all four related structures are disordered in a similar way.

This work was supported by grant HL-20350 from the National Institutes of Health. We are grateful to Dr R. Shiono for the use of computer programs and also to Dr John Ruble and Mrs Joan Klinger for technical assistance.

References

BRISSE, F. & SANGIN, J.-P. (1982). *Acta Cryst.* B**38**, 215–221.

BUSING, W. R. & LEVY, H. A. (1964). *Acta Cryst.* **17**, 142–146.

CRAVEN, B. M., HE, X. M. & WEBER, H.-P. (1986). *Programs for Thermal Motion Analysis* (updated). Tech. Rep. Department of Crystallography, Univ. of Pittsburgh, USA.

CRAVEN, B. M., WEBER, H.-P. & HE, X. M. (1987). *The POP Refinement Procedure*. Tech. Rep. Department of Crystallography, Univ. of Pittsburgh, USA.

CREMER, D. & POPLE, J. A. (1975). *J. Am. Chem. Soc.* **97**, 1354–1358.

CROMER, D. T. & WABER, J. T. (1965). *Acta Cryst.* **18**, 104–109.

DUNITZ, J. D. & WHITE, D. N. J. (1973). *Acta Cryst.* A**29**, 93–94.

GILMORE, C. G. (1983). *MITHRIL. Computer Program for the Automatic Solution of Crystal Structures from X-ray Data.* Univ. of Glasgow, Scotland.

HAMILTON, W. A. (1974). *International Tables for X-ray Crystallography*, Vol. IV, p. 288. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

JOHNSON, C. K. (1976). *ORTEPII*. Report ORNL-5138. Oak Ridge National Laboratory, Tennessee, USA.

JOHNSON, C. K. & LEVY, H. A. (1974). *International Tables for X-ray Crystallography*, Vol. IV, p. 316. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

KUHS, W. F. (1983). *Acta Cryst.* A**39**, 148–158.

LIDE, D. R. (1960). *J. Chem. Phys.* **33**, 1514–1518.

POLLARD, C. B., ADELSON, D. E. & BAIN, J. P. (1934). *J. Am. Chem. Soc.* **56**, 1759–1760.

SANGIN, J.-P. & BRISSE, F. (1984). *Acta Cryst.* C**40**, 2094–2096.

SCHOMAKER, V. & TRUEBLOOD, K. N. (1968). *Acta Cryst.* B**24**, 63–76.

SHELDRICK, G. M. (1976). *SHELX*. Program for crystal structure determination. Univ. of Cambridge, England.

STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175–3187.

TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1984). *Acta Cryst.* B**40**, 280–288.

Acta Cryst. (1991). B**47**, 975–986

Structure of Porcine Insulin Cocrystallized with Clupeine Z

BY P. BALSCHMIDT AND F. BENNED HANSEN

Novo-Nordisk A/S, Niels Steensensvej 1, 2820 *Gentofte, Denmark*

AND E. J. DODSON, G. G. DODSON AND F. KORBER*

*Department of Chemistry, University of York, York YO*1 *5DD, England*

(*Received* 19 *September* 1989; *accepted* 9 *July* 1991)

Abstract

The crystal structure of NPH-insulin, pig insulin cocrystallized with zinc, *m*-cresol and protamine, has been solved by molecular replacement and refined using restrained least-squares refinement methods. The final crystallographic *R* factor for all reflections between 2 and 10 Å is 19·4%. The insulin molecules are arranged as hexamers with two tetrahedrally coordinated Zn atoms in the central channel and one *m*-cresol bound to each monomer near His *B*5. One protamine binding site has been unequivocally identified near a dimer–dimer interface, although most of the polypeptide is crystallographically disordered. The conformation of the insulin moiety and the structural differences between the three unique monomers have been analysed. The zinc and *m*-cresol environments are described and the nature of the protamine binding site is outlined.

* To whom all correspondence should be addressed.

Introduction

Microcrystalline preparations of insulin cocrystallized with protamines, basic polypeptides of the cell nucleus which are rich in arginine and lysine, (NPH-insulin) were obtained as long as 40 years ago (Krayenbuhl & Rosenberg, 1946) and have been used successfully in the treatment of diabetes ever since. Owing to its slow release properties after subcutaneous injection, protamine–insulin has been a very popular pharmaceutical formulation. Our aim is to investigate how this delayed action is brought about, how and to what extent the protamine is incorporated into the crystal structure, and to find and describe possible changes in the insulin structure in this particular crystal form. Preliminary data from two different structural studies on salmine–insulin have been reported (Baker & Dodson, 1970; Eggena, Magdoff-Fairchild, Rudko, Fullerton & Low, 1969; Simkin, Cole, Ozawa, Magdoff-Fairchild, Eggena, Rudko & Low, 1969; Fullerton & Low, 1970),

0108-7681/91/060975-12$03.00

revealing that salmine–insulin can crystallize in two different space groups, $P2_12_12_1$ and $P4_12_12$, with almost identical unit-cell dimensions. The diffraction pattern at low resolution for the orthorhombic form has a strong tetragonal appearance (Baker & Dodson, 1970). Furthermore, there is chemical evidence for the presence of approximately 0·15 protamine molecules per insulin monomer and the presence of at least two Zn atoms per six insulin molecules (Fullerton & Low, 1970; Nordisk Gentofte AS, unpublished results). We investigated the structure of pig insulin cocrystallized with clupeine Z, a homogeneous protamine subcomponent of herring sperm consisting of 31 amino acids (Iwai & Nakahara, 1971). In this paper we present our findings on the structure of the insulin component and preliminary findings on the protamine. More detailed investigations on the protamine and the water structure are underway and will be reported separately.

Experimental

Crystal growth

All materials for the preparation of the solutions used for the crystallization were of analytical grade and purchased from Merck. Pig insulin was prepared at Nordisk Gentofte AS by standard production methods as a highly purified crystalline zinc precipitate. Clupeine (protamine from Atlantic Herring) was prepared at Nordisk Gentofte AS by a process very similar to that reported by Michael (1967) and fractionated by ion-exchange chromatography according to the method of Ando & Watanabe (1969). The last eluting protein peak containing the Z component was pooled and desalted and the protamine was precipitated as a sulfate salt by acidified ethanol. Pig insulin is cocrystallized with clupeine Z in batch by preparing a standard buffer of 4 *M* urea, 27 m*M* phosphoric acid, 21 m*M* *m*-cresol and 120 m*M* NaCl. From this a protein solution (solution 1) is made by dissolving 3 mg zinc-free pig insulin per ml of buffer and adding 0·275 m*M* $ZnCl_2$ (0·55 Zn atoms per insulin monomer). A protamine solution (solution 2) is prepared by dissolving approximately 0·5 mg of clupeine Z in each ml of standard buffer. The exact amount of clupeine to be added has to be determined beforehand by measuring the clupeine to insulin ratio at which maximum precipitation of both insulin and clupeine Z occurs upon mixing insulin and clupeine test solutions (isophane ratio) as described by Krayenbuhl & Rosenberg (1946). The pH of both solution 1 and solution 2 is adjusted to 7·3 with NaOH, equal amounts of the solutions are mixed and left at 277 K. Useful crystals appear within a few days. Crystal size and the number of crystals can be varied by changing the urea concentration.

Crystal characterization

Insulin crystals cocrystallized with clupeine Z are bipyramidally tipped parallelepipeds with maximum dimensions of 0·2 × 0·2 × 1 mm. Precession X-ray photographs show that the crystals belong to space group $P4_12_12$ or its enantiomorph with $a = b = 62{\cdot}9\ (3)$, $c = 85{\cdot}9\ (3)$ Å, very similar though not isomorphous to those described by Eggena *et al.* (1969). Owing to the hexameric nature of the insulin moiety as established from the self-rotation function, the asymmetric unit must contain three monomers with a V_M of 2·40 $Å^3$ dalton^{-1} including Zn and bound *m*-cresols but excluding the protamine. This corresponds to a calculated solvent content of 50% and a density of 1·22 g cm^{-3}, in good agreement with the measured value for salmine–insulin crystals (Fullerton & Low, 1970). The crystals diffract to 2·8 Å on a sealed-tube generator with a graphite monochromator and to 1·8 Å on the Wiggler Station of the SERC Synchrotron Radiation Source at Daresbury, England.

Data collection and processing

The rotation camera data used for the structure determination were collected on the wiggler station of the Synchrotron Radiation Source at Daresbury using a wavelength of 0·878 Å and flat cassettes. Two complete data sets were obtained: the first one extending to 2·5 Å resolution was taken from one crystal; the second one to better than 2 Å resolution from three crystals. The data were recorded on CEA Reflex 25 film by rotating the crystals about *c* in steps of 1·5°. Care was taken to obtain most of the data normally contained in the blind region by missetting the crystals deliberately in the arcs by 2 to 8°.

Films were scanned using a Joyce–Loebl Scandig 3 densitometer at a sampling interval of 50 μm. The intensities were integrated by a local adaptation (Korber, unpublished) of the *MOSFLM* program suite which contains the Rossmann profile-fitting algorithm (Wonacott, Dockerill & Brick, 1980; Rossmann, 1979; Greenhough & Suddath, 1986). Profile fitting was used throughout on both fully recorded reflections and partials (Greenhough, 1987). The high-resolution data set which theoretically could provide data to 1·8 Å was curtailed to 2 Å because a rise in the merging $R_I = \sum_{hkl}(|I_o - I_{mean}|)_{hkl}/\sum_{hkl} I_{mean_{hkl}}$ to well over 20% was observed beyond 2·1 Å (Fig. 1). From the high-resolution data set we obtained 11331 unique reflections which are 91% of the theoretically observable data to 2 Å. The average multiplicity for each Miller index was 4·5,

with a ratio of fully observed reflections to partially observed reflections of 3:1. Partially recorded reflections were included as the partial bias was much smaller than the standard deviation for all resolution shells. The total partial bias is −0·02. The overall merging R factor to 2 Å resolution is 7·1%. Merging this data set with the data set measured to 2·5 Å, which has similar processing statistics, added another 600 unique reflections so that data for 96% of the theoretically observable reflections was obtained. The overall merging R_I remained at 7·1% with 75% of the reflections greater than three standard deviations.

Molecular replacement

Both self-rotation and cross-rotation functions (program *ALMNFR*, Dodson, 1985) and an R-factor search (program *SEARCH*, Dodson, 1985) in conjunction with a packing search (program *PRJANG*, Dodson, unpublished) were used to determine the positional parameters for the insulin moiety in the unit cell.* The self-rotation function on data between 3 and 10 Å gave one strong peak with 0·52 of the height of the origin peak at $\omega = 54{\cdot}60$, $\theta = -45{\cdot}20$ and $\chi = 119{\cdot}80°$. The appearance of this local threefold axis perpendicular to the [1,1,0] crystallographic twofold is a strong indication for the hexameric nature of NPH-insulin. In this case the asymmetric unit is a trimer as already deduced from density measurements by Fullerton & Low (1970). The two other diads perpendicular to the hexamer axis then give peaks which merge into symmetry equivalents of the local threefold peak: the values for the direction cosines for the triad obtained from the self-rotation function are (0·5774, −0·5774, 0·5774) and those of the two non-crystallographic diads are (−0·7071, 0, 0·7071) and (0, −0·7071, −0·7071). A further examination of the self-rotation function in different resolution shells gave values between 45 and 55° for the angle between the triad and the z axis.

The cross-rotation function was initially computed with calculated structure factors from the 2Zn hexamer in a 100 × 100 × 80 Å primitive cell with the triad orientated along the z axis. Data between 10 and 3 Å resolution were used throughout. The strongest maximum, more than 30% higher than the next-highest peak, was found at $(\alpha, \beta, \gamma) = (-45, 50, -15)$ which confirms the findings from the self-rotation function.

The threefold symmetry was confirmed by using a single dimer as a search unit: three peaks appeared separated by approximately 120° in γ at the appropriate α and β. The tilt angle, β, for the search with this model was less well defined: β sections between $\beta = 40$ and 55° have maxima with similar heights. One of the hexamer diads coincides with the crystallographic diad. Therefore the three-dimensional translation problem is reduced to one dimension for the tetragonal form of NPH-insulin, *i.e.* to find the displacement of the hexamer along the [1 1 0] diad. The R-factor translational search was carried out with a trimer derived from the 2Zn monomer 1 as an initial model, searching exclusively along the diagonal in the $z = 0$ plane. Owing to the space-group symmetry a search from (0,0,0) to $(\frac{1}{2},\frac{1}{2},0)$ is sufficient. The search was performed in steps of 0·5 Å with data between 8 and 3 Å resolution for space groups $P4_12_12$ and $P4_32_12$ with a variety of tilt angles β. α and γ were fixed to −45 and −150° respectively. Fig. 2 shows the result of the search for a tilt angle of 51°: for both space groups the minimum could be found for a translation at or between 26·5 and 27 Å. The associated R values are 2·71% below average for space group $P4_32_12$ and 1·49% below average for space group $P4_12_12$. The preference for $P4_32_12$ was confirmed by an R-factor search with odd values of l only. Also, the translation by (26·5, 26·5, 0) is excluded by the packing search for space group $P4_12_12$.

This solution was then further refined by looking for the lowest agreement factor R_F in a grid search in which both the tilt angle β and the translation t along the [1,1,0] direction were varied systematically. A round of restrained least-squares refinement (Hendrickson & Konnert, 1980) was performed for each possible solution to establish which starting values produced the largest decrease in R. The initial R value depends critically on the translation vector t, whilst the magnitude of the tilt angle is less well defined. To be within 0·2% of the best obtainable R value, t has to be known with accuracy of 0·05 Å and β has to be within 0·4°. The best values were determined as $\beta = 51{\cdot}6°$ and $t = 26{\cdot}75$ Å giving an

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory, and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37051 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

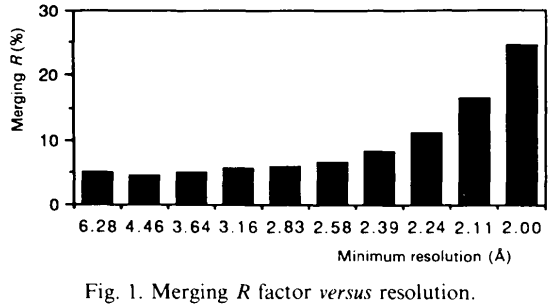

Fig. 1. Merging R factor *versus* resolution.

initial R factor of 57·9% between 2·1 and 4·5 Å. Using a model based on the 4Zn-insulin monomer 2 provided a better starting model for refinement with an additional drop in R_F of 2·4%.

Refinement

The structure was refined using the restrained least-squares refinement program *PROLSQ* (Hendrickson & Konnert, 1980) for the refinement of the atomic coordinates coupled with cycles of unrestrained B-value refinement (Agarwal, 1978) and 13 stages of model building using *FRODO* (Jones, 1978) on a PS300 or a PS390 graphics system (Evans & Sutherland). All computations were performed on a Vax 11-750 computer (DEC). Initially the protamine moiety and solvent structure were not included in the model and consequently the low-resolution data were excluded from the calculations. As can be seen from Fig. 3, rebuild 9, even a minor extension towards lower resolution induced a sharp increase in R value. It can also be seen that in spite of frequent model building the R factor remained on a plateau for a considerable time; the drop between rebuilds 4 and 5 is due to B-value refinement rather than improvement in the geometry of the model. Only when model building for the tightly bound waters progressed and protamine residues had been tentatively fitted into regions of continuous electron density was it possible to include the resolution range between 10 and 7 Å and get a drop in R factor coupled with an improvement in geometry. The current R factor is 19·4% for all data between 10 and 2 Å (11 522 unique reflections) with an average B value of 33·7 Å^2 including solvent and protamine. As the protamine is crystallographically disordered only part of it can be modelled properly, *i.e.* the R factor does not reflect fully the correctness of the model for the insulin moiety. At 2 Å resolution even isotropic B-value refinement is often divergent due to the low cutoff in the Fourier terms. Temperature factors were therefore restricted to between 5 and 60 Å^2 and B refinement was stopped at an average B of around 30 Å^2. Wilson statistics (Wilson, 1949) suggest an average B of 23 Å^2. The discrepancy may well be due to the incomplete modeling of the protamine and water structure.

Results

Accuracy of the model

The weights for the parameters of the restrained refinement and the associated deviations are listed in Table 1. From a list of deviations from ideality for each residue it can be seen that most of the errors are in the very first and last residues of the insulin chains, especially the B chains, which were not even detectable in the electron density map for the first few model building cycles and had to be rebuilt several times. There are only very few disordered side chains with ill-fitting density or no density at all, notably the Tyr $A14$ of monomers 1 and 3 and Lys $B29$ of monomer 3. The electron density identified with protamine cannot be fitted unambiguously presumably because of movement, partial occupancy and the resulting confusion with water molecules.

Another method of estimating the probable error in a structure with non-crystallographic symmetry is to

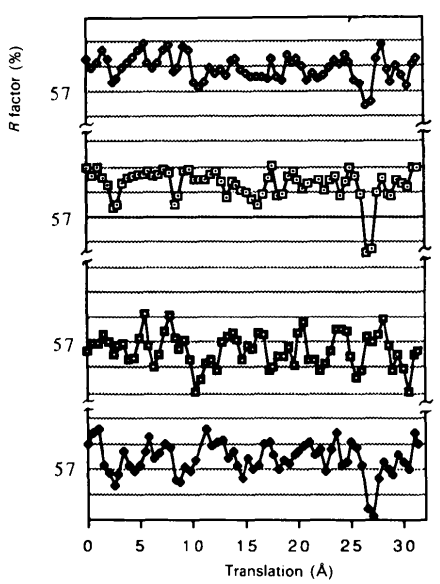

Fig. 2. R-factor search. Data from top to bottom: space group 92, all data; space group 96, all data; space group 92, l odd; space group 96, l odd. Gridline spacing 1%.

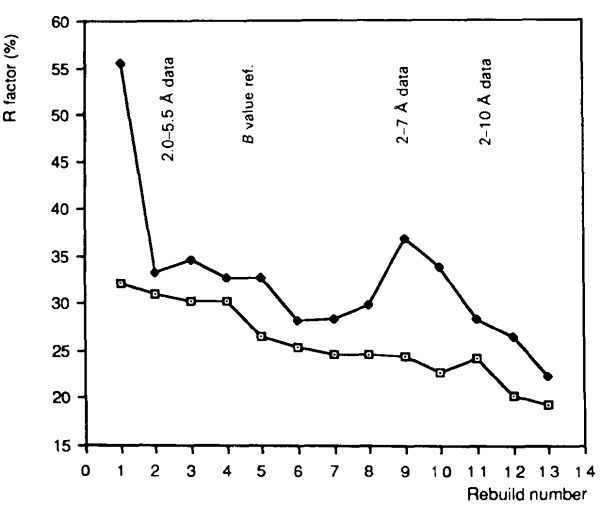

Fig. 3. Refinement progress of NPH-insulin. Upper curve: R factor before restrained refinement. Lower curve: R factor after restrained refinement.

Table 1. *Refinement weights and errors in the parameters*

	Square root of inverse weight	Observed deviation
Bonded contacts (Å)		
Bond lengths	0·02	0·03
Angle-related distance	0·04	0·09
Interplanar distance	0·05	0·10
Planar groups (Å)	0·02	0·02
Chiral centres (Å)	0·15	0·34
Non-bonded contacts (Å)		
Single torsion	0·5	0·25
Multiple torsion	0·5	0·47
Hydrogen bond	0·5	0·36
Torsion angles ( )		
Planar	3	4·2
Staggered	15	25·7
Transverse	20	28·4

Table 2. *Correlation coefficients for main-chain-atom temperature factors*

	Chain	
Between monomers	*A*	*B*
NPH 1 NPH 2	0·55	0·71
NPH 1 NPH 3	0·52	0·79
NPH 2–NPH 3	0·67	0·87
NPH_{mean} $2Zn\text{-}insulin_{mean}$	0·87	0·77
NPH_{mean} $2ZnI_{mean}$ *B*6 *B*30		0·87

compare the r.m.s. deviation for a least-squares superposition of equivalent parts of the structure. A comparison of the main-chain atoms of the 'invariant' part of single chains (*i.e.* *A*6 to *A*20 and *B*5 to *B*25) gives the smallest r.m.s. deviation as 0·25 Å, the average deviation for these single-chain comparisons is 0·45 Å for the *A* chains and 0·28 Å for the *B* chains. The larger deviations for the *A* chains can be explained by crystal contacts.

We aligned the main-chain atoms of *B*12 to *B*20 of pairs of different monomers and observed r.m.s. deviations in the positions of the side-chain aromatic groups of *B*16, *B*24 and *B*26 as low as 0·2 Å. Thus it may be assumed that the true crystallographic error is of the order of 0·2 Å and that the probable error obtained from a Luzzati plot (Fig. 4) (Luzzati, 1952), which is between 0·25 and 0·3 Å, is an overestimate for the majority of insulin residues. A comparison of the normalized *B* values of each of the NPH monomers with the average *B* values of the 2Zn-insulin monomers shows that the *B* factors follow a similar pattern with only minor local differences (Fig. 5). The correlation coefficients in Table 2 confirm this qualitative impression.

The variations for the main-chain *B* factors are greater for the *A* chains, reflecting both the less well defined secondary structure of the *A* chains and the distortion of the local environment through crystal contacts. Also, with the exception of the less well defined termini, the *B*-chain temperature factors are generally lower, both for side-chain and main-chain atoms. The results of the *B*-factor analysis thus seem to reflect true conformational flexibility rather than crystallographic errors. The quality of the model is further underlined by a main-chain torsional angle analysis. The Ramachandran plot (Fig. 6) shows that all but nine residues have torsional angles inside or nearly inside the allowed region of the map. The residues outside the allowed region are Lys *B*29, Gly *B*20 and Gly *B*23. Tyr *A*19, Pro *B*28 and some cysteines show small deviations from the permitted values.

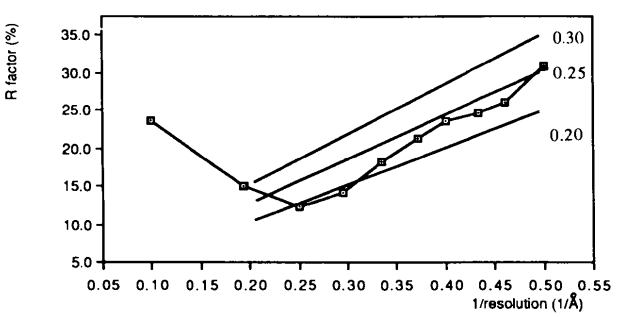

Fig. 4. Luzzati plot for NPH-insulin.

Overall description of the structure

The insulin molecules in NPH-insulin are organized into hexamers (Fig. 7). There are four hexamers per unit cell making two different kinds of crystal contacts. The asymmetric unit consists of a trimer which together with a crystallographically twofold-related trimer forms one hexamer. The three monomers within each trimer are related by an approximate triad. Fig. 8 shows the asymmetric unit, a trimer, viewed down the threefold axis and the two different kinds of crystal contact formed within the NPH-insulin crystals. Dimers within each hexamer are formed by two monomers from two different asymmetric units. The monomer which will subsequently be named monomer 1 forms a dimer with its symmetry-related counterpart, the other monomers named 2 and 3 (sequence by clockwise rotation about the triad seen from the centre of the hexamer) dimerize with the symmetry equivalent of the crystallographically different monomer.

Each hexamer contains two tetrahedrally coordinated zinc ions on the local threefold axis. The alternative zinc binding sites found in 4Zn-insulin near His *B*5 are occupied here by a *m*-cresol molecule. The hexamer is very tightly packed and has a smaller radius of gyration (r.m.s. distance from threefold axis) than the hexameric 2Zn- and 4Zn-insulin structures (Baker, Blundell, Cutfield, Cutfield, Dodson, Dodson, Crowfoot Hodgkin, Hubbard, Isaacs, Reynolds, Sakabe, Sakabe & Vijayan, 1988;

Bentley, Dodson, Dodson, Hodgkin & Mercola, 1976) (Table 3). Further, an analysis of the r.m.s. distances shows that on average the main-chain atoms are 0·5 Å closer to the centre of the hexamer than for 2Zn- and 4Zn-insulin. Both the radius of gyration and the r.m.s. distance are comparable to those for the monoclinic crystal structure (Derewenda, Derewenda, Dodson, Dodson, Reynolds, Smith, Sparks & Swenson, 1989). The comparison of individual monomers below shows that this tightening of the structure is due to relative shifts of whole insulin molecules rather than to shifts of atoms within individual monomers. The monomers themselves all have α-helical conformation from *B*1 to *B*8 and thus resemble 4Zn monomer 2 very closely. With the protein truncated to the 'core', *i.e.* *A*5 to *A*20 and *B*10 to *B*26, the least-squares superposition of the main-chain atoms is best between 4Zn monomer 1 and any of the NPH-insulin monomers. The r.m.s. difference for this superposition is comparable to that obtained by superimposing the three different NPH monomers onto each other (Tables 4 and 5). The typical NPH-insulin monomer thus resembles a 4Zn monomer 1 which has an extended *B* helix.

Comparison of the three monomers

A qualitative analysis of the conformational differences between the monomers was performed by

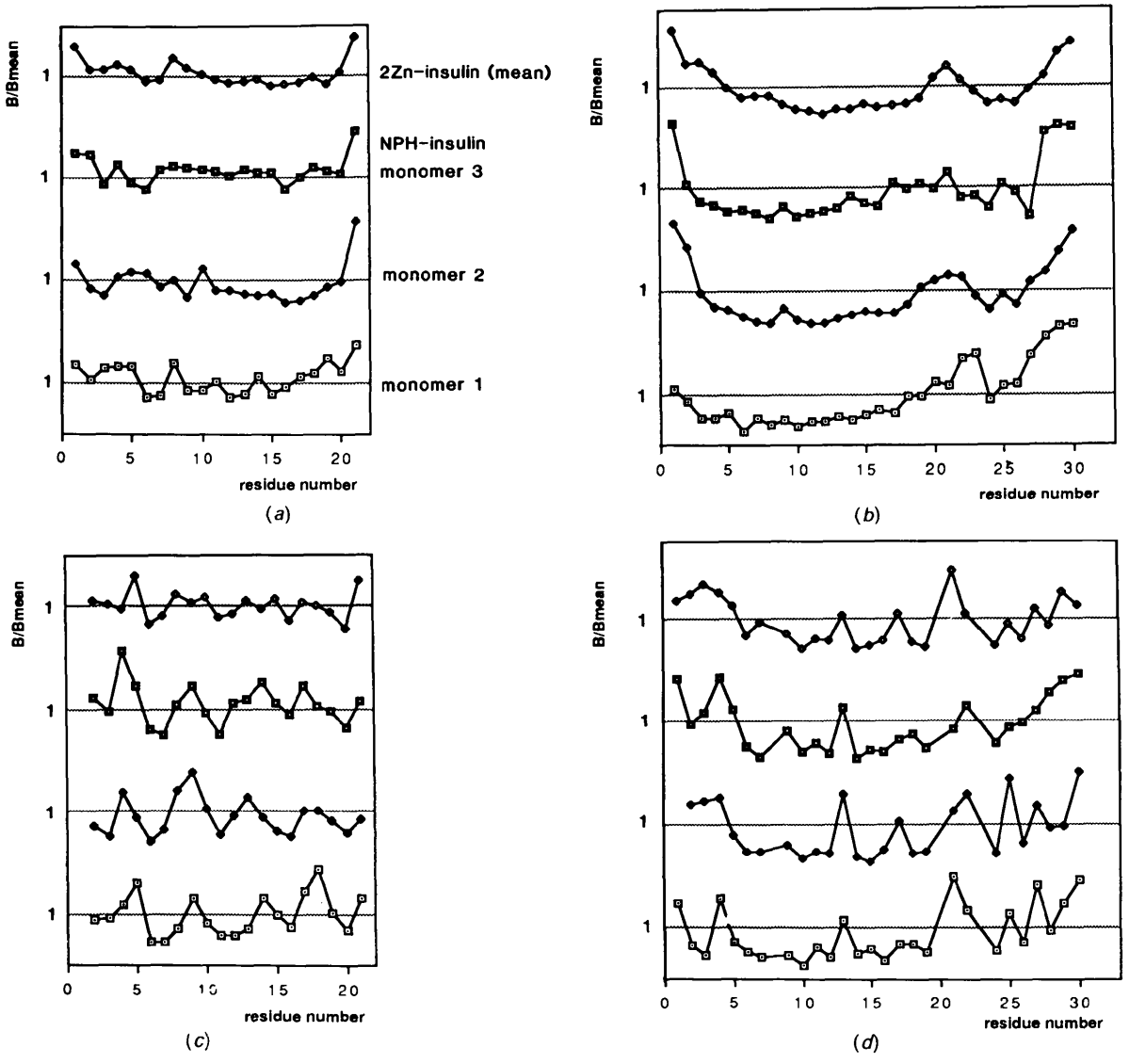

Fig. 5. Relative *B* factors for main-chain atoms of: (*a*) *A* chains, (*b*) *B* chains. Relative *B* factors for side-chain atoms of: (*c*) *A* chains, (*d*) *B* chains.

superimposing the three unique monomers using a least-squares fit on the main-chain atoms of the helix *B*12–*B*20 to define their relative orientation. The superposition shows that the conformational differences are localized and that the relative orientation of the two chains is almost identical for all three monomers. The deviations are largest for the residues in the centre of the *A* chains which participate in crystal contacts. Especially the large shifts of up to 1·5 Å of *A*11 to *A*15 of monomer 2 relative to the other monomers can be attributed directly to a close crystal contact at *A*14 (see Fig. 8).

Differences in the conformation of the *B* chains are generally smaller and restricted to the beginning and end of the chains and to the loop region around *B*21 and *B*22. For the greater part of the *B*-chain helix and the extended chain region up to *B*26 the differences in the atomic coordinates are of the order of the expected crystallographic error, even for the side chains. The NPH-insulin monomer 2 is unique in that residue *B*1 is very ill defined and there is density to indicate that residues *B*29 and *B*30 can be in two different conformations. For *B*1 this phenomenon seems to be associated with electron density which can be ascribed to protamine nearby. The two conformations at the end of the *B* chain, which occur with approximately equal probability, vary mostly in the ψ angle of Pro *B*28. In the conformation of *B*29 and *B*30 observed only in monomer 2 the two residues are swung towards *B*21 of its dimer-related monomer enabling the formation of a very weak hydrogen bond between NH *B*29 and O *B*21 (N—O distance 3·88 Å). Density for the side chain of Lys

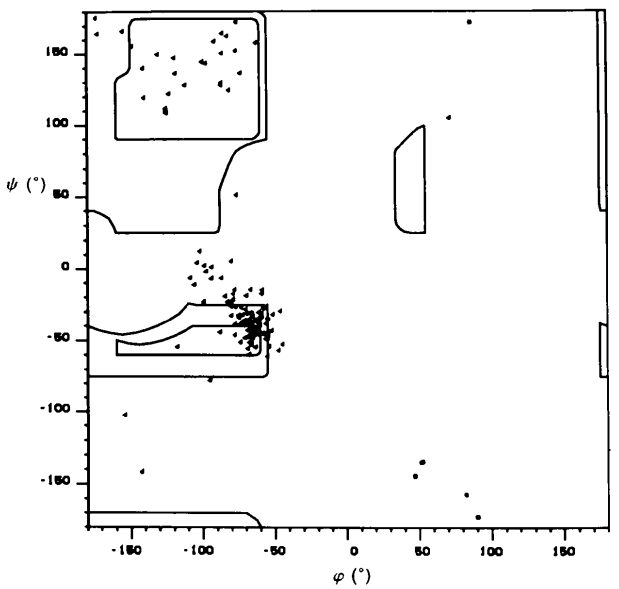

Fig. 6. Ramachandran plot.

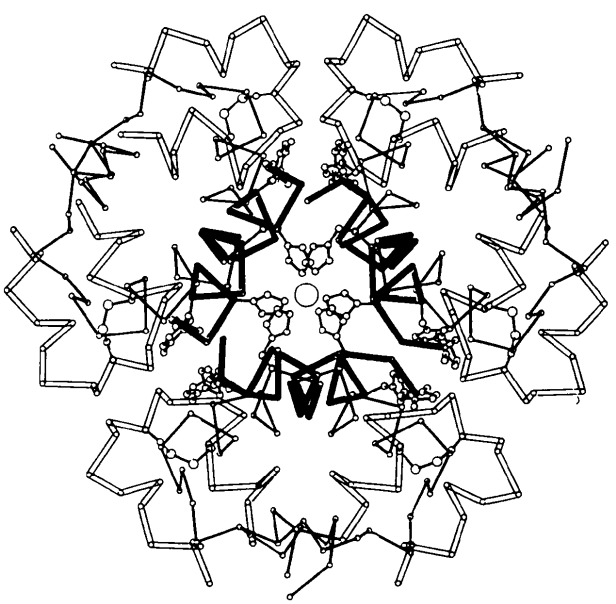

Fig. 7. C_α plot of the NPH-insulin hexamer with extended *B* helix emphasized for both trimers. Imidazoles of His *B*5 and His *B*10, and the positions of zinc ions and *m*-cresols are shown. *A* chains: double lines. *B* chains: single lines.

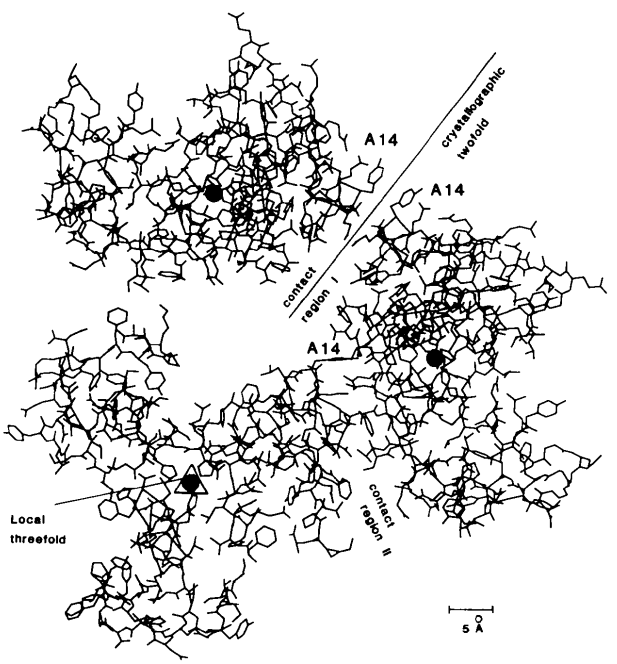

Fig. 8. Crystal contacts in tetragonal NPH-insulin crystals. The two different types of contact between trimers from different hexamers are shown as described in the text. The trimers related by the crystallographic twofold [1, −1, 0] at $z = \frac{1}{4}$ are viewed at a general projection angle. The third trimer is orientated with its local threefold perpendicular to the plane of the paper.

Table 3. *Radius of gyration and r.m.s. distances from centre of hexamer for main-chain atoms*

Crystal form	Radius of gyration (Å)	R.m.s. distance (Å)
NPH-insulin	15·38	17·98
Monoclinic insulin	15·47	18·08
2Zn-insulin	16·35	18·53
4Zn-insulin	16·38	18·53

Table 4. *R.m.s. differences* (Å) *between NPH-insulin monomers least-squares-fitted to* 2Zn- *and* 4Zn-*insulin monomers* (*main-chain atoms A*5 *to A*20, *B*10 *to B*26 *only*)

NPH-insulin	Monomer 1	Monomer 2	Monomer 3
2Zn-insulin monomer 1	0·86	0·87	0·89
2Zn-insulin monomer 2	0·74	0·76	0·82
4Zn-insulin monomer 1*	0·43	0·47	0·42
4Zn-insulin monomer 2	0·77	0·65	0·82

* Monomer 1 is the 4Zn monomer *without* the helix at the beginning of the *B* chain.

Table 5. *R.m.s. differences* (Å) *between the NPH-insulin monomers*

	Monomer 2	Monomer 3
Monomer 1	2·59[a]	1·18[a]
	1·64[b]	0·77[b]
	0·49[c]	0·46[c]
Monomer 2		2·57[a]
		1·63[b]
		0·43[c]

Notes: (*a*) for all atoms, (*b*) main-chain atoms only, (*c*) main-chain atoms *A*1–*A*21, *B*4–*B*26.

*B*29 is not observed but the residue is moved close enough to Gln *A*15 and Asn *A*18 of a different hexamer to enable hydrogen-bond formation between side chains in a crystal contact. The carboxy terminus is also within hydrogen-bonding distance of density which has been identified as protamine.

Crystal contacts

Each insulin hexamer makes essentially two different types of crystal contact (Fig. 8): contact region I is between two staggered hexamers with centres at (0·43, 0·43, 0) and (0·57, 0·57, 0·5) which share the twofold axis along the [1, −1, 0] diagonal. Contact region II is between the 'top end' of one hexamer and the side of its equivalent rotated by −90 and translated by (0, 0, 0·25).

Contact region I is related by the diagonal diad through (0·5, 0·5, 0·25) and therefore involves the same residues on both neighbouring hexamers. The contact region consists mostly of residues of the *A* chain of monomer 1: Thr *A*8 to Ser *A*12 are packed against their symmetry-related equivalents. The contact is mediated by water molecules and possibly protamine. Direct contact with possible hydrogen bonding can be observed between Ser *A*12 (monomer 1) or Gln *A*15 (1) and Glu *B*21 (3) close by. Hydrogen bonding between Gln *A*15 (1) or Asn *A*18 (1) and Lys *B*29 (2) seems feasible but has not been verified from the electron density. The phenol of Tyr *A*14 (1), although near the contact region, is completely disordered and does not contribute to the interhexamer contact. Contact region II is close to contact region I and involves the residues between Tyr *A*14 and Tyr *A*19 of monomer 2. The residues pack against residues *A*7 and *A*8 and the beginning of the *B* chain of monomer 1. The contact is strengthened by possible hydrogen bonds between the carbonyl oxygen of Tyr *A*14 (2) and the side-chain NH_2 of Asn *B*3 (1). Oδ1 of Asn *A*18 (2) can bind to the main-chain NH groups of *B*2 (1) and *B*3 (1). The side chain of Tyr *A*14 (2) is crystallographically well defined. It extends into a shallow pocket formed by *B*3 (1), *B*4 (1) and *A*7 (1) of the symmetry-related trimer. Although there is van der Waals contact the side-chain OH does not seem to be involved in any hydrogen bonding. Indirect hydrogen bonding *via* a network of water molecules can also be detected. This crystal contact wedges in the beginning of the *B* chain of monomer 1 which explains its lower *B* values for *B*1 and *B*2 compared with the other two monomers (Fig. 5*b*).

Zinc coordination geometry

The zinc ions in NPH-insulin are tetrahedrally coordinated. Three of the coordination sites are provided by the imidazole nitrogens of His *B*10, the fourth site is occupied by an unidentified ligand (Fig. 9). The threefold nature of this arrangement is not preserved exactly. A least-squares superposition of the three monomers based on the main-chain atoms of *B*6 to *B*20 shows that there is a large difference in the relative positions of the imidazole rings of His *B*10 whose atoms have an average r.m.s. difference of 0·43 Å, which is more than twice than those of the main-chain atoms of *B*10 (0·20 Å). It is highly improbable that these deviations are artefacts, especially as the electron density for the imidazoles is well defined with average *B* values of 11·45 $Å^2$ comparable to those of the main-chain atoms of this residue. Recently, almost identical differences have also been observed for the *B*10 imidazoles in insulin cocrystallized with salmine which has been refined independently in this laboratory (Xiao, Korber & Dodson, to be published). As can be seen from Fig. 10 the bond lengths from the imidazole nitrogens to the zinc vary from 1·74 to 2·18 Å and the central angles range from 97 to 118°. These distortions can be traced to the crystal contact at the top of the

hexamer which through a push on the *B*-chain helix brings the side chain of Leu *B*6 into van der Waals contact with the imidazole of His *B*10 of monomer 3, lengthening the distance to the zinc.

The nature of the fourth ligand has not yet been established. The examination of both the density of a $2F_o - F_c$ map at 0·5 e Å^{-3} and an omit difference map where protamine and water were excluded from the calculation of the structure factors at 0·25 e Å^{-3} suggests that the ligand fits into an elongated pear-like shape which has space to accommodate at least two non-H atoms. The density is compatible with a two-atomic arrangement with one atom at a distance of 2·25 Å from the zinc and the other at 4·2 Å in van der Waals contact with side-chain atoms of two of the Leu *B*6. The Zn atom and the two ligand atoms proposed lie on a straight line. The observed distance between the zinc and its nearest ligand atom is compatible with the binding of either a water molecule or a chloride ion. For the binding of chloride to zinc, distances between 2·2 and 2·3 Å have been observed in small molecule structures (Jacobi & Brehler, 1969; Follner, 1970). The average bond length of water to four-coordinated zinc is 2·006 Å [Orpen, Brammer, Allen, Kennard, Watson & Taylor, 1989; ligand class identifier (l.c.i.) 5.22.1]. The distance between the atom directly liganded to the zinc and the second atom in the density, however, is too close to account for the coordination of a water molecule to either chloride or water: water–chloride distances in the $ZnCl_2$ crystal structure, where the zinc is also tetrahedrally coordinated, range from 3·2 to 3·8 Å (Follner, 1970) and the water–water distance is commonly assumed to be 2·8 Å (Bondi, 1964; Lee & Richards, 1971). Alternatively, the observed pear-shaped electron density can be interpreted as a small organic molecule, such as an acetate ion. Urea, which is present in the crystallization medium, is known to attach itself to zinc-like ions *via* its carbonyl group. Listed oxygen–metal distances are close to 2·1 Å (Orpen *et al.*, 1989; l.c.i. 5.4.2), in agreement within the error limits with the observed value for the ligand closest to the zinc. There is, however, no observed electron density for the two nitrogens near the side chains of Leu *B*6 to support this hypothesis.

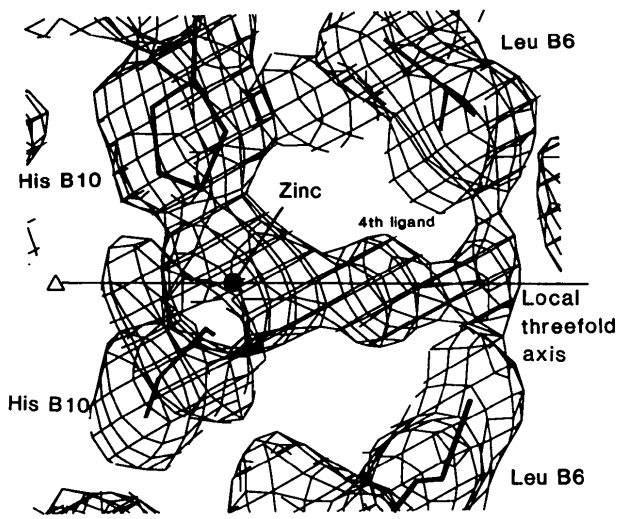

Fig. 9. Zinc environment. View perpendicular to non-crystallographic triad. $2F_o - F_c$ map at 0·5 e Å^{-3}.

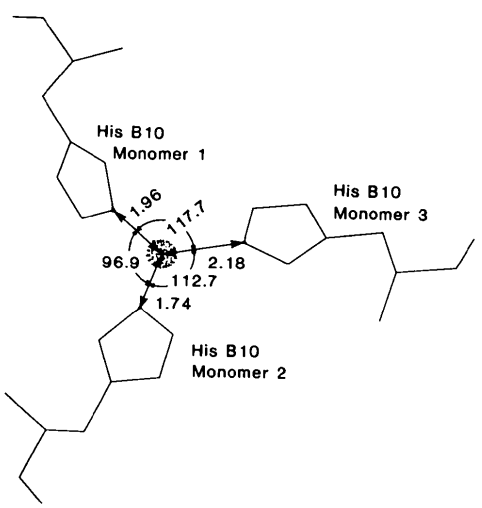

Fig. 10. The coordination of His *B*10 imidazoles to the zinc.

Binding of the m-cresol

One fully occupied *m*-cresol binding site is situated on each monomer near the imidazole of His *B*5. The *m*-cresol pocket is also bounded by *A*6 to *A*11 and *B*10 of a different monomer within the same trimer (Fig. 11). The electron density for the ring system and the hydroxyl is well defined for all *m*-cresol molecules but density for the methyl group can only be detected for two of the three *m*-cresols. The interactions of the *m*-cresol with the insulin hexamer are made *via* two hydrogen bonds to the CO of *A*6 and the NH of *A*11 as is the case for phenol in the monoclinic structure (Derewenda *et al.*, 1989). Van der Waals contacts to insulin residues are weak if only the unsubstituted ring (phenol) is considered. For the *meta*-substituted ring system one of the possible positions for the additional methyl group is in fact blocked by a severe collision of the methyl with Cβ of *B*10 (Table 6). Owing to the coordination of the imidazole of *B*10 to the zinc these clashes cannot be easily removed. For the second possible *meta* position the clashes are not so severe: The offending atoms are different for all three *m*-cresols and belong to side chains that can be moved out of

the way relatively easily. In case of the very short distances to Cδ2 of *A*16 there the electron density map is smeared out more than for comparable side chains. This indicates that the side-chain atoms possibly move by up to 1 Å to accommodate the *m*-cresol. The *m*-cresol near *B*5 of monomer 2 is unique in that it appears to adopt two different positions for its hydroxyl group, one in which it makes hydrogen bonds with both O *A*6 and NH *A*11, the other with just a single hydrogen bond to O *A*6. Judging from the electron density, both arrangements occur with approximately equal probability. The density for the methyl group in the *meta*-substituted position is of the same magnitude as that for the atoms of the phenol ring, indicating full occupancy of this position. As we shall outline below we explain this fact by the presence of protamine which also appears to force the *m*-cresol to give up one of its hydrogen bonds and assume its second position.

Table 6. *Modelled close contacts* (Å) *of cresols in different conformations*

Ideal distances are given in parentheses.

Conformation	Coordination to *B*5 of			
contact	Monomer 1	Monomer 2	Monomer 3	
meta 1	2·65 (3·4)	2·88 (3·4)	2·75 (3·4)	Cβ *B*10
			3·07 (3·4)	Cδ *B*6
meta 2	3·19 (3·4)			Cδ *B*17
		2·42* (3·4)		Cδ *A*16
			3·25 (3·4)	Cβ *B*14
ortho 1	2·29 (3·05)	2·22 (3·05)	1·94 (3·05)	O *A*11
ortho 2	2·85 (3·35)		3·06 (3·35)	Sγ *B*7
	2·50 (3·05)		2·94 (3·03)	O *A*6
para		2·72 (3·4)		Cβ *B*14
			2·77 (3·05)	O *B*10
			3·18 (3·4)	C *B*10

* Electron density indicates that close contact can be alleviated by movement of the side-chain atoms.

The main protamine binding site

The clupeine Z molecule has 31 residues of which 21 are arginines. It is not likely to have an inherently well defined three-dimensional structure. The interactions of clupeine Z with insulin are assumed to be through binding of the clupeine Z guanidinium groups to carboxylic acid side chains or carboxy termini on the insulin. The interactions between the polypeptide and insulin are sufficiently specific to establish a unique and well defined stable crystal lattice. Inspection of the electron density and difference electron density reveals several continuous threads which are consistent with peptide structure. It has not been possible, however, to identify any fragment of sequence of the clupeine Z molecule unambiguously. The stoichiometry of clupeine Z to the insulin hexamer is 0·85, which suggests that the molecule if present in a single binding site will have reasonably high occupancy. If, however, the peptide is distributed among more than one site its density will be confused with that of water molecules and a confident interpretation of the peptide's structure and interactions will be impossible. Pending a detailed investigation of the isomorphous structure of pig insulin cocrystallized with salmine we shall here only discuss the protamine density near the dimer–dimer interface for which we have not only direct and persistent evidence from the electron density maps but also supportive evidence from its influence on the protein structure and *m*-cresol orientation.

A narrow tube of density compatible with that expected for a polypeptide chain can be observed at the dimer–dimer interface between monomer 1 and the symmetry equivalent of monomer 3 (Fig. 12). Positive but diffuse difference electron density can be found at a very low level (1·5 r.m.s.) near crystal contact region 1. The density passes the disordered side chain of Tyr *A*14 and reaches a level of more than 0·5 e Å $^{-3}$ in the $2F_o - F_c$ map at Leu *A*13, having now acquired its tubular shape. The feature

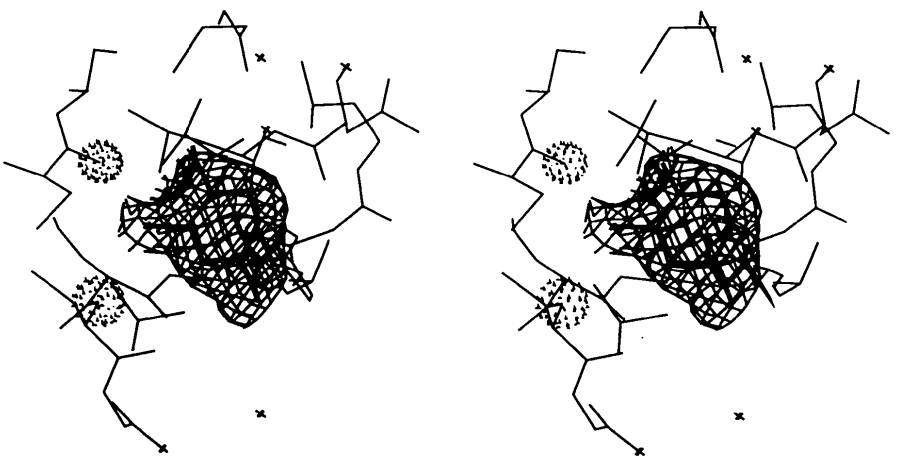

Fig. 11. The environment of the unique *m*-cresol. $2F_o - F_c$ map at 0·5 e Å $^{-3}$. O *A*6 and N *A*11 are highlighted.

stretches to the imidazole of His *B*5 of monomer 2 and at slightly lower level extends beyond that towards the methyl group of the bound *m*-cresol. There is apparently a connection to a guanidinium shaped piece of electron density closer to the central cavity near His *B*10 but this cannot be identified unambiguously at 0·4 e Å $^{-3}$. Near His *B*5 we found two branches off the main string of density described above, one of which is within contact distance of the carboxy terminus of the *B* chain of monomer 2. The other branch has density near Phe *B*24. At present similar features have not been detected at the other dimer–dimer interfaces. Although the density for the protamine is too disordered to define the interactions with the insulin at atomic level, our interpretation is supported by the unique disorder observed for the *m*-cresol near His *B*5 of monomer 2, which seems in contact with the protamine. It is consistent with the observed electron density that the energetically less favourable conformation of the *m*-cresol results from the collision of its methyl group with the clupeine Z apparently extending towards His *B*10. Similarly, the second conformation observed for the carboxy terminus of monomer 2 could result from favourable association with the protamine. Lastly, the density associated with clupeine Z extends to the outside of the hexamer into the crystal contact region and thus fulfils the condition imposed by the continuity of the polypeptide chain.

Discussion

The NPH-insulin crystal structure and the monoclinic crystal form of insulin are the only crystalline insulins analysed so far which have an extended helix at the beginning of the *B* chain for all the insulin monomers. The helical fold of the *B* chain is evidently brought about by the binding of chloride, phenol, *m*-cresol or related agents. Retardation of the action of insulin is caused by slowing the entry of insulin from the subcutaneous compartment into the bloodstream, basically by lowering the solubility of the peptide at physiological pH. In the case of NPH-insulin this is provided by admixing protamine, together with low amounts of zinc salt, phenol or *meta*-substituted phenol. The initially amorphous precipitate is then gradually transformed into microcrystals, and there is circumstantial evidence that this crystallization leads to a further retardation of the uptake of insulin, as demonstrated for the Lente-type insulins (Binder, 1969). It has also been shown that *p*-cresol and *o*-cresol are inferior crystallizing agents compared with *m*-cresol (Krayenbuhl & Rosenberg, 1946). It is therefore of particular interest to define the interactions between zinc, *m*-cresol, protamine and insulin, to understand better the forces that stabilize the insulin moiety in the crystals and hence advance our insight into the mechanisms of retardation. Although the exact mode of interaction of the protamine molecule with the insulin moiety could not be established, predominantly due to disorder in the protamine peptide, we can nevertheless put forward suggestions for the enhanced stability of the NPH crystals: one contributary factor is the binding of *m*-cresol near His *B*5 which results in the formation of an extended *B* helix and hence a tightening of the secondary structure. As the binding and subsequent helical fold is also observed in solution (Wollmer, Rannefeld, Johansen, Hejnaes, Balschmidt & Hansen, 1987) it is evident that it is associated with a drop in free energy and the complex is stabilized. The helical structure at *B*1 to *B*8 also covers the zinc binding site and greatly reduces zinc exchange. Consequently, the hexamers' stability is increased significantly. It is not clear whether these effects operate in the trimers.

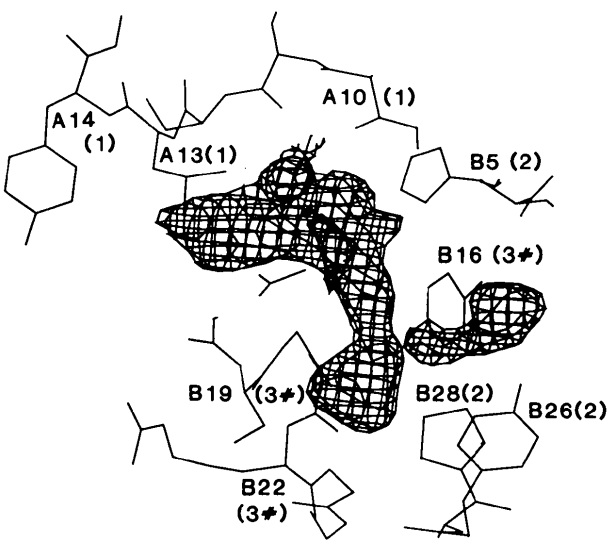

Fig. 12. The protamine density at the dimer–dimer interface. Difference density at 0·25 e Å $^{-3}$.

From model building studies (Table 6) no compelling reason could be found why *p*-cresol and *o*-cresol should be inferior crystallization agents. In fact, crystals of insulin with 4-iodophenol have been grown in this laboratory recently and crystallize in an orthorhombic space group similar to the NPH-insulin crystals (Whittingham, Dodson & Korber, unpublished results). At least for the *m*-cresol binding site where we have located protamine density nearby, binding of *p*-cresol should actually be preferential, in that both hydrogen bonds, the one to O *A*6 and NH *A*11 can then be maintained all of the time. Molecular dynamics studies are planned on this particular problem in the near future.

It is not yet clear whether the clupeine Z is involved in the further stabilization of the zinc ions

although there are indications that the protamine can be found close to the central cavity. Further model building and comparison with the salmine NPH insulin structure will hopefully elucidate this point in particular. However, we are convinced that protamine mediates the interactions at one dimer–dimer interface and extends into a crystal contact region. Thus it may also be involved in hexamer–hexamer linkage. The expected ion-pair formation between the guanidinium groups of the protamine and carboxylic acid groups of the insulin is likely with the carboxy terminus of one of the *B* chains although the electron density does not provide detail at atomic resolution. This lack of detail concerning the protamine can partly be attributed to crystallographic disorder, partly by disorder through the conformational flexibility of the protamine itself. Furthermore, the high arginine content of the clupeine Z creates the possibility of different alignments of the protamine whilst still meeting the requirements of most of the possible binding sites. It is thus obvious that partial disorder of the clupeine Z moiety is inevitable, which makes it difficult to observe specific interactions.

We are therefore investigating the structures of insulins cocrystallized with salmine and dekalysine in the hope that the comparison of these structures will provide a clearer picture of the protamine–insulin interactions.

We are grateful to the SERC and their staff at SRS, Daresbury, for the use of their synchrotron radiation facilities. We should also like to thank Dr Bruno Hansen of the Hagedorn Research Laboratory for suggesting the project initially and maintaining a lively and active interest in it ever since.

References

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
ANDO, T. & WATANABE, S. (1969). *Int. J. Protein Res.* **1**, 221–224.
BAKER, E. N., BLUNDELL, T. L., CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., CROWFOOT HODGKIN, D. M., 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.
BAKER, E. N. & DODSON, G. G. (1970). *J. Mol. Biol.* **54**, 605–609.
BENTLEY, G., DODSON, E., DODSON, G., HODGKIN, D. & MERCOLA, D. (1976). *Nature (London)*, **261**, 166–168.
BINDER, C. (1969). *Acta Pharmacol. Toxicol.* **27**, 1–87.
BONDI, A. (1964). *J. Phys. Chem.* **68**, 441–451.
DEREWENDA, U., DEREWENDA, Z., DODSON, E. J., DODSON, G. G., REYNOLDS, C., SMITH, G. D., SPARKS, C. & SWENSON, D. (1989). *Nature (London)*, **338**, 594–596.
DODSON, E. J. (1985). *Molecular Replacement*, edited by P. A. MACHIN, pp. 33–45. Daresbury, England: SERC.
EGGENA, P., MAGDOFF-FAIRCHILD, B., RUDKO, A., FULLERTON, W. W. & LOW, B. W. (1969). *Acta Cryst.* A**25**, S-184.
FOLLNER, H. (1970). *Acta Cryst.* B**26**, 1544–1547.
FULLERTON, W. W. & LOW, B. W. (1970). *Biochim. Biophys. Acta*, **214**, 141–147.
GREENHOUGH, T. J. (1987). *Computational Aspects of Protein Crystal Data Analysis*, edited by J. R. HELLIWELL, P. A. MACHIN & M. Z. PAPIZ, pp. 51–57. Daresbury, England: SERC.
GREENHOUGH, T. J. & SUDDATH, F. L. (1986). *J. Appl. Cryst.* **19**, 400–409.
HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.
IWAI, K. & NAKAHARA, C. (1971). *J. Biochem.* **69**, 493–509.
JACOBI, H. & BREHLER, B. (1969). *Z. Kristallogr.* **128**, 390–405.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
KRAYENBUHL, C. & ROSENBERG, T. (1946). *Rep. Steno Mem. Hosp. Nord. Insulinlab.* **1**, 60–73.
LEE, B. & RICHARDS, F. (1971). *J. Mol. Biol.* **55**, 379–400.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
MICHAEL, S. E. (1967). GB Patent 1233540.
ORPEN, G. A., BRAMMER, L., ALLEN, F. H., KENNARD, O., WATSON, D. G. & TAYLOR, R. (1989). *J. Chem. Soc. Dalton Trans.* pp. S1–S83.
ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225–258.
SIMKIN, R. D., COLE, S. A., OZAWA, H., MAGDOFF-FAIRCHILD, B., EGGENA, P., RUDKO, A. & LOW, B. W. (1969). *Biochim. Biophys. Acta*, **200**, 385–394.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
WOLLMER, A., RANNEFELD, B., JOHANSEN, B. R., HEJNAES, K. R., BALSCHMIDT, P. & HANSEN, F. B. (1987). *Biol. Chem. Hoppe-Seyler*, **368**, 903–911.
WONACOTT, A. J., DOCKERILL, S. & BRICK, P. (1980). *MOSFLM*. Unpublished notes.