Vanderbilt high-field NMR facilities, and to the Southern Regional Education Board for a Small Grants Program Award. DvdH and XJ acknowledge partial support from a grant (CA 17562) from the National Cancer Institute, Department of Health and Human Services.

References

ABRAHAM, R. J. & FISHER, J. (1985). *Magn. Reson. Chem.* **23**, 856–861.

ALTONA, C. & SUNDARALINGAM, M. (1970). *J. Am. Chem. Soc.* **92**, 1995–1999.

AUE, W. P., BARTHOLDI, E. & ERNST, R. R. (1976). *J. Chem. Phys.* **64**, 2229–2246.

BARRETT, A. G. M., BARTON, D. H. R. & NAGUBANDI, S. (1980). *J. Chem. Soc. Perkin Trans.* 1, pp. 237–239.

BECKHAUS, H., KIMURA, M., WATSON, W. H., VENIER, C. G. & KOJIC-PRODIC, B. (1979). *Acta Cryst.* B**35**, 3119–3122.

DAY, R. O., KINGSBURG, C. A. & DAY, V. W. (1981). *J. Org. Chem.* **46**, 1004–1009.

GLASS, R. S., REINEKE, K. & SHANKLIN, M. (1984). *J. Org. Chem.* **49**, 1527–1533.

HOYOS GUERRERO, M. A., MARTÍNEZ-CARRERA, S. & GARCÍA-BLANCO, S. (1983). *Acta Cryst.* C**39**, 118–119.

JACKMAN, L. M. & STERNHELL, S. (1969). *Applications of Nuclear Magnetic Resonance Spectroscopy in Organic Chemistry*, 2nd ed. Oxford: Pergamon Press.

JOHNSON, C. K. (1965). *ORTEP*. Report ORNL-3794. Oak Ridge National Laboratory, Tennessee, USA.

KIMURA, M., WARD, M. A., WATSON, W. H. & VENIER, C. G. (1979). *Acta Cryst.* B**35**, 3122–3124.

MAIGNAN, C. & RAPHAEL, R. A. (1983). *Tetrahedron*, **39**, 3245–3249.

MARCHAND, A. P. (1982). *Stereochemical Applications of NMR Studies in Rigid Bicyclic Systems*. Florida: Verlag Chemie International.

PAQUETTE, L. A., MOERCK, R. E., HARIRCHIAN, B. & MAGNUS, P. D. (1978). *J. Am. Chem. Soc.* **100**, 1597–1599.

SHELDRICK, G. M. (1976). *SHELX*76. Program for crystal structure determination. Univ. of Cambridge, England.

SORIANO-GARCIA, M., TOSCANO, R. A., GARCIA, G., LARRAZA, M. I. & SANCHEZ, I. H. (1984). *Acta Cryst.* C**40**, 887–889.

Acta Cryst. (1989). B**45**, 99–107

Structure of Insulin: Results of Joint Neutron and X-ray Refinement

BY ALEXANDER WLODAWER* AND HUGH SAVAGE†

Center for Chemical Physics, National Bureau of Standards, Gaithersburg, MD 20899, *USA, and Laboratory of Molecular Biology, National Institute of Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD* 20892, *USA*

AND GUY DODSON

Department of Chemistry, University of York, York, England

(*Received* 18 *April* 1988; *accepted* 19 *September* 1988)

Abstract

Neutron diffraction data for porcine 2Zn insulin were collected to 2·2 Å resolution from a single crystal deuterated by slow exchange of mother liquor. A joint neutron/X-ray restrained-least-squares refinement was undertaken using the neutron data, as well as the 1·5 Å resolution X-ray data collected previously. The final R factors were 0·182 for the X-ray data and 0·191 for the neutron data. Resulting atomic coordinates were compared with the initial X-ray model, showing a total r.m.s. shift of 0·36 Å for the protein and 0·6 Å for the solvent. Protonation of a number of individual amino acids was investigated by analysis of the neutron maps. No D atoms were found between the carboxylates of Glu *B*13 which make an intermolecular contact, suggesting nonbonded interaction rather than the predicted hydrogen bond. Amide hydrogen exchange was investigated in a refinement of their atomic occupancies. Regions of unexchanged amide groups were found in the center of the *B* helices. The results of this study emphasize the limited amount of information available in neutron diffraction studies of proteins at resolution lower than 2 Å.

Introduction

Insulin was one of the first polypeptide hormones to be crystallized (Abel, 1926) and characterized by X-ray diffraction (Crowfoot, 1935). The first report of its structure based on multiple isomorphous replacement methods was published by Adams *et al.* (1969). The principal difficulty in determining the crystal structure was in finding and solving heavy-atom derivatives. At one stage it was suggested by S. Ramaseshan (unpublished) that neutron diffraction and the use of a strong anomalous scatterer such as ^{157}Gd, ^{149}Sm or ^{113}Cd might alleviate the problems (also see Schoenborn,

* Current address: National Cancer Institute, Crystallography Laboratory, BRI-FCRF, PO Box B, Frederick, MD 21701, USA.

† Current address: Department of Chemistry, University of York, York, England.

0108-7681/89/010099-09$03.00

1975). While the insulin structure was finally solved without resorting to the technique of neutron diffraction, such investigations were continued and 4·5 Å data were collected using crystals substituted by $^{157}Gd^{3+}$ at the Harwell reactor (S. Mason, personal communication). Low neutron flux from that source presented substantial experimental difficulties and this line of research was discontinued.

The X-ray structure of rhombohedral crystals of 2Zn insulin was initially solved at 2·8 Å resolution (Adams *et al.*, 1969; Blundell *et al.*, 1971). These crystals belong to the space group $R3$ ($a = b = 82{\cdot}5$ Å, $c = 34{\cdot}0$ Å for the hexagonal setting) and contain a hexamer in the rhombohedral unit cell (dimer in the asymmetric unit). Atomic coordinates of both molecules, consisting of 51 amino acids each, were traced, and it was noted that while two molecules looked similar, some marked differences existed. The multiple isomorphous phasing was extended to 1·9 Å, and new maps were calculated by Blundell, Dodson, Hodgkin & Mercola (1972). The resolution of the X-ray diffraction data was later extended to 1·5 Å (Dodson, Dodson, Hodgkin & Reynolds, 1979) and the model deposited with the Protein Data Bank (Bernstein *et al.*, 1977) in 1980 was the result of refinement at that resolution.

Several other groups have also refined the X-ray structure of porcine insulin. Isaacs & Agarwal (1978) utilized the same reflection data to develop a fast Fourier least-squares algorithm. The refinement was completed after 67 cycles, with the geometry of the model periodically regularized using the method of Dodson, Isaacs & Rollett (1976). The final R factor was 0·113 for the data with $F > 2\sigma(F)$, or 0·148 for all data between 14 and 1·5 Å. Sakabe. Sakabe & Sasaki (1980, 1981) refined a 1·1 Å data set collected independently at 277 K using the same algorithm. Their final R factor was 0·11 and the resulting model was used to describe the water structure associated with the protein, as well as in the investigation of the hydrogen bonding. The authors reported that they observed 90% of the H atoms directly in difference Fourier maps. Another model of insulin was obtained by the Beijing group (Chang *et al.*, 1984; Liang, 1982). They refined 1·2 Å data (also collected at 277 K) using an anisotropic temperature factor version of the program *PROLSQ* (Hendrickson, 1985) to an R of 0·12 and also observed directly more than half of the H atoms. The coordinates of the models described in this paragraph have not been deposited in publicly accessible databases. However, comparisons have recently been made between the Beijing 1·2 Å resolution and the original 1·5 Å resolution electron density maps. No significant differences were detected in the two molecules and the electron density for the solvent regions in general looked very similar (D. Hodgkin, personal communication).

Neutron diffraction has been shown to provide additional information about protein structure not easily obtained in X-ray diffraction studies (Wlodawer, 1982). Although many H atoms were reported to be observed in different X-ray Fourier maps of insulin, neutron diffraction should make their location more precise, since the magnitudes of scattering lengths for H and D are of the order of those for C, O and N. Since the scattering lengths of H and D are of opposite signs, neutron diffraction was shown to be a powerful tool for the investigation of hydrogen exchange in crystals (Wlodawer & Sjölin, 1982; Kossiakoff, 1982). This technique can provide a detailed description of the bound solvent (Teeter & Kossiakoff, 1984; Wlodawer, Walter, Huber & Sjölin, 1984). Simultaneous refinement with X-ray and neutron data was shown to work well and to provide more information than refinement of either type of data alone (Wlodawer & Hendrickson, 1982). For these reasons, we initiated a joint neutron/X-ray study of insulin crystals, results of which are presented here.

Experimental procedures

An insulin crystal used to collect neutron diffraction data was grown by the method of Schlichtkrull (1956) in the laboratory of D. C. Liang, Beijing, People's Republic of China, and was carried to Washington in its mother liquor, at room temperature. The maximum dimensions of the irregularly shaped crystal were approximately $1 \times 1{\cdot}5 \times 2{\cdot}3$ mm (volume approximately 3 mm^3) and its faces were not well developed. The crystal was deuterated by slow addition of a synthetic mother liquor containing 30% d_6-acetone (Merck), 70% D_2O (Sigma, gold label), 0·1 M sodium citrate and 0·005% zinc chloride. The final pH of this solution was adjusted to 6·5 by addition of sodium hydroxide (uncorrected pH-meter reading). Exchange took four and a half months with slow addition in the first several weeks, followed by three complete exchanges of the mother liquor. Upon completion of the exchange the crystal was mounted in a quartz tube and was immobilized with quartz wool (Wlodawer, 1980).

Neutron diffraction data were measured using the flat-cone diffractometer (Prince, Wlodawer & Santoro, 1978) located at the National Bureau of Standards Reactor. The neutrons were monochromatized by reflection from a 50×100 mm graphite crystal that had a mosaic spread of 40′. The wavelength of the neutrons was 1·68 Å and the flux was 6×10^6 neutrons cm^{-2} s^{-1}. The neutron diffraction data set was collected in two parts. First, the data extending to a nominal resolution of 2 Å for each level were measured in flat-cone geometry. In this mode the crystal was rotated around its rhombohedral a axis while the data for one level were measured. Next, the detector was moved to a position appropriate for a new level, and the

process was repeated. Altogether data from 13 levels ($h = 0$–12) were collected. We realized that rotation around the *c* axis of the hexagonal crystal system would have been preferable owing to its shorter length (34 *versus* 49 Å for the rhombohedral *a*), but this was precluded by the orientation of the crystal in the quartz tube. The number of reflections measured in this mode and integrated using the dynamic mask procedure (Sjölin & Wlodawer, 1981) was 6253, which yielded 3142 independent reflections of which 2253 were observed [$I > 1{\cdot}5\sigma(I)$]. The symmetry *R* factor was 10·4%. The reflection indices were transformed to the hexagonal setting, which was used from then on. A list of missing reflections in the 2·2 Å shell was prepared after the analysis of the flat-cone data and they were measured in equatorial geometry by ω step scans (61 steps, reflection width 1·5°, ~20 min per reflection). In addition, data between 10 and 3 Å resolution for all reflections with *l* positive, *h* between 1 and 4 (hexagonal setting) were collected for the purpose of scaling. The total number of equatorial reflections was 2334, only half of them observed. Scaling of the flat-cone and equatorial data was based on 309 common reflections ($R = 5{\cdot}7\%$). The final 2·2 Å data set consisted of 4757 reflections, with 2843 of them (≃60%) considered observed (only 35% in the 2·25 Å shell). Thus, the quality of the neutron data was clearly inferior compared to the data previously measured in this laboratory for ribonuclease A (Wlodawer & Sjölin, 1983) and bovine pancreatic trypsin inhibitor (Wlodawer *et al.*, 1984), most probably as a result of the much smaller size of the crystal. Neutron structure factors have been deposited.*

X-ray data used in this work were collected in England by G. Dodson and coworkers. It was the same data set that was previously used for X-ray structure refinement, both at York (Dodson, unpublished) and by Isaacs & Agarwal (1978). The number of reflections in the 10–1·5 Å shell was 13476 (including all data with $F > 0$), and the number of reflections with $F > 2\sigma(F)$ was 12877.

Computational procedures

All structure refinements were accomplished on a VAX 11/780 computer with the least-squares program *PROLSQ* of Hendrickson (1985), modified for use with neutron data by Wlodawer & Hendrickson (1982). The refinement program utilized a space-group-specific structure-factor-calculation routine in which isotropic temperature factors were included. A limited number of disordered side-chain groups could be handled. The program was capable of using either X-ray or neutron structure factors, or of joint refinement of the coordinates with both types of data. The contribution of the scattering factors of H atoms was excluded from X-ray structure-factor calculations, but was always included for the neutron data. Some refinement runs were performed with specifically modified versions of the program, such as a version that would not apply coordinate shifts to the solvent. The positions of both Zn atoms in each asymmetric unit were taken from the initial model and were held constant.

Difference Fourier maps were examined using interactive computer graphics. The program *FRODO* (Jones, 1978), implemented for the Evans and Sutherland Multi-picture System attached to a VAX 11/780 by B. Bush, was used for this purpose. The types of maps examined by us included $(2F_o - F_c)$, $(F_o - F_c)$ and 'fragment ΔF' maps. Maps of the latter type were calculated after excluding parts of the structure from the phase calculations.

Refinement with the neutron data

The initial model of insulin from York was refined with the X-ray data, in order to optimize it before further use. After 10 cycles of refinement with *PROLSQ* the *R* factor was lowered from 0·171 to 0·148, while the deviation of bond lengths from ideality became 0·023 Å. H atoms were added to the rerefined model using the program *HAFFIX*, written by W. Hendrickson (unpublished). Their positions were determined by chain geometry, while in the absence of such information (in terminal hydroxyls, methyls, solvents, *etc.*) the H atoms were oriented in an arbitrary way. The neutron *R* factor for the hydrogen-containing model was 0·309, and was lowered to 0·175 in ten cycles of refinement with the neutron data only. While the overall geometry of the resulting model was largely unchanged from the starting point defined above and the *R* factor was much lower, the two models were quite different (r.m.s. deviation 0·532 Å). Large movements of some side chains were probably caused by the paucity of information in the neutron data about groups such as long aliphatic chains (Wlodawer & Hendrickson, 1982) and by the very poor ratio of observations to parameters for the neutron structure factors. While separate neutron refinement was capable of yielding a model with good geometry and with an acceptably low *R* factor, we questioned whether the amount of information was sufficient in the absence of other restraints such as those provided by the X-ray data to give credence to this model. In the end this

* Atomic coordinates and neutron structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 3INS, R3INSSF), and are available in machine readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37027 (as microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

particular model was not used, and only convinced us about the feasibility of refining the neutron data.

Joint refinement with X-ray and neutron data

Using both the X-ray and neutron data, further refinement was undertaken in several steps as described below.

(a) Refinement of protein and solvent

Joint refinement was initiated using the refined 1·5 Å X-ray model described above. The model included 1942 protein atoms (including six disordered side chains) and 325 solvent sites. The initial R factors were $R_X = 0{\cdot}147$ and $R_N = 0{\cdot}293$. After fifteen cycles of refinement the resulting R factors were $R_X = 0{\cdot}168$ and $R_N = 0{\cdot}201$ and the r.m.s. deviation of bond lengths from ideality was 0·027 Å. Nine further cycles of refinement were performed in which the restraint weights were altered to improve the geometry of the model. The R factors were $R_X = 0{\cdot}176$ and $R_N = 0{\cdot}199$ with an r.m.s. deviation from bond-length ideality of 0·020 Å. The total r.m.s. shift for all the atoms from the 1·5 Å X-ray model was 0·402 Å (protein = 0·352 and solvent = 0·592 Å). As discussed by Wlodawer & Hendrickson (1982) the higher values of the R factor obtained for joint refinement than for individual refinements with one type of data are due to both real differences between the structures and to the need for one model to fit simultaneously two sets of structure factors.

$2F_o - F_c$ neutron and X-ray maps were calculated to analyze the agreement of densities and the model. No major differences of the neutron and X-ray density distributions were found in the areas of the maps covering the main-chain and the side-chain atoms. The disordered side chains of *B*12, *B*21, *B*22 and *B*27 of molecule 1 and *B*22 and *B*29 of molecule 2 in the X-ray model were also seen in the neutron density. No estimation of their relative occupancies was attempted. Overall, the neutron model of the protein was seen to be essentially the same as the X-ray model.

However, this was not the case for all of the refined solvent sites. Analysis of the $2F_o - F_c$ maps, revealed that nearly all of the more ordered solvent sites (mainly those forming hydrogen bonds to the protein) in the neutron and X-ray densities were very similar, but many of the remaining sites were located either in X-ray density alone, or in regions of overlapping X-ray and neutron densities for which the centroids (centers of the respective peaks) of the two densities were far apart (> 1·0 Å). Moreover, several of the solvent sites appeared to move into areas of no density, similar to the 1·5 Å X-ray refinement carried out previously (see above). The non-bonded restraints on the solvent sites tended to push these sites apart where they were less than ~2·8 Å from each other. These problems related to the refinement of solvent positions severely limited the amount of new information provided on this subject by the neutron data.

(b) Amide H/D *occupancy refinement*

In this step, the occupancies of the amide H/D-exchanged sites were estimated from refinement using the neutron data only. The positional and thermal parameters for all the atoms were kept constant, while their occupancies were allowed to vary (Wlodawer & Sjölin, 1982). During the occupancy refinement, the amide H atoms were treated as D atoms using the scattering length for this atom type.

Three different starting models were used with all amide hydrogens assigned as (1) D atoms, (2) dummy atoms with zero occupancies, or (3) H atoms. For technical reasons, this was achieved by always assuming that the atom was a deuterium and by changing its occupancy to 1·0, 0·0 or −0·55 for the three models mentioned above, while in the following discussion the occupancies are translated to refer to H and D atoms individually. The occupancies of the non-amide H or D atoms were constrained to have maxima of 1·0. Although the occupancy values can absorb errors in the model and data (scaling, temperature factors, *etc.*), the amide H/D occupancies appeared to show a consistent trend in all three of the individual refinements, with a maximum difference in individual occupancy values of ~0·3 between refinements 1 and 3. Fig. 1 shows a histogram of the final amide occupancies obtained from the third refinement run. The final occupancies were not corrected for any of the expected errors and were assigned either as fully occupied D sites (occupancies > 0·75 for D), partially occupied H/D sites (occupancies < 0·75 for H or D) or fully occupied H sites (occupancies > 0·75 for H).

In all, four amide H atoms were assigned as full hydrogens, 32 amide H atoms were assigned as partially occupied sites and 64 as full deuteriums. Most of the partially occupied and three of the fully occupied amide H atoms occur in the center of the helical region of the *B* chains (see Fig. 1): between residues 14 and 19 in molecule 1 and residues 12 to 19 in molecule 2. These regions of the helices do not appear to be directly exposed to the solvent. The fourth fully occupied amide H atom is located at the end of the second helix of the *A* chain of molecule 2 where it forms a hydrogen bond to the carbonyl group of *A*17 in the helix, but is nevertheless adjacent to a water position which also forms an hydrogen bond to the same carbonyl group.

With the amide-exchange occupancy values obtained as described above, further joint X-ray/neutron refinement was performed. After ten cycles of *PROLSQ*, the R factors were reduced as follows: R_X from 0·190 to 0·182 and R_N from 0·230 to 0·191.

During this refinement, the solvent positions were not varied, but were held fixed in the positions of the York model. The final r.m.s. deviation from the ideal bond lengths was 0·019 Å and the r.m.s. shift in the positions of protein atoms from the previous joint X-ray/neutron model was 0·264 Å. This model will be further discussed in the next section, and has been deposited with the Protein Data Bank (see deposition footnote).

Results and discussion

(*a*) *Structure*

The insulin molecule consists of 51 amino acids, which make up two chains: the *A* chain of 21 residues and the *B* chain of 30 residues. It is a compact globular structure with an α helix in the center of the *B* chain (*B*10–*B*19) and two short helices in the *A* chain (*A*1–*A*6 and *A*16–*A*21) linked by a small extended chain (*A*7–*A*13). The *A* chain lies on the *B* chain and is largely constrained by the extended N and C terminal groups of the *B* chain. The molecule has two hydrophobic surfaces; these are buried by the formation of dimers, which in turn aggregate into hexamers with two Zn ions on the center axis. The secondary structure of insulin is shown in Fig. 2.

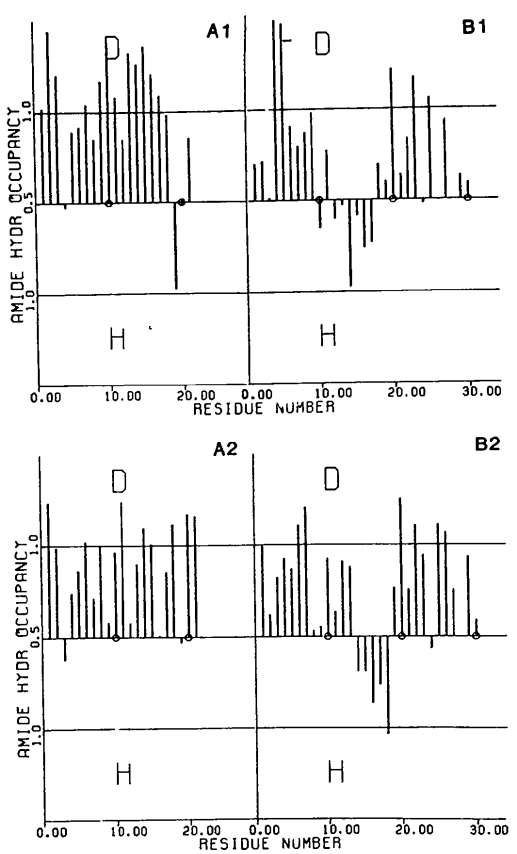

Fig. 1. Occupancies of the amide H atoms in 2Zn pig insulin for the *A* and *B* chains of molecules 1 and 2. Atoms that are predominantly of deuterium character are plotted above the central line, hydrogens are plotted below the line. Every tenth residue is marked by an open circle. The scatter of occupancies beyond 1·0 is an indication of errors in their determination.

In the original 1·5 Å X-ray refinement (Dodson *et al.*, 1979), six of the side chains were observed to be disordered and each of the alternate positions was assigned half occupancy. These were arginine *B*22 and lysine *B*29 of molecule 1 and glutamine *B*4, valine *B*12, aspartic acid *B*21, arginine *B*22 and threonine *B*27 of molecule 2. About 342 water sites (many disordered and less than 2·3 Å apart) were also refined, but the details of the observed water structure have not yet been reported. In the present neutron/X-ray refinement the same six side chains were observed to be disordered in the neutron maps. The r.m.s. deviation in the protein coordinates between the original X-ray model from York and the presently refined joint X-ray/neutron model is 0·36 Å and about 0·6 Å for the solvent.

The agreement between the 1·5 Å X-ray and the 2·2 Å neutron Fourier maps ($F_o - F_c$) was quite good over almost all of the two insulin molecules and most of the ordered solvent. The main areas of discrepancy were observed in the more disordered regions, such as, for example, around the disordered side chains and solvent not located next to polar groups – that is, in hydrogen-bonding contact.

(*b*) *Hydrogen positions*

The hydrogens of the side-chain methylene groups were very poorly visible in the neutron maps at this resolution owing to the overlap and almost total cancellation of the positive scattering densities of carbon ($0{\cdot}665 \times 10^{-12}$ cm) and the negative hydrogen-scattering densities ($-0{\cdot}374 \times 10^{-12}$ cm for each of the two H atoms). Positions of the H atoms can, however, be deduced from chain geometry. Some negative peaks were observed for several of the terminal methyl groups but could not be resolved into preferred identifiable H-atom positions. Further analysis of methyl groups in terms of averaging the densities (Kossiakoff & Shteyn, 1984) will be required to ascertain their orientations.

The densities at the exchanged positions of a number of residues have helped to clarify the nature of some close contacts made within both the hexamer and the crystal.

(i) In molecule 1 the ND1 and NE2 positions of the histidyl group *B*5 make contacts of 3·0 Å to carbonyl O atoms of residues *A*7 (of molecule 1) and *A*9 (of symmetry-related molecule 2) respectively, suggesting that the imidazole group is protonated. In molecule 2, only one contact, through ND1, is made to the carbonyl O atom of *A*7 of the same molecule. In Fig. 3 the densities for the His *B*5 imidazole groups of molecules 1 and 2 are shown. The extensions of density

at ND1 and NE2 positons in both molecules demonstrate that both histidines are fully protonated, even though the presence of the D atom bound to ND1 of molecule 2 could not have been ascertained without reference to the neutron map, since no clear hydrogen-bond partner for this group could be established.

(ii) The six *B*13 glutamic acid side chains cluster tightly at the hexamer center. There is a particularly close contact between the two glutamates within the dimer, their carboxylate O atoms being found to be 2·44 Å apart in the joint refinement analysis (2·6 Å in the X-ray refinement). In the structure of 4Zn insulin, the corresponding distance was found to be 2·95 Å, a slightly more likely value (Smith, Swenson, Dodson, Dodson & Reynolds, 1984). Other nonbonded distances, however, are too close in that structure as well. As seen in Fig. 4, the neutron map shows that there is no density between carboxylate oxygens, which would be expected if a deuterium were present and a hydrogen bond were formed. While the map in this region is not very clear, we cannot confirm the existence of a hydrogen bond (unlikely as it is at a pH of 6·5) and can suggest a nonbonded interaction instead.

(iii) The side-chain amide groups of the surface glutamines *A*5 and *A*15 are involved in close contacts at both the carbonyl oxygens and the amide nitrogens, but it is not clear from the X-ray analysis which alternative conformation, N–H···O=C or C=O···H–N, between *A*5 and *A*15 is to be preferred. Electron density maps of proteins usually cannot distinguish between the oxygen and nitrogen in side-chain amides. Thus their assignment can be only indirect. Neutron density, on the other hand, is markedly asymmetric due to the presence of two D atoms bound to an N and the orientation of an amide can be quite easily determined with neutron data (Wlodawer *et al.*, 1984). We were indeed able to make the assignments for the two residues in question and found that the most probable interactions differ in the two insulin molecules. It appears that NE2 of Gln *A*15 is a deuterium donor in molecule 1, while OE1 of Gln *A*5 is an acceptor. The opposite is true for molecule 2, where it is NE2 of Gln *A*5 which is a donor for a rather poor hydrogen bond to OE1 of Gln *A*15. The latter atom also accepts another hydrogen bond from the main-chain amide of residue A12, with the latter hydrogen bond appearing to be much better than the former.

(*c*) H/D *exchange of the amide groups*

As seen in Fig. 2, the pattern of protection of amide H-atoms from exchange is strikingly different in insulin

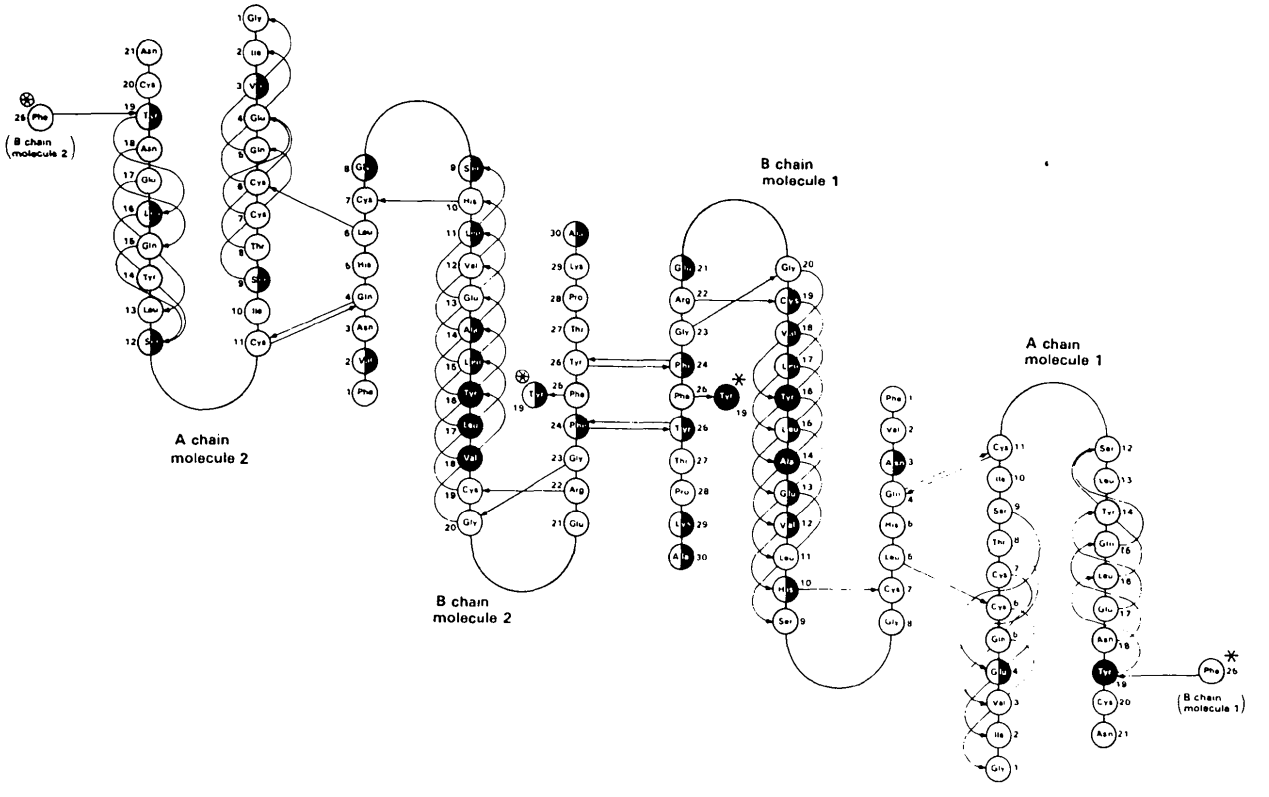

Fig. 2. Secondary structure of insulin, showing all main-chain to main-chain hydrogen bonds in both molecules. Fully exchanged amide H atoms are shown as open circles, partially protected as half-filled circles and fully protected as filled circles.

than in other proteins studied so far. The only extended region of significant protection under the experimental conditions employed by us (over four months soak of crystals at pH = 6·5) was found in the center of a helical region in each of the *B* chains (residues 14–19 in molecule 1 and 12–19 in molecule 2). No significant protection of amide H atoms was observed in the areas of β sheet (Fig. 2). This is in contrast with the observations in bovine pancreatic trypsin inhibitor (BPTI), in which only the amides involved in forming β sheets were protected, while all amides in α helices were fully exchanged (Wlodawer *et al.*, 1984). Only those amide protons in α helices that were adjacent to the areas of β sheet were protected in ribonuclease A (Wlodawer & Sjölin, 1982) and in trypsin (Kossiakoff, 1982) while extensive protection was observed in both proteins in the areas of β sheet. The latter result is usually attributed to the large degree of deformation allowed for β sheets without disruption of their hydrogen-bonding networks (Salemme, 1981). Slightly more amide protons belonging to β sheets than to α helices are observed among 35 most highly protected groups in lysozyme (Mason, Bentley & McIntyre, 1984), while only six amide protons, two of them in the center of an α helix but not adjacent, were reported to be protected in carbon monoxymyoglobin (Hanson & Schoenborn, 1981).

The errors associated with occupancy values in insulin were estimated to be quite large, in the region of 0·2 occupancy units. Thus, very little can be said about the variations of the individual values. For example, the *A*19 amide in molecule 1 is seen to be unexchanged, while in molecule 2 the exchange value is ~0·5. The environment around this amide group is very similar in both molecules, with no easy access for water molecules. With a 2σ level of ~0·4, no real difference in protection of the *A*19 amides in the two molecules can be seen from these results.

Hvidt & Pedersen (1974) observed that under identical conditions of pH the exchange of peptide groups in insulin is much faster than in BPTI, and it is likely that if the same mechanism of protection were present in both proteins, it would be conceivable that full exchange could have occured in insulin under our experimental conditions. Insulin is much less rigid than BPTI and thus is more susceptible to processes leading to exchange. Our interpretation of the results in insulin is that the protection is offered by very tight packing in the hexamers forming the crystals, and the behavior in solution may be very different. Results available for lysozyme (Mason, Bentley & McIntyre, 1984) and BPTI (Wlodawer *et al.*, 1984) show a large degree of similarity in the pattern of protection in crystals and in solution, even though the actual rates of exchange in these two phases may vary by up to several orders of magnitude. This, however, may not hold true for

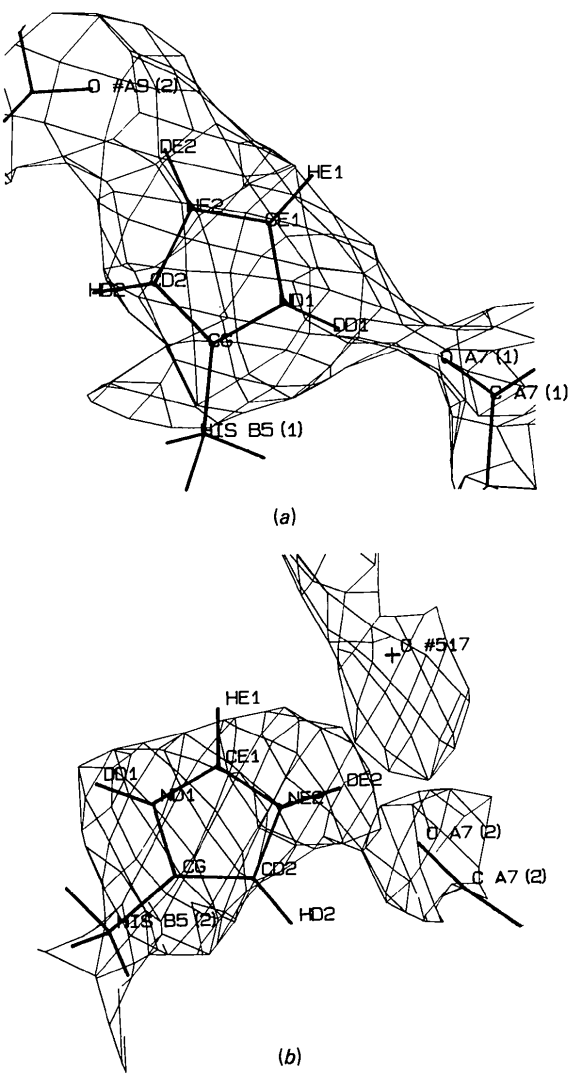

Fig. 3. Neutron density for the side chains of *B*5 histidines of molecules (*a*) 1 and (*b*) 2 of insulin.

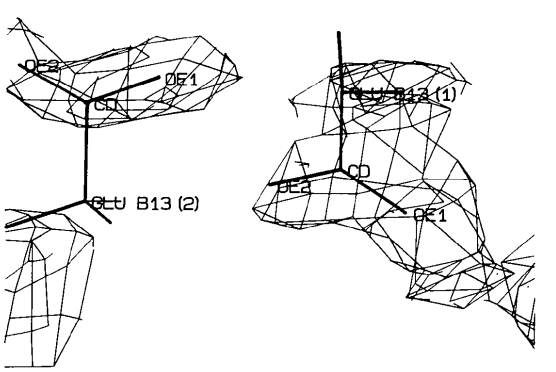

Fig. 4. Neutron density for the side chains of Glu 13 of molecules 1 and 2. Note that no density that would correspond to a D atom is present between oxygens OE2 (1) and OE1 (2), suggesting lack of protonation of these groups.

insulin, and further studies are indicated. Only a combination of neutron diffraction measurements utilizing much shorter soaking times with two-dimensional NMR can answer if it is possible to explain a different exchange behavior of insulin by the differences in protection offered by a crystal, as opposed to the easier accessibility of amide groups in solution, or if another mechanism is operating.

(*d*) *Solvent positions*

The positions of solvent molecules resulting from this refinement were not analyzed in detail since the number of neutron structure factors was much smaller than the number of X-ray data, with the quality of neutron maps in the solvent regions being inferior. Some difference in the solvent positions may also be expected owing to small changes in pH (6·0 *versus* 6·5 respectively for the crystals used to collect X-ray and neutron data). On the other hand, even though hydrogen-bonding affinities differ somehow for H_2O and D_2O, the topological water structure is not very dependent on the nature of the hydrogen isotope. We found that about 25% of the 342 assigned solvent sites in the York model coincided with solvent peaks in the neutron $2F_o - F_c$ maps. These sites were mainly the more ordered ones that readily form hydrogen bonds to the insulin molecule. Fig. 5 shows two solvent sites that are highly occupied in both the X-ray and neutron maps. For a further 25% of the X-ray sites, neutron peaks were observed that either just overlapped these sites or were very close to them (< 1·0 Å away). For the remaining 50% of the X-ray sites, no significant neutron density was observed above the noise level. It is clear from these results that higher resolution neutron data are required to provide significant contribution to the assignment of solvent positions in insulin.

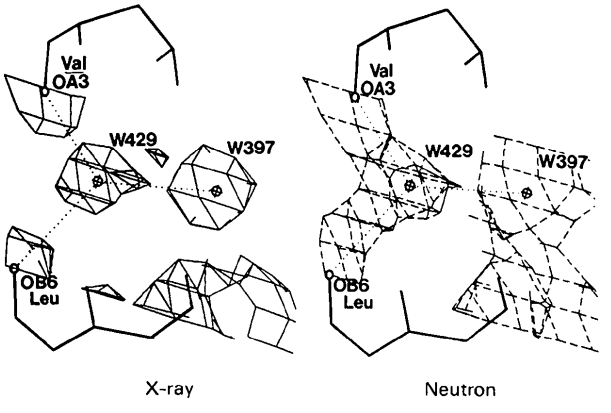

Fig. 5. X-ray (solid contours) and positive neutron density (in dashed contour lines) in the region of waters 429 and 397. These waters are found on the surface of the insulin molecule. Continuous density of the neutron map is due to unresolved D atoms and indicates the directions of hydrogen bonds.

We would like to thank Professor Liang Dong-Cai of Academia Sinica, Beijing, People's Republic of China, for the insulin crystal used to collect the neutron data. Certain commercial equipment, instruments and materials are identified in this paper in order to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

ABEL, J. J. (1926). *Proc. Natl Acad. Sci. USA*, **12**, 132–136.

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.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BLUNDELL, T. L., CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., HODGKIN, D. C., MERCOLA, D. A. & VIJAYAN, M. (1971). *Nature (London)*, **231**, 506–511.

BLUNDELL, T. L., DODSON, G. G., HODGKIN, D. C. & MERCOLA, D. A. (1972). *Adv. Protein Chem.* **36**, 279–402.

CHANG, W. R., STUART, D., DAI, J. B., TODD, R., ZHANG, J. P., XIE, D. L., KUANG, B. & LIANG, D. C. (1984). 13th Int. Congr. Crystallogr. Abstracts, p. C-16.

CROWFOOT, D. (1935). *Nature (London)*, **135**, 591–592.

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.

HANSON, J. C. & SCHOENBORN, B. P. (1981). *J. Mol. Biol.* **153**, 117–146.

HENDRICKSON, W. A. (1985). *Methods in Enzymology*, Vol. 115, edited by H. W. WYCKOFF, C. H. W. HIRS & S. N. TIMASHEFF, pp. 252–270. Orlando: Academic Press.

HVIDT, A. & PEDERSEN, E. J. (1974). *Eur. J. Biochem.* **48**, 333–338.

ISAACS, N. W. & AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 782–791.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KOSSIAKOFF, A. A. (1982). *Nature (London)*, **296**, 713–721.

KOSSIAKOFF, A. A. & SHTEYN, S. (1984). *Nature (London)*, **311**, 582–583.

LIANG, D. C. (1982). *Proteins in Biology and Medicine*, edited by R. A. BRADSHAW, R. L. HILL, J. TANG, C. LIANG, T. TSAO & C. TSOU, pp. 25–42. New York: Academic Press.

MASON, S. A., BENTLEY, G. A. & MCINTYRE, G. J. (1984). *Neutrons in Biology*, edited by B. P. SCHOENBORN, pp. 323–334. New York: Plenum Press.

PRINCE, E., WLODAWER, A. & SANTORO, A. (1978). *J. Appl. Cryst.* **11**, 173–178.

SAKABE, N., SAKABE, K. & SASAKI, K. (1980). *Water and Metal Cations in Biological Systems*, edited by B. PULLMAN & K. YAGI, pp. 117–128. Tokyo: Scientific Societies Press.

SAKABE, N., SAKABE, K. & SASAKI, K. (1981). *Structural Studies on Molecules of Biological Interest*, edited by G. DODSON, J. P. GLUSKER & D. SAYRE, pp. 509–526. Oxford: Clarendon Press.

SALEMME, R. (1981). *J. Mol. Biol.* **146**, 143–156.

SCHLICHTKRULL, J. (1956). *Acta Chem. Scand.* **10**, 1455–1458.

SCHOENBORN, B. (1975). *Anomalous Scattering*, edited by S. RAMASESHAN & S. C. ABRAHAMS, pp. 407–416. Copenhagen: Munksgaard.

SJÖLIN, L. & WLODAWER, A. (1981). *Acta Cryst.* A37, 594–604.
SMITH, G. D., SWENSON, D. C., DODSON, E. J., DODSON, G. G. & REYNOLDS, C. D. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 7093–7097.
TEETER, M. M. & KOSSIAKOFF, A. A. (1984). *Neutrons in Biology*, edited by B. P. SCHOENBORN, pp. 335–348. New York: Plenum Press.
WLODAWER, A. (1980). *Acta Cryst.* B36, 1826–1831.
WLODAWER, A. (1982). *Prog. Biophys. Mol. Biol.* **40**, 115–159.
WLODAWER, A. & HENDRICKSON, W. A. (1982). *Acta Cryst.* A38, 239–247.
WLODAWER, A. & SJÖLIN, L. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 1418–1422.
WLODAWER, A. & SJÖLIN, L. (1983). *Biochemistry*, **22**, 2720–2728.
WLODAWER, A., WALTER, J., HUBER, R. & SJÖLIN, L. (1984). *J. Mol. Biol.* **180**, 301–331.

Acta Cryst. (1989). B45, 107–112

Static Disorder in (–)-(1*R*,5*R*,9*R*,13*S*)-2′-Hydroxy-5,9-dimethyl-2-(2-methyltetrahydrofurfuryl)-6,7-benzomorphan,* $C_{20}H_{29}NO_2$. Crystal Structure and *MM*2 Pucker Analysis of the Tetrahydrofuran Ring†

BY C. L. VERLINDE, C. J. DE RANTER,‡ N. M. BLATON AND O. M. PEETERS

Laboratorium voor Analytische Chemie en Medicinale Fysicochemie, Instituut voor Farmaceutische Wetenschappen, Katholieke Universiteit Leuven, Van Evenstraat 4, B-3000 Leuven, Belgium

(*Received* 23 *February* 1988; *accepted* 28 *September* 1988)

Abstract

$M_r = 315{\cdot}454$, monoclinic, $P2_1$, $a = 8{\cdot}942\ (4)$, $b = 13{\cdot}719\ (6)$, $c = 7{\cdot}313\ (3)$ Å, $\beta = 91{\cdot}80\ (3)°$, $V = 896{\cdot}7\ (7)$ Å^3, $Z = 2$, $D_m = 1{\cdot}15\ (2)$, $D_x = 1{\cdot}168$ Mg m^{-3}, λ(Mo $K\alpha$) = 0·71069 Å, μ(Mo $K\alpha$) = 0·069 mm^{-1}, $F(000) = 344$, $T = 291$ K, final $R = 0{\cdot}041$ for 1010 observed reflections. The side-chain torsion angles about the exocyclic nitrogen bond are eclipsed while the N–C–C–O torsion angle is (–)-synclinal. Static disorder of the tetrahydrofuran ring is analyzed through *MM*2 force-field calculations of the pseudorotation pathway. Two equienergetic mirror-image puckering forms are identified as minima. A straightforward interpretation of the principles governing their selection is given. The ability of the molecular-mechanics approach to scrutinize this kind of problem in a successful way is verified through a Cambridge Structural Database search. Other cases of similar disorder are pointed out, and a warning is issued against using databases without consulting the original papers.

Introduction

Initially the structure of the title compound was determined as part of a structure–activity study on 6,7-benzomorphans. The drug is a mixed opioid agonist–antagonist *in vivo* (Merz & Stockhaus, 1979), and is about as potent as the prototype κ-opioid agonist (±)-U-50 488 (Lahti, Von Voightlander & Barsuhn, 1982) in the rabbit vas deferens twitch inhibition test (Verlinde & De Ranter, 1988).

The main purpose of this determination was to provide insight into the conformational behaviour of the *N*-side chain which is believed to be essential for opioid κ activity (De Ranter, Verlinde, Blaton & Peeters, 1984). Most likely, the oxygen in β-position engages as an acceptor in a hydrogen bond with the receptor. In this respect, the conformational restriction of the oxygen in a tetrahydrofuran ring should affect the intramolecular entropy in a favourable way compared with the virtually inactive opened-ring analogues.

However, upon solving the structure a major drawback of the conformational restriction in a five-membered ring became apparent: the presence of more than one puckering mode of the tetrahydrofuran. Full crystallographic elucidation through group refinement of this phenomenon proved to be very hard, mainly because of large correlations. Accordingly a molecular-mechanics study of the puckering was undertaken, using the *MM*2 force field (Allinger & Flanagan, 1983). Although *MM*2 is widely used, and has been found to be applicable to large numbers of compounds and chemical problems (Allinger & Lii, 1987), a confrontation with a large sample of experimental data was considered to be necessary. For this purpose use was made of tetrahydrofuryl fragments obtained through a search of the Cambridge Structural Database (CSD) (Allen, Bellard, Brice, Cartwright, Doubleday, Higgs, Hummelink, Hummelink-Peters, Kennard, Motherwell, Rodgers & Watson, 1979).

* *Chemical abstracts* name: (–)-(2*R*,6*R*,11*R*,2′*S*)-1,2,3,4,5,6-hexahydro-6,11-dimethyl-3-(2-methyltetrahydrofurfuryl)-2,6-methano-3-benzazocin-8-ol.

† Structural Studies of Substituted 6,7-Benzomorphan Compounds. XIII. Part XII: Verlinde, Blaton, Peeters & De Ranter (1988).

‡ To whom correspondence should be addressed.

0108-7681/89/010107-06$03.00