Acta Cryst. (1983). B**39**, 90–98

Molecular-Replacement Studies on Crystal Forms of Despentapeptide Insulin

BY BI RU-CHANG*

Department of Chemistry, University of York, Heslington, York YO1 5DD, England

S. M. CUTFIELD

Biochemistry Department, University of Otago, Box 56, Dunedin, New Zealand

E. J. DODSON AND G. G. DODSON

Department of Chemistry, University of York, Heslington, York YO1 5DD, England

F. GIORDANO

Università di Napoli, Via Mezzocannone 4, 80134 Napoli, Italy

AND C. D. REYNOLDS AND S. P. TOLLEY

Department of Chemistry, University of York, Heslington, York YO1 5DD, England

(*Received* 15 *April* 1982; *accepted* 27 *July* 1982)

Abstract

Despentapeptide insulin, prepared by removing the five C-terminal B-chain residues, occurs in several crystal forms. The relationship between these and the known pig 2-Zn insulin structure is analysed by the fast rotation function and *R*-factor searches. Reasonably good models were obtained even in the cases when there are two molecules in an asymmetric unit. The 2-Zn insulin coordinates were positioned in the *C*2 sheep cell and refinement commenced using the fast Fourier least-squares method. The packing of this modified insulin is quite different from that of all other insulins studied so far but the molecular structure is similar to that of 2-Zn insulin.

Introduction

The relationships between insulin's three-dimensional structure and its biological properties is being analysed through chemical and structural studies of different species and crystalline forms. Insulin analogues with B-chain C-terminal residues removed are particularly interesting owing to the important role these residues have in aggregation and activity. Among these modified insulins despentapeptide insulin (DPI), obtained by enzymatic removal of the five B-chain terminal residues (Fig. 1), retains considerable biological activity (The Shanghai Insulin Research Group, 1976*a*) and a high affinity for insulin antibodies (Reeves, 1981). In general, removing more residues from the C terminus has been shown to reduce greatly the activity of insulin (Kikuchi, Larner, Freer, Doy, Morris & Dell, 1980; Meyts, Obberghen, Roch, Wollmer & Brandenburg, 1978).

Insulin molecules in the 2-Zn pig insulin crystals and other crystals studied so far are organized as aggregates such as dimers and hexamers (Adams *et al.*, 1969;

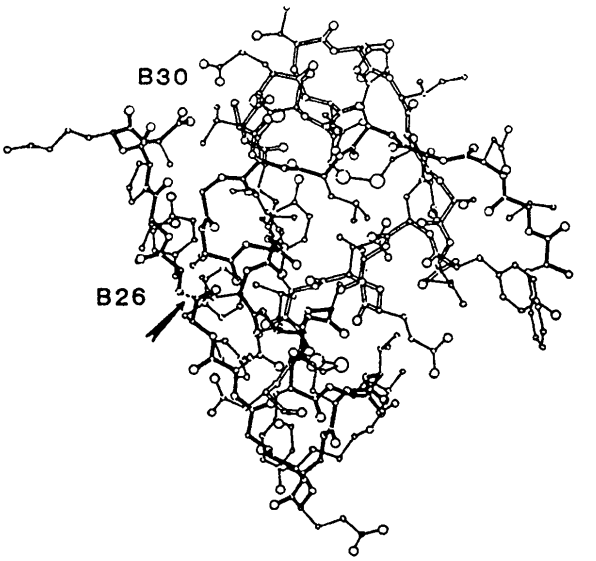


Fig. 1. The insulin molecule. The five C-terminal B-chain residues and their point of cleavage from insulin are indicated.

* Permanent address: Institute of Biophysics, Academia Sinica, Peking, China.

0567-7408/83/010090-09$01.50

Bentley, Dodson, Dodson, Hodgkin & Mercola, 1976; Bentley, Dodson, Dodson & Levitova, 1978; Cutfield, Cutfield, Dodson, Dodson, Emdin & Reynolds, 1979). The geometry of the dimer interaction is largely directed by the antiparallel β sheet between B24–B26 peptide segments of two monomers which are related by a local twofold axis. As a result of losing these five residues, the DPI molecules cannot form this β sheet and, therefore, remain as monomers in solution (The Shanghai Insulin Research Group, 1976*b*). There is persuasive evidence that insulin acts as a monomer in the body, and thus studies on DPI are an important step in relating the hormone's structure and function.

Crystals of DPI insulin have been obtained in several laboratories (Gattner, 1975; Peking Insulin Structure Research Group, 1976), and a low-resolution map of pig DPI phased with a single isomorphous heavy-atom derivative has been reported (Li Jia-yao, Song Shih-yan, Li Jun, Wang Jia-hai & Yau Tsi-ho, 1980). In York we have prepared good crystals of beef DPI in two forms, and of sheep DPI.

The usual methods for determining phases by isomorphous replacement with heavy atoms has proved difficult with DPI (Cutfield, 1981; Wang Jia-hai *et al.*, 1980). The very low solvent volume (about 20%) implies a lack of suitable available reacting surfaces, and this is probably responsible for the failure to bind heavy atoms. The molecular-replacement method in which the detailed knowledge of the insulin structure could be expoited was therefore applied.

The rotation function (Rossmann, 1962), used to determine the relative orientations of different molecules, overlaps two Patterson maps, $\rho(\mathbf{u}_0)$ and $\rho(\mathbf{u}_1)$ over a sphere centred on the origin, and should show a maximum when $\mathbf{x}_0, \mathbf{x}_1$, the coordinates which generate $\mathbf{u}_0$ and $\mathbf{u}_1$, satisfy

$$\mathbf{x}_0 = [R]\,\mathbf{x}_1 + \mathbf{t}.$$

Only $[R]$, defined by three angular variables, can be found from this Patterson overlap.

The three translational parameters were obtained by searching for a minimum R factor between observed intensities and the calculated ones obtained from a molecule with correct orientation which is translated over the unit cell.

In other insulin studies this procedure has worked well (Bentley *et al.*, 1978; Cutfield, Cutfield, Dodson, Dodson & Sabesan, 1974), but there were particular difficulties with DPI. First it was possible that the removal of the five B-chain residues had significantly disturbed its molecular structure and secondly the close packing of the molecules greatly weakens the approximation that the intramolecular and intermolecular vectors are separable by choice of an appropriate sphere of integration.

Methods and materials

(1) *Crystallization and data collection*

The DPI samples from beef and sheep insulin were prepared following the method of Gattner (1975). The DPI crystals were grown from a solution prepared by dissolving the DPI in a buffer with 0·05 *M* sodium citrate and 0·02 *M* tris (2-amino-2-hydroxymethyl-1,3-propanediol), to which were added by volume a 20% solution of $(NH_4)_2SO_4$ at 40% saturation and 25% acetone by volume. The pH was adjusted to a value of about 6·1. The concentration of the DPI in the final solution ranged from 4 to 5 mg ml^{-1}. The solution was filtered using a millipore filter, warmed to 323 K and then allowed to cool slowly in a Dewar flask.

Until recently only small crystals have been grown with typical dimensions of about 0·5 × 0·2 × 0·1 mm. Several crystal forms were obtained; their crystallographic details are given in Table 1.

In addition, the Peking Insulin Structure Research Group kindly sent 4 Å data collected on a *C*2 pig DPI form.

X-ray diffraction data to 2·1 Å resolution were collected for the *C*2 sheep, *C*2 beef and $P2_12_12_1$ beef forms on a four-circle Nonius CAD-4 diffractometer controlled by a PDP-8 computer, using Cu $K\alpha$ radiation. The ω-scan technique was used and, because of the large number of weak reflections, the intensities were estimated by the profile-fitting method. Radiation damage was monitored by measuring three standard reflections after each batch of fifty. When the intensities of the standards dropped by about 10% the crystal was replaced. The intensities were corrected for Lorentz and polarization effects, radiation damage, and for absorption by the empirical method of North, Phillips & Mathews (1968). Fig. 2, which is a plot of $\langle F/\sigma \rangle$ against 4 $\sin^2\theta/\lambda^2$, illustrates the quality of the data sets.

Table 1. *Crystallographic data for the DPI crystal forms*

Species	Space group	*a* (Å)	*b* (Å)	*c* (Å)	β (°)	Volume of asymmetric unit (Å^3)	No. of molecules per asymmetric unit	Resolution (Å)
Beef	*C*2	52·74	26·11	51·64	93·41	17 750	2	2·1
Beef	$P2_12_12_1$	27·81	43·52	57·77	90·00	17 477	2	2·15
Sheep	*C*2	53·90	25·90	25·60	93·80	8606	1	2·05
Pig	*C*2	58·70	27·90	24·00	100·6	9659	1	4·0
Model	*P*1	36·00	42·00	60·00	90·00	—	1	—

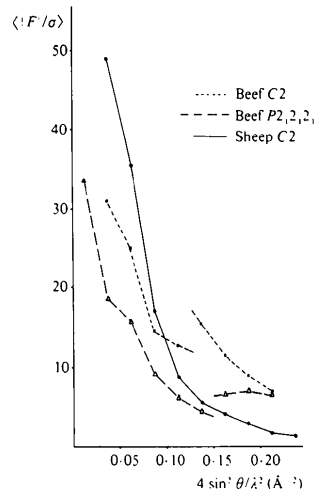


Fig. 2. Distribution of F/σ for beef and sheep DPI $C2$ crystals. The higher-angle diffraction data were collected using larger crystals.

(2) *Calculation of the C2 beef Patterson function*

There is a marked similarity between the X-ray intensity patterns of the $C2$ sheep and $C2$ beef forms, allowing for the doubling of the c axis in the $C2$ beef crystals. At a resolution of less than 5 Å the reflections with odd l indices were all very weak in the $C2$ beef diffraction pattern.

Patterson maps were calculated using all the data between 10–2 Å (Fig. 3), and also with limited data over the ranges 10–4·5 Å and 4·5–3 Å. All these maps gave a strong 'translation' peak, one third the origin in height, at $u = 0{\cdot}0$, $v = 0{\cdot}08$, and $w = 0{\cdot}5$. This suggested that the two molecules present in the $C2$ beef asymmetric unit are related by a translation only and that the translation is fairly exact, since the peak persists at high resolution.

(3) *Calculation of the rotation functions*

The relative orientations of all pairs of DPI molecules to each other were investigated by the Crowther fast rotation function program (Crowther, 1972), as was the match between each DPI form and a model structure based on the 2-Zn pig insulin molecule. (This model was constructed by removing B26–B30, B1–B4, which have very different conformations in 2-Zn and 4-Zn insulin, the termini of the A chain, plus many of the side chains on the surface of the molecule. CB atoms were retained for all these residues. This molecule was placed in a $P1$ unit cell with cell dimensions about twice the molecular dimensions, so that no intermolecular vectors would hamper the rotation search.)

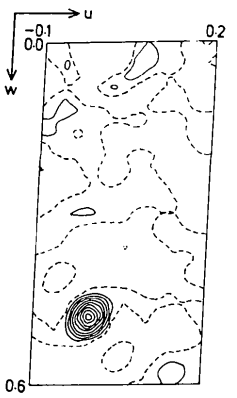


Fig. 3. The beef DPI $C2$ Patterson map, section $v = 0{\cdot}083$, calculated with 10–3 Å data.

If the two sets of coordinates are defined relative to an orthonormal axial system **I**, **J**, **K**, the matrix $[R]$ can be written as a function of three Eulerian angles, α, β, and γ.

If α is a rotation about the initial direction of $\mathbf{K}_0$, β is a rotation about the subsequent direction of $\mathbf{J}_1$, and γ is a rotation about the final direction of $\mathbf{K}_2$, then

$$[R] = \begin{bmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix}.$$

The crystal symmetry will obviously be reflected in the rotation-function solutions.

Say $[R]$ and $[R']$ are rotation matrices relating symmetrically equivalent molecules.

Then, if

$$\mathbf{X}_0 = [R]\,\mathbf{X}_1,$$

there is a matrix $[R']$ such that

$$[S_{0_i}]\,\mathbf{X}_0 = [R'][S_{1_j}]\,\mathbf{X}_1$$

and

$$[R] = [S_{0_i}]^{-1}[R'][S_{1_j}],$$

where the $[S_{0_i}]$ are the symmetry operations to apply to the orthogonal coordinates $\mathbf{X}_0$, and $[S_{1_j}]$ are applied to the orthogonal $\mathbf{X}_1$.

The program has been modified to generate these symmetry matrices from the input crystal-symmetry cards and the orthogonalizing matrices.

The large number of possible comparisons for DPI means that any result can be tested for consistency with all the others, that is, if $\mathbf{X}_0 = [R_{01}]\mathbf{X}_1$, $\mathbf{X}_1 = [R_{12}]\mathbf{X}_2$ and $\mathbf{X}_0 = [R_{02}]\mathbf{X}_2$ we would hope that $[R_{02}]$ would equal $[R_{01}][R_{12}]$. However, this will only be true for one of each of the symmetry equivalents of the $[Rij]$ matrices. We modified the program to list all equivalent $(\alpha\beta\gamma)$ sets for each rotation maximum, thus avoiding many potential user errors!

These comparisons proved to be vital in drawing conclusions from the rotation functions. The quality of many of the maps was very disappointing; maxima

were different in maps using different shells of intensity data, with slightly different radii for the spheres of integration, and with different sharpening factors for the intensities.

Table 2 lists the observed and derived values of $(\alpha\beta\gamma)$ for the various $[Rij]$ matrices. Only those symmetry-equivalent angles which relate the different molecules are listed. Details for the calculations are given in Table 3.

(4) *Comments on individual rotation functions*

(i) *C*2 *beef*/*C*2 *beef*. The cell dimensions of this form and the *C*2 sheep form predicted that there would be an approximate repeat along the *z* axis. This was confirmed by the Patterson function. As expected, the rotation maps had no consistent maxima other than those due to crystal symmetry. The translation of one molecule from the other by (0·0, 0·08, 0·5) generates pseudo 2_1 symmetry which is illustrated in Fig. 8.

(ii) *C*2 *beef*/$P2_12_12_1$ *beef*. This rotation map was one of the few convincing results. The maxima persisted through all the resolution shells, and was well above background. There was only one such peak, which suggested that the two molecules in the $P2_12_12_1$ beef cell must also be related by some sort of twofold axis.

(iii) $P2_12_12_1$ *beef*/$P2_12_12_1$ *beef*. If this cell had two molecules in the asymmetric unit related by a twofold screw axis it was reasonable to hope that the self-rotation map would contain a convincing peak. However, the only possible maximum was a shoulder on the origin streak which extended from $\beta = 0$ to $\beta = 10$ (Fig. 4). This could not be pinpointed with any exactness. But, if we assume that the *C*2 beef/$P2_12_12_1$ beef result is right, and that there is a twofold screw axis relating the molecules in the $P2_12_12_1$ form we can deduce the expected position of the self-rotation maximum; this is marked on Fig. 4 and is consistent with the observed shoulder.

(iv) *C*2 *beef*/*model*. These maps with one large peak on section $\beta = 85°$ were also quite good. There was, however, alarming variation in peak position and height as the calculation parameters were changed slightly (Fig. 5).

(v) $P2_12_12_1$ *beef*/*model*. This map had one clear and persistent peak, where we had hoped to find two defining the orientation of the two molecules exactly. This good peak was beautifully consistent with the *C*2 beef results (see Table 2). The position of the second predicted peak could be calculated, and Fig. 6 shows the best match we obtained for it. The good peak

Table 2. *The rotation matrices and Eulerian angles relating the various DPI molecules*

The derived values for $(\alpha\beta\gamma)$ are obtained from appropriate matrix products (see text). Rotation matrices are applied to orthogonal coordinates relative to axes **I**o, **J**o, **K**o, defined using the following convention. For space group $P2_12_12_1$ **I**o is parallel to **a**, **J**o is parallel to **b**, and **K**o is parallel to **c**. For space group *C*2 **I**o is parallel to **c**, and **K**o is parallel to **b***.

$$\begin{bmatrix} Xi \\ Yi \\ Zi \end{bmatrix} = [Rij] \begin{bmatrix} Xj \\ Yj \\ Zj \end{bmatrix}$$

	Beef *C*2, molecule 1 ($j = 1$)	Beef *C*2, molecule 2 ($j = 2$)	Beef $P2_12_12_1$, molecule 2 ($j = 4$)	Model ($j = 7$)
Beef *C*2, molecule 1 ($i = 1$)				
$\alpha\beta\gamma$ (obs.)	(0 0 0) (defines $[R11]$)	(180 0 0) (defines $[R12]$)	(83 −85 −162) (defines $[R14]$)	(120 85 −70) (defines $[R17]$)
Beef *C*2, molecule 2 ($i = 2$)				
$\alpha\beta\gamma$ (obs.)	(0 0 180) ($[R21] = [R12]$)	(0 0 0) ($[R22] = [R11]$)		(120 −85 110) ($[R27] \approx [R17]$)
Beef $P2_12_12_1$, molecule 1 ($i = 3$)				
$\alpha\beta\gamma$ (obs.)	(162 85 97) ($[R31]$)	(162 85 −83) ($[R32] = [R14]^{-1}$)	(? ? ?)	(? ? ?)
$\alpha\beta\gamma$ (derived)			(162 170 18) (from $[R31][R14]$)	(74 37 −158) (from $[R31][R17]$)
Beef $P2_12_12_1$, molecule 2 ($i = 4$)				
$\alpha\beta\gamma$ (obs.)	(162 85 −83) ($[R41] = [R14]^{-1}$)	(162 85 97) ($[R42] = [R31]$)		(60 −145 −175) ($[R47]$)
$\alpha\beta\gamma$ (derived)				(57 −142 −175) (from $[R41][R17]$)
Pig *C*2 ($i = 5$)				
$\alpha\beta\gamma$ (obs.)	(158 22 20) ($[R51]$)		(85 90 0) ($[R54]$)	(122 −68 121) ($[R57]$)
$\alpha\beta\gamma$ (derived)			(82 90 −3) (from $[R51][R14]$)	(114 −68 125) (from $[R51][R17]$)
Sheep *C*2 ($i = 6$)				
$\alpha\beta\gamma$ (obs.)	(0 0 0)†			

† This result is expected from cell dimensions and is verified by the rotation function. All other results are the same as for beef *C*2.

Table 3. *Parameters used in calculating rotation functions, and resulting Eulerian angles*

For all the calculations listed the outer radius of the spherical shell of integration was 13 Å; the inner radius was equal to the limit of the data resolution.

A sharpening factor of exp ($B \sin\theta/\lambda$) was applied to each set of amplitudes to give a reasonably equal distribution of large intensities against (sin θ/λ).

The ratio listed is that of the height of this peak to that of the remaining highest in the map. For the $P2_12_12_1$ beef self search the result lies on the shoulder of the origin peak, and no value can be given for the ratio.

				Observed peak (°)			
Number	Function	B (Å^2)	d (Å)	α	β	γ	Ratio
(1)	$C2$ beef/$C2$ beef	10,10	12–5	150	85	30	1·06
			5–3	125	175	55	1·05
			12–3	130	50	50	1·03
(2)	$C2$ beef/$P2_12_12_1$ beef	10,15	12–5	75	−85	−160	1·06
			5–3	83	−85	−162	1·31
(3)	$P2_12_12_1$ beef/$P2_12_12_1$ beef	15,15	12–5	125	40	120	—
			5–3	160	180	20	—
			12–3	160	180	20	—
			3·5–2·5	160	180	15	—
(4)	$C2$ beef/model	10,20	11–4	116	−81	105	1·14
			8–3	120	−85	110	1·40
(5)	$P2_12_12_1$ beef/model	10,20	11–4	65	−140	−170	1·21
			8–3	60	−145	−175	1·70
		(second molecule)	8–3	90	30	−165	1·20
(6)	$C2$ pig/$C2$ beef	10,10	8–4	158	22	20	0·77
(7)	$C2$ pig/$P2_12_12_1$ beef	10,10	11–4	85	90	0	1·60
(8)	$C2$ pig/model	10,10	9–4	122	−68	121	1·09
(9)	$C2$ sheep/$C2$ beef	10,10	11–4	0	0	0	2·60
			6–4	0	0	0	2·41
			8–3	0	0	0	3·11
(10)	$C2$ sheep/$P2_12_12_1$ beef	10,15	9–4	90	−95	−155	0·80
			8–3	85	−80	−165	1·37
(11)	$C2$ sheep/model	10,10	8–4	120	−85	110	1·21

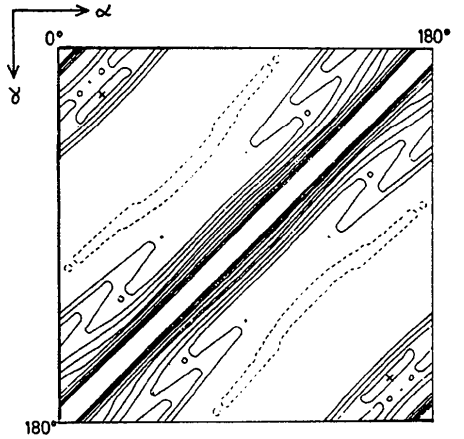


Fig. 4. The $\beta = 10°$ section of the map of the beef DPI $P2_12_12_1$ self-rotation function. The predicted peak (×) lies on the shoulder of the origin peak.

extends in the expected direction but there is no convincing second maximum.

There are several possible explanations for this poor result; perhaps the model molecule is not so similar to the second $P2_12_12_1$ beef molecule (4-Zn insulin shows that two chemically identical molecules may have very different conformations) (Bentley *et al.*, 1976). Alternatively, the errors inherent in all rotation functions may obscure the result. For the translational search we assumed that there were two molecules related by a twofold screw axis.

(vi) (vii) (viii) *C2 pig/other forms.* The limited intensity data made us wary of placing too much weight on the result. In all the other functions the best results were obtained using 3 Å data. However, the given set of three maxima are consistent with each other and the predicted results generated by functions (ii) and (iv).

(ix) *C2 sheep/C2 beef.* As expected from the inspection of the intensity-data sets, there is an excellent match between these two forms when the crystal axes are aligned. The pseudo 2_1 axis in $C2$ beef becomes an exact crystal axis in the sheep form. A further confirmation of the result is that all results between $C2$ beef and the other forms were nearly duplicated for the sheep data.

(5) *Positioning of the molecules*

The translation parameters were found by looking for a minimum R value between the observed intensities and the different sets of calculated amplitudes generated as the molecule was moved through the unit cell. This method was used by Cutfield *et al.* (1974) and described fully by Nixon & North (1976). This

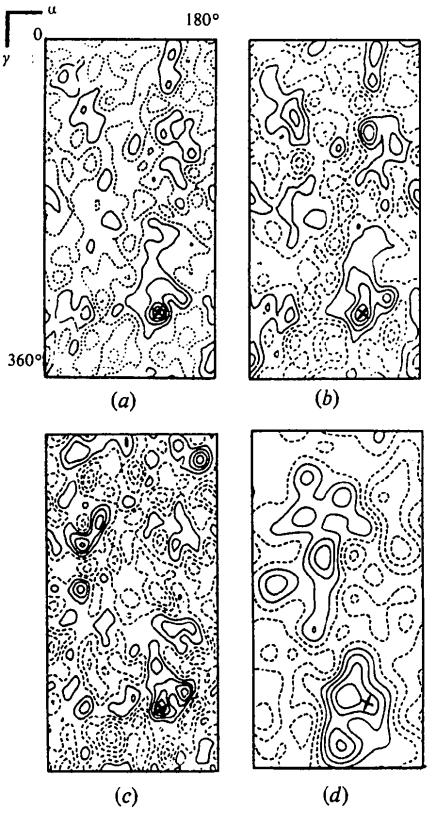


Fig. 5. The $\beta = 85°$ section for various beef DPI $C2$/model rotation functions. (*a*) The model contains 70% of the expected atoms; the sphere of integration extends to 13 Å; sharpening factor $B = 20$ Å; data range 8–3 Å. This map gave the clearest maxima. (*b*) as (*a*) but with the sphere of integration limited to 11 Å. Reducing the volume of integration has reduced the height of the true peak relative to spurious ones. (*c*) as (*a*) but with the model limited to main-chain atoms only. Again, spurious peaks are now the same height as the true one. (*d*) as (*a*) but with data limited to 12–5 Å. The maximum peak is now seriously misplaced.

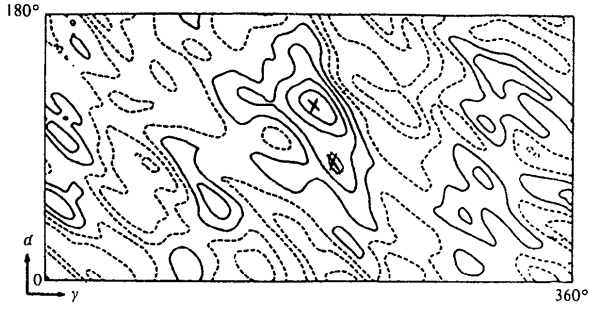


Fig. 6. The beef DPI $P2_12_12_1$/model rotation-function map; section $\beta = 35°$ using 8–3 Å data. The maximum gives a clear peak marking the orientation of one of the two molecules in the asymmetric unit. If there is a non-crystallographic twofold screw axis the second peak will lie on this section and is marked (×). This is not a sharp maximum, but the first peak is unusually extended in this direction.

program uses the fact that if the partial F_c's, $F_c1, F_c2, \ldots$, are calculated for all the symmetrically equivalent model molecules, orientated according to the appropriate rotation matrix, then the value of F_c for any set of translations, $\mathbf{t}_1, \mathbf{t}_2, \ldots$, where the $\mathbf{t}i$ are symmetrically equivalent, will be given by

$$F_c(hkl) = F_{c1}(hkl) \exp(-2i\pi \mathbf{h}.\mathbf{t}_1) + F_{c2}(hkl) \exp(-2i\pi \mathbf{h}.\mathbf{t}_2) + \ldots$$

So, once the F_ci are calculated it is only necessary to sum them together with appropriate phase modifications to generate the different sets of F_c's. It is possible to use the inverse fast Fourier transform for space group $P1$ using the cell dimensions of the observed intensity set to calculate all the required F_ci relatively inexpensively. The scale factor between F_o and F_c varies by as much as 20% depending on the position of the molecule in the unit cell, and the amount of overlap in the symmetry-related positions. The scale factor was determined by comparing $\langle F_o \rangle$ and $\langle F_c \rangle$, with some allowance for the incompleteness of the model. Earlier calculations had shown that the positions of correct minima were not very sensitive to the changes in the scale.

Initial searches on a 1 Å grid with 4 Å data were used to position molecules approximately, then a finer search on a 0·2 Å grid using 3 Å data was carried out. The method is limited by the fact that the rotation parameters are fixed throughout the search.

(i) *C2 sheep.* In this space group the origin along the y axis is arbitrary, so it was only necessary to search one quarter of one xz section. The minimum was very sharp (Fig. 7).

In any space group with only two symmetry operations it is important to exclude reflections where either F_{c1} or F_{c2} is weak from the R-factor calculation. For any such reflection $|F_c|$ is almost independent of the values of $\mathbf{t}_1$ and $\mathbf{t}_2$. In general, such space groups have an R value for a correctly orientated molecule in any position in the unit cell considerably below that of a randomly wrong structure.

Formally it should be possible to find the translation parameters from the $h0l$ projection intensities alone. But for these crystals there were so few of these

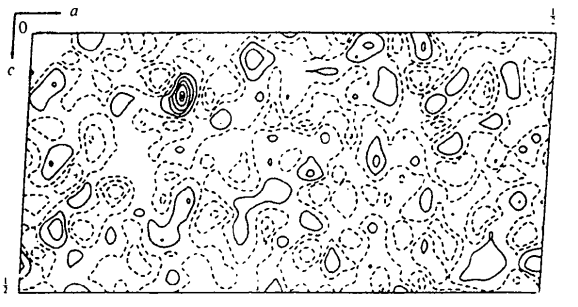


Fig. 7. The 7–3 Å R-factor map for sheep $C2$ DPI. The minimum R value is contoured as a peak.

reflections that we preferred to use the three-dimensional data.

(ii) *C*2 *beef*. The first search for *C*2 beef was made using only one molecule of the two known to be in the asymmetric unit. Despite this, there was a clear minimum which checked with the observed sheep position. A second search, adding the translated F_c's for the second molecule to those generated by the first reduced the *R* value to 48% (Table 4), at a relative translation consistent with the Patterson result. The second search had to be made over half the whole unit cell; the origin had been fixed by positioning the first molecule.

(iii) $P2_12_12_1$ *beef*. This *R*-factor search function was the least successful of the three. The space group requires an initial search volume of one eighth of the unit cell for the first molecule, and of the whole unit cell for the second. Projection searches did not give consistent results, and were very dependent on the choice of scale factor between the F_c and observed intensities.

This meant it was essential to work on a coarse grid for the three-dimensional search. It is clear that the *R* value is very sensitive to small translations, and we fear that we may have missed true minima. The search was made with each of the model molecules orientated independently. Each search had several minima of around 54% with a background *R* value around 60%. The computing involved in testing all these possibilities with the second molecule was prohibitive, so we decided to use a similar pair of molecules to those found in the *C*2 beef form. In *C*2 beef there were close contacts between the molecule at *x*,*y*,*z* and each molecule at: x,y,z; $0{\cdot}51 - x$, $y + 0{\cdot}59$, $1{\cdot}485 - z$; $0{\cdot}51 - x, y - 0{\cdot}41, 1{\cdot}485 - z$; $\frac{1}{2} - x, y + \frac{1}{2}, -z$; $\frac{1}{2} - x, y - \frac{1}{2}, 1 - z$. Each pair in turn was rotated into the $P2_12_12_1$ beef orientation, and translated over the asymmetric volume on a 1 Å grid. The pair generated from x,y,z and $0{\cdot}51 - x, y + 0{\cdot}59, 1{\cdot}485 - z$ gave the lowest *R* value of 51% at 4 Å resolution. Fine adjustment of the two molecules independently on a 0·2 Å grid reduced the *R* value to 49% (Table 4). But this result is not truly independent of the *C*2 beef result.

It is possible that the errors in the Eulerian angles used to define this rotation have reduced the power of the search function.

(6) *Verifying the results*

There are two procedures we have used to check these parameters. Once the preferred molecular position was found, all symmetry-equivalent positions were generated, and checked for impossible packing where one molecule was colliding with another. Of course, this is only a necessary prerequisite for a correct solution; it does not really verify the result. For all the results given in Table 3 the packing was acceptable. Fig. 8 illustrates the molecular packing for the *C*2 beef form.

The second check we used was to omit the disulphide bonds from the phasing and to calculate Fourier maps phased on the other atoms. In the *C*2 beef case the omitted atoms were clearly visible, suggesting that this solution is essentially correct. The $P2_12_12_1$ beef result was, however, not so clear-cut. Fig. 9 shows the *C*2 beef difference Fourier maps.

(7) *Refinement of the C*2 *sheep model*

Refinement was begun for the *C*2 sheep form using the fast Fourier least-squares refinement technique

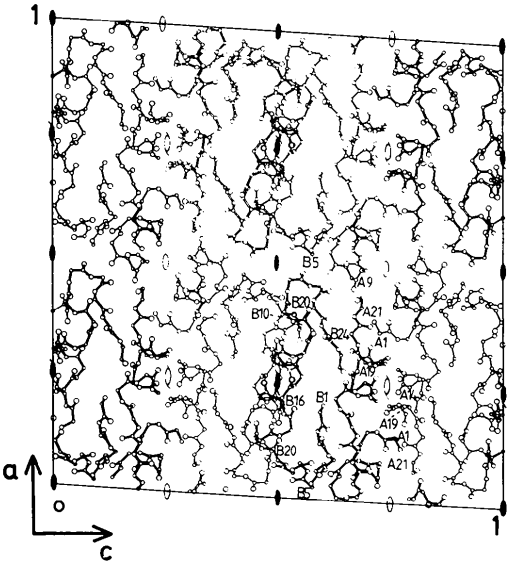
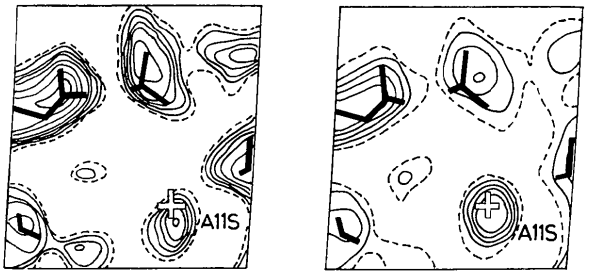


Fig. 8. The packing of the beef *C*2 DPI crystals, showing the pseudo twofold and twofold screw axes 0, ◊. Close contacts occur between B12, B24 and B17 about the twofold screw axis, and between A14 and A19 about the pseudo twofold screw axis.

Table 4. *Details of the search for a minimum R value in the DPI crystals*

For the beef forms the two translation shifts apply to the two molecules of the asymmetric volume. The average *R* value is only an estimate; it is very dependent on the scale factor chosen.

Structure	*d* (Å)	Reflections used	Reflections total	Minimum *R* value	Average *R* value	Translation shift			Close contacts	Eulerian angles (°)		
Beef (*C*2)	7–3	890	1370	0·48	0·59	0·155	0·00	0·31	93	120	85	−68
						0·145	0·09	0·825		120	85	−68
Beef ($P2_12_12_1$)	7–3	1251	1427	0·49	0·60	0·110	0·29	0·232	133	84	32	−158
						0·050	0·32	0·242		65	−146	−173
Sheep (*C*2)	7–3	634	675	0·48	0·59	0·150	0·00	0·119	68	120	85	−68



(*a*) (*b*)

Fig. 9. Part of the F_{obs} Fourier map showing the electron density at the sulphur atom (A11) excluded from the phasing. Map (*a*) is unweighted and map (*b*) was calculated using modified Sim weights (Bricogne, 1976).

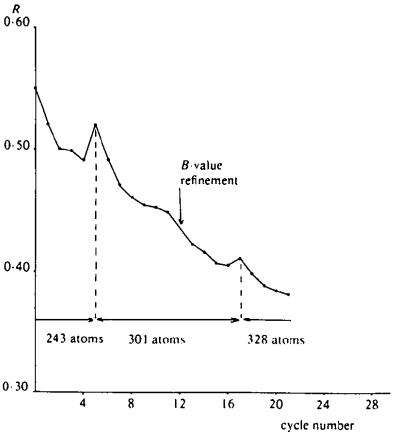


Fig. 10. Plot of R values during the preliminary sheep DPI $C2$ refinement.

(Agarwal, 1978). The following points were taken into account.

(1) All available reflections were used. In the early cycles higher weight was given to the low-order terms.

(2) The coordinates were fitted to standard geometry by model fitting (Dodson, Isaacs & Rollett, 1976) after each cycle. The calculated CA H atoms were included in each pass; this was sufficient to hold the peptides on the right hand despite the large average shifts for each cycle.

(3) Atoms were added to the model as they became visible in difference Fourier maps. Initially 243 of the 357 atoms were included.

In spite of the poor data the model has been substantially improved. The R factor has fallen from 53 to 36% (Fig. 10), and the difference Fourier maps are much less noisy, allowing residues to be placed with confidence.

Discussion

(1) *Methods*

This work has revealed some of the strengths and weaknesses of molecular-replacement studies. All these rest on the assumption that the molecular structure will remain much the same from one crystal form to another, but the 2-Zn and 4-Zn insulin comparison shows that this is not necessarily so (Bentley *et al.*, 1976). In all cases crystal-packing forces will cause some alteration to the structures but it is not clear how different two molecules can be before the technique will fail. Obviously, the rotation maps are better when the model contains most of the atoms of the required molecule; Fig. 5(*c*) shows how much worse the contrast was when side chains were left off the model molecule.

A second unavoidable problem arises from the use of a spherical volume of integration for the Patterson product. Without this the Crowther fast Fourier summation cannot be performed; it requires that the data are explained in spherical harmonics (Crowther, 1972), but for DPI the actual molecule is far from spherical, and there is very close packing between molecules, so any chosen sphere of integration either excludes many vectors within the molecule or includes a high proportion of intermolecular vectors, which will not be matched by the non-crystallographic rotation. Either error increases the noise level significantly (Fig. 5).

The third problem is simply a matter of luck; if the non-crystallographic axes of rotation are approximately parallel to any of the true crystallographic axes any peak will tend to merge with its symmetry equivalent. This study was particularly bedevilled by this difficulty. Peaks with β near 0 or 90° were the norm, and even the $P2_12_12_1$/model map had two predicted peaks almost on the same β section. In general, of course, the higher the symmetry of the space groups the more likely it is that this problem will arise.

(2) *Structure*

The crystal structure of a monomeric insulin is important since all the evidence indicates that the molecule is active as a monomer. It has been assumed that the hormone's structure does not change significantly when it dimerizes but it is difficult to prove this unequivocally. Indeed, spectroscopic studies have suggested that the insulin monomer partially unwinds in solution (Pocker & Biswas, 1980), which would invalidate the structure–function arguments (Blundell, Dodson, Hodgkin & Mercola, 1972; Peking Insulin Structure Research Group, 1974). However, the crystal structure of DPI shows that this modified monomer is closely similar to the insulin structure present in 2-Zn insulin and in the dimer forms of hagfish insulin and pig insulin (Cutfield, Cutfield, Dodson, Dodson, Reynolds & Vallely, 1981; Dodson, Dodson, Hodgkin & Reynolds, 1979). The packing of the DPI molecules in the crystal shows no resemblance to that in the insulin dimer structure. It appears, therefore, that the DPI crystal structure represents the solution structure of the

molecule reasonably well, and that on dimer formation there is little change in the molecule's folding and conformation.

Recently, large crystals have been obtained which diffract well to beyond 1·5 Å spacing. The refinement calculations are now being carried out at 1·7 Å spacing with newly collected data; these calculations will be further extended when the higher-resolution data become available.

We are grateful to the Wellcome Trust, the Kroc Foundation, the British Diabetic Association and the Medical Research Council for financial support and to Dr H. Gattner for material.

References

ADAMS, M. J., BLUNDELL, T. L., DODSON, E. J., DODSON, G. G., VIJAYAN, M., BAKER, E. N., HARDING, M. M., HODGKIN, D. C., RIMMER, B. & SHEAT, S. (1969). *Nature (London)*, **224**, 491–495.

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–800.

BENTLEY, G. A., DODSON, E. J., DODSON, G. G., HODGKIN, D. C. & MERCOLA, D. A. (1976). *Nature (London)*, **261**, 166–168.

BENTLEY, G. A., DODSON, E. J., DODSON, G. G. & LEVITOVA, A. (1978). *J. Mol. Biol.* **125**, 387–396.

BLUNDELL, T. L., DODSON, G. G., HODGKIN, D. C. & MERCOLA, D. A. (1972). *Adv. Protein Chem.* **26**, 279–402.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon and Breach.

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. G. DODSON, J. P. GLUSKER & D. SAYRE. Oxford Univ. Press.

CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G. & SABESAN, M. (1974). *J. Mol. Biol.* **87**, 23–30.

CUTFIELD, S. M. (1981). Private communication.

DODSON, E. J., DODSON, G. G., HODGKIN, D. C. & REYNOLDS C. D. (1979). *Can. J. Biochem.* **57**, 469–479.

DODSON, E. J., ISAACS, N. W. & ROLLETT, J. S. (1976). *Acta Cryst.* A**32**, 311–315.

GATTNER, H. G. (1975). *Hoppe–Seyler's Z. Physiol. Chem.* **386**, 1397–1404.

KIKUCHI, K., LARNER, J., FREER, R., DOY, A., MORRIS, H. & DELL, A. (1980). *J. Biol. Chem.* **255**, 9281–9288.

LI JIA-YAO, SONG SHIH-YAN, LI JUN, WANG JIA-HAI & YAU TSI-HO (1980). *Adv. Biochem. Biophys.* **4**, 43–45 (in Chinese).

MEYTS, P. D., OBBERGHEN, E. V., ROCH, J., WOLLMER, A. & BRANDENBURG, D. (1978). *Nature (London)*, **273**, 504–509.

NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A**32**, 320–325.

NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 35–59.

PEKING INSULIN STRUCTURE RESEARCH GROUP (1974). *Sci. Sin.* **17**, 752–777.

PEKING INSULIN STRUCTURE RESEARCH GROUP (1976). *Sci. Sin.* **19**, 358–361.

POCKER, Y. & BISWAS, S. B. (1980). *Biochemistry*, **19**, 5043–5049.

REEVES, W. G. (1981). Private communication.

ROSSMANN, M. G. (1962). *Acta Cryst.* **15**, 24–31.

THE SHANGHAI INSULIN RESEARCH GROUP (1976*a*). *Sci. Sin.* **19**, 351–357.

THE SHANGHAI INSULIN RESEARCH GROUP (1976*b*). *Sci. Sin.* **19**, 475–485.

WANG JIA-HAI, LI JIA-YAO, YAU TSI-HO, KWEI LOO-LOO, LOU MAI-JAING, YU JUN-MENG & SONG SHIH-YAN (1980). *Sci. Bull.* **8**, 369–371 (in Chinese).

Acta Cryst. (1983). B**39**, 98–104

The Structure of the Ribodinucleoside Monophosphate Guanylyl-3′,5′-cytidine as its Ammonium Octahydrate Salt

BY A. AGGARWAL, S. A. ISLAM, R. KURODA, M. R. SANDERSON* AND S. NEIDLE†

Cancer Research Campaign Biomolecular Structure Research Group, Department of Biophysics, King's College, 26–29 Drury Lane, London WC2B 5RL, England

AND H. M. BERMAN

Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Fox Chase, Philadelphia, Pennsylvania 19111, USA

(*Received* 30 *June* 1982; *accepted* 19 *July* 1982)

Abstract

The crystal and molecular structure of an ammonium octahydrate salt of the ribodinucleoside monophosphate guanylyl-3′,5′-cytidine ($C_{19}H_{24}N_8O_{12}P$) has been determined by X-ray methods, and refined to a final R of 0·111 for 1197 observed reflections. The salt crystallizes in space group $C2$ with cell dimensions a = 20·987 (6), b = 16·470 (4), c = 9·566 (1) Å and β = 94·36 (2)°; Z = 4 for a calculated density of 1·50 Mg m^{-3} and μ(Cu $K\alpha$) = 1·19 mm^{-1}. The unit-cell

* Present address: Gorlaeus Laboratories, The State University, Leiden, The Netherlands.

† To whom correspondence should be addressed.

0567-7408/83/010098-07$01.50