN R. K. (1976). Unpublished work.
MEULENAER, J. DE & TOMPA, H. (1965). *Acta Cryst.* **19**, 1014–1018.
NYBURG, S. C., FAERMAN, C. H. & PRASAD, L. (1987). *Acta Cryst.* **B43**, 106–110.
SAWZIK, P. & CRAVEN, B. M. (1979). *Acta Cryst.* **B35**, 895–901.

SCHOMAKER, V. & TRUEBLOOD, K. N. (1968). *Acta Cryst.* **B24**, 63–76.
TEMPLETON, L. K. & TEMPLETON, D. H. (1973). Am. Crystallogr. Assoc. Meet., Storrs, CT, Abstracts, p. 143.
WEBER, H. P. & CRAVEN, B. M. (1983). *NOOT.* Tech. Rep. Department of Crystallography, Univ. of Pittsburgh, USA.
WEBER, H. P., CRAVEN, B. M. & McMULLAN, R. K. (1983). *Acta Cryst.* **B39**, 360–366.

Structure of the Pig Insulin Dimer in the Cubic Crystal

BY J. BADGER*

Department of Physics, University of York, York Y01 5DD, England

M. R. HARRIS

Astbury Department of Biophysics, University of Leeds, Leeds LS2 9JT, England

C. D. REYNOLDS†

Department of Chemistry, University of York, York Y01 5DD, England

A. C. EVANS‡

Astbury Department of Biophysics, University of Leeds, Leeds LS2 9JT, England

E. J. DODSON AND G. G. DODSON

Department of Chemistry, University of York, York Y01 5DD, England

AND A. C. T. NORTH

Astbury Department of Biophysics, University of Leeds, Leeds LS2 9JT, England

Abstract

Atomic coordinates for pig insulin in the cubic crystal have been refined by reciprocal-space methods to an *R* factor of 0·173 for data between 10·0 and 1·7 Å resolution with structure-factor amplitudes greater than two standard deviations. Stereochemical parameters for the refined model are close to standard values and the estimated error in the positions of well-ordered atoms is about 0·1 Å. Residues directly involved in the formation of the exact (crystallographic) cubic insulin dimer are oriented similarly to those in the non-crystallographic 2Zn insulin dimer. Other residues, which make different molecular contacts in the different crystal forms, have locally altered conformations. The cubic insulin molecule is significantly more similar to one of the two independent molecules in the 2Zn insulin dimer than the other. This more similar molecule is expected to be the more stable conformer.

1. Introduction

Insulin is a small (51 amino acid) hormone important for regulating the cellular uptake of glucose and which, over longer time periods, is necessary for normal cell growth and proliferation. In solution, pig insulin forms monomers, dimers or hexamers depending on the concentration. Insulin molecules are stored as hexameric arrays in the pancreatic B

* Current address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254, USA.

† Current address: Biophysics Laboratory, Liverpool Polytechnic, Liverpool L3 3AF, England.

‡ Current address: Department of Neurology and Neurosurgery, Montreal Neurological Institute, 3801 University Street, Montreal H3A 2B4, Quebec, Canada.

cells but at concentrations prevalent in the bloodstream the predominant aggregation state is the monomer (Blundell, Dodson, Hodgkin & Mercola, 1972). The crystal structure of pig insulin has been refined at high resolution (Dodson, Dodson, Hodgkin & Reynolds, 1979) and analysed in some detail in two hexameric forms that crystallize in space group R3. In 2Zn insulin (Baker *et al.*, 1988) three dimers are organized about two axial zinc ions to form the hexamer. The two monomers that form the dimeric asymmetric unit are generally similar but not identical. When the salt concentration in the crystallization medium is increased, the so-called 4Zn insulin (Cutfield, Cutfield, Dodson, Dodson, Reynolds & Valley, 1981) is obtained. In this form the two monomers in the dimer differ more extensively and there is an additional off-axial zinc site between dimers. Most recently the structure of a symmetric insulin hexamer has been obtained by growing crystals in the presence of phenol (Derewenda *et al.*, 1989) whilst preliminary studies of a number of other variants have been reported (De Graaff, Lewit-Bentley & Tolley, 1981). A further insulin form which has been analysed at high resolution is the monomeric despentapeptide insulin which lacks five resides from the C terminus of the B chain (Bi, Dauter, Dodson, Dodson, Giordano & Reynolds, 1984).

This paper reports the structure of dimeric insulin in the cubic crystal (space group $I2_13$ with $a = 78.9$ Å) which is obtained in the absence of zinc at about pH 9. There are 24 molecules contained within the unit cell forming 12 exact (crystallographic) dimers (Dodson, Dodson, Lewitova & Sabesan, 1978). The symmetric dimer in this cubic crystal provides a standard for characterizing the departures from symmetry in the two molecules within the 2Zn insulin dimer. By comparing different crystal forms information may be obtained on the effect of packing forces on insulin conformation and the degree of rigidity in different parts of the molecule. Atomic coordinates for a dimeric pig insulin will provide a better basis for structural comparisons with the dimeric hagfish insulin structure (Cutfield, Cutfield, Dodson, Dodson, Emdin & Reynolds, 1979) than the hexameric forms.

The development and application of refinement methods for protein structures has made it possible to obtain reliable and detailed atomic models. Accurately determined coordinates are particularly important for making close comparisons between similar structures. The structure of 2Zn insulin has been very extensively refined and analysed at 1.5 Å resolution (Baker *et al.*, 1988). The refinement of cubic insulin at 1.7 Å resolution by reciprocal-space least-squares procedures is outlined in this report and the resulting model discussed with reference to the 2Zn form.

2. Experimental

2.1. Overall refinement strategy

In least-squares refinements of protein structure two differing strategies have been commonly employed. One approach has been to refine atomic parameters subject to stereochemical restraints (Konnert, 1976; Jack & Levitt, 1978). The alternative method is to intersperse cycles of structure-factor refinement with separate steps of model idealization (Isaacs & Agarwal, 1978). Parallel but independent refinements of the cubic insulin structure were carried out by both these methods. A comparison of the two sets of refined coordinates against an electron density map was then made so as to locate errors in the resulting models and in order to construct an improved combined model for further refinement.

2.2. Previous work

X-ray data extending to 1.7 Å resolution were processed (Oatley & French, 1982) to a final merging R factor of 0.07 (Dodson *et al.*, 1978). Coordinates for molecule (1) of 2Zn insulin from which residues B1–B4, B29–B30 and side chains B5, B6 and B10 had been removed were used in the calculation of rotation–translation functions. Placement of this model at the function minima gave an R factor of 0.51. Preliminary map fitting and refinement cycles reduced the R factor to 0.38 and provided a complete protein model (Evans, 1979). This model was the basis for the refinement studies described here.

2.3. Refinement by restrained least squares, refinement (1)

The progams PROTIN (to prepare the input file of atomic coordinates) and PROLSQ (to perform the minimization) were used for this refinement (Hendrickson & Konnert, 1980). Standard geometry restraints (Table 1) were used and the structure-factor data were weighted by the standard deviations $\sigma = 0.5\langle|F_o - F_c|\rangle$. Cycles of least-squares refinement were interspersed with examination of maps (calculated from coefficients $[(2F_o - F_c)\exp i\alpha$ or $(F_o - F_c)\exp i\alpha]$ displayed on an Evans and Sutherland PS300 running the FRODO program package (Jones, 1978). Occasionally 'omit maps', from which uncertain portions of the model had been removed, were calculated to check particular atomic positions. The entire refinement required only 18 cycles of structure-factor calculations during which the R factor fell from 0.386 to 0.202. Refinement of the protein structure was largely automatic. The major changes made by hand involved the insertion of solvent molecules and building second conformations for side chains A5 Gln and B29 Lys. The final

Table 1. *Refinement statistics for cubic insulin models*

Target restraint values used in the final refinement cycles are given. Distances are given in Å and angles are in °.

	Refinement (1)	Refinement (2)	Final	Target
R factors				
All data	0·202	0·204	0·199	–
10–1·7 Å data with $F/\sigma > 2$	0·184	0·189	0·173	–
Geometric terms				
Bonded distances (1–2)	0·027	0·019	0·012	0·02
Angle distances (1 3)	0·081	0·043	0·037	0·04
Planar distances (1–4)	0·080	0·085	0·044	0·05
Chiral volumes	0·218	0·745	0·138	0·15
Planes	0·020	0·017	0·010	0·02
van der Waals radii				
Single torsion	0·176	0·184	0·162	0·50
Multiple torsion	0·472	0·259	0·285	0·50
Planar ω angles	15·1	4·6	2·1	3·0
χ_1 angles	15·1	19·3	12·3	15·0
χ_2 angles	28·5	22·4	27·6	20·0

model resulting from this refinement (Harris, 1985) contained all protein atoms and 102 solvent atoms.

2.4. *Refinement by the method of alteration, refinement* (2)

Since the electron density for the side chain of residue *B*29 Lys and residue *B*30 Ala was weak the workers carrying out this refinement chose to remove these atoms from the initial model. Otherwise the starting model was the same as for refinement (1). Structure-factor calculations were carried out by Fourier inversion of model electron density (Agarwal, 1978) and model idealization by the *MODELFIT* program (Dodson, Isaacs & Rollett, 1976). Typically three or four cycles of structure-factor refinement with respect to atomic positions would be followed by one or two cycles of temperature-factor refinement and then model regularisation. For the geometry idealization step standard deviations in atomic positions were associated with the individual atomic temperature factors such that $\sigma = 0.2(B/8\pi^2)^{1/2}$ (Isaacs & Agarwal, 1978). Standard deviations used in weighting the geometry terms were 0·02 Å for covalent bond distances, 2° for angles and 0·01 Å for deviations from planarity for peptide planes (except for C*A*'s for which a target value of 0·04 Å was used). During the early refinement cycles C*A* hydrogens were included in the geometry optimization routine to ensure that amino acids did not change hand from L to D. Electron density maps calculated from Fourier coefficients $(2F_o - F_c)\exp i\alpha$ and $(F_o - F_c)\exp i\alpha$ were plotted on paper after every few refinement cycles. Interactive computer graphics was not used in this work. Virtually the entire task of manual intervention was in placing and adjusting solvent molecules following examinations of these maps. After 140 cycles the *R* factor had fallen to 0·20. The final coordinate list contained 164 solvent molecules and, with the excep-

tion of the side chain *B*29 Lys, all protein atoms (Badger, 1986).

2.5. *Model rebuilding and final refinement cycles*

The restrained least-squares refinement [refinement (1)] was completed after far fewer cycles than refinement (2). Increases in the *R* factor after the model idealization steps in refinement (2) offset the decrease during the structure-factor refinement cycles, resulting in a fairly slow overall convergence. Since the computing time per cycle for both methods is similar (introducing geometric restraints adds little extra computing time), the restrained least-squares method would appear more efficient. The use of an interactive graphics system for model corrections and checks in refinement (1) helped to accelerate the convergence of that refinement but probably not by enough to account for the differences in overall efficiency. It was quicker and more accurate to use an interactive graphics device than to work from plotted map sections for these adjustments although the more global, multicontour visualizations provided by hard-copy plots may contribute useful additional information.

The *R* factor agreement with the observed data is similar for the models resulting from refinements (1) and (2) (Table 1). Most geometry indices [including van der Waals contacts which were not restrained during refinement (2)] are of comparable quality in both refinements. Only the chiral volumes index appears significantly worse in refinement (2) whilst most other indices are slightly better. Coordinate comparisons showed that the two models differed by an average of 0·24 Å over all atoms and only 0·18 Å for main chain atoms alone. In terms of r.m.s. differences the values are 0·40 and 0·30 Å. A histogram of the coordinate differences reveals that a few large terms account for most of the discrepancy (Fig. 1). In order to evaluate the significance of these differences a new phasing model was constructed in which the positions of atoms in the two models were average where they differed by less than 0·4 Å and omitted where they differed more. This involved the omission of 14 main chain and 31 side chain atoms. Solvent molecules were not included in the set of average coordinates. An 'average' electron density map was then calculated from Fourier coefficients $(2F_o - F_c)\exp i\alpha$ where the calculated structure-factor amplitudes and phases were obtained from the averaged phasing model. Comparisons of the two refinement models against this map revealed that the six N-terminal *A*-chain residues were slightly misplaced in refinement (1) as was residue *B*28 Pro and the side chains of *A*8 Thr and *B*12 Val. With the exception of *B*28 Pro these atoms refined with average temperature factors and were situated in quite

well-defined electron density in the final model. These model biases may have arisen from the tendency of groups of atoms in restrained-parameter refinements to become trapped in false minima. The disordered residues B29–B30 (the mean final temperature factor over all atoms in these amino acids is 69 Å^2) were better fitted by the coordinates resulting from refinement (1). The availability of an interactive computer graphics system for fitting this region had clearly been helpful in this case. The side chain A15 Gln (mean temperature factor 53 Å^2) appeared to be slightly misorientated in both refinements. The remaining discrepancies were in the last 'atom or two' of side chains A2 Ile, A4 Glu, A14 Tyr, A18 Asn, A21 Asn, B4 Gln, B22 Arg and most of side chains B21 Glu and B25 Phe. In these regions both models appeared to be equally satisfactory. All but one of these atoms finally refined with temperature factors greater than 38 Å^2 and the weaker electron density does not allow the positions to be defined with great precision. Changes were made to the model resulting from refinement (2) in these regions in order to accommodate these observations. Ordered solvent structure was completely rebuilt at this point. A total of 62 water molecules visible in the 'average' map were entered into the coordinate file.

For the final 28 refinement cycles the restrained least-squares approach was used (Table 1). All data in the resolution range 10–1·7 Å were included in the calculations. FFT methods were used to speed up the calculation of structure factors and derivatives. The balance between X-ray and geometry terms was obtained by applying a scale factor that set the r.m.s. values of terms down the long diagonal of the normal matrix of equations equal. Apart from the insertion of additional solvent molecules and the re-introduction of the alternative conformation for the side chain A5 Gln no further manual intervention was carried out. Occupancies for atoms in the side chains of residues B21 Glu and B29 Lys were set

equal to a half since the electron density and temperature factors refined to very high values but at the termination of this refinement is was not possible to build convincing additional conformations. The final model contained all protein atoms, a second conformation for the side chain of A5 Gln and 108 solvent molecules. Differences between this model and the conformation of 2Zn insulin (with coordinates close to those used in the initial rotation–translation search) are described in §3.5.

3. Discussion

3.1. Reliability of the final model

The final R factor values (Table 1) and the form of the variation of R factor with resolution (Fig. 2) are typical for refined protein structures. The poor agreement between observed and calculated structure-factor amplitudes at low resolution is largely a result of bulk solvent scattering effects which are not represented by the atomic model. After termination of the refinement, the effect of adding a simple continuum model for the solvent, in which solvent electron density outside the protein volume was represented by a uniform value which is smoothed at the protein/solvent boundary (Blake, Pulford & Artymiuk, 1983) was investigated. These bulk solvent models improved the R factor in the lowest resolution shell from $R = \sim 0.6$ to $R = \sim 0.3$ but had no obvious effect on the position of electron density peaks at the protein atom sites. At high resolution the combined effects of weaker data, small errors in the atomic positions and motional effects outside the isotropic Gaussian approximation impair the agreement between measured and calculated structure factors. The final Fourier difference map computed from data in the resolution interval 10–1·7 Å shows no features stronger than 0·2 e Å^{-3}. This value may be compared to peak heights of $\sim 1·1$ e Å^{-3} for 'typical' protein atoms in the final
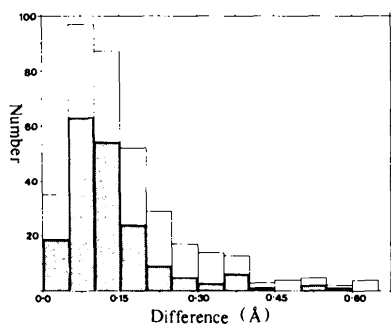

Fig. 1. Histogram of the positional differences between the models resulting from refinements (1) and (2). The shaded bars represent the differences over only main chain atoms.
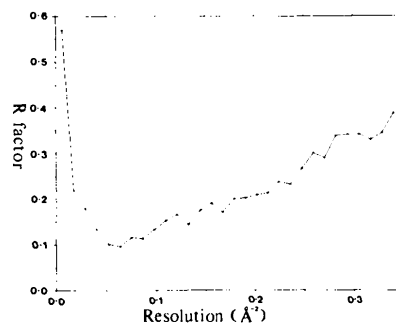
Fig. 2. Crystallographic reliability index as a function of resolution ($1/d^2$) for the final refined cubic insulin model against all measured data.

electron density map (calculated as described in the caption to Fig. 6).

The approximately linear region of the R factor plot provides an estimate of about 0·2 Å for the average error in the atomic positions (Luzzati, 1952). Comparisons of the models resulting from refinements (1) and (2) (§2.5.) give an error estimate of about 0·1 Å for the well-ordered majority of atoms and an overall average difference of 0·24 Å (Fig. 1). This value is similar to that arrived at from refinement studies, also at 1·7 Å resolution, of structures disturbed by energy minimizations (Kuriyan, Karplus & Petsko, 1987). Other comparisons of identical structures independently refined against separate sets of diffraction data gave similar error estimates (Chambers & Stroud, 1979; Wlodawer, Borkakoti, Moss & Howlin, 1986).

Statistics relating to the protein geometry indicate a close level of agreement with the standard values used in the restraint function (Table 1). The almost complete absence of van der Waals violations between molecules related by crystallographic symmetry and the close to ideal hydrogen-bond lengths across the dimer (§3.2.) are a further indication of a reliably determined structure.

3.2. Crystal contacts

The 12 insulin dimers are arranged in continuous non-intersecting rows parallel to the cubic crystal cell edges with each monomer making close (<4·0 Å) contacts with five neighbours (Fig. 3). A total of 26 different residues are involved in intermolecular packi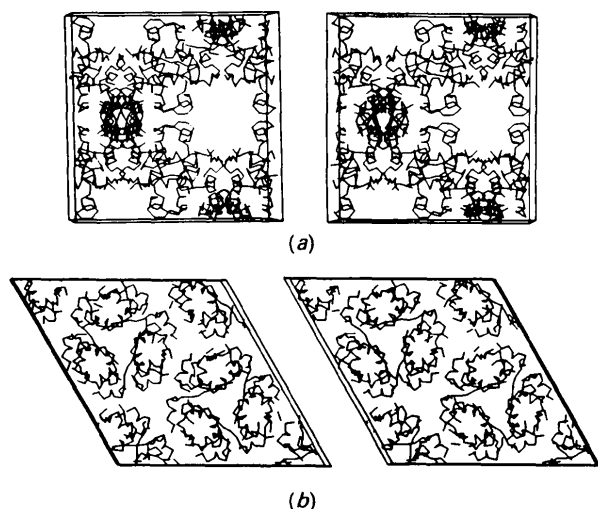ng contacts. When analysed (Richards, 1977) as a function of contact surface area (van der Waals radii 1·9, 1·7, 1·4, 1·8 Å for C, N O and S atom types with a 1·4 Å probe radius) 45% of the surface area of the free monomer is lost to crystal contacts. The large number of packing constraints provide an explanation for the high degree of order in the molecule (Bragg reflections were measured to 1·7 Å) and reasonable resistance to radiation damage despite a solvent volume of ~65%. In effect the cubic crystal cell is partioned into solvent and closely packed protein spaces.

The most extensive contacts, involving residues A21, B8, B9, B12, B16, B21, B23–B29, are between molecules in the molecular dimer (Fig. 4a). The antiparallel β-sheet hydrogen-bonding network between the main chain atoms of B24 Phe and B26 Tyr seen in 2Zn and 4Zn insulins is also found in the cubic crystal form (see §3.5. for discussion). An additional hydrogen bond is formed between the side chain of A21 Asn and the carbonyl of B26 Tyr and a salt bridge may link the flexible side chains B21 Glu and B29 Lys. The closest non-bonded interactions are between the CG atoms of B12 Val (2·99 Å) and the B23 Gly CA with the B26 Tyr carbonyl (2·96 Å).

The $I2_13$ unit cell contains a threefold symmetry axis running along one cube diagonal. The molecular contact point is with residue B1 Phe from two neighbouring molecules (Fig. 4b). Further contacts are made with five additional amino acids (A13, A17, B3, B13, B16) from one of these molecules. Side chains A17 Glu–B16 Tyr and B3 Asn–B13 Glu probably form a hydrogen bond and there are no van der Waals violations. The most disordered residues in the molecule (B21, B29, B30) lie on the opposite side of the molecule from B1 Phe, very loosely arranged around the threefold axis, but are much too widely separated to engage in intermolecular contacts.

Between dimers that are packed in rows, end-to-end, there are seven residues (A5, A9, A10, A12, A14, A15, B3) involved in intermolecular contacts. One conformation of side chain A5 Gln and the main chain nitrogen of A10 Ile represents the only likely hydrogen bond and no salt bridges are indicated. The fifth molecule, which completes the crystal packing, belongs to a dimer perpendicular to the reference molecule (Fig. 4c). Residues A13, A14, A17, A18, B8–B10 make contacts with this molecule. Hydrogen bonds may be formed between A17 Glu–B9 Ser and A14 Tyr–B10 His. The only van der Waals violation in these last two sets of packing contacts is 3·3 Å for a single C—C interaction.

3.3. Temperature factors and discrete disorder

The diffraction data available for cubic insulin extended to high enough resolution to permit an

(a)

(b)

Fig. 3. Crystal packing diagram for insulin molecules in (a) the cubic crystal cell and (b) the 2Zn crystal cell. For clarity only CA positions are shown.

unrestrained refinement of individual isotropic thermal parameters. Temperature factors refined to physically realistic values with an overall average of 26 Å^2 (Fig. 5). Main chain temperature factors are mostly lower than side chain values and the values within individual side chains generally increase with distance from the protein backbone.

Examples of other small proteins refined at high resolution suggest that the disordering of protein structure into recognizably separate conformations may account for 6–13% of amino acids (Smith,

Hendrickson, Honzatko & Sheriff, 1986). This would correspond to between three and seven side chains in cubic insulin. In this refinement the practical difficulty of distinguishing between alternative side chain conformations and bound water molecules led us to treat evidence for static disorder in a fairly conservative fashion. Only the side chain of $A5$ Gln has been described by alternative positions in our current coordinate file. In this case the likelihood for an alternative conformation is strengthened by the possibility of the formation of hydrogen bonds with protein and well-ordered water molecules and the analogy with 2Zn insulin where a conformation like that of the alternative is found. Nevertheless, the crystallographic evidence (Fig. 6a) is not conclusive since at electron density contour levels at which the 'extra' electron density connects with the $A5$ Gln density some of the more tightly bound water molecules also become connected to the protein electron density surface.

The side chain $B4$ Gln is disordered in molecule 2 of 2Zn insulin but appears to be represented by only one conformation in the cubic crystal (Fig. 6b). The side chain $B12$ Val makes a close contact across the molecular dimer axis which results in discrete disorder in molecule 2 of 2Zn insulin. In the cubic crystal the contact atom has a slightly higher temperature factor than the other CG (30 Å^2 compared to 23 Å^2) but modeling an additional conformation appears unwarranted (Fig. 6c). This side chain $B22$ Arg exhibits discrete disorder in both molecules 1 and 2 of 2Zn insulin. In the cubic crystal the spreading electron density implies some disorder but there are no well-defined independent conformations (Fig. 6d). Temperature factors for the aromatic ring in $B25$ Phe are markedly asymmetric in that they are about 20 Å^2 larger on one side of the ring than the other. The electron density around the aromatic ring is oddly shaped suggesting a twisting motion or an alternative, slightly displaced conformation (Fig. 6e). In 2Zn insulin the symmetry of the dimers is broken
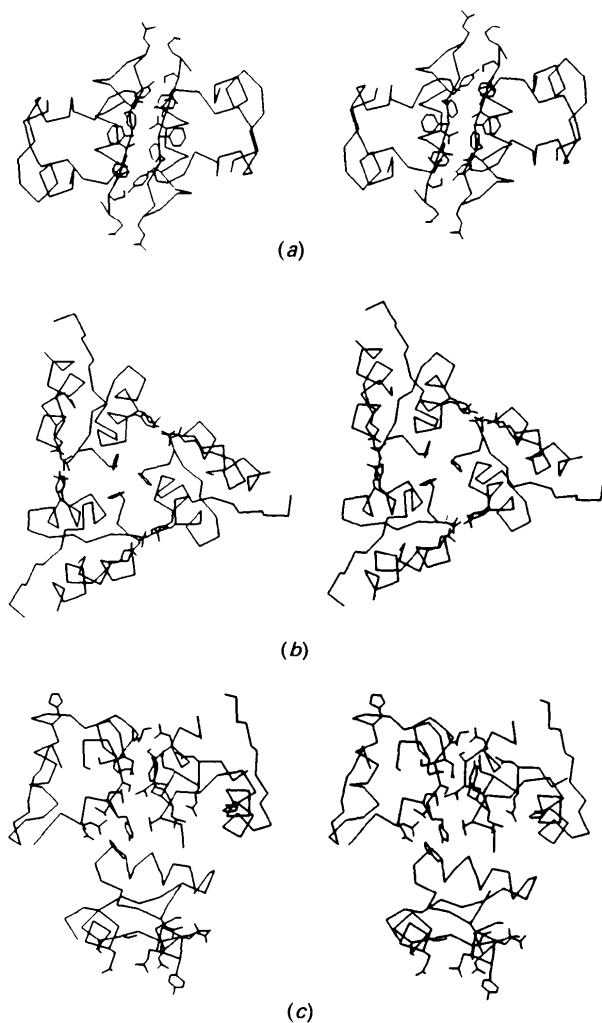

(a)

(b)

(c)

Fig. 4. Crystal packing diagram for (a) the molecular dimer (X, Y, Z and X, $1 - Y$, $\frac{1}{2} - Z$), (b) insulin molecules related by the crystallographic threefold axis (X, Y, Z; Y, Z, X; and Z, X, Y) and (c) insulin molecules packed end-to-end and perpendicular to the reference molecule (X, Y, Z; $\frac{1}{2} - X$, Y, $1 - Z$; and $\frac{1}{2} - Z$, X, $1 - Y$). In each case the CA backbone and side chains of residues involved in intermolecular contacts are shown. Only the monomer from the dimer making particular contacts is illustrated in (b) and (c).
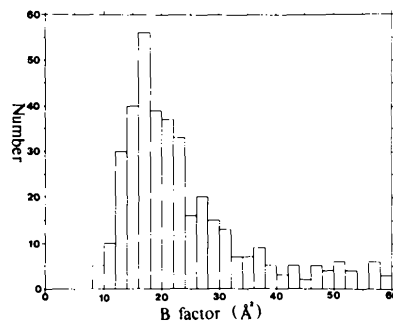

Fig. 5. Histogram of the distribution of temperature factors for protein atoms in the final refined cubic insulin model.

by the side chain of B25 from molecule 2 folding across the twofold axis. In cubic insulin this conformation violates the crystallographic symmetry and the refined conformation is virtually identical to molecule 1 of the 2Zn crystal form. Nevertheless, tiny fragments of electron density on the dimer axis provide a hint that a folded-over conformation may be present for one molecule in the cubic crystal for a small fraction of the time. The side chain of B27 Thr is disordered in molecule 2 of 2Zn insulin but appears to be well-ordered in the cubic crystal (Fig. 6f).

The temperature factors for side chains of B21 Glu and B29 Lys were ultimately refined with half occupancies since otherwise very high values were obtained. Electron density for B21 Glu is low but clearly present (Fig. 6g). The side chain of B29 Lys (Fig. 6h) is only the part of the cubic insulin structure which cannot be positioned with confidence. The electron density becomes extremely weak beyond the CB atom and side chain amino group is placed in a small volume of electron density so as to form a salt bridge with B21 Glu. An alternative position for the side chain of B29 Lys had been built into the model for part of the refinement but has not been included in the final model because both the presence of bound water molecules and the overall level of disorder make this interpretation uncertain. In molecule 1 of 2Zn insulin B29 Lys shows alternative conformations while B21 Glu exhibits discrete disorder in molecule 2. Since these atoms are found in different positions in 2Zn dimers and are relatively free from packing constraints in the cubic crystal either static or dynamic disorder is possible.
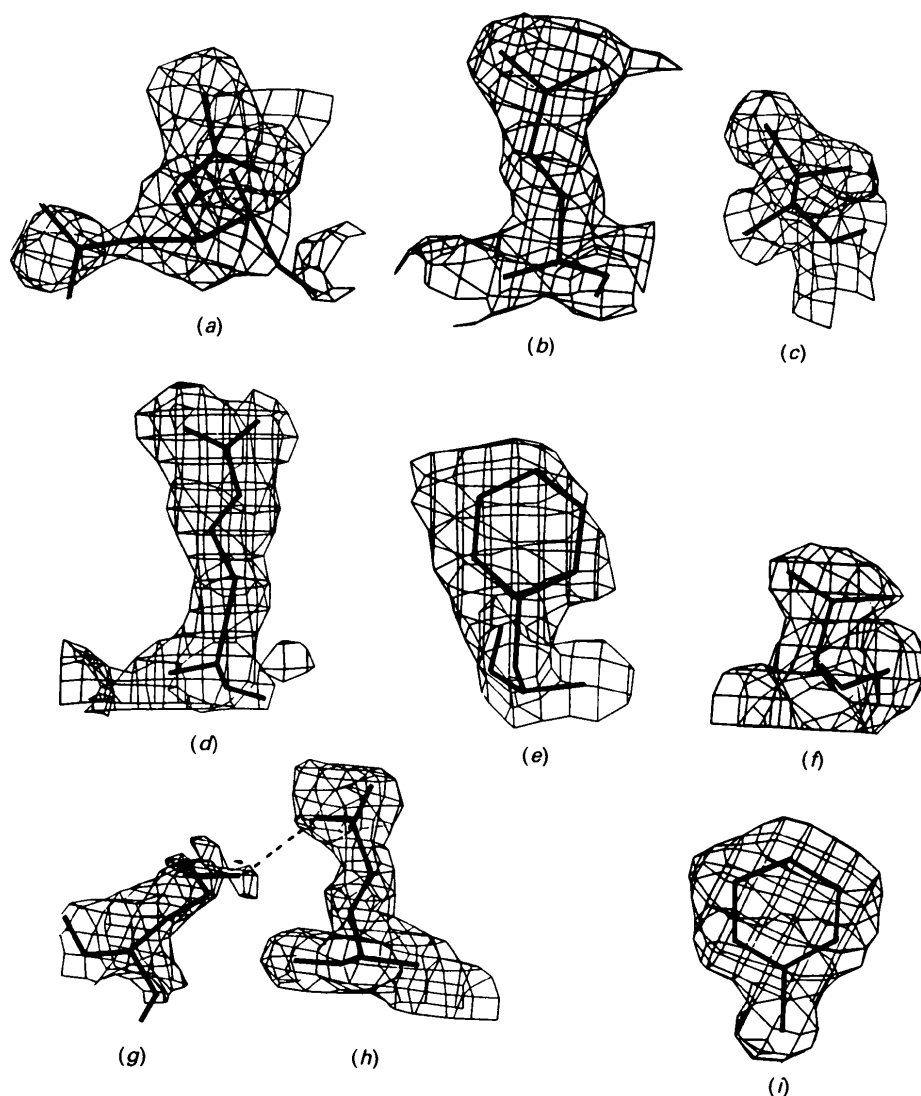

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

Fig. 6. Atomic models and electron density for (a) A5 Gln, (b) B4 Gln, (c) B12 Val, (d) B22 Arg, (e) B25 Phe, (f) B27 Thr, (g) B21 Glu and (h) B29 Lys. The electron density surface is contoured at 0.53 e Å^{-3} except in (g) and (h) where the contour level is 0.50 e Å^{-3} in order to increase the connectivity between atoms. The electron density map was calculated from Fourier coefficients $(2F_o - F_c)\exp i\alpha$ for data in the range $10.0 > d > 1.7$ Å with an $F(000)$ term equal to 169 139 electrons. Electron density for (i), the well-ordered residue B24 Phe, is shown for comparison with the less-ordered side chains. At higher contour levels the central hole in the aromatic ring of B24 Phe becomes visible.

3.4. *Solvent structure*

Electron density peaks due to solvent molecules range from well-defined spherical features to weak irregular volumes of density. Since a separation of the effects of occupancy and temperature factor is difficult to achieve in practice (Kundrot & Richards, 1987; Bhat, 1989) water molecules are parameterized with fixed unit occupancy values and only the positions and isotropic temperature factors were treated as refinable parameters. Occupancy factors for water molecules situated on symmetry axes were set to the appropriate reduced values. Neither xylene molecules nor the salt ions that were present in the crystallization media were recognized in the electron density maps obtained at the termination of this refinement.

Of the 108 water molecules which have been positioned only nine refined with temperature factors less than 40 Å². These and other more strongly defined water molecules are invariably linked to the protein surface by multiple hydrogen bonds. At places where the protein environment is similar these sites generally correspond to water sites identified in 2Zn insulin crystals. One particular feature of interest is the solvent structure around side chain B13 Glu. In 2Zn insulin copies of this residue within the dimer are linked by a network of water molecules while the glutamate groups from neighbouring dimers make direct hydrogen-bonding contact about the hexamer central axis (Baker *et al.*, 1988). Calcium ion binding between copies of B13 Glu in neighbouring dimers has been implicated in the assembly of the insulin hexamer (Palmieri, Lee & Dunn, 1988) and would help balance the negative charge from the glutamate groups. In cubic insulin the arrangment of dimers differs and the orientation of the side chain B13 Gln is slightly altered but an ordered network of linking solvent molecules is present.

While the vast majority of water molecules for which model coordinates have been obtained are bonded directly to the protein surface, other less well-defined features extending out into the solvent volume are discernable. The extensive solvent channels running through cubic insulin crystals make this a useful system for investigating solvent structure around protein molecules. A survey of the entire solvent volume in the electron density map will be described in a later report (Badger & Caspar, 1990).

3.5. *Comparisons with 2Zn insulin*

The overall fold of the cubic insulin monomer is more similar to molecule 1 of 2Zn insulin than to molecule 2. A least-squares superposition of main chain atoms A1–A21 and B5–B28 gives an r.m.s. difference of 0·60 Å for molecule 1 and a difference of 0·92 Å for molecule 2. Since the conformation of molecule 1 is also present in 4Zn insulin and in

despentapeptide insulin it may be argued that this is the most stable conformation. It might also be supposed that the structure of the insulin monomer in solution most resembles molecule 1. More detailed comparisons based on the overlap with molecule 1 reveal that major displacements of the entire structure (Figs. 7 and 8) occur in the regions B1–B4 and A9–A14.

The dimer-forming residues (B12, B16, B23–B26) appear with similar conformations in both the 2Zn and cubic species (Fig. 9a). In 2Zn insulin the symmetry of the dimer interface is broken most obviously by side chain B25 Phe from molecule 2 folding across the dimer axis and to a lesser extent by the disorder in one copy of the side chain of B12 Val. In contrast to the dimer-forming residues, residues involved in the formation of the 2Zn insulin hexamer (A12–A14, A17, B1–B4, B6, B14, B17, B18) through dimer packing show much larger positional differences between the 2Zn and cubic crystal forms (Fig. 9b).

Some distinction may be made between residues in which atomic positions are simply shifted in response to crystal packing forces and those residues which adopt radically altered conformations in the different crystal forms. Particularly striking are the
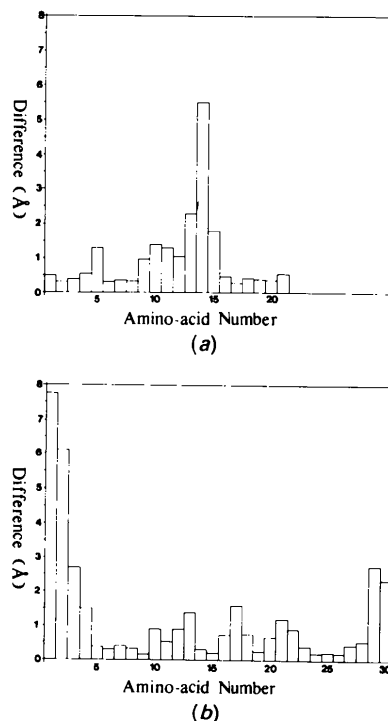

Fig. 7. Average positional differences per amino acid between the cubic insulin molecule and molecule (1) of 2Zn insulin. The A-chain differences are shown in (a) and the B-chain differences in (b).
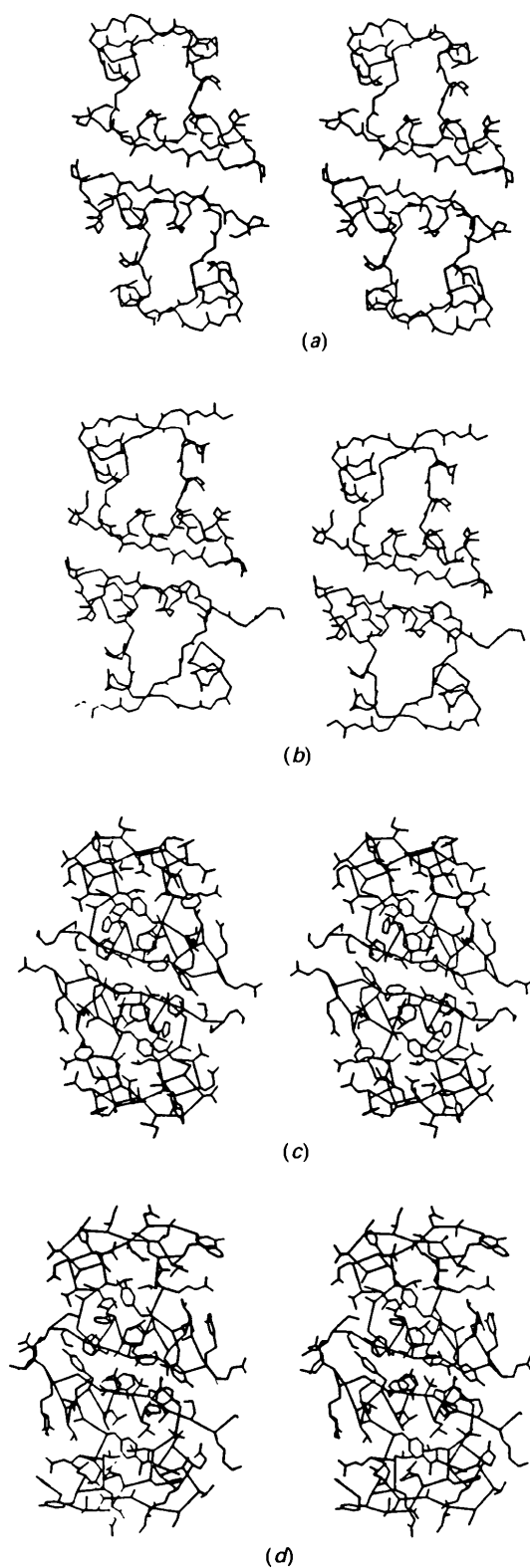
(a)

(b)

(c)

(d)

Fig. 8. Main chain trace for (a) the cubic insulin dimer and (b) the 2Zn insulin dimer. Side chain positions superimposed on a CA plot are shown for (c) cubic insulin and (d) 2Zn insulin.
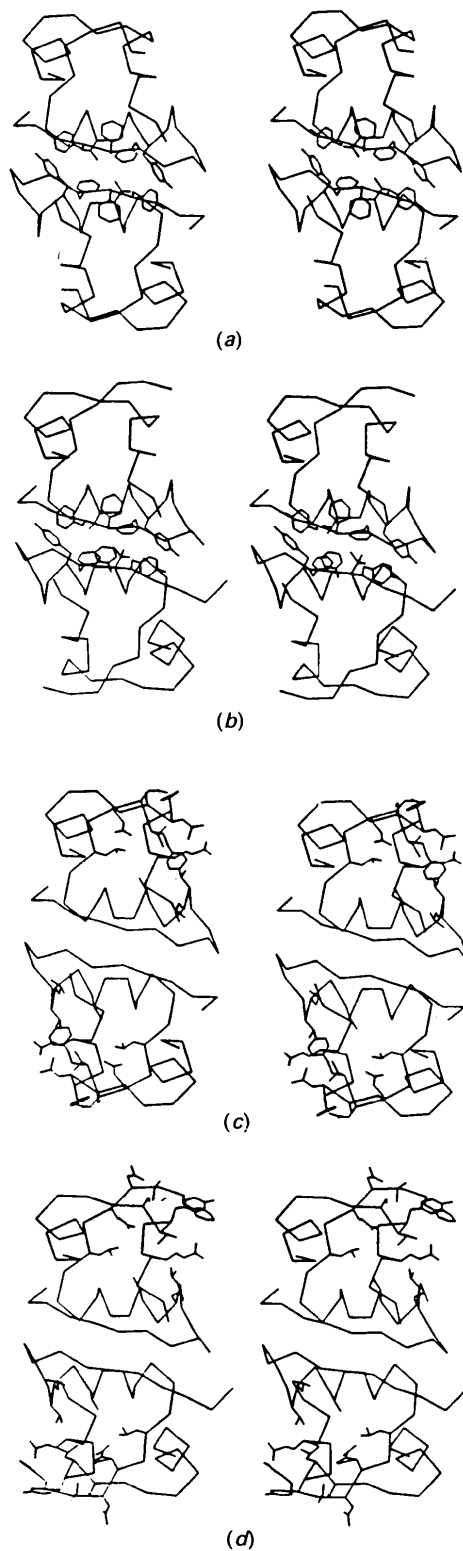

(a)

(b)

(c)

(d)

Fig. 9. CA trace and side chains on the molecular dimer surface for (a) cubic insulin and (b) 2Zn insulin. The CA trace and side chains involved in hexamer formation are also shown for (c) cubic insulin and (d) 2Zn insulin.

rearrangements of residues B1–B4 which also adopt differing orientations between dimers within the 2Zn and 4Zn hexamers. Side chains A5 Gln, A9 Ser, A10 Ile, A14 Tyr and B17 Leu are re-oriented in response to altered packing constraints in the cubic crystal. The side-chain ring of B10 His which is involved in Zn binding in 2Zn insulin is rotated by about 90° in the cubic crystal. All of these observations, coupled with the heterogeneity already found in other insulin structures, indicate that crystal packing forces may play a critical role in the determination of local structure.

Atomic coordinates resulting from this work have been deposited with the Protein Data Bank.*

---

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 5INS, R5INSSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37039 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

References

AGARWAL, R. C. (1978). Acta Cryst. A34, 791–809.
BADGER, J. (1986). DPhil Thesis, Univ. of York, England.
BADGER, J. & CASPAR, D. L. D. (1990). Proc. Natl Acad. Sci. USA. In the press.
BAKER, E. N., BLUNDELL, T. L., CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., HODGKIN, D. M. C., HUBBARD, R. E., ISAACS, N. W., REYNOLDS, C. D., SAKABE, K., SAKABE, N. & VIJAYAN, N. M. (1988). Philos. Trans. R. Soc. London Ser. B, 319, 369–456.

BHAT, T. N. (1989). Acta Cryst. A45, 145–146.
BI, R. C., DAUTER, Z., DODSON, E., DODSON, G., GIORDANO, F. & REYNOLDS, C. (1984). Biopolymers, 23, 391–395.
BLAKE, C. C. F., PULFORD, W. C. A. & ARTYMIUK, P. J. (1983). J. Mol. Biol. 167, 693–723.
BLUNDELL, T., DODSON, G., HODGKIN, D. & MERCOLA, D. (1972). Adv. Protein Chem. 26, 279–402.
CHAMBERS, J. L. & STROUD, R. M. (1979). Acta Cryst. B35, 1861–1874.
CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., EMDIN, S. O. & REYNOLDS, C. D. (1979). J. Mol. Biol. 132, 85–100.
CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., REYNOLDS, C. D. & VALLELY, D. (1981). Structural Studies on Molecules of Biological Interest, edited by G. DODSON, J. GLUSKER & D. SAYRE, pp. 527–546. Oxford: Clarendon Press.
DE GRAAFF, R. A. G., LEWIT-BENTLEY A. & TOLLEY, S. P. (1981). Structural Studies on Molecules of Biological Interest, edited by G. DODSON, J. GLUSKER & D. SAYRE, pp. 553–556. Oxford: Clarendon Press.
DEREWENDA, U., DEREWENDA, Z., DODSON, E. J., DODSON, G. G., REYNOLDS, C. D., SMITH, G. D., SPARKS, C. & SWENSON, D. (1989). Nature (London), 338, 594–596.
DODSON, E. J., DODSON, G. G., HODGKIN, D. C. & REYNOLDS, C. D. (1979). Can. J. Biochem. 57, 469–479.
DODSON, E. J., DODSON, G. G., LEWITOVA, A. & SABESAN, M. (1978). J. Mol. Biol. 125, 387–396.
DODSON, E. J., ISAACS, N. W. & ROLLETT, J. S. (1976). Acta Cryst. A32, 311–315.
EVANS, A. C. (1979). PhD Thesis, Univ. of Leeds, England.
HARRIS, M. R. (1985). PhD Thesis, Univ. of Leeds, England.
HENDRICKSON, W. A. & KONNERT, J. H. (1980). Computing in Crystallography, edited by R. DIAMOND, S. RAMASESHAN & D. VENKATESAN, pp. 1–25. Bangalore: Indian Academy of Sciences.
ISAACS, N. W. & AGARWAL, R. C. (1978). Acta Cryst. A34, 782–791.
JACK, A. & LEVITT, M. (1978). Acta Cryst. A34, 931–935.
JONES, T. A. (1978). J. Appl. Cryst. 11, 268–272.
KONNERT, J. H. (1976). Acta Cryst. A32, 614–617.
KUNDROT, C. E. & RICHARDS, F. M. (1987). Acta Cryst. B43, 544–547.
KURIYAN, J., KARPLUS, M. & PETSKO, G. A. (1987). Proteins, 2, 1–12.
LUZZATI, V. (1952). Acta Cryst. 5, 802–810.
OATLEY, S. & FRENCH, S. (1982). Acta Cryst. A38, 537–549.
PALMIERI, R., LEE, R. W.-K. & DUNN, M. F. (1988). Biochemistry, 27, 3387–3397.
RICHARDS, F. M. (1977). Annu. Rev. Biophys. Bioeng. 6, 151–176.
SMITH, J. L., HENDRICKSON, W. A., HONZATKO, R. B. & SHERIFF, S. (1986). Biochemistry, 25, 5018–5027.
WLODAWER, A., BORKAKOTI, N., MOSS, D. S. & HOWLIN, B. (1986). Acta Cryst. B42, 379–387.