High-Resolution Structure of an Engineered Biologically Potent Insulin Monomer, B16 Tyr → His, As Determined by Nuclear Magnetic Resonance Spectroscopy

Svend Ludvigsen,* Melinda Roy,‡ Henning Thøgersen, and Niels C. Kaarsholm

Novo Research Institute, Novo Nordisk A/S, Novo Allé, DK-2880 Bagsværd, Denmark

Received November 30, 1993; Revised Manuscript Received March 28, 1994®

ABSTRACT: Site-directed mutagenesis is used in conjunction with ^{1}H nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy in order to find an insulin species amenable for structure determination in aqueous solution by NMR spectroscopy. A successful candidate in this respect, i.e., B16 Tyr → His mutant insulin, is identified and selected for detailed characterization by two-dimensional ^{1}H NMR. This mutant species retains 43% biological potency and native folding stability, but in contrast to human insulin it remains monomeric at millimolar concentration in aqueous solution at pH 2.4. The resulting homogeneous sample allows high-quality 2D NMR spectra to be recorded. The NMR studies result in an almost complete assignment of the ^{1}H resonance signals as well as identification of NOE cross peaks. NOE-derived distance restraints in conjunction with torsion restraints based on measured coupling constants, $^3J_{H^NH^\alpha}$, are used for structure calculations using the hybrid method of distance geometry and simulated annealing. The calculated structures show that the major part of the insulin monomer is structurally well-defined with an average rms deviation between the 20 calculated structures and the mean coordinates of 0.89 Å for all backbone atoms, 0.46 Å for backbone atoms (A2–A19 and B4–B28), and 1.30 Å for all heavy atoms. The structure of the A-chain is composed of two helices from A2 to A7 and from A12 to A19 connected by a short extended strand. The B-chain consists of a loop, B1–B8, an α-helix, B9–B19, a β-turn, B20–B23, and an extended strand from B24 to B30. Residues A1–A6, B1–B8, and B28–B30 are less well defined than the rest of the polypeptide. The solution structure resembles structures in crystals, in particular, molecule 1 of T_6 pig insulin.

Insulin is a small protein hormone consisting of two polypeptide chains, the A-chain (21 residues) and the B-chain (30 residues), which contain two interchain (A7–B7 and A20–B19) cystine linkages and one intrachain (A6–A11) cystine linkage. The three-dimensional structure of the crystalline zinc insulin hexamer has been extensively studied by X-ray crystallography (Adams et al., 1969; Smith et al., 1984; Derewenda et al., 1989; Bentley et al., 1992). These structures have been designated T_6 (formerly 2Zn), T_3R_3 (formerly 4Zn), and R_6 (formerly 2Zn phenol-induced) species as proposed by Kaarsholm et al. (1989), but also a Zn-free hexamer and a dimer of DPI[1] (Bi et al., 1983) have been described. In solution insulin exhibits a complex pattern of aggregation and precipitation dependent on protein concentration, pH, temperature, metal ions, ionic strength, and solvent composition; see, e.g., Blundell et al. (1972) and Baker et al. (1988) for reviews of early literature. These properties complicate high-resolution NMR studies of the 5808-Da monomer, the physiologically active species. In the millimolar concentration range in aqueous solution, insulin dimers and higher oligomers predominate at pH 2.0–3.5. As upscale pH titration proceeds, aggregation increases, which in turn leads to precipitation in the pH range between 4.2 and 6.6. Above pH 6.6 insulin dissolves, presumably as a mixture of dimers and higher aggregates which then gradually dissociate as the pH increases. As the pH is raised above 10–11, the monomeric state is finally reached [see Mark et al. (1987), Kaarsholm et al. (1990), and references cited therein].

In the past few years, several NMR studies have been performed wherein insulin aggregation is modified at the acidic side of the precipitation zone. These studies employ either modified insulins, organic cosolvents, or both (Boelens et al., 1990; Kline & Justice, 1990; Knegtel et al., 1991; Hua & Weiss, 1991a,b; Hua et al., 1991, 1992, 1993; Kristensen et al., 1991; Jørgensen et al., 1992). A general result of these studies has been solution structures that are rather poorly defined either as a result of lack of NOEs due to broad resonance lines (reflecting, e.g., exchange between monomers and dimers, the nature of the organic cosolvent, etc.) or simply because of the inherent dynamic nature of the insulin molecule. Indeed, the apparent lack of structural detail has recently been proposed to be important for insulin biological activity (Hua et al., 1991, 1992, 1993).

In order to examine the origins of apparent lack of structural definition, we designed and studied an insulin mutant that remains monomeric in millimolar concentration in aqueous solution at low pH while maintaining near-native biological activity and folding stability. We here describe a structure determination using NMR for one such mutant, i.e., B16 Tyr → His insulin. This species retains 43% biological potency, exhibits native folding stability, and is monomeric at pH 2.5 at concentrations up to 8 mM. It is shown herein that the major part of the molecule is structurally well-defined as evidenced by the high number of sequential NOEs and indeed many long-range NOEs. In particular, the C-terminal part of the B-chain is shown to have a well-defined set of sequential NOEs as well as many NOEs to the main body of the molecule, with the only exception being the last two residues, B29 and B30, which are disordered.

‡ Present address: Control Division, The Upjohn Co., Kalamazoo, MI 49001.

® Abstract published in *Advance ACS Abstracts,* June 1, 1994.

[1] Abbreviations: NMR, nuclear magnetic resonance; CD, circular dichroism; COSY, two-dimensional correlated spectroscopy; DQF, double quantum filtered; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional nuclear Overhauser enhanced spectroscopy; TOCSY, two-dimensional total correlation spectroscopy; DPI, des-(B26–B30) pentapeptide insulin (native insulin lacking residues B26–B30); rms, root mean square.

0006-2960/94/0433-7998$04.50/0

MATERIALS AND METHODS

Materials. Native and mutant insulins were constructed by oligonucleotide-directed mutagenesis, fermented in yeast, and purified as described (Markussen et al., 1987; Brange et al., 1988).

CD Spectroscopy. CD spectra were recorded with a Jobin Yvon Mark V dichrograph calibrated with (+)-10-camphorsulfonic acid. Near-UV CD spectra were recorded between 250 and 350 nm using an appropriate combination of cell path length and protein concentration so that the absorbance is less than 1. Protein concentrations were determined by UV absorbance using $\epsilon_{276} = 6.2 \times 10^3$ M^{-1} cm^{-1}. The same extinction coefficient was used for estimation of concentration of mutant species with the assumption that each of the four tyrosines in human insulin contributes 25% to ϵ_{276}. Guanidine hydrochloride induced reversible unfolding of native and mutant insulins was studied by far-UV CD detection as described (Kaarsholm et al., 1993).

NMR Spectroscopy. Samples for NMR were prepared by dissolving the lyophilized protein powder in 10/90 D_2O/H_2O or 99.8% D_2O and adjusting the pH as desired by addition of small volumes of 1 M DCl or NaOD. All pH meter readings are without correction for isotope effects. Samples of B16 Tyr → His insulin for two-dimensional NMR spectra were prepared at concentrations of 2–5 mM at pH 2.4, whereas samples for one-dimensional NMR were prepared at 1 mM concentration. The temperature used was 302 K, and only freshly prepared samples were used (not older than 6 days) since deamidation of AsnA21 severely limits the lifetime of the samples at this pH.

Two-dimensional ^1H–^1H NMR spectra, DQF-COSY (Piantini et al., 1982; Rance et al., 1983), TOCSY (Braunschweiler & Ernst, 1983; Bax & Davis, 1985), and NOESY (Jeener et al., 1979; Anil-Kumar et al., 1980, 1981) were recorded at 600 MHz on a Bruker AMX 600. For each phase-sensitive two-dimensional NMR spectrum 1024 t_1 increments were acquired each with 2048 or 4096 real data points, using the time-proportional phase incrementation scheme (TPPI) (Marion & Wüthrich, 1983). Spectral widths of 8024 Hz in both dimensions were used, with the carrier placed exactly on the water resonance which was attenuated by using saturation between scans for 1.5 s. For TOCSY spectra mixing times between 30 and 80 ms were used and for NOESY mixing times between 50 and 200 ms.

The processing of the two-dimensional NMR spectra was performed on a SGI Personal Iris 4D/35 using the processing package MNMR (PRONTO Software Development and Distribution, Copenhagen, Denmark). Prior to Fourier transformation window functions were applied, sine bells shifted $\pi/4$ in t_2 and squared sine bells shifted $\pi/2.6$ in t_1, and zero filling was performed once in each dimension. Baseline corrections were applied if necessary using the algorithm by Dietrich et al. (1991). Processing of one-dimensional NMR spectra was performed using the X32 computer and BRUKER software.

The spectral assignment, cross-peak integration, sequence-specific assignment, stereospecific assignment, and all other bookkeeping were performed using the program PRONTO (PRONTO Software Development and Distribution, Copenhagen Denmark) (Kjær et al., 1991). Chemical shifts are measured in parts per million as observed relative to dioxane (3.75 ppm). The applied algorithm for measurements of $^3J_{H^NH^\alpha}$ coupling constants built into the PRONTO software package uses a combined analysis of NOESY and COSY peaks (Ludvigsen et al., 1991). Stereospecific assignments for side-chain protons were obtained by identifying one of three staggered rotamer conformations using NOESY and a high ω_2 resolved (1.96 Hz/point) COSY (Hyberts et al., 1987). Amide protons observed in spectra of freshly prepared samples of B16 Tyr → His in 99.8% D_2O were characterized as having slow exchange.

Structure Calculations. Distance restraints for the subsequent structure calculation were obtained from integrated NOESY cross peaks classified as either weak, medium, or strong, corresponding to upper distance restraints of 5.5, 3.3, and 2.7 Å, respectively. For distance restraints involving methyl groups, an additional 0.5 Å was added to the upper limit (Wagner et al., 1985). ϕ-angle restraints were obtained from $^3J_{H^NH^\alpha}$ coupling constants: $-60° \pm 30°$ (2–4 Hz), $-70° \pm 30°$ (4–6 Hz), $-120° \pm 60°$ (6–8 Hz), and $-120° \pm 35°$ (8–10 Hz). Structure calculations were performed using the hybrid method combining distance geometry (Crippen & Havel, 1988; Kuszewski et al., 1992) and simulated annealing based on the ideals of Nilges et al. (1988) using X-PLOR 3.0 (Brünger, 1992) according to the examples given by the X-PLOR manual (dg_sub_embed.inp, dgsa.inp, refine.inp). The used parameter file "parallhdg.pro" in X-PLOR 3.0 was expanded by one improper angle entry for N-terminal glycine residues.

RESULTS

Design of Monomeric Insulin Suitable for 2D NMR at Low pH. In solution, the aggregation of metal-free insulin into dimers and higher oligomers is known to be a complex function of protein concentration, pH, ionic strength, temperature, and solvent composition (Jeffrey & Coates, 1966; Pekar & Frank, 1972; Lord et al., 1973; Goldman & Carpenter, 1974; Jeffrey et al., 1976; Pocker & Biswas, 1981, Strazza et al., 1985; Mark et al., 1987; Roy et al., 1990; Kadima et al., 1992). Accordingly, the crystal structure of the zinc insulin hexamer shows that distinct protein–protein interactions are involved in insulin dimer and hexamer packing (Baker et al., 1988). While the published equilibrium constants for aggregation along the monomer/monomer interface are similar at pH 2 and 7 (i.e., $\approx 10^5$ M^{-1}), the corresponding aggregation along the dimer/dimer interface appears much weaker at pH 2 than at pH 7 (Mark et al., 1987). Hence, at the acidic side of the precipitation zone, destabilization of the dimer interface was employed in search of an insulin which remains monomeric at millimolar concentration in aqueous solution.

Aggregation properties of various insulin mutants were evaluated by near-UV circular dichroism (CD) and NMR resonance line widths. The insulin CD spectrum in the 350–250-nm range reflects the environment of the tyrosine chromophore with unknown contributions from the disulfide linkages. The signal is very sensitive to aggregation and may be used to monitor the formation of the insulin dimer interface (Morris et al., 1968; Goldman & Carpenter, 1974; Wood et al., 1975; Strickland & Mercola, 1976). NMR spectra are sensitive to insulin aggregation because longer rotational correlation times lead to line broadening and because dynamic equilibria between oligomeric states can lead to line broadening in the intermediate-exchange regime.

Figure 1 shows the aromatic and amide proton resonances of the 1D ^1H-NMR spectra for human insulin (panel A) and selected mutants (panels B–D) at 1 mM protein in 10/90 D_2O/H_2O, pH 2.4. Under these conditions human insulin is mainly dimeric with a significant contribution from higher oligomers. The figure illustrates the effect of destabilizing the dimer interface relative to native human insulin: (B)

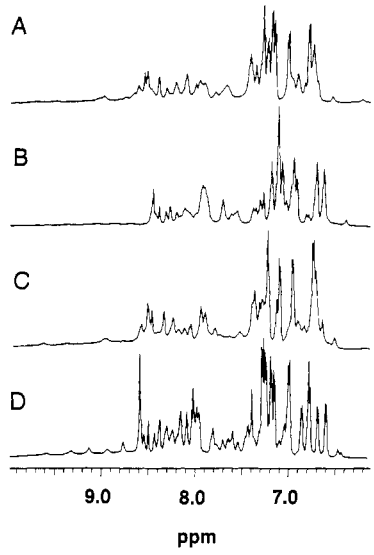

FIGURE 1: 1D NMR spectra of human insulin and three insulin mutants showing the obtained increased resolution. Panels A, native human insulin; B, human insulin with B26–B30 deletion and amidation of B25; C, human insulin with B25 Phe → His mutation; and D, B16 Tyr → His human insulin. All spectra are recorded at pH 2.2–2.3 in 10/90 D_2O/H_2O solution at 302 K.

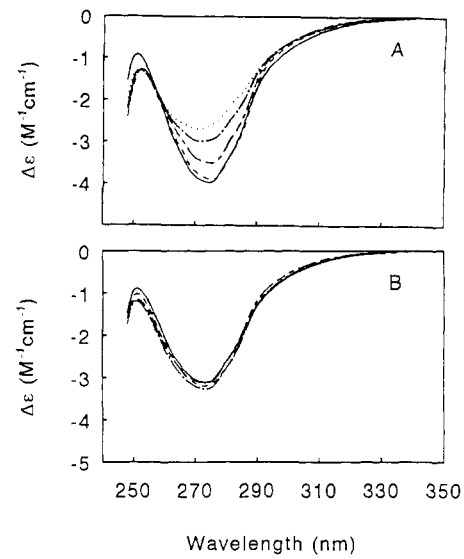

FIGURE 2: Near-UV CD spectra as a function of protein concentration in unbuffered aqueous solution, pH 2.4. Panel A: human insulin at (···) 0.004, (– · –) 0.02, (– – –) 0.2, (- - -) 2.0, and (—) 6 mM. Panel B: B16 Tyr → His insulin at (– · –) 0.02, (– – –) 0.2, (- - -) 2.0, and (—) 8 mM.

deletion of the B26–B30 peptide segment and amidation of the C-terminal PheB25 residue, i.e., des-(B26–B30) B25 Phe-NH_2 insulin; (C) introduction of positive charge central in the nonpolar dimer forming interface, i.e., B25 Phe → His insulin; and (D) introduction of positive charge closer to the edge of this interface, i.e., B16 Tyr → His insulin. As shown in Figure 1, the three modifications (B–D) lead to progressive enhancement of the resolution of the ^{1}H-NMR spectrum while retaining the dispersion of resonance lines characteristic for globular proteins. Surprisingly, however, the improvement of spectral resolution is clearly better for the B16 Tyr → His species than for the others. This result suggests that both the B25 Phe → His and the des-(B26–B30) B25 Phe-NH_2 species exhibit broad lines due to intermediate exchange between monomeric and dimeric states, while the B16 Tyr → His species is monomeric under these conditions since line widths are reasonable compared to the size of the molecule. Accordingly, the number of observed cross peaks in a 2D NOESY spectrum of HisB16 insulin is much higher than is observed in the corresponding spectra of either HisB25 or des-(B26–B30) B25 Phe-NH_2 insulins (data not shown). The crystal structure of the T_6 insulin hexamer (Baker et al., 1988) suggests two different explanations for the insufficient dissociation observed for the des-(B26–B30) B25 Phe-NH_2 and the B25 Phe → His insulins. In the first case part of the native dimer interface is removed; however, the underlying nonpolar surface (IleA2, LeuB15, etc.) is simultaneously exposed as a starting point for additional nonspecific aggregation. In the second case, the insulin hexamer structure in the crystal shows that the PheB25 side chain can exist in two distinct conformations. In molecule 1, it turns toward the A-chain touching TyrA19, while in molecule 2 it turns away from the core of the molecule. This flexibility together with the inherent thermodynamic drive for dimer formation appears sufficient to accommodate a protonated His residue in the monomer–monomer interface at low pH (Kaarsholm et al., 1990).

Figure 2 shows near-UV CD spectra as a function of protein concentration for native insulin and the B16 Tyr → His mutant at pH 2.4. For human insulin, the progressive increase in intensity of the negative signal around 274 nm indicates that association increases with the protein concentration as expected. In contrast, the near-UV CD of B16 Tyr → His insulin is independent of the protein concentration in the 20 μM to 8 mM range. Although this mutant is missing one tyrosine reporter group, TyrB26 is expected to be strongly affected by monomer–monomer interactions. Hence, these results indicate that the mutant remains monomeric under these conditions. Furthermore, far-UV CD studies of the guanidine hydrochloride induced reversible unfolding showed that the folding stability of B16 Tyr → His insulin is indistinguishable from that of the native species at pH 2.4 (data not shown). Finally, the ability of B16 Tyr → His insulin to enhance the incorporation of [2-^{3}H]glucose into lipids in isolated mouse adipocytes was studied according to Moody et al. (1974) and showed that this insulin species exhibits 43% the biological potency of native human insulin. On the basis of these results the B16 Tyr → His mutant was identified as a relevant species for solution structural studies.

Resonance Assignments. The resonance lines in the ^{1}H-NMR spectra of B16 Tyr → His insulin are mainly characterized by having a good dispersion characteristic for globular proteins. Resonance assignments (Table 1) and spin system identification were performed using the two-dimensional spectra DQF-COSY and TOCSY, occasionally supported by the NOESY spectra. The assignment strategy employed was essentially that outlined by Wüthrich and coworkers (Wüthrich, 1986). The fingerprint region of the DQF-COSY spectrum (Figure 3) exhibits cross peaks of all expected H^N–H^α correlations except those of GlnA5, CysA11, and the N-terminal amides of GlyA1 and PheB1. The H^N–H^α cross peak of GlnA5, however, is observable in spectra with lower contour levels. The resonance position of the H^N–H^α correlation of CysA11 could only be verified once all other sequential assignments were obtained. As has been reported for several other insulin mutants, the line width of the amide proton resonances sequentially close to CysA11 is extraordinary broad. For CysA11, in particular, this explains the absence of a correlation cross peak between the amide and the α-proton.

Sequential and Stereospecific Assignments. B16 Tyr → His insulin has several amino acid types that appear only once in the sequence (AlaB14, ArgB22, ProB28, and LysB29).

Table 1: ^{1}H Chemical Shifts in B16 Tyr → His Insulin at pH 2.4 and 302 K Measured in ppm As Observed Relative to Dioxane (3.75 ppm)

residue	H^N	H^α	H^β	others
GlyA1	N/A	3.97, 3.97		
IleA2	8.44	3.87	1.15	$H^{\gamma 1}$ 0.84, 1.08; $H^{\gamma 2}$ 0.66; H^δ 0.51
ValA3	8.06	3.60	1.93	$H^{\gamma 1}$ 0.84; $H^{\gamma 2}$ 0.91
GluA4	8.10	4.19	2.06, 2.06	H^γ 2.46, 2.46
GlnA5	8.19	4.04	2.06, 2.01	H^γ 2.43, 2.36; $H^{\delta 2}$ 6.84, 7.44
CysA6	8.26	4.86	*3.28, 2.79*	
CysA7	8.25	4.78	*3.26, 3.70*	
ThrA8	8.23	3.98	4.35	$H^{\gamma 2}$ 1.20
SerA9	7.33	4.73	3.99, 3.84	
IleA10	7.82	4.30	1.53	$H^{\gamma 11}$ 1.06; $H^{\gamma 12}$ 0.40; $H^{\gamma 2}$ 0.60; H^δ 0.46
CysA11	9.58	4.96	3.11, 3.24	
SerA12	8.62	4.54	*3.94, 4.23*	
LeuA13	8.59	3.76	1.34, 1.25	H^γ 1.35; H^δ 0.68, 0.71
TyrA14	7.54	4.09	2.94, 2.88	H^δ 7.01; H^ϵ 6.76
GlnA15	7.47	3.92	1.97, 2.25	H^γ 2.30, 2.38; $H^{\epsilon 2}$ 6.90, 7.43
LeuA16	7.98	4.08	*1.47, 1.86*	H^γ 1.66; $H^{\delta 1}$ 0.75; $H^{\delta 2}$ 0.70
GluA17	7.98	4.14	*2.00, 2.04*	$H^{\gamma 1}$ 2.51; $H^{\gamma 2}$ 2.29
AsnA18	7.35	4.46	2.58, 2.50	$H^{\delta 2}$ 6.49, 7.10
TyrA19	7.85	4.37	*3.33, 2.86*	H^δ 7.27; H^ϵ 6.72
CysA20	7.32	5.00	*3.22, 2.78*	
AsnA21	8.21	4.68	2.82, 2.61	$H^{\delta 2}$ 7.39, 6.49
PheB1	N/A	4.21	3.12, 3.12	H^δ 7.18; H^ϵ 7.31; H^ζ 7.21
ValB2	8.06	4.03	1.83	$H^{\gamma 1}$ 0.77; $H^{\gamma 2}$ 0.80
AsnB3	8.40	4.62	2.67, 2.67	$H^{\delta 2}$ 6.82, 7.46
GlnB4	8.35	4.39	1.82, 2.02	H^γ 2.15, 2.07; $H^{\epsilon 2}$ 7.31, 6.73
HisB5	8.59	4.40	*3.52, 3.22*	$H^{\delta 2}$ 7.29; $H^{\epsilon 1}$ 8.53
LeuB6	8.95	4.43	*0.78, 1.70*	H^γ 1.54; $H^{\delta 1}$ 0.82; $H^{\delta 2}$ 0.65
CysB7	8.30	4.95	*3.21, 2.92*	
GlyB8	9.35	3.98, 3.81		
SerB9	9.12	4.11	3.85, 3.82	
HisB10	8.00	4.47	*3.25, 3.52*	$H^{\delta 2}$ 7.42; $H^{\epsilon 1}$ 8.62
LeuB11	7.04	3.93	1.83, 1.14	H^γ 1.17; H^δ 0.61, 0.61
ValB12	7.14	3.23	2.04	H^γ 0.94, 0.94
GluB13	7.84	4.08	2.13, 2.06	H^γ 2.52, 2.52
AlaB14	7.63	4.06	1.41	
LeuB15	8.06	3.75	*0.41, 1.01*	H^γ 1.31; $H^{\delta 1}$ 0.30; $H^{\delta 2}$ 0.56
HisB16	8.12	4.49	3.35, 3.35	$H^{\delta 2}$ 7.31; $H^{\epsilon 1}$ 8.62
LeuB17	7.73	4.05	*1.82, 1.71*	H^γ 1.67; $H^{\delta 1}$ 0.85; $H^{\delta 2}$ 0.89
ValB18	8.32	3.80	1.99	$H^{\delta 1}$ 0.80; $H^{\gamma 2}$ 0.92
CysB19	8.60	4.76	*3.26, 2.86*	
GlyB20	7.65	3.96, 3.89		
GluB21	8.75	4.22	2.19, 2.06	H^γ 2.52, 2.52
ArgB22	8.02	4.11	1.96, 1.96	H^γ 1.76, 1.76; H^δ 3.25, 3.25; H^ϵ 7.03
GlyB23	7.44	4.01, 3.74		
PheB24	7.58	4.97	*3.21, 2.87*	H^δ 6.81; H^ϵ 7.03; H^ζ 7.07
PheB25	8.39	4.76	*3.14, 3.11*	H^δ 7.21; H^ϵ 7.26
TyrB26	8.03	4.57	*2.81, 2.95*	H^δ 6.90; H^ϵ 6.64
ThrB27	7.68	4.52	4.00	$H^{\gamma 2}$ 1.13
ProB28		4.24	*1.87, 2.16*	$H^{\gamma 1}$ 1.84; $H^{\gamma 2}$ 1.94; $H^{\delta 1}$ 3.44; $H^{\delta 2}$ 3.61
LysB29	8.17	4.30	1.81, 1.72	H^γ 1.41, 1.41; H^δ 1.64, 1.64; H^ϵ 2.93, 2.93; H^ζ 7.42
ThrB30	8.04	4.37	4.33	$H^{\gamma 2}$ 1.13

[a] Chemical shifts of H^βs in italic type indicate stereospecific assignments listing $H^{\beta 1}$ first.

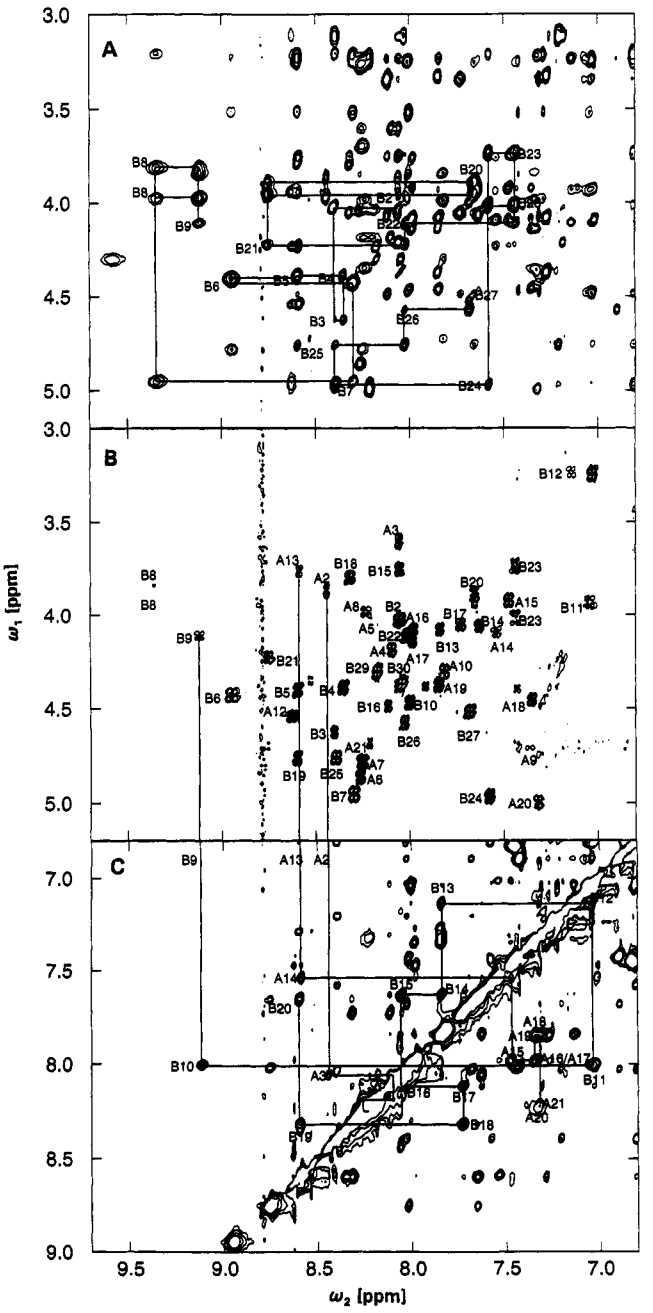

FIGURE 3: Regions of NOESY and DQF-COSY spectra at 600 MHz of B16 Tyr → His insulin dissolved in 10/90 D_2O/H_2O at 302 K. The upper spectrum (panel A) shows the fingerprint region in NOESY with selected sequential assignments; residue numbers follow the chain name. The middle spectrum (panel B) shows the same region of the DQF-COSY spectrum. The broad amide resonance of CysA11 (9.58 ppm) is absent in the DQF-COSY spectrum. The NOESY spectrum at the bottom (panel C) shows the amide resonances characteristic of helices. The sequential pathway of the three helices is drawn by lines.

These are easily found in the NMR spectra and served as anchor points for the sequential spin system assignment. The sequential NOEs are summarized in Figure 4, and parts of the backbone sequential assignments are sketched in Figure 3. The sequential pathway from A4 to A12 was the most difficult to achieve for two main reasons: (a) amide proton resonances of residues A5–A8 are all between 8.19 and 8.26 ppm and (b) amide resonances of residues A5, A8, A9, A10, and in particular A11 are extraordinary broad, ranging from 15 to 20 Hz and almost 50 Hz for A11 [see leftmost NOESY cross peak (A10 H^α to A11 H^N) observed in Figure 3, panel A]. Additional unusually broad amide proton resonances were observed for residues B6, B8, B9, B11, and B12. Desite overlap for backbone amides of residues A16 and A17, the sequential assignments were easily confirmed using NOESY spectra recorded at similar conditions except the temperature was raised to 307 K. For several residues side-chain stereospecific assignment could be uniquely determined, indicating the conformation of these must be reasonably rigid. Of 45 prochiral C^β atoms in B16 Tyr → His insulin, stereospecific assignments were achieved for 18, most of which are buried inside the protein. The stereospecific assignments allowed a classification of 23 side chains with respect to which of the three staggered conformations were occupied (see Figure 4) by measurement of NOE intensities and comparisons of COSY cross-peak patterns with those presented by Hyberts et al. (1987). Zero quantum coherence contributions to NOEs

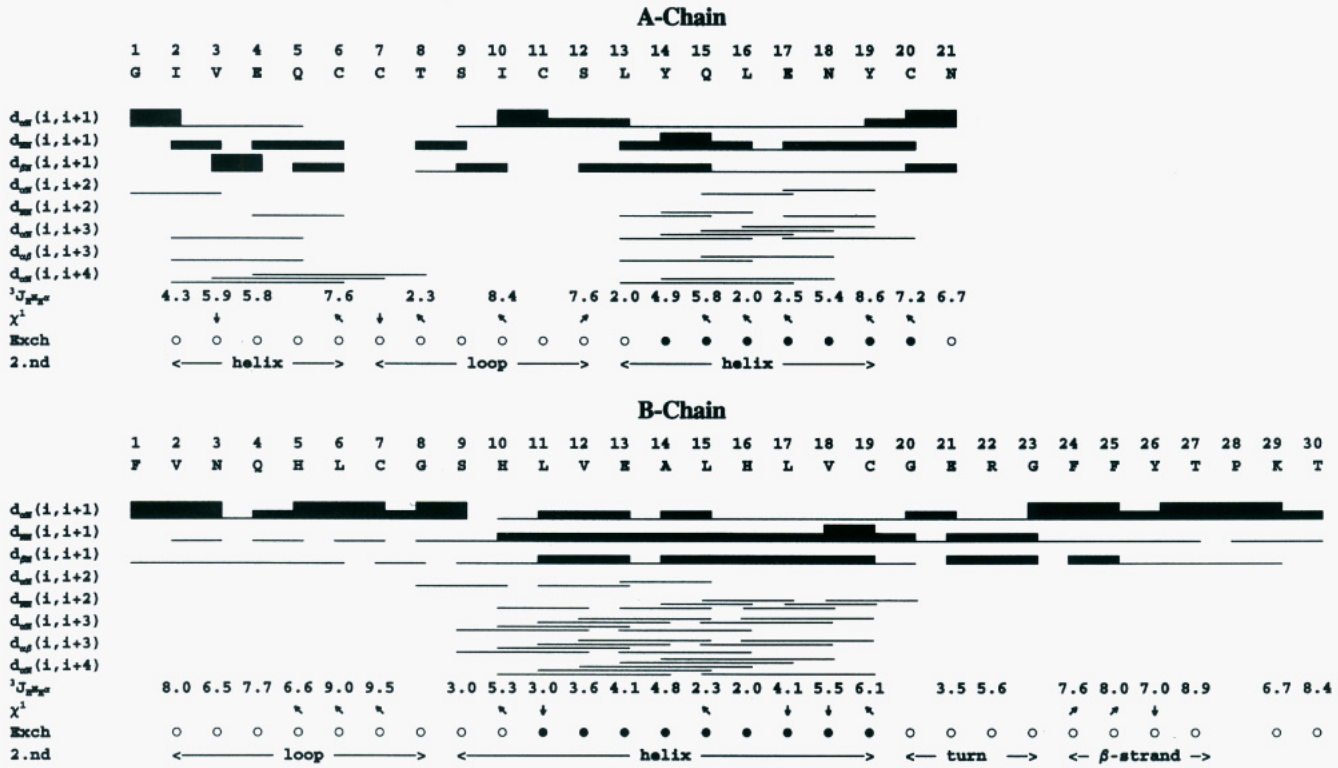

FIGURE 4: Schematic representation of the sequential NOEs assigned for B16 Tyr → His human insulin. The size of the bars indicate the measured intensity according to weak, medium, and strong. Coupling constants are measured in hertz; slowly exchanging amide protons are labeled by a filled circle and fast exchanging with a open circle. Arrows indicate which of three possible staggered conformations for side chains were found: arrows pointing northwest indicate –60° χ^1 angle; arrows pointing northeast, 60°; and arrows pointing south, 180°. For NOEs including ProB28, H^δ protons are considered instead of H^N.

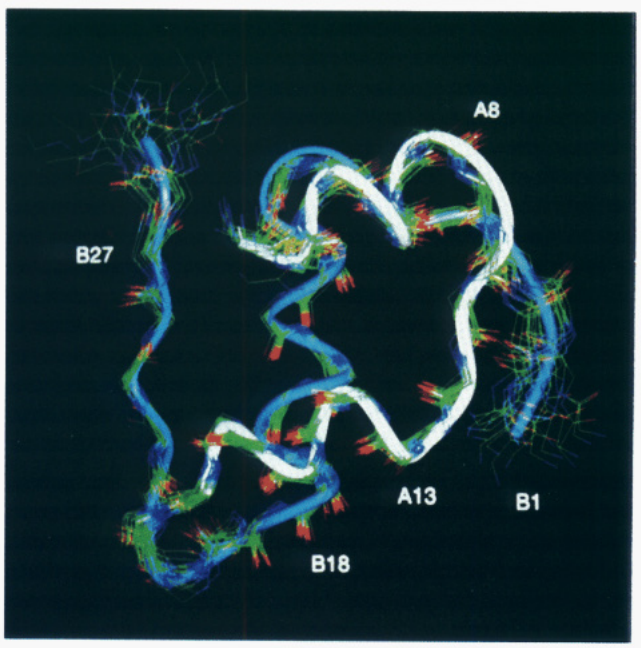

FIGURE 5: Twenty converged structures (backbone representation) colored according to atom type. The folding geometry is enhanced by a ribbon at average coordinates; the A-chain is gray and B chain light blue.

between H^α and H^β were only observed for residues B1, B21, and B29 for which no side-chain orientation nor stereospecific assignment could be achieved.

Structure Calculations. Forty structures were calculated in X-PLOR version 3.0 using distance geometry and simulated annealing, followed by 200 cycles of restrained Powell minimization. These calculations were initially performed without side-chain torsion angle restraints and then combined with further analysis of the NMR spectra in an iterative manner. Distance restraints for hydrogen bonds were not included in the calculations at any stage. On the basis of fulfillment of distance and torsion restraints, 20 converged structures (Figure 5) were selected for further analysis. The structural statistics are summarized in Table 2, and an overview of the distance restraints employed is shown in Figure 6. The NOE matrix in Figure 6 is not symmetrical because some NOEs close to the water resonance were only seen on one side of the diagonal in the NOESY spectra; however, for calculational purposes all distance restraints were only included once. The main body of the hormone (A2–A19 and B4–B28) is well-defined and described in further detail in Figure 7, illustrating the atomic rms color-coded for the average coordinates in the three-dimensional structure.

Description of Structures. The calculated structures agree well with the secondary structure predicted from the pattern of sequential NOEs and $^3J_{H^NH^\alpha}$ coupling constants described in detail in Figure 4. From A2 to A8 the sequential NOEs imply a helical structure although a true α- or 3_{10}-helix normally would have a more complete set of sequential NOEs than is seen here. However, the presence of $d_{\alpha N}(i,i+4)$ effects clearly indicate an α-helical structure despite relatively large $^3J_{H^NH^\alpha}$ for A3 and A4 and despite the absence of slowly exchanging amide protons. These findings are reflected in the structures as a poorly defined helix with ambiguous hydrogen-bonding pattern. Due to overlap and the presence of broad amide resonance lines, the list of sequential NOEs from A7 to A12 is incomplete, but this part of the structure is only a little less well defined compared to the rest of the molecule. The structure from A13 to A19 is well-defined, starting up as an α-helix but ending as a 3_{10}-helix having amide protons with slow exchange in good agreement with a helical structure. The remainder of the A-chain, TyrA19 to

Table 2: Structural Statistics for 20 B16 His Insulin Structures Obtained by DISGEO and Simulated Annealing

No. of Distance and Torsion Angle Restraints			
distances		torsion angles	
short[a] range ($\|i-j\|<5$)	325	ϕ	44
long-range intrachain	56	χ^1	23
cross chain	98	χ^2	6

Energy Statistics after Simulated Annealing (kcal/mol)						
total	bond	angle	repel	NOE[b] restraint	torsion[b] restraint	improper
35.6 ±4.0	1.8 ±0.2	21.4 ±2.0	5.8 ±1.4	4.2 ±0.9	0.1 ±0.1	2.4 ±0.4

Deviations from Ideal Geometry				
bond (Å)	angle (deg)	improper (deg)	NOE restraint (Å)	torsion restraint (deg)
0.0015 ±0.0001	0.32 ±0.01	0.19 ±0.01	0.0095 ±0.0011	0.115 ±0.068

Average No. of NOE Restraint Violations			
0.0–0.1 Å	44.6 ± 5.4	0.2–0.3 Å	0.1 ± 0.2
0.1–0.2 Å	1.6 ± 1.0	>0.3 Å	0

van der Waal Energy Measured with CHARMM[c] Potential	
vdW (kcal/mol)	−90 ± 11

Atomic rms Values for 20 Calculated Structures vs Average Coordinates			
atoms included	rms (Å)	atoms included	rms (Å)
backbone (all)	0.89 ± 0.17	backbone (B4–B28)	0.46 ± 0.07
heavy atoms (all)	1.30 ± 0.15	backbone (A2–A19 + B4–B28)	0.46 ± 0.07
backbone (A2–A19)	0.44 ± 0.12		

[a] Ranges are defined between residues i and j in the sequence within one chain. [b] Force constants for distance and torsion angle restraints are 50 kcal mol^{-1} Å^{-2} and 200 kcal mol^{-1} radian^{-2}. [c] CHARMM potential (Brooks et al., 1983) used for van der Waal energy calculation.

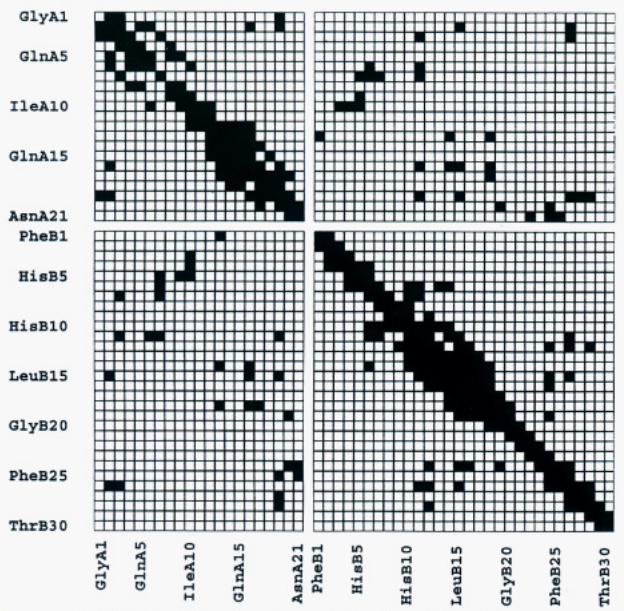

FIGURE 6: NOE matrix showing residue by residue the presence of NOEs. The matrix is nonsymmetric since some NOEs are only present in the spectra once due to partial overlap with the residual solvent signal.

AsnA21, occupies an extended structure. The first eight residues of the B-chain exhibit $^3J_{H^NH^\alpha}$ coupling constants and sequential NOEs characteristic of extended strands, but this part of the structure is not perfectly defined. After GlyB8 a typical α-helix is formed by residues B9–B19, and most backbone amides of these residues exhibit slow exchange rates, indicating the presence of regular hydrogen bonds. GlyB20 to GlyB23 forms a type I turn, and from B24 to B27 the chain

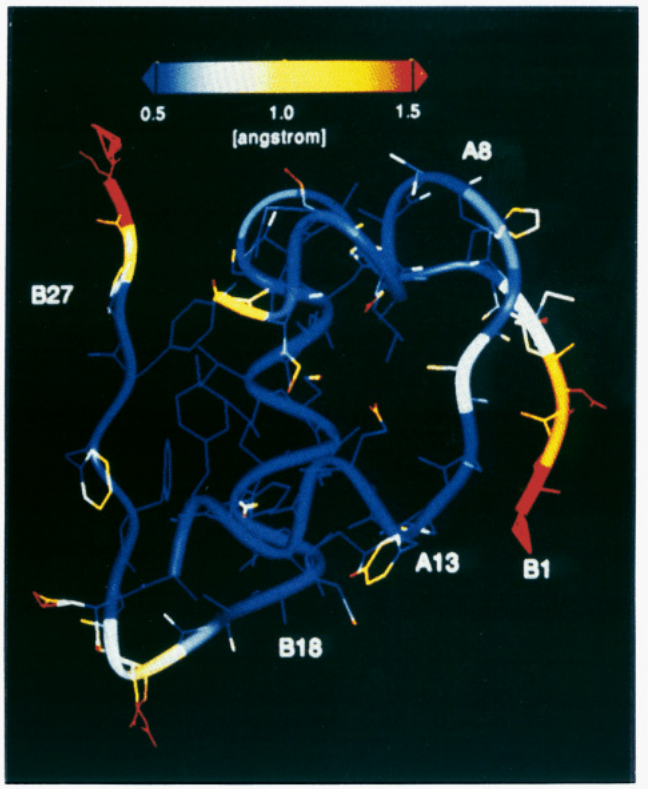

FIGURE 7: Average structure of B16 His human insulin color-coded by atomic rms values according to the color scale at the top. Side chains with high rms values are obscured since average coordinates are physically meaningless and bond connectivities are scrambled.

is in extended conformation until ProB28 for which sequential NOEs describing only the trans peptide bond configuration are observed. Finally, residues B29 and B30 are disordered in the structure.

DISCUSSION

The B16 Tyr → His mutant is the first example of a full-length insulin species that remains monomeric at millimolar concentration in aqueous solution at low pH. Previous studies have employed shortened insulins, organic cosolvents, or both to counteract protein self-association. Consequently, these studies have been met with the challenge of determining the origins of broad lines in the NMR spectra. In principle, any combination of the organic cosolvent, the overlay with insulin self-association processes, and/or the inherent dynamics of the insulin molecule(s) could give rise to broad lines in the spectra. We have circumvented these problems by the use of site-directed mutagenesis to effectively prevent insulin aggregation in the millimolar concentration range at low pH. This approach allows the direct observation of the insulin monomer in aqueous solution by conventional NMR techniques and serves as the appropriate starting point for the determination of a high-resolution three-dimensional structure of the molecule.

The list of chemical shift values (Table 1) shows several differences compared to earlier NMR studies. Most differences fall in the 0.1–0.2 ppm range, but the effect of organic solvents, differences in the mutations employed, and notably the overlay with insulin dimerization probably limit the conclusions that can be made on the basis of such differences.

The three-dimensional structure of the B16 Tyr → His mutant is well-defined by many NOE-derived distance restraints and torsion angle restraints derived from the $^3J_{H^NH^\alpha}$

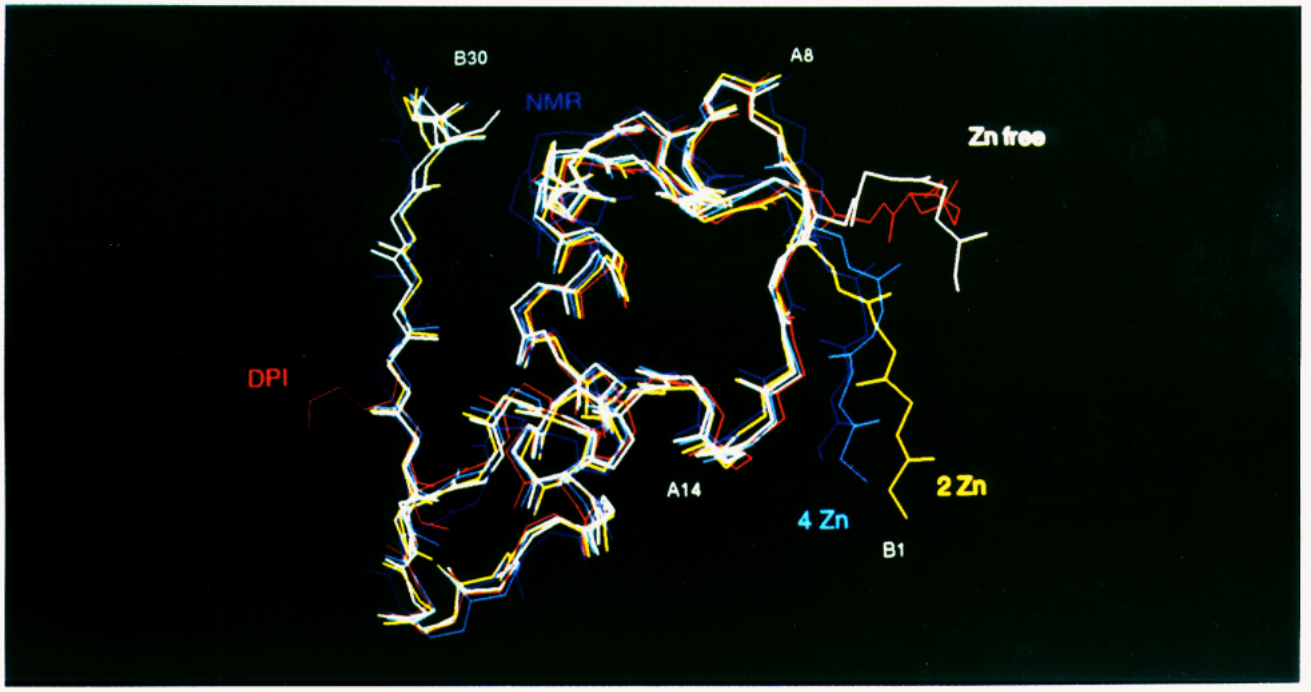

FIGURE 8: Average structure of B16 Tyr → His (blue) and crystal structures of molecule 1 of 2Zn (T_6) (green), 4Zn (T_3R_3) (light blue), Zn-free low pH (white), and beef DPI (red), respectively.

coupling constant. Herein, hydrogen bonds are only described as slowly exchanging amide protons in aqueous solution at pH 2.4, and these are only observed for amides in the helices. However, analysis of the structures indicates that the helices in the A-chain are not regular because the pattern of sequential NOEs and hence the hydrogen bonds differs from those derived from ideal α-helices. On the other hand, the central helix of the B-chain more closely resembles a regular α-helix. Conclusions concerning the geometry of the hydrogen bonds in the helices are therefore not warranted for structure calculations but should merely be based on analysis of the calculated structures.

The structure of B16 Tyr → His human insulin has a general fold equivalent to most other crystal and NMR structures reported previously, but minor significant differences clearly exist. Structures of DPI in the crystal (Bi et al., 1983) as well in solution (Knegtel et al., 1991) have been determined. A prominent difference between DPI structures in the crystal and solution is the orientation of the N-terminal segment B1–B4, which points away from the main body of the molecule in the crystal, while in solution contacts between B1 and A13 are present. Such contact is also seen in the B16 Tyr → His monomer, in the T monomer of the T_3R_3 B13Gln insulin crystal structure (Bentley et al., 1992), and to some extent in the T_6 crystal structure but clearly not in the Zn-free structure (see Figure 8). The PheB25 side-chain conformation in the B16 Tyr → His mutant agrees with molecule 1 of the crystal structures as manifested by several NOEs from B25 to A21 and A19. The side chain of TyrB26 deviates a little from the orientations seen in the crystal structures of Zn-free, T_6, and T_3R_3 insulins probably as a result of missing crystal packing forces. Similarly, the side chains of B26 and A19 are approximately 3 Å closer in the NMR structure compared to the crystal structures. Another difference in side-chain conformation is observed for TyrA14 which in solution adopts the distinct conformation observed in molecule 1 of the Zn-free and T_3R_3 crystal structures. The hydrogen bonds in the second helix of the A-chain and the helix of the B-chain are essentially the same in solution as those reported for T_6 pig insulin except for a few differences observed in the N-terminal part of both helices. The hydrogen bonds, H^N A15–CO A12 and H^N B11–CO B8, are observed both in the crystal and in solution, while the bifurcated hydrogen bonds in the crystal, H^N A16–CO (A12, A13) and H^N B12–CO (B8, B9), are absent in the solution structure where only the first of each hydrogen bond is formed. We propose that the observed slow exchange of A14 H^N is explained by a bifurcated hydrogen bond to the oxygen atom of the hydroxo group of SerA12. In the B16 Tyr → His structure cross-chain hydrogen bonds, H^N B6–CO A6, H^N A21–CO B23, and H^N B25–CO A19, are present as they are in the crystal; however, the H^N B4–CO A11 and H^N A11–CO B4 hydrogen bonds present in the crystal phase are absent in the solution structure. These two latter hydrogen bonds are formed in the crystal as a result of the crystal packing forces that push these residues close together. One of the three tyrosine side chains, i.e., A19, forms a hydrogen bond from the oxygen to the amide proton of IleA2. Note that no energy terms for hydrogen bonds nor electrostatic energy are included at any stage of the calculations. In essence molecule 1 of T_6 insulin is the crystalline structure that is most closely related to the solution structure. The stereospecific assignments of β-methylene protons and γ-methyl groups of valines are generally obtained for residue side chains that are packed inside the protein. When compared to χ^1 angles found in the T_6 pig insulin crystal structures, no deviations were found with respect to which of the three staggered conformations exist.

In the C-terminal part of the B-chain (B24–B30), B16 Tyr → His insulin exhibits prominent differences in comparison with the insulin molecules described by Hua and Weiss (1991b) and Hua et al. (1991, 1992, 1993). Residues B24–B28 behave as a single well-defined part of a single β-strand, as demonstrated by the pattern of sequential and long-range NOEs. Side chains of PheB24 and TyrB26 exhibit many long-range NOEs to the main body of the molecule, particularly to ValB12 and LeuB15. A similar observation has been made at near-neutral pH (Roy et al., 1990). We do not observe any NOEs from PheB25 to LeuB15 similar to those reported

by Hua and Weiss (1990). Our proposed orientation of the side chain of PheB25 can explain the difference in chemical shift for the H^{α} resonance of CysA20 (5.00 ppm) in B16 Tyr → His insulin compared to human insulin (4.73 ppm) in 20% acetic acid (Hua & Weiss, 1991b) and in B24 Phe → Gly insulin (4.68 ppm) in 20% acetic acid. For the ProB28 residue, our NOE data indicate the presence of the trans conformation only, while Kline and Justice (1990), working with human insulin in 35% acetonitrile, reported the detection of NOEs consistent with both cis and trans conformations. In full agreement with the structures of native insulin in the crystal and in solution, residues B29 and B30 in B16 Tyr → His insulin are disordered. Jørgensen et al. (1992) reported several NOEs for B9 Ser → Asp human insulin consistent with a dimer, but no such NOEs were observed in the present study and no overlap could bury such NOEs.

The observation of extraordinarily broad amide resonances for residues A5, A8, A9, A10, A11, B6, B8, B9, B11, and B12 is in agreement with earlier studies on human insulin (Kline & Justice, 1990) and may be ascribed to the occurrence of multiple states that exchange on a intermediate NMR time scale. We note that these broad resonances are confined to a particular region of the molecule.

Finally, it should be noted that only freshly prepared samples of B16 Tyr → His insulin (or any insulin) were used for the present NMR experiments at low pH. This is due to the well-known phenomenon of deamidation at the AsnA21 residue (Sundby, 1962) which occurs with a half-life of about 1 week at pH 2.0–2.5 and 37 °C. The result of deamidation is seen in the NMR spectra (not depicted) as an extra set of resonances which all can be traced to residues in the proximity of AsnA21. Besides the potential for destabilizing the C-terminal β-strand of the B-chain per se, partial deamidation of AsnA21 obviously leads to degraded geometric interpretations of the NOEs. It is therefore essential that the side-chain amides of A21 are present in the spectra, indicating that the correct molecule is under investigation.

CONCLUSION

Aggregation problems for insulin at millimolar concentrations in aqueous solution can be successfully avoided by a single amino acid substitution, B16 Tyr → His. This mutation has little effect on the biological potency and the folding stability of the molecule. The structure of this monomeric molecule in solution is an important prerequisite for the understanding of interactions between insulin and its receptor. The present study has shown that the major part of the molecule is structurally well-defined with atomic rms for backbone atoms (B4–B28, A2–A19) relative to average coordinates of 0.46 Å as a consequence of the large number of sequential NOEs and indeed many long-range NOEs. Amide proton resonances from residues in a small part of the molecule are unusually broad, which can be explained by some structural mobility on the NMR time scale. The overall structure in solution is mainly equivalent to those known from X-ray crystallographic studies, but small differences are observed, e.g., for residues A14, A19, and B26. Most differences are likely due to the inherent effect of crystal packing forces. Collectively, our results emphasize that a homogeneous and monomeric insulin sample is crucial for the quality of the NMR-derived structure. In particular, the structure of the C-terminal part of the B-chain, B23–B28, is well-defined due to a high number of sequential NOEs as well as many NOEs to the main body of the molecule. In agreement with X-ray structures, residues B29 and B30 are disordered in solution. Hence, it appears that high biological potency is compatible with a distinct structure of the C-terminal part of the B-chain.

ACKNOWLEDGMENT

We thank S. Havelund, I. Diers, and A.-M. Kolstrup for fermentation and purification of the B16 Tyr → His insulin; A. Blom for sample preparation and maintenance of the NMR instrument; A. R. Sørensen for biological potency measurements; M. Kjær for valuable help with the MNMR processing software; and G. Dodson and colleagues for insulin crystal structure coordinates.

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, R., & Sheet, S. (1969) *Nature 224*, 491–496.

Anil-Kumar, Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–5.

Anil-Kumar, Wagner, G., Ernst, R. R., & Wüthrich, K. (1981) *J. Am. Chem. Soc. 103*, 3654–3658.

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

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bentley, G. A., Brange, J., Derewenda, Z., Dodson, E. J., Dodson, G. G., Markussen, J., Wilkinson, A. J., Wollmer, A., & Xiao, B. (1992) *J. Mol. Biol. 228*, 1163–1176.

Bi, R. C., Dauter, Z., Dodson, E., Dodson, G., Gordiano, F., Hubbard, R., & Reynolds, C. (1983) *Proc. Indian Acad. Sci. (Chem. Sci.) 92*, 473–483.

Blundell, T., Dodson, G. G., Hodgkin, D. C., & Mercola, D. A. (1972) *Adv. Protein Chem. 26*, 279–402.

Boelens, R., Ganadu, M. L., Verheyden, P., & Kaptein, R. (1990) *Eur. J. Biochem. 191*, 147–153.

Brange, J., Ribel, U., Hansen, J. F., Dodson, G., Hansen, M. T., Havelund, S., Melberg, S. G., Norris, F., Norris, K., Snel, L., Sørensen, A. R., & Voigt, H. O. (1988) *Nature 333*, 679–682.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Brooks, B. R., Bruccoleri, R., Olafson, B., States, D., Swaninathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Brünger, A. T. (1992) *X-PLOR Manual Version 3.0*, Yale University, New Haven, CT.

Crippen, G., & Havel, T. (1988) *Distance Geometry and Molecular Conformation*, Research Studies Press, Taunton, Somerset, England.

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

Dietrich, W., Rüdel, C. H., & Neumann, M. (1991) *J. Magn. Reson. 91*, 1–11.

Goldman, J., & Carpenter, F. H. (1974) *Biochemistry 13*, 4566–4574.

Hua, Q. X., & Weiss, M. A. (1990) *Biochemistry 29*, 10545–10555.

Hua, Q. X., & Weiss, M. A. (1991a) *Biochim. Biophys. Acta 1078*, 101–110.

Hua, Q. X., & Weiss, M. A. (1991b) *Biochemistry 30*, 5505–5515.

Hua, Q. X., Shoelson, S. E., Kochoyan, M., & Weiss, M. A. (1991) *Nature 354*, 238–241.

Hua, Q. X., Shoelson, S. E., & Weiss, M. A. (1992) *Biochemistry 31*, 11940–11951.

Hua, Q. X., Shoelson, S. E., Inouye, K., & Weiss, M. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 582–586.

Hybert, S. G., Märki, W., & Wagner, G. (1987) *Eur. J. Biochem. 164*, 625–635.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Jeffrey, P. D., & Coates, J. H. (1966) *Biochemistry 5*, 489–498.
Jeffrey, P. D., Milthorpe, B. K., & Nichol, L. W. (1976) *Biochemistry 15*, 4660–4665.
Jørgensen, A. M., Kristensen, S. M., Led, J. J., & Balschmidt, P. (1992) *J. Mol. Biol. 227*, 1146–1163.
Kaarsholm, N. C., Ko, H.-C., & Dunn, M. F. (1989) *Biochemistry 28*, 4427–4435.
Kaarsholm, N. C., Havelund, S., & Hougaard, P. (1990) *Arch. Biochem. Biophys. 283*, 496–502.
Kaarsholm, N. C., Norris, K., Jørgensen, R. J., Mikkelsen, J., Ludvigsen, S., Olsen, O. H., Sørensen, A. R., & Havelund, S. (1993) *Biochemistry 32*, 10773–10778.
Kadima, W., Roy, M., Lee, R. W.-K., Kaarsholm, N. C., & Dunn, M. F. (1992) *J. Biol. Chem. 267*, 8963–8970.
Kjær, M., Andersen, K. V., Shen, H., Ludvigsen, S., Windekilde, D., Sørensen, B., & Poulsen, F. M. (1991) *NATO ASI Series* (Hoch, J. C., Redfield, C., & Poulsen, F. M., Eds.) Plenum, New York.
Kline, A. D., & Justice, R. M., Jr. (1990) *Biochemistry 29*, 2906–2913.
Knegtel, R. M. A., Boelens, R., Ganadu, M. L., & Kaptein, R. (1991) *Eur. J. Biochem. 202*, 447–458.
Kristensen, S. M., Jørgensen, A. M., Led, J. J., Balschmidt, P., & Hansen, F. (1991) *J. Mol. Biol. 218*, 221–231.
Kuszewski, J., Nilges, M., & Brünger, A. T. (1992) *J. Biomol. NMR 2*, 33–56.
Lord, R. S., Gubensek, F., & Rupley, J. A. (1973) *Biochemistry 12*, 4385–4392.
Ludvigsen, S., Andersen, K. V., & Poulsen, F. M. (1991) *J. Mol. Biol. 217*, 731–736.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 486–492.
Mark, A. E., Nichol, L. W., & Jeffrey, P. D. (1987) *Biophys. Chem. 27*, 103–117.
Markussen, J., Diers, I., Engesgaard, A., Hansen, M. T., Hougaard, P., Langkjaer, L., Norris, K., Ribel, U., Snel, L., Sørensen, A. R., Sørensen, E., & Voigt, H. O. (1987) *Protein Eng. 1*, 215–223.
Moody, A. D., Stan, M. A., Stan, M., & Gliemann, J. (1974) *Hormone Metab. Res. 6*, 12–16.
Morris, J. W. S., Mercola, D., & Arquilla, E. R. (1968) *Biochim. Biophys. Acta 160*, 145–155.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 239*, 317–324.
Pekar, A. H., & Frank, B. H. (1972) *Biochemistry 11*, 4013–4016.
Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.
Pocker, Y., & Biswas, B. (1981) *Biochemistry 20*, 4354–4361.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Roy, M., Lee, R. W. K., Brange, J., & Dunn, M. F. (1990) *J. Biol. Chem. 265*, 5448–5452.
Smith, G. D., Swenson, D. C., Dodson, E. J., Dodson, G. G., & Reynolds, C. D. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7093–7097.
Strazza, S., Hunter, R., Walker, E., & Darnall, D. W. (1985) *Arch. Biochem. Biophys. 238*, 30–42.
Strickland, E. H., & Mercola, D. (1976) *Biochemistry 15*, 3875–3884.
Sundby, F. (1962) *J. Biol. Chem. 237*, 3406–3411.
Wagner, G., Braun, W., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295–315.
Wood, S. P., Blundell, T. L., Wollmer, A., Lazarus, N. R., & Neville, R. W. J. (1975) *Eur. J. Biochem. 55*, 531–542.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.