Crystal Structure of a Prolonged-Acting Insulin with Albumin-Binding Properties†

Jean L. Whittingham,*,‡ Svend Havelund,§ and Ib Jonassen§

Department of Chemistry, University of York, Heslington, York YO1 5DD, England, and Novo Nordisk Research Institute, Novo Alle, DK-2880 Bagsvaerd, Denmark

Received October 7, 1996; Revised Manuscript Received December 20, 1996⊗

ABSTRACT: The fatty acid acylated insulin, Lys^{B29}-tetradecanoyl, des-(B30) human insulin, has been crystallized and the structure determined by X-ray crystallography. The fatty acid substituent on residue B29 Lys binds reversibly to circulating albumin protein *in vivo*, and by this mechanism the hormone's action is prolonged. Crystals of the fatty acid insulin grow in space group *R*3, with two dimers in the asymmetric unit, and diffract to 1.8 Å spacing. The structure has been solved by molecular replacement and refined using a maximum likelihood method. The crystal structure consists of R6 zinc insulin hexamers which contain phenol. The fatty acids can be seen bound between the hexamers, making specific interactions with the side chains of residue B1 Phe; however, the lysine side chains to which the fatty acids are covalently attached are mostly disordered. The mode of binding of the fatty acids appears to be determined by crystal packing, and whether or not they interact with the protein in this way in solution remains uncertain.

Insulin-dependent diabetics rely on a combination of rapid-acting and long-acting insulin preparations in order to maintain near normoglycemia during periods of high and low insulin demand. Unfortunately, insulin delivery from the pancreas is difficult to reproduce by subcutaneous injection, and consequently most diabetics suffer from short-term and long-term complications associated with hypo- and hyperglycemia, including retinal nerve damage and kidney disorders. Over the years insulin preparations have been modified in an attempt to reduce these complications. At the heart of these experiments is an understanding of the structure of insulin and its complex aggregation properties.

The insulin molecule consists of two polypeptide chains, an A chain of 21 amino acids and a B chain of 30 amino acids. Although circulating insulin is active as a monomer, it is synthesized and stored as hexamers, and in many insulin crystal structures, three insulin dimers aggregate around two zinc ions to form a globular, hexameric structure (Adams et al., 1969; Baker et al. 1988). The classical strategy to produce long-acting insulins is formulation of crystalline or amorphous suspensions which can be injected subcutaneously to form a slowly dissolvable depot (Scott & Fisher, 1936; Krayenbühl & Rosenberg, 1946; Hallas-Møller, 1956). Although these preparations fulfill their role, they display variable absorption rates giving rise to inadequate blood glucose control. A more recent strategy is to change the p*I* of insulins to neutral by amino acid substitutions. This type of insulin is formulated as acid solutions and injected subcutaneously where it forms an amorphous or crystalline depot which is slowly dissolved (Markussen et al., 1987, 1988; Obermeier et al., 1986). In contrast, Lys^{B29}-tetradecanoyl, des-(B30) human insulin (NN304 insulin) is a neutral, soluble insulin preparation in which residue B29 Lys, which is not required for activity, has been covalently bound to a 14-carbon fatty acid, as shown in Figure 1 (Kurtzhals et al., 1995).

The fatty acid extension on NN304 insulin facilitates its reversible binding to albumin, a circulatory protein whose role is to bind and transport various small molecules including fatty acids (Peters, 1985). In this way the hormone's action is delayed (Kurtzhals et al., 1996; Markussen et al., 1996). Although fatty acids of various lengths have been attached to residue B29 Lys, the myristoyl derivative was found to have the optimum binding properties (Kurtzhals et al., 1995). Furthermore, it has been shown that NN304 insulin has a more reproducible absorption rate than NPH (neutral protamine hagedorn) insulin, and owing to its nonsolid nature it does not stimulate macrophage attack at the site of injection (Markussen et al., 1996).

Structural studies on NN304 insulin have been carried out in order to assess the effect of the fatty acid on the insulin structure and perhaps gain insight into how it binds to albumin, the structure of which is known (Carter & Ho, 1994). The insulin fold, already well established from many crystallographic and NMR studies, consists of two α-helices, residues A1−A8 and A13−A19, an α-helix, residues B9−B19, and two extended chain regions at the B-chain N- and C-termini. One intrachain and two interchain disulfide bonds fix the relative positions of the two chains. In hexameric insulin structures the conformation of the B-chain N-termini changes depending on ambient conditions, and in the presence of 0.07 M phenol (used as a preservative in insulin preparations) they are all fully α-helical (Derewenda et al., 1989; Smith & Dodson, 1992). This conformational change results in the movement of residue B1 Phe from a buried, hydrophobic pocket between the dimers to a position on the surface of the hexamer. The terms T and R are used to describe the insulin molecules with residues B1−B8 in an extended conformation and α-helical conformation respectively, and hence the phenol-containing hexamer is said to have an R6 conformation (Kaarsholm et al., 1989). Both monoclinic and rhombohedral forms of the native, phenol-containing R6 insulin hexamer have been refined, to 2.0 and 2.5 Å, respectively. In this paper the monoclinic form has

† The coordinates for the NN304 insulin structure have been deposited in the Brookhaven Protein Data Bank (reference 1XDA).

* Author to whom correspondence should be addressed.
‡ University of York.
§ Novo Nordisk Research Institute.

⊗ Abstract published in *Advance ACS Abstracts,* February 15, 1997.

S0006-2960(96)02510-X CCC: $14.00

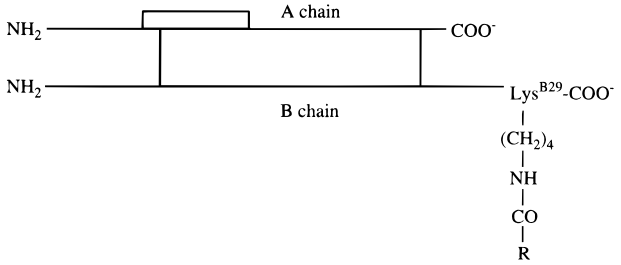

FIGURE 1: Schematic representation of NN304 insulin showing the peptide link between the ϵ-amino group of the side chain of B29 Lys and the myristoyl fatty acid. R denotes the fatty acid side chain consisting of 13 carbon atoms with associated hydrogen atoms. By removal of B30 Thr the negative charge of the C-terminal gets closer to the aliphatic side chain, whereby the modified site may mimic a nonesterified fatty acid better than in the presence of residue B30 Thr.

been used for detailed comparisons with the NN304 insulin crystal structure since the two structures are refined to similar resolutions.

MATERIALS AND METHODS

Crystallization. NN304 insulin was prepared as previously described (Kurtzhals et al., 1995), and all chemical reagents used were of analytical grade. Crystallization trials for NN304 insulin were based on a well-established protocol in which crystals were grown by the vapor diffusion hanging drop method in the presence of zinc ions and phenol (Hu et al., 1992). Phenol was found to be essential for the crystallization of the fatty acid insulin; in its absence the protein formed a glassy precipitate even in the presence of zinc ions. Initially, the crystals were large and deformed and appeared in 2−3 days, however, with the addition of ethanol the rate of crystal growth was reduced and the resulting crystals were smaller, forming chunky hexagonal rods. The optimum crystallization conditions for NN304 insulin are as follows. The protein solution consisted of 7.5 mg/mL NN304 insulin dissolved in 0.02 M HCl. The reservoir solution contained 0.1 M trisodium citrate, 6% (w/v) Tris, and 0.02% (w/v) zinc acetate, adjusted to pH 8.2. To each 1 mL reservoir was added 10 μL of 5% (w/v) aqueous phenol solution and 10−60 μL of ethanol. Each hanging drop contained 1 μL of the protein solution and 1 μL of the reservoir solution.

Data Collection and Data Processing. X-ray diffraction data were collected from a single crystal of NN304 insulin (dimensions 0.3 × 0.3 × 0.2 mm) mounted in a rayon loop and flash-frozen to 120 K. The cryosolvent in which the crystal was initially soaked had a composition of 30% glycerol plus 70% reservoir solution. A 20.0−2.0 Å data set was measured on an MSC R-AXIS II imaging-plate system mounted on a Rigaku RU200 rotating-anode generator with Cu Kα radiation. The data set was processed using DENZO version 1.2.1 and scaled with SCALEPACK version 1.2.3 (Otwinowski, 1993), from which it was determined that NN304 insulin crystals are in space group *R*3 with equivalent hexagonal unit cell dimensions $a = b = 79.14$ Å and $c = 79.29$ Å, corresponding to two dimers in the asymmetric unit. The length of the crystallographic *c* axis is double that of other rhombohedral insulin crystals. Data statistics are shown in Table 1.

Toward the end of the refinement it became obvious that higher resolution X-ray data were needed in order to clarify some ambiguities in the structure. Crystals were taken to the DESY synchrotron source at the EMBL outstation in Hamburg, where a 1.8 Å data set was collected from a single crystal using a MAR imaging plate detector on beamline X31 ($\lambda = 0.93$ Å). The crystal was soaked in a 40% glycerol-based cryosolvent and flash-frozen as before. All the data between 15.0 and 1.8 Å were measured, and the data were processed as above. Data statistics for the second data set are also shown in Table 1.

Table 1: Data Collection and Processing Statistics for NN304 Insulin

	data set	
	2.0 Å	1.8 Å
space group	*R*3	*R*3
cell axes *a*, *c* (Å)	79.14, 79.29	78.75, 79.20
diffraction limit (Å)	19.76−2.00	14.88−1.80
no. of observations	43230	40024
no. of independent reflections	12492	16624
highest resolution shell (Å)	2.11−2.00	1.83−1.80
completeness (%): all data	99.9	98.1
highest resolution shell	100.0	72.7
R_{merge} (on I)[a] (%)		
all data	5.2	5.5
highest resolution shell	23.0	27.8
data > 3σ (%)		
all data	86.0	79.8
highest resolution shell	62.3	57.6

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$.

Structure Solution and Refinement. The crystal structure of NN304 insulin was solved by molecular replacement using the AMoRe package (Navaza, 1994). The search model for molecular replacement was the dimer of a 4-iodophenol insulin hexamer which also crystallizes in space group *R*3 (unpublished results). This was chosen on the basis that the NN304 insulin structure contains phenol, which, like 4-iodophenol, stabilizes the R6 insulin hexamer conformation (Derewenda et al., 1989; Smith et al., 1992). There were two solutions to the molecular replacement calculation corresponding to the two expected insulin dimers in the asymmetric unit. The positions of the two dimers are related by a translation of half a unit cell in the direction of the crystallographic *c* axis and a rotation of approximately 11° about the same axis. The almost symmetrical arrangement of the two dimers in the asymmetric unit introduces pseudosymmetry into the crystal lattice, which is manifested in the diffraction pattern by alternating strong and weak layers of data along *c**.

The NN304 insulin model was refined using the maximum likelihood refinement suite of programs REFMAC (Murshudov et al., 1996). A 5% sample of reflections was excluded from the refinement in order to give some estimate of cross-validation in the form of a free *R* value (Brünger, 1993a). The progress of the refinement is shown in Table 2. Between rounds of refinement, $2F_o - F_c$ and $F_o - F_c$ electron density maps were calculated using all the data and displayed using the molecular graphics program XFIT (MSI, San Diego, CA). Zinc ions, chloride ions, phenol ligands, fatty acid side chains, and water molecules were built into the difference map as the refinement proceeded (Table 2). The nature of the electron density for the fatty acids indicated that the central sections of the hydrocarbon chains were well ordered, while the ends and the B29 Lys residues were disordered.

The refinement continued until successive cycles produced only minimal improvements in the *R* factors and the electron density maps, after which it was decided to include all the

Table 2: Summary of Refinement Procedure for NN304 Insulin

cycle	*R*-factor[a] (*R* free)	comments
0−8	0.519 → 0.402 (0.439)	refinement with 2.0 Å data set (5% *R* free data set); initial model contained no ions or ligand or water molecules, and all atomic *B*-factors were set to 20.0 Å^2; four zinc ions and two chloride ions were added after refinement
9−17	0.399 → 0.329 (0.391)	two phenol molecules added
18−35	0.327 → 0.254 (0.326)	38 water molecule, two chloride ions, and two phenols molecules added
36−53	0.253 → 0.226 (0.307)	11 waters and two fatty acid side chains added
54−67	0.207 → 0.196 (0.273)	two fatty acids and a few water molecules added
68−74	0.203 → 0.179	refinement now included the whole 2.0 Å data set (no data excluded for *R*-free calculation); 25 water molecules added
75−133	0.195 → 0.174	refinement converged with a total of 110 water molecules
134−243	0.283 → 0.174	after rigid-body refinement using XPLOR (Brünger, 1993b), *xyzb* refinement continued with new 1.8 Å data set; initially, the atomic *B*-factors were reset to 20.0 Å^2; during refinement more water molecules were added and the positions of the fatty acid became more well-defined; in the final model there were 153 water molecules

[a] Crystallographic *R* factor $= \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure amplitudes.

Table 3: Refinement and Model Statistics for NN304 Insulin

parameter	target rms	rms Δ NN304
bond lengths (1−2) (Å)	0.02	0.01
angle related distances (1−3) (Å)	0.04	0.03
intraplanar distances (1−4) (Å)	0.05	0.03
planar groups		
peptide plane (Å)	0.03	0.02
aromatic plane (Å)	0.02	0.01
chiral volumes (Å^3)	0.10	0.08
nonbonded contacts		
single torsion contacts (Å)	0.30	0.18
multiple torsion contacts (Å)	0.30	0.28
possible hydrogen bonds (Å)	0.30	0.27
torsion angles		
planar (0, 180°)	7.0	4.5
staggered (60°/120°)	15.0	15.7
orthonormal (±90°)	20.0	24.7
thermal factors		
main-chain bond (1−2) (Å^2)	2.00	1.63
main-chain angle (1−3) (Å^2)	3.00	2.37
side-chain bond (Å^2)	2.00	1.99
side-chain angle (Å^2)	3.00	3.03

data in the refinement. As a result the *R* factor fell by about 0.02, and the next set of electron density maps contained new peaks for water molecules and some disordered side chains. Nevertheless, there was still some difficulty in determining the correct positions of the fatty acid chains and the B29 Lys side chains, and at this point the 1.8 Å data set was introduced to complete the refinement. Since the unit cell for these data was slightly different from that of the 2.0 Å data (Table 1), the two sets of reflections were not merged. With the additional data the positions of three of the fatty acids were clarified, while it was apparent that the fourth fatty acid in the asymmetric unit was disordered. Refinement parameters and statistics are given in Table 3. Coordinates for the NN304 insulin structure have been deposited in the Brookhaven Protein Data Bank (reference 1XDA).

RESULTS

Structure Description. The NN304 insulin structure contains four molecules of insulin in the asymmetric unit plus four zinc ions, four chloride ions, four phenol molecules, four fatty acid side chains, and 153 water molecules. The accuracy of the structure was assessed by examining the protein geometry, the rms values for the geometrical parameters of the structure being in good agreement with the refinement restraints (Table 2). From the program PROCHECK (Laskowski et al., 1993), a Ramachandran plot showed that all the residues have main-chain dihedral angles within the allowed regions. In addition, the atomic *B* factors within the structure averaged 16.2 Å^2 for the main-chain atoms, 21.5 Å^2 for the side chain atoms, and 33.0 Å^2 for the water molecules and are in agreement with the overall temperature factor of 17.3 Å^2 obtained from a Wilson plot of the data. The overall precision of the structure was estimated using Cruickshank's formula, which calculates an expected positional error for an atom and is dependent on *R* factor, number of refined parameters, resolution, and completeness of data (Dodson et al., 1996). For NN304 insulin this equation gives an atom coordinate error of 0.1 Å for an atom with average *B* factor.

The four molecules of NN304 insulin aggregate to form two dimers packed side by side in the asymmetric unit. The molecules are numbered 1−4 and the two peptide chains in each molecule are labeled A and B. By convention, molecules 1 and 3 are equivalent to molecule 1 in the 2-zinc insulin dimer (Baker et al., 1988). The structures of the two crystallographically independent dimers are very similar, the rms difference of the main-chain atoms being 0.4 Å. This calculation does not include residues B28 and B29, which are connected to the fatty acid side chain and show small variations from molecule to molecule, or B30, which is absent in NN304 insulin.

Each of the dimers forms a component of a typical R6 hexamer centered around the crystallographic 3-fold axis (Figure 2). The hexamer contains two zinc ions, both having a tetrahedral coordination sphere comprising three B10 His side chains and a chloride ion, similar to the zinc ions in the native R6 phenol insulin hexamer (Derewenda et al., 1989; Smith & Dodson, 1992). At each of the three dimer−dimer interfaces in the hexamer there are two phenol molecules bound in identical hydrophobic pockets. The interactions made by the phenol molecules are the same as those in the native insulin hexamer, with the hydroxyl group of the phenol hydrogen-bonding to A6 Cys O and A11 Cys N and the phenolic ring in van der Waals contact with the side chain of B5 His of the adjacent dimer.

The hexamers derived from the two dimers of the asymmetric unit stack on top of one another with their 3-fold axes collinear with the crystallographic 3-fold axis. Unlike other rhombohedral insulin hexamer structures, however, the orientation of adjacent NN304 insulin hexamers is not the same. Instead, one hexamer is rotated by approximately 11° about its 3-fold axis relative to the neighboring hexamer. As a result the length of the crystallographic *c* axis in NN304 insulin is approximately double that of similar native rhombohedral insulin crystal structures. The interactions between the hexamers along the 3-fold axis constitute the principal crystal contacts in this structure.

The solvent content of the NN304 insulin crystals is relatively low, in the region of 40%, and the observed water

FIGURE 2: Stereoview of a hexamer of NN304 insulin showing the position of the zinc ions (white spheres, one eclipsing the other) at the center and the phenol molecules (ball and stick) at the dimer−dimer interfaces. The fatty acid side chains (ball and stick) are located at the front and back of the hexamer. The figure was made using MOLSCRIPT (Kraulis, 1991).

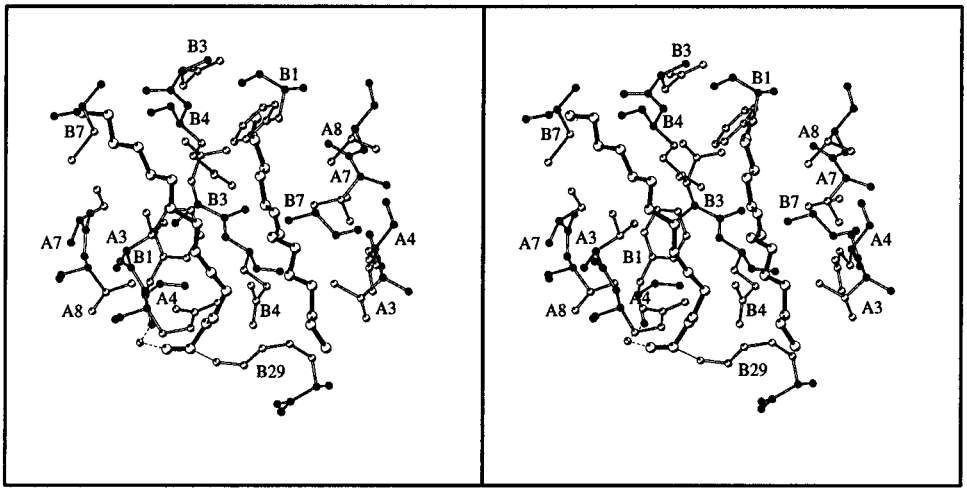

FIGURE 3: Stereoview of one of the fatty acid binding sites. The fatty acids are in the center of the figure with bonds colored in black. Main-chain protein atoms are also shaded black while the side-chain atoms are white. Adjacent fatty acids make a few van der Waals contacts although part of the right-hand fatty acid is disordered (see Figure 4). The peptide link between residue B29 Lys and the first carbon atom of the fatty acid is shown at the bottom of the figure by a thin line, and the hydrogen bonding between the carbonyl oxygen of the fatty acid, a water molecule (white sphere), and B1 Phe N is represented by dashed lines. The figure was made using MOLSCRIPT (Kraulis, 1991).

structure consists mainly of first-shell water molecules interacting directly with the protein. Most of the waters interact with the hydrophilic surface of the hexamer; however, there are some water molecules inside the hexamer that solvate the zinc ion channel and others that hydrogen-bond to residues B10 His and B13 Glu at the bottom of the phenol binding pockets.

Fatty Acid Binding. Each of the four molecules of NN304 insulin has a 14-carbon fatty acid joined by a peptide linkage to the NZ atom of B29 Lys. In the crystal structure the fatty acids are involved in the crystal contacts between adjacent hexamers, the aliphatic side chains radiating out from the crystallographic 3-fold axis toward the solvent continuum (Figure 2). The binding site for each of the four fatty acid side chains (Figure 3) is very similar and consists of the following residues (with closest approach to the fatty acid in parentheses): B1 Phe (3.6 Å) and B4 Gln (3.3 Å) of one dimer; A3 Val (4.1 Å), A4 Glu (4.4 Å), A7 Cys (3.6 Å), A8 Thr (3.7 Å), B1 Phe (4.0 Å), B3 Asn (3.6 Å), B4 Gln (3.9 Å), B7 Cys (4.3 Å), and B29 Lys (4.7 Å) from a dimer in the adjacent hexamer. A fatty acid side chain from the adjacent hexamer also forms part of the binding site. The principal noncovalent interactions between the fatty acid side chain and the protein are van der Waals contacts between the central nonpolar atoms of the fatty acid and the side chain of B1 Phe from the same insulin molecule. In fact, each pair of fatty acids forms a compact hydrophobic bundle with two B1 Phe side chains from two different hexamers. In total there are six fatty acids and six B1 Phe side chains involved in the hydrophobic interactions between two adjacent hexamers. A cluster of four water molecules fills the void between the fatty acids at the hexamer−hexamer interface, but otherwise the fatty acid side chains have very little contact with water molecules. The only significant interaction is a 2.4 Å hydrogen bond between the carbonyl oxygen of the fatty acid attached to molecule 2 and a water molecule. This water in turn hydrogen-bonds to a nearby nitrogen atom of B1 Phe (Figure 3).

The electron density for the fatty acids shows that two of the four in the asymmetric unit are well-ordered in their binding sites and have atomic *B*-factors ranging from 26 $Å^2$ (end nearest 3-fold axis) to 45 $Å^2$ (end bound to B29 Lys), as shown in Figure 4. However, the fatty acid covalently bound to molecule 1 is ordered only as far as halfway along

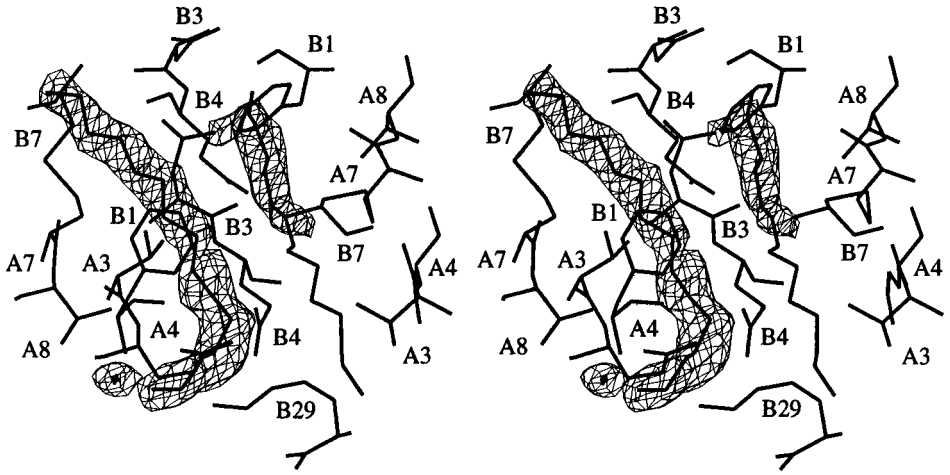

FIGURE 4: Stereoview of a $2F_o - F_c$ electron density map contoured at a level of just below 1σ, showing the variation in map quality for the fully ordered fatty acid side chain attached to molecule 2 and the disordered fatty acid side chain attached to molecule 3. Residues which constitute the fatty acid binding site include A3 Val, A4 Glu, A7 Cys, A8 Thr, B1 Phe, B3 Asn, B4 Gln, B7 Cys, and B29 Lys.

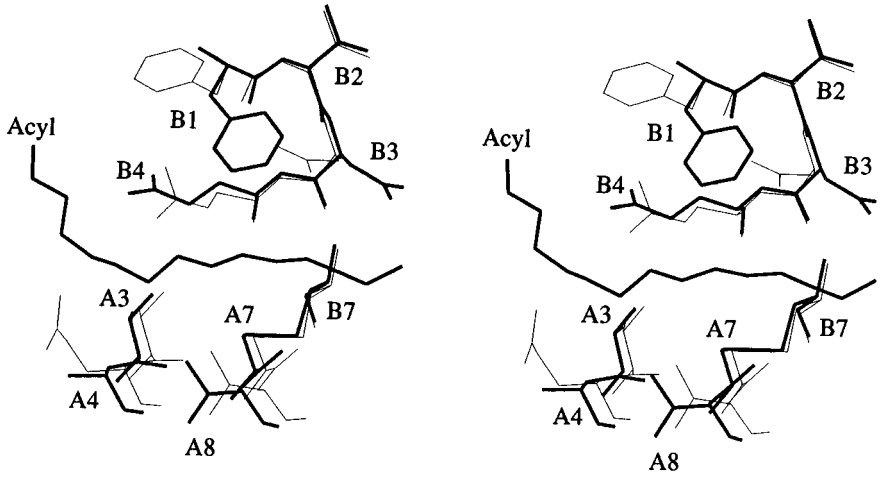

FIGURE 5: Stereoview of a superposition of molecule 1 of NN304 insulin (heavy lines) with the equivalent molecule of monoclinic phenol insulin (light lines) (Derewenda et al., 1989) made by least-squares fitting the atoms of residues B9 Ser−B19 Cys of both molecules. The fatty acid is shown in a heavy line with the acylated end labeled. Most of the site residues are displaced by only 0.1−0.2 Å; however, the side chain of B1 Phe undergoes a more significant conformational change, and consequently side chains of B3 Asn and B4 Gln are displaced. In this fatty acid binding site the side chain of residue A4 Glu is disordered and has been truncated in the figure.

from the aliphatic end (*B* factors ranging from 24−30 Å^2), and the remaining fatty acid, joined to molecule 3 has only a small amount of associated electron density (Figure 4). Consequently, the fatty acid side chain of molecule 3 was assigned an occupancy of 0.5 and its refined B-factors are in the region of 18−20 Å^2. This fatty acid presumably has an alternative conformations in the solvated regions between the hexamers. For all of the fatty acids the covalently linked B29 Lys residues are disordered. Furthermore, in molecule 3 the position of the side chains of B3 Asn is different from those of equivalent side chains in the other binding sites, such that it interferes with the fatty acid side chain. This caused some confusion during the refinement until it was decided that this side chain, as well as its associated fatty acid, is disordered.

DISCUSSION

While lipid modifications to proteins occur routinely in natural systems, NN304 insulin is the first example of a protein with an engineered fatty acid linkage. This ability to link the chemistry of lipids and proteins is opening up new possibilities in macromolecular interactions, with useful applications in drug therapy. The fatty acid modification in NN304 insulin has been introduced in order to facilitate the temporary binding of insulin to circulating albumin. The lipid has been placed in an ideal position at the end of the B chain, where it neither disrupts the aggregation properties of the molecule nor abrogates its activity, which is no different than that of human insulin (Markussen et al., 1996). It is clear from the X-ray analysis that the position of the fatty acid makes it available on the surface of the monomer for interaction with albumin. Albumin is a flexible protein and versatile in the types of ligands it binds, transporting synthetic drugs such as acetylsalicylic acid, as well as many naturally occurring ions and small organic molecules for which it has specific binding sites (Peters, 1985). Hence, it is capable of interacting with a protein molecule such as the fatty acid acylated insulin. The capacity of albumin for binding acylated insulin (NN304) exceeds 5 monomers (Kurtzhals et al., 1995).

Reminiscent of many protein structures containing lipid molecules, the fatty acid side chains in crystalline NN304 insulin occupy a hydrophobic environment in between the hexamers. Along the crystallographic *c* axis adjacent hexamers are staggered in order to optimize the burial of the fatty acids, and there are small alterations in some side chain conformations around the fatty acids, particularly B1 Phe (Figure 5). Otherwise the structure of the hexamer is not

seriously affected by the binding of the fatty acids, such that the rms difference between the main-chain atoms of NN304 insulin hexamer and native monoclinic R6 phenol insulin hexamer is 0.6 Å excluding residues B1 and B28−B30. It has been seen that phenol is an essential component of the crystallization medium. The crystal structure shows that this is because phenol promotes the T → R transformation, thereby placing the side chains of residues B1 Phe on the outside of the hexamer available for interactions with the fatty acids.

The important crystal contacts between the fatty acids and the side chains of residues B1 Phe are likely to constitute a considerable driving force in the crystallization of the protein. It is clear that there is mutual stabilization of fatty acids and the phenlyalanine side chains, since in other R6 insulin hexamer structures these side chains are often disordered (Derewenda et al., 1989; Smith & Dodson, 1992). Furthermore, a comparison of the NN304 insulin hexamer with that of an isomorphous phenol-containing insulin (Smith & Dodson, 1992) suggests that there might be a significant entropy gain when the fatty acids bind during crystal formation, owing to the displacement of several water molecules which in the native rhombohedral R6 insulin hexamer structure mediates crystal contacts at the hexamer−hexamer interface. It is difficult to estimate a precise number of water molecules displaced by the fatty acids, first because the resolutions of the two structures are significantly different (the native structure has data to 2.5 Å) and second because the crystals of NN304 insulin were soaked in a cryosolvent which will have altered their water structure relative to the native form. This is evident from a comparison of the length of the crystallographic *c* axes in the two structures. In the native rhombohedral phenol insulin structure the cell axis length is 40.39 Å while in NN304 insulin it is 79.20 Å, which is slightly less than twice that of the native structure. Hence, it appears that the fatty acid insulin structure is dehydrated relative to the native form. With the 14-carbon fatty acid, as opposed to any other length, the largest entropic gain is achieved since a longer fatty acid would not fit in the structure. Even so, the disorder of one of the fatty acids perhaps reflects that the system is too strained to bind all the fatty acids satisfactorily between the hexamers.

Finally, in considering NN304 insulin as a therapeutic molecule, this study, as well as solution experiments, shows that the addition of a myristoyl side chain at the B-chain C-terminus does not alter the aggregation properties of the molecule, an important result since insulin preparations are generally more stable as hexamers. In the crystal structure the fatty acid achieves a mode of binding to satisfy its hydrophobic requirements, as it does when binding to albumin; it now remains to be seen exactly how albumin accommodates the fatty acid-acylated insulin.

ACKNOWLEDGMENT

We thank Judy Clarkson for assistance with crystallizations, and Richard Tyrrell for collecting and processing the 1.8 Å data set at the EMBL Hamburg Outstation, based at DESY, Hamburg, Germany. Tom Oldfield and David Edwards gave advice on the XFIT modeling program, and Garib Murshudov and Eleanor Dodson provided support during trials of the REFMAC program.

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 224*, 491−495.

Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Crowfoot Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N., & Vijayan, N. M. (1988) *Philos. Trans. R. Soc. (London) 319*, 369−456.

Brünger, A. T. (1993a) *Acta Crystallogr. D49*, 24−36.

Brünger, A. T. (1993b) X-PLOR, version 3.1. Yale University, New Haven, CT.

Carter, D. C., & Ho, J.-X. (1994) *Adv. Protein Chem. 45*, 153−203.

Ciszak, E., Beals, J. M., Frank, B. H., Baker, J. C., Carter, N. D., & Smith, G. D. (1995) *Structure 3*, 615−622.

Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D50*, 760−763.

Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Smith, G. D., Sparks, C., & Swenson, D. (1989) *Nature 338*, 594−596.

Dodson, E., Kleywegt, G. J., & Wilson, K. (1996) *Acta Crystallogr. D52*, 228−234.

Hallas-Møller, K. (1956) *Diabetes 5*, 7−14.

Hu, Y.-L., Wu, B.-M., Zhang, Y., Wang, D.-C., Kaarsholm, N. C., & Norris, R. (1992) *Chin. Sci. Bull., 37*, 1390−1393.

Kaarsholm, N. C., Hui-Chong K., & Dunn, M. F. (1989) *Biochemistry 28*, 4427−4435.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Krayenbühl, C., & Rosenberg, T. (1946) *Rep. Steno. Mem. Hosp. Nord. Insulinlab 1*, 60−73.

Kurtzhals, P., Havelund, S., Jonassen, I., Kiehr, B., Larsen, U. D., Ribel, U., & Markussen, J. (1995) *Biochem. J. 312*, 725−731.

Kurtzhals, P., Havelund, S., Jonassen, I., Kiehr, B., Ribel, U., & Markussen, J. (1996) *J. Pharm. Sci. 85*, 304−308.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.

Markussen, J., Diers, I., Engesgaard, A., Hansen, M. T., Hougaard, P., Langkjaer, L., Norris, K., Riebel, U., Sørensen, A. R., Sørensen, E., & Voigt, H. O. (1987) *Protein Eng. 1*, 215−223.

Markussen, J., Diers, I., Hougaard, P., Langkjaer, L., Norris, K., Snel, L., Sørensen, E., & Voigt, H. O. (1988) *Protein Eng. 2*, 157−166.

Markussen, J., Havelund, S., Kurtzhals, P., Andersen, A. S., Halstrøm, J., Hasselager, E., Larsen, U. D., Ribel, U., Schäffer, L., Vad, K., & Jonassen, I. (1996) *Diabetologia 39*, 281−288.

Murshudov, G. N., Dodson, E. J., & Vagin, A. A. (1996) in *Proceedings of the CCP4 Study Weekend*, pp 93−104, Daresbury Laboratory, Warrington, England.

Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.

Obermeier, R., Geier, R., & Grau, U. (1986) Ger. Offen. DE 3333640.

Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend*, pp 56−62, Daresbury Laboratory, Warrington, England.

Peters, T. (1995) *Adv. Protein Chem. 37*, 161−246.

Scott, D. A., & Fisher, A. M. (1936) *J. Pharmacol. Exp. Ther. 58*, 78−92.

Smith, G. D., & Dodson, G. G. (1992) *Biopolymers 32*, 441−445.

BI9625105