Structure of Rhombohedral 2 Zinc Insulin Crystals

by
M. J. ADAMS* E. N. BAKER
T. L. BLUNDELL M. M. HARDING†
E. J. DODSON D. C. HODGKIN
G. G. DODSON B. RIMMER‡
M. VIJAYAN S. SHEAT§
Chemical Crystallography Laboratory
and Molecular Biophysics Laboratory,
University of Oxford

An electron density map has been calculated at a resolution of 2·8 Å which shows many details of the arrangement of the atoms in rhombohedral crystals of insulin.

We here report the three-dimensional arrangement of the atoms in the peptide hormone, insulin, as observed in an electron density map derived by X-ray analysis at a resolution of 2·8 Å. The crystals studied were of pig insulin, which has the chemical structure shown in Fig. 1, found by Ryle, Sanger, Smith and Kitəi[1]. They were of the rhombohedral form first obtained by Abel in 1925[2] and first characterized by X-ray diffraction in 1935[3]. They have been shown—by earlier crystallographic studies—to contain two zinc atoms and six insulin molecules of weight 5,780 in the rhombohedral unit cell, the two insulin molecules in the crystal asymmetric unit being related approximately by a non-crystallographic two-fold axis[4,5]. These units correspond in weight with the insulin hexamer observed in solutions containing zinc and the insulin dimer found in other solutions, particularly at acid *p*H.

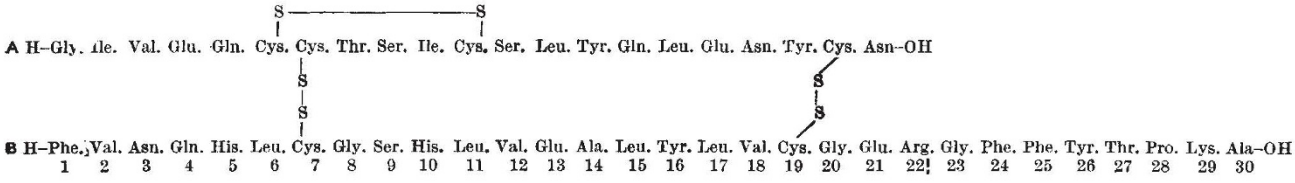

Fig. 1. The primary structure of pig insulin.

The X-ray Analysis

The methods used in our analysis were largely similar to those developed in the study of other proteins.

The crystals of the native insulin were grown as flat rhombohedra, 0·2–0·4 mm across, by methods described by Schlichtkrull[6]. Heavy atoms were introduced into them by soaking in appropriate solutions, after a number of experiments varying and controlling the conditions of concentration, buffer and temperature. Five of the derivatives made, listed in Table 1, were used for the phase determination required to compute the present electron density map. Of them, the first was obtained, following a suggestion made by Strandberg and Tilander, by removing zinc from the native crystals by soaking in EDTA solution and then inserting lead by soaking in 0·01 M lead acetate. The lead ions entered into three sites additional to those occupied by zinc. These led us to the making of derivative 2, obtained from zinc insulin soaked in more concentrated lead acetate solution. Potassium uranyl fluoride entered in two sites in derivative 3, but these were somewhat disordered; insulin and uranyl acetate (derivative 4, one of the earliest derivatives found) is also the most complicated. Mercury substituted *m*-hydroxybenzaldehyde (derivative 5) was found to occupy usefully different sites from any other derivative. It was introduced as a result of experiments made by Liang Tung Tsai on cross-linking insulin with dialdehydes (private communication).

The patterns of heavy atom substitution in all these derivatives were low and generally complicated; often the sites occupied did not obey the two-fold axis. They were not fully unravelled until we were able to make improved intensity measurements with the Hilger and Watts 4-circle diffractometer. X-ray data on all the crystals were

Table 1. PARAMETERS OF FIVE DERIVATIVES

Insulin heavy atom derivative	Occupancy of metal atom sites in electrons					Limit of data collection (Å)	*R* in ranges in Å 4·5	4·5–3·5	3·5–2·8
1 Zinc-free insulin acetate, 0·01 M lead	56	23	24	45*	36*	2·8	34	47	48
2 Zinc insulin acetate, 0·1 M lead	52	45	12	14		2·8	39	45	48
3 Zinc insulin citrate, 0·02 M uranyl fluoride	28	17				4·5	16	—	—
4 Zinc insulin acetate, 0·01 M uranyl acetate	64	24	9	20		2·8	31	45	50
5 Zinc insulin citrate, mercuribenzaldehyde, saturated solution	14	7	6	10		2·8		42†	
Figure of merit							0·92	0·8	0·72

The residual $R = \Sigma|F_{HIR} - F_{H}\text{calc.}|/\Sigma F_{HIR}$ where, F_{HIR} is the experimentally obtained value for the heavy atom contribution. The figure of merit is the mean value of the cosine of error in the phase angle.

* These two sites are at the zinc positions; their occupancy is different.

† Refinement of derivative 5 was on selected terms extending out to 2·4 Å spacing.

Present addresses:
* Purdue University, Lafayette, USA; † University of Edinburgh; ‡ West Kilbride; Massey University, New Zealand.

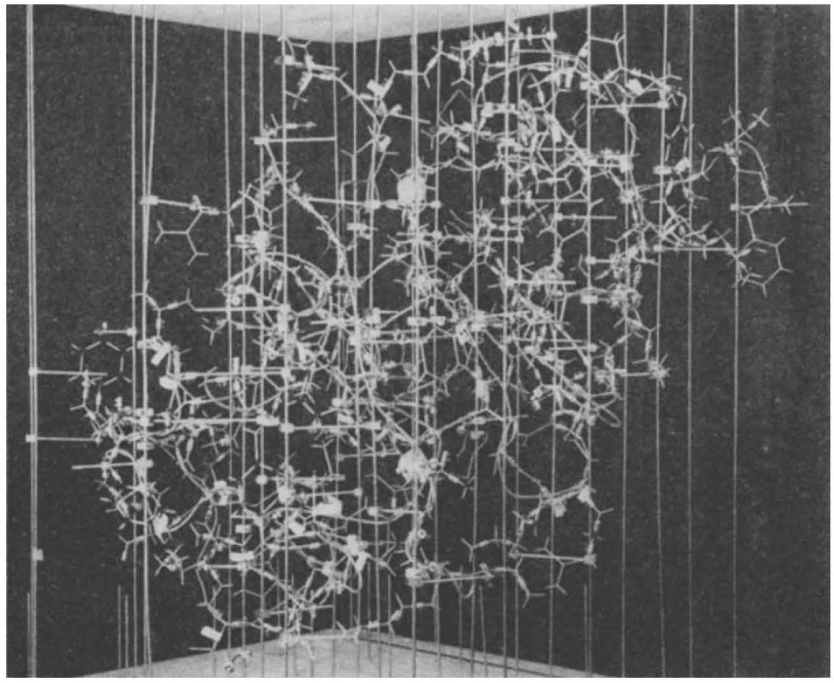

Fig. 2. Photograph of a model of the insulin dimer, viewed along the non-crystallographic two-fold axis. The β-pleated sheet and two phenylalanine B25 residues are in the centre of the model. (Taken by Mrs W. Brown and Mr N. Fogden.)

collected in shells on the diffractometer, ensuring that all equivalent reflexions and their Friedel pairs were measured on the same crystal. In the case of 2 zinc insulin, the data collected extended to 2·4 Å, at least two equivalent reflexions being measured for each index type. Heavy atom parameters were determined and refined by three-dimensional Patterson, Fourier and least-squares techniques. The observed heavy atom contributions, F_{HLE}, here a complex quantity for all terms, were derived from combined isomorphous and anomalous intensity differences as discussed by Singh and Ramaseshan[7]. Some data bearing on the structure factor and phasing calculations are included in Table 1.

The electron density was computed at intervals of approximately 1 Å parallel with a and 0·75 Å parallel with c, and was printed out on sheets parallel with the c plane on a scale very close to 1 cm to 1 Å. Contours were drawn at intervals of 0·2 Å (lowest contour at 0·1 Å) and traced onto 'Mylar' sheets. These, stacked between 'Perspex' sheets at the correct interval, gave a three-dimensional map which could be compared directly with peptide chains and side chains built from Beevers models at the same scale.

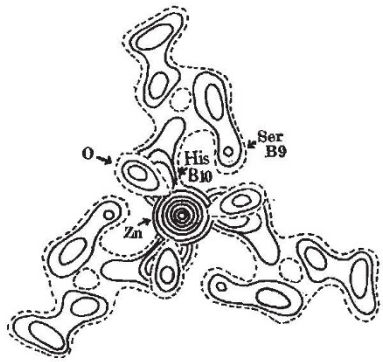

Fig. 3. Composite electron density map taken from sections 9/48, 12/48 and 16/48 c. The sections show the zinc ions in contact with B10 histidine and a solvent molecule, O.

The electron density map proved to be very readily interpretable, given the sequence of residues known in pig insulin. The first features recognized were two stretches of α-helix; at one end of each of these was density connected with one of the two peaks clearly representing the zinc atoms on the three-fold axis. That this density represented the B10 histidine residues was soon found by model building, guided by the insulin sequence. The peptide chains proved easy to follow throughout their whole length and the disulphide bonds were readily identifiable both by their spatial geometry and large electron density. All the aromatic residues were very well defined by individual ellipsoidal peaks. There are a few regions which are difficult to interpret, particularly where there are long side chains on the outside of the molecule.

The model built from the map is illustrated in Fig. 2 while Figs. 3 and 4 illustrate parts of the observed electron density. The three-dimensional arrangement of the α carbon atoms in the crystal asymmetric unit is shown in Figs. 5 and 8.

The Insulin Molecule

As a consequence of the complexity of the asymmetric unit we see in the electron density map two crystallographically independent insulin molecules. They are very similar in internal organization but not quite identical, approximately related to one another by two-fold axes, at P and Q. Figs. 6 and 7 illustrate the separate conformations of the A and B chains within each molecule and, diagrammatically, our evidence for the individual residue positions. In each molecule the A chain is a compact unit around which the B chain is wrapped, an arrangement which readily explains the advantages observed in the chemical synthesis of insulin in closing first the A6–11 disulphide bond[8–10].

The A chain has an involved, much folded structure with short stretches of near helical conformation between

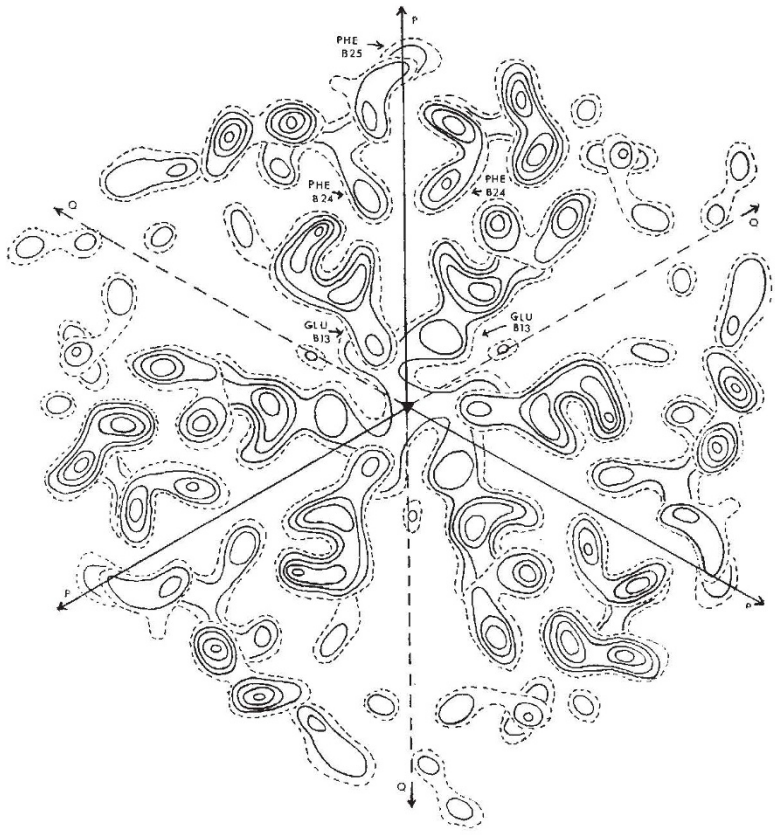

Fig. 4. Composite electron density map showing sections 1/48 and $\bar{1}/48$ c to illustrate density related by the non-crystallographic two-fold axis at $z=0$. Note that the phenylalanine residues, B24, are symmetrical about the axis; one phenylalanine residue, B25, lies on it, the other being away from this section.

residues A2–6 and A13–19. The A chain loop, A6–11, lies with the residues A8, 9 and 10 directed out from the surface of the molecule; these three vary most between mammalian species and their arrangement corresponds nicely with clinical and immunological effects. The disulphide bond, A6–11, is within the body of the molecule. There are a number of structurally important internal hydrogen bonds which serve to support the A chain internally; both A19 and A14 tyrosines, for example, seem to be hydrogen bonded to residues A1 and A5, respectively.

In the central part of the B chain, there are three turns of α-helix, slightly opened out at each end. Both the interchain disulphide bonds, B7–A7 and B19–A20, are at the ends of the helix which provides a rigid backbone. B1–B6 and B21–B30 are largely extended and loosely packed around the A chain. The two residues A4 glutamic acid and B29 lysine appear to be in contact as expected by Zahn (private communication).

The Insulin Dimer

The closest series of contacts between two insulin molecules occurs between the extended B chains, residues B23 B28; these are arranged antiparallel and appear to form over at least part of their lengths, a hydrogen-bonded pleated sheet. They surround the non-crystallographic two-fold axis, P, of Figs. 4 and 5, and this axis is taken to be the dimer two-fold axis. It is clear from the electron density map, part of which is shown in Fig. 4, that the two-fold axis is not exactly obeyed in this region, probably as a consequence of the very congested packing of residues which is here observed. Thus there are close contacts between two-fold axis related non-polar residues such as the valines at B12. There is also the very interesting development of an aromatic cage which is formed by tyrosine B26 and phenylalanine B24 and their two-fold axis related partners. The two phenyl groups at B25, on the other hand, form part of this system only by turning towards one another in a way which violates the two-fold symmetry relation (compare Fig. 4). Various adjustments follow in the positions of surrounding residues, which destroy exact two-fold symmetry.

The dimer as a whole, an elongated cylinder roughly 20 Å across and 40 Å long, is shown in Fig. 5. It is a consequence of the A and B chain arrangement in the monomer that only interactions between the B chain are involved in the formation of the dimer.

The Insulin Hexamer

The insulin hexamer (Fig. 9) is a compact, oblate spheroid, formed by the coordination of three insulin dimers around the two zinc ions, one 8·9 Å above the two-fold axis, the other 8·9 Å below it. Each zinc ion makes contact with the three B10 histidine residues and also with three atoms, probably the oxygens of water molecules, represented by the peaks in Fig. 3. The metal coordination is six-fold but rather far from being regular octahedral. A density suggesting system of hydrogen bonded water molecules can be traced connecting the atoms in the zinc coordination sphere with tyrosine B16 and histidine B5.

The organization of the hexamer involves close contacts between the molecules around the two-fold axis which appears at Q in Fig. 4. Indeed, they are almost "hooked" together by the B1 phenylalanine residues. There is also a close, probably hydrogen bonded, contact between the six glutamic acid residues, B13, which enclose a space surrounding the three-fold axis. Many of the heavy atoms used to form isomorphous derivatives enter this space; others occupy regions on the hexamer surface, particularly near other glutamic acid residues. None penetrates the dimer.

Future Work

We hope that this structure, on further study, will supply clues relating to the nature of the biological activity of insulin. So far we can say little except that certain residues that seem necessary for activity, for example asparagine, A21[11], are on the surface of the hexamer. We plan to extend both the accuracy of our analysis and the investigation of insulin in contact with other molecules.

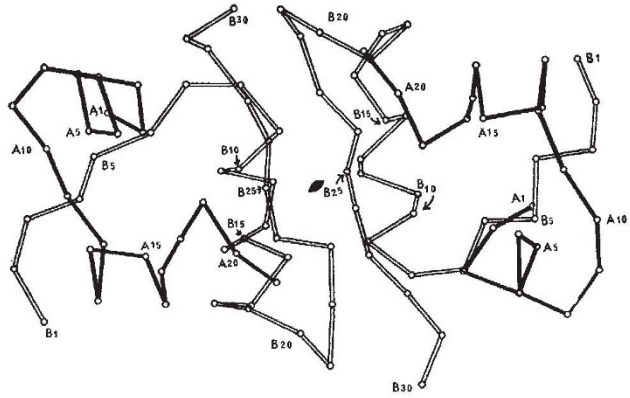

Fig. 5. View of the α carbon positions in the peptide backbone of the insulin dimer. This is seen projected along the non-crystallographic two-fold axis. A chain, black: B chain, open. This and Figs. 8 and 9 are taken from a programme written by Dr A. C. T. North.

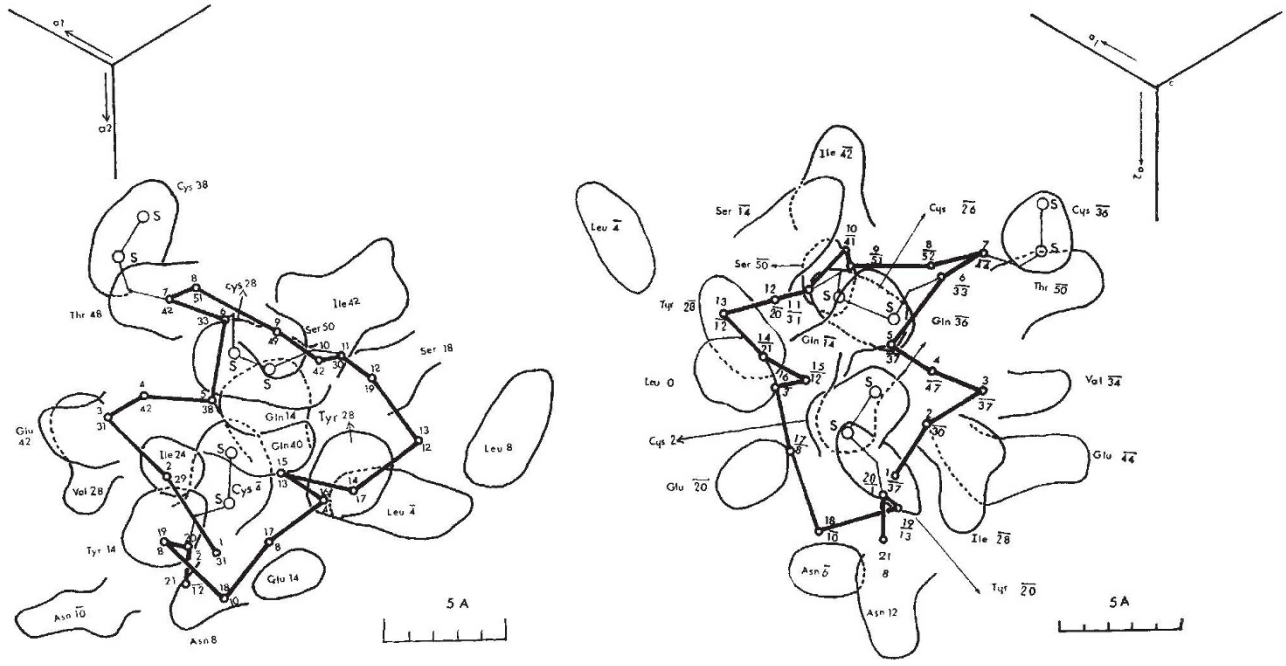

Fig. 6. Diagram to show the residue arrangement in the A chains of the two observed insulin molecules. Here and in Fig. 7 the peptide backbone is projected along the three-fold axis over contours which outline the regions in the electron density map occupied by particular side chains. Numbers show the approximate levels in hundredths. The levels of α carbons are given below residue numbers. The two chains are not exactly related by a two-fold axis. Some of the differences between them are real; some may arise from our present inaccuracy in atom placing.

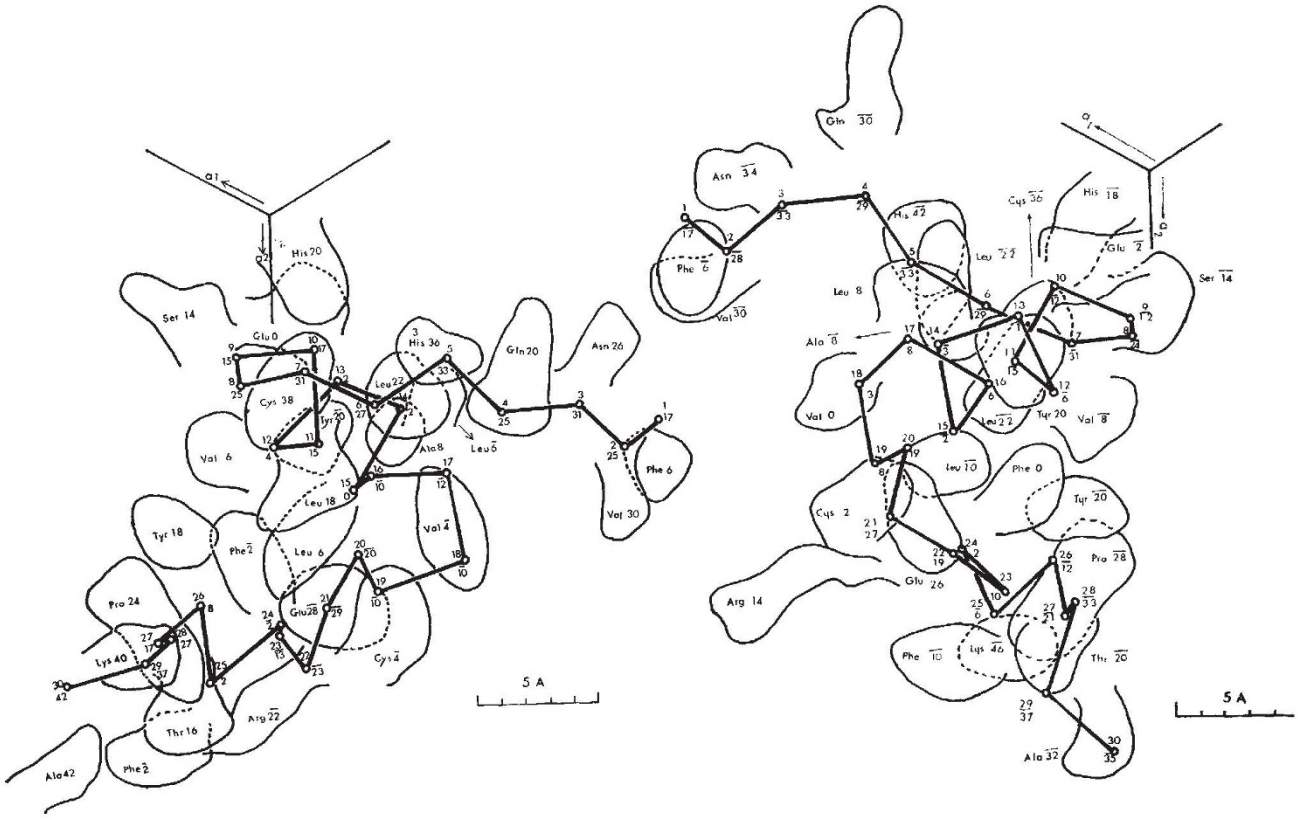

Fig. 7. The two B chains shown in the same convention as in Fig. 6.

Experiments are already in progress to measure X-ray data to higher resolution and to obtain by chemical substitution simple and fully substituted heavy atom derivatives.

All of the crystals used for this study were grown from insulin supplied by Dr J. Schlichtkrull, Novo Terapeutisk Laboratorium, Copenhagen, to whom our first thanks are due. Other insulin samples were kindly sent by Boots Pure Drug Company and Eli Lilly and Company. Almost all the intensity measurements, of some 70,000 reflexions, on which the latest calculations were made, were carried out on the Hilger and Watts 4-circle diffractometer during the last eighteen months. The major computations required were made at the Oxford University Computing Laboratory. Our research on insulin has continued for so long that the line we draw between authors and those who contributed to our solution is necessarily arbitrary. We are most grateful for scientific advice, guidance in computing and many useful experiments to K. Dornberger Schiff, S. Ramaseshan, John Hodder, J. S. Rollett, A. F. Kennedy, Liang Tung Tsai and P. D. J. Weitzmann; for growing crystals and derivatives to F. Bertinotti, J. Knox, M. Le Quesne and A. Tench; for photography, drawing and model making to L. Blundell, S. Cole, A. Cropper, G. Humphries and P. Pancharatnam; for technical assistance to F. Bowers, C. Grimshaw, M. Pickford, J. Marsh, A. Renshaw and F. Welch. The research was generously supported by grants from the Rockefeller Foundation, the Science Research Council, the Medical Research Council and the Royal Society.

Received October 10, 1969.

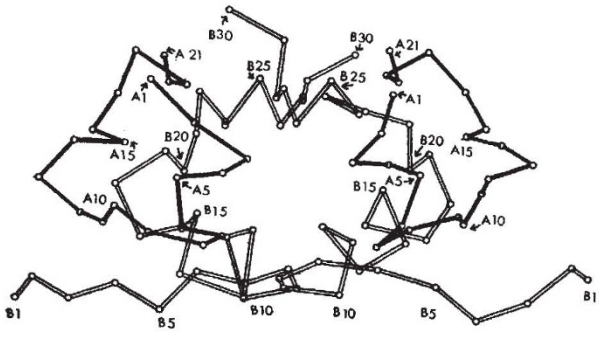

Fig. 8. View of the insulin dimer as seen projected along the three-fold axis showing α carbon atoms and peptide backbone. A chain, black; B chain, open.

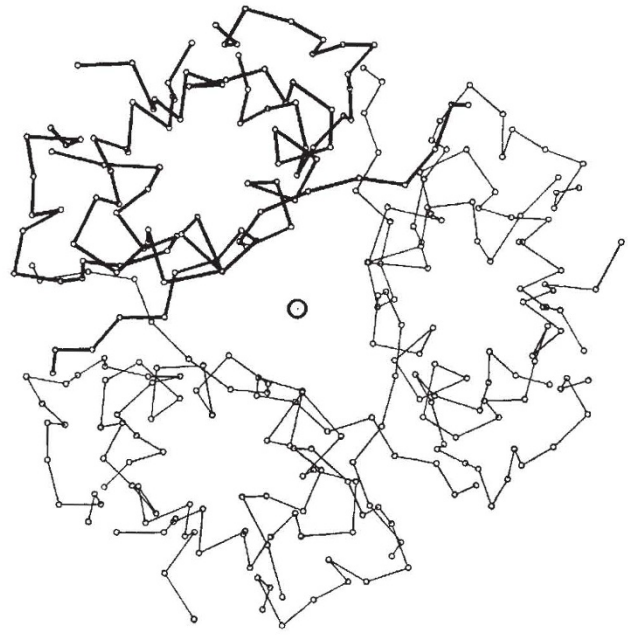

Fig. 9. View of the hexamer projected down the three-fold axis showing α carbon atoms and the peptide backbone. One dimer is in bold lines.

[1] Ryle, A. P., Sanger, F., Smith, L. F., and Kitai, R., *Biochem. J.*, **60**, 541 (1955).
[2] Abel, J. J., *Proc. US Nat. Acad. Sci.*, **12**, 132 (1926).
[3] Crowfoot, D. M., *Nature*, **135**, 591 (1935).
[4] Harding, M. M., Hodgkin, D. C., Kennedy, A. F., O'Connor, A., and Weitzmann, P. D. J., *J. Mol. Biol.*, **16**, 212 (1966).
[5] Dodson, E. J., Harding, M. M., Hodgkin, D. C., and Rossmann, M. G., *J. Mol. Biol.*, **16**, 227 (1966); Low, E. W., and Einstein, J. R., *Nature*, **186**, 470 (1960).
[6] Schlichtkrull, J., *Acta Chem. Scand.*, **10**, 1455 (1956); thesis, Copenhagen University (1958).
[7] Singh, A. K., and Ramaseshan, S., *Acta Cryst.*, **21**, 279 (1966).
[8] Kung *et al.*, *Scientia Sinica*, **15**, 544 (1966).
[9] Katsoyannis, P. G., and Tometsko, A., *Proc. US Nat. Acad. Sci.*, **65**, 1554 (1966).
[10] Zahn, H., *et al.*, *Z. Naturforsch.*, **20B**, 666 (1965); *Liebig Ann. d'Chem.*, **691**, 225 (1966).
[11] Carpenter, F. H., *Amer. J. Med.*, **40**, 750 (1966).

Excision of Thymine Dimers and Other Mismatched Sequences by DNA Polymerase of *Escherichia coli*

by
REGIS B. KELLY*
MAURICE R. ATKINSON †
JOEL A. HUBERMAN
ARTHUR KORNBERG

Department of Biochemistry,
Stanford University School of Medicine,
Stanford, California

A nuclease function of DNA polymerase excises mismatched regions of DNA. Exercise of its multiple functions *in vivo* might enable this enzyme to participate in the excision and repair of DNA lesions, as well as in replication of DNA.

DNA polymerase from *E. coli* catalyses both polymerization and hydrolysis in the 5′→3′ direction and it has been proposed that adjacent sites function concurrently, translating a nick in a 5′→3′ direction when nicked double-stranded DNA is used as primer[1,2]. The 5′→3′ exonuclease activity also functions in the absence of synthesis[3,4]; it requires a double-stranded template[5], and is characterized by production not only of mononucleotides, but also of di, tri and larger oligonucleotides[2] (compare Table 1). This distribution of oligonucleotides may result from failure of DNA polymerase to cleave terminal bonds, resulting in exposure of the second or subsequent phosphodiester linkages to hydrolysis at the 5′→3′ exonuclease site. An earlier model of the active centre of DNA

* Present address: Department of Neurobiology, Harvard Medical School, Boston, Massachusetts.

† On leave from the School of Biological Sciences, Flinders University, Bedford Park, South Australia.