Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2379–2383, March 1992
Biophysics

Structure and dynamics of des-pentapeptide-insulin in solution: The molten-globule hypothesis

QING-XIN HUA*†‡, MICHEL KOCHOYAN*§, AND MICHAEL A. WEISS*†¶

*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115; and ¶Department of Medicine, Massachusetts General Hospital, Boston, MA 02114

Communicated by Elkan Blout, December 16, 1991 (received for review September 4, 1991)

ABSTRACT **Structures of insulin in different crystal forms exhibit significant local and nonlocal differences, including correlated displacement of elements of secondary structure. Here we describe the solution structure and dynamics of a monomeric insulin analogue, des-pentapeptide-(B26–B30)-insulin (DPI), as determined by two-dimensional NMR spectroscopy and distance geometry/restrained molecular dynamics (DG/RMD). Although the solution structure of DPI exhibits a general similarity to its crystal structure, individual DG/RMD structures in the NMR ensemble differ by rigid-body displacements of α-helices that span the range of different crystal forms. These results suggest that DPI exists as a partially folded state formed by coalescence of distinct α-helix-associated microdomains. The physical reality of this model is investigated by comparison of the observed two-dimensional nuclear Overhauser enhancement (NOE) spectroscopy (NOESY) spectrum with that predicted from crystal and DG/RMD structures. The observed NOESY spectrum contains fewer tertiary contacts than predicted by any single simulation, but it matches their shared features; such "ensemble correspondence" is likely to reflect the effect of protein dynamics on observed NOE intensities. We propose (*i*) that the folded state of DPI is analogous to that of a compact protein-folding intermediate rather than a conventional native state and (*ii*) that the molten state is the biologically active species. This proposal (the molten-globule hypothesis) leads to testable thermodynamic predictions and has general implications for protein design.**

We and others have recently described the two-dimensional NMR resonance assignment of a monomeric insulin analogue, des-pentapeptide-(B26–B30)-insulin (DPI) (1, 2), which provides a favorable system for NMR studies due to reduced aggregation (3). DPI retains sufficient structural information for proper function and folding (4–13), and its crystal structure is similar to the corresponding portion of the intact protein (6, 7). Crystal structures of DPI and native insulin have been determined in a variety of forms (6–11). Individual protomers exhibit distinctive features, which may be classified as *local and uncorrelated* (alternative configurations of surface side chains), *transitions in secondary structure* (the N-terminal α-helical extension of the B-chain α-helix in the phenol-related hexamer), and *nonlocal and correlated* (rigid-body displacement of elements of secondary structure). The latter differences require coordinate adjustment in the configurations of multiple side chains at points of helix–helix contact (14).

We present here the NMR structure of DPI.‖ Overall features of the crystal structure (6, 7) are retained in solution. Interestingly, the range of NMR structures is similar to that observed among inequivalent crystallographic protomers: individual members of the distance geometry/restrained molecular dynamics (DG/RMD) ensemble differ by displacement of α-helices. The physical reality of the NMR ensemble is investigated by analysis of simulated two-dimensional nuclear Overhauser enhancement (NOE) spectroscopy (NOESY) spectra predicted from the crystal and NMR structures (15). We propose that DPI may be viewed as the coalescence of distinct microdomains as envisaged by models (16) of protein folding intermediates (the molten-globule hypothesis).

Materials and Methods

NMR Methods. DPI was prepared for NMR study in 20% (vol/vol) perdeuterated acetic acid (2). Two-dimensional NMR spectra were obtained at 25°C as described (2, 17). NOE and *J*-coupling (dihedral angle) restraints were used in DG/RMD calculations as described (17).

DG/RMD. DG models were obtained with the program DG-II (18) and refined by restrained energy minimization/restrained molecular dynamics (19) using the program XPLOR (A. T. Brunger, Yale University). Statistical parameters are given in Table 1.

Model Calculations. The crystal coordinates of DPI [1.2-Å resolution (7)] were kindly provided by D. C. Liang (Institute of Biophysics, Academica Sinica, Beijing, China). Back-calculated NOESY spectra (15) were calculated by using XPLOR (20).

Statement of the Problem: The "NOESY Paradox"

DPI exhibits two key characteristics of a native protein (2): (*i*) secondary structure (as monitored by circular dichroism and two-dimensional NMR) whose stability is demonstrated by observation of slowly exchanging amide resonances in ^2H_2O; and (*ii*) tertiary structure, as demonstrated by long-range NOEs and inaccessibility of an internal tyrosine to an NMR probe (photochemical dynamic nuclear polarization). The NOESY spectrum of such a protein would ordinarily be expected to define its solution structure. The observed NOESY spectrum of DPI is shown in Fig. 1*A*, and that predicted from the crystal structure (7) is shown in Fig. 1*B*. Although predicted sequential and medium-range NOEs (i.e., α-helix and β-turn-associated contacts) are observed,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: DG, distance geometry; DPI, des-pentapeptide-(B26–B30)-insulin; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional NOE spectroscopy; SA, simulated annealing; RMD, restrained molecular dynamics.

†Permanent address: Institute of Biophysics, Academia Sinica, Beijing, China.
‡Permanent address: Centre National de la Recherche Scientifique, Unite de Recherche Associée, 32, Biop Polytechnique, 91128 Palaiseau, France.
§To whom reprint requests should be addressed at: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115.

‖The coordinates of DPI and the NMR restraint list have been deposited in the Brookhaven Data Bank.

Table 1. Statistical parameters describing NMR ensembles

rms deviations*	NMR	Crystal
Main chain		
A chain	0.89 Å	0.79Å
B chain	1.81 Å	1.45Å
α-Helices†	0.81 Å	0.62Å
Side chains		
A chain	1.74 Å	1.60Å
B chain	2.72 Å	2.45Å
α-Helices†	1.64 Å	1.51Å
Average NOE restraint violations‡		
Upper-bound violations	0.032 Å	
Deviations from ideal covalent geometry		
Bond lengths	0.010 Å	
Bond angles	2.90°	
Restrained empirical energy function§		
Constrained dihedral angles	1.3 kcal/mol	
NOE restraint energy	16.1 kcal/mol	
van der Waals	−194.0 kcal/mol	
Hydrogen bonds	−25.2 kcal/mol	
Improper dihedral angles	3.3 kcal/mol	
Dihedral angles	161.2 kcal/mol	
Covalent bond lengths	12.1 kcal/mol	
Bond angles	156.0 kcal/mol	
Total	129.9 kcal/mol	

*The ensemble was aligned according to the C^{α} positions of residues A2–A19 and B4–B20; rms values shown correspond to these regions.

†α-Helices as defined by NMR studies of native insulin include A2–A8, A12–A17, and B9–B19.

‡NOE restraint violations represent rms upper-bound violations; no lower bounds were assumed.

§Empirical energies were calculated with XPLOR; the restrained NOE force constant was 40 kcal/Å^2 and the restrained dihedral force constant was 40 kcal/radian^2 (1 kcal = 4.18 kJ).

remarkably fewer long-range NOEs are observed than predicted (87 of 343 predicted to be resolved). The absence of such tertiary contacts suggests loss of ordered structure in solution. This is paradoxical: how can a protein be at once folded and unfolded?

Results and Interpretation

NMR-Derived Restraints. (*i*) Secondary structure may be determined from the pattern of sequential (total 187) and medium-range (total 94) main-chain NOEs. The A chain contains an N-terminal α-helix (residues A2–A8) and a C-terminal α-helix (residues A12–A17); the B chain contains a central α-helix (residues B9–B19) and β-turn (residues B20–B23). (*ii*) Stereospecific assignments have been obtained for the β resonances of 10 residues (A5–A7, A12, A14, A19, B4–B6, B10, B11, B17, B19, and B24); stereospecific assignment of the Val-A3 methyl groups and side-specific assignment of the A19 and B24 aromatic protons were obtained after initial DG/simulated annealing (SA) modeling. (*iii*) Long-range NOEs (total 87) are observed within the A chain, within the B chain, and between chains; these are consistent in each case with the crystal structure (2, 7).

DG/RMD Calculations. An ensemble of 14 backbone structures, shown in Fig. 2*B*, encompasses the crystal structure (7, 8). The superposition shown is based on least-squares C^{α} alignment of residues B9–B19 (B-chain α-helix); this orientation has previously been used to analyze differences among crystallographic protomers (14) as shown in Fig. 2*A*. Although the A-chain α-helices are well ordered in *individual* NMR or crystal structures, in each collection of structures the orientation of the A chain is not precisely defined relative to the B chain. Among different crystal forms interchain displacements of 1–3 Å are observed and attributed to the effects of crystal packing interactions (14). A similar range of structures is observed in the NMR ensemble and presumably is due to absent restraint information (the NOESY paradox).

Comparison with Crystal Structures. rms deviations among NMR and crystal structures are strikingly similar (Table 1). In making such comparisons we emphasize that NMR- and crystal-related rms deviations represent fundamentally different calculations: precision of an underdetermined NMR ensemble in the first case and structural differences among well-defined but distinct crystal forms in the second. There is no *a priori* relationship between these calculations. Imprecision of an NMR ensemble also contains contributions from technical loss of restraint information (chemical-shift overlap, incomplete stereospecific assignments, and coarse interpretation of NOE intensities as strong, medium, or weak).

In Fig. 3 *a–d* the two sets of rms deviations are shown by residue (NMR ensemble in solid line; crystal structures in broken line). In each case the α-helices are defined with greatest precision (NMR) or consistency (x-ray). In the A chain (Fig. 3 *a* and *b*) the N- and C-terminal residues (A1 and A21) are disordered in solution. The position of the N-terminal α-helix (A2–A8) is less well defined among crystal forms than in the NMR ensemble, presumably reflecting effects of crystal lattice interactions. The A6–A11 loop and A11–A15 segment exhibit comparable rms deviations in the two cases. The C-terminal region (A16–A20), whose internal side chains pack against those of the B-chain α-helix, is less well ordered in solution than among crystal structures. We speculate that this difference reflects an indirect effect of the absence of the B-chain dimer interface in solution. In the B chain (Fig. 3 *c* and *d*) the N-terminal arm (B1–B4) is partly disordered: among the NMR structures the proximity of Phe-B1 to the A chain varies from 3 to 8 Å (Fig. 2*B*). A similar range is observed among crystal structures in the T state (Fig. 2*A*). The B-chain α-helix (B9–B19) is comparably well defined in the two cases. The B-chain β-turn is less well defined in solution than among crystal forms. Its endpoints are stabilized by the internal packing of Cys-B19 and Phe-B24. The B-chain α-helix and adjoining β-turn are locally well ordered (Fig. 2*C*): the side chains of Leu-B11, Val-B12, Leu-B15, and Phe-B24 (green) exhibit well-defined configurations. These side chains are near but not correlated with A-chain hydrophobic residues Ile-A2, Val-A3, Leu-A16, and Tyr-A19 (red). In Fig. 3 *e–h* the x-ray structure of DPI is compared with the NMR ensemble (solid line) and the set of crystal structures of native insulin (broken line). The crystal structure of DPI is consistent with either group of structures, and the resulting pattern of rms deviations is similar to that described above.

Physical Interpretation of DG/RMD Ensemble. The ensemble of NMR structures and range of crystal structures each suggest that the insulin core consists of coalescence of discrete helix-associated microdomains with variation allowed in the details of side-chain packing (21, 22). We propose that subnanosecond fluctuations in interproton vectors lead to quenching of long-range NOEs (NOESY paradox).

Agreement between predicted and observed NOESY spectra provides a criterion for evaluating the accuracy of an NMR structure (15, 20). Although each of the DPI DG/SA structures satisfies the experimental restraints (see above), *none satisfies this criterion*. In each case simulated NOESY spectra back-calculated from DG/RMD structures contain additional interresidue NOEs that are absent from the experimental data (Fig. 1*C*). The additional NOEs—interpreted as strong, medium, or weak with the same limitations in resolution and stereospecific assignment as in the experimental data (2)—are in each case sufficient to specify a DG/RMD structure with the precision expected of a high-resolution NMR structure [i.e., main-chain rms deviation < 0.3 Å (23, 24)]. Further control calculations demonstrate that

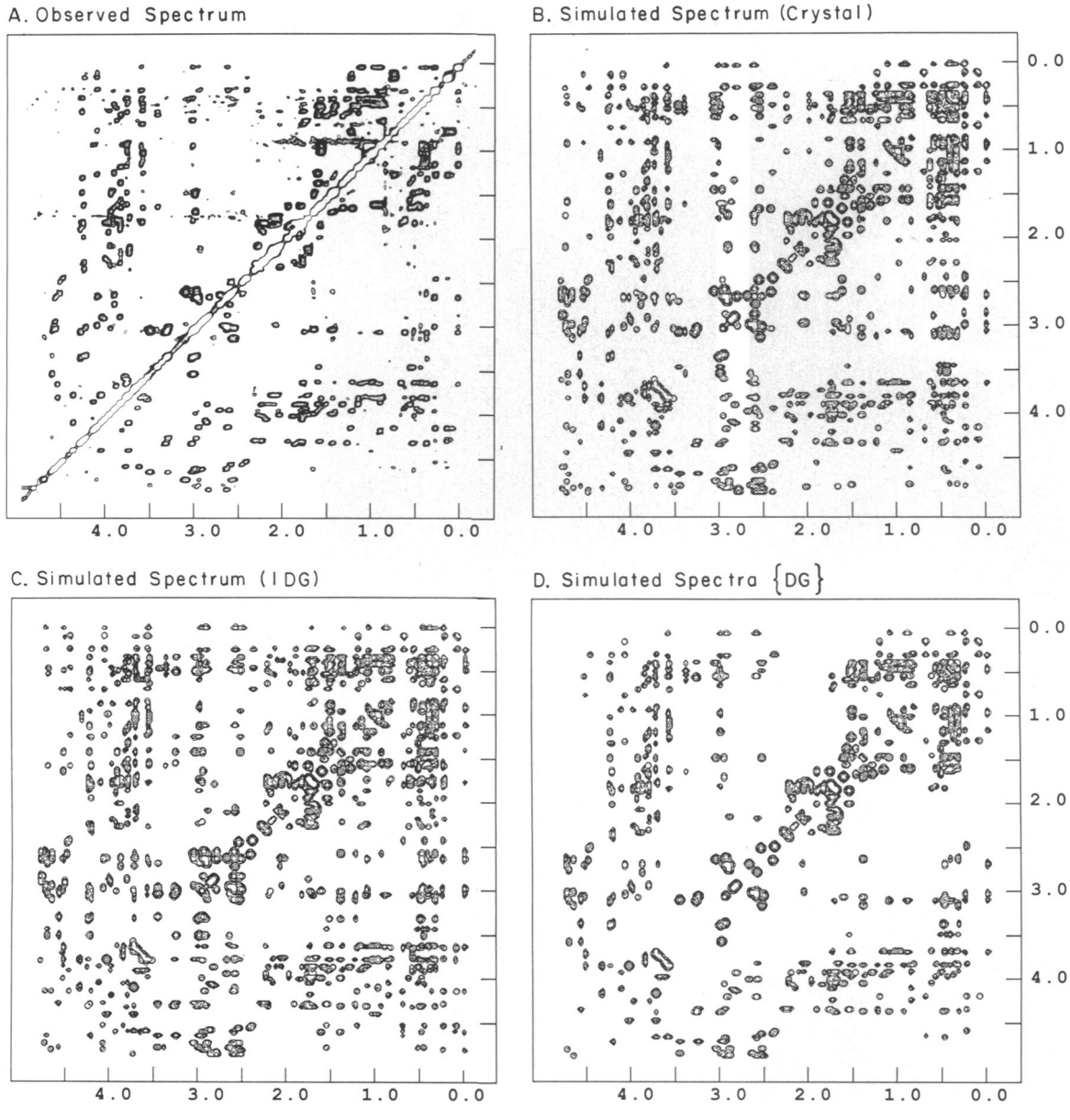

FIG. 1. (*A*) Aliphatic region of experimental NOESY spectrum. (*B*) Corresponding region of simulated spectrum back-calculated from the DPI crystal structure (7). (*C*) Corresponding region of simulated spectrum back-calculated from one representative DG/RMD structure. (*D*) Set of NOESY crosspeaks predicted in common by all members of DG/RMD ensemble (designated {DG}) closely matches experimental data. Experimental and simulated data were normalized with respect to the *ortho–meta* crosspeak of Tyr-A14; similar contour levels are shown in *A–D*, reflecting experimentally obtained signal-to-noise ratio (2). A uniform rotational correlation time of 2.5 ns was assumed.

there is no consistent DG structure that simultaneously satisfies the observed (positive) restraints and an additional set of (negative) restraints: that absent NOEs should represent interproton distances > 4.5 Å.

The Molten-Globule Hypothesis

What is the relationship between the observed and predicted NOESY spectra? Remarkably, the experimental data are closely matched by the subset of NOEs that are predicted in common by the DG/RMD ensemble (Fig. 1*D*). This "ensemble concordance" suggests that the range of NMR structures has an underlying physical meaning—i.e., that a dynamic mechanism must exist to quench nonlocal NOEs that would otherwise be observed. Whereas the shared NOEs contain information that specifies the general insulin fold (Fig. 2*A*), the absent NOEs contain information that specifies particular structural relationships that distinguish one DG/SA structure (or one crystal form) from another. We propose that the dynamics of DPI in 20% acetic acid—stably folded elements of secondary structure with fluctuation in tertiary orientation—is analogous to that of compact protein-folding intermediates. We further propose a global correspondence between the existence of different crystallographic (T) states of insulin and a similar range of conformations in solution.

Evidence for flexibility in DPI has previously been provided by observation of limited dispersion of chemical shifts, temperature dependence of chemical shifts, and resonance linewidths (1, 2). Of particular interest is the conformational broadening of amide and selected aliphatic ^{1}H NMR resonances (1–3, 23). Such variation in linewidth reflects an equilibrium among conformational substrates related by millisecond motions. Motions on this time scale are distinct from rapid fluctuations required for quenching of NOEs. The subset of broadened amide resonances includes those of α-helices (2, 3). Since the latter are slowly exchanging in

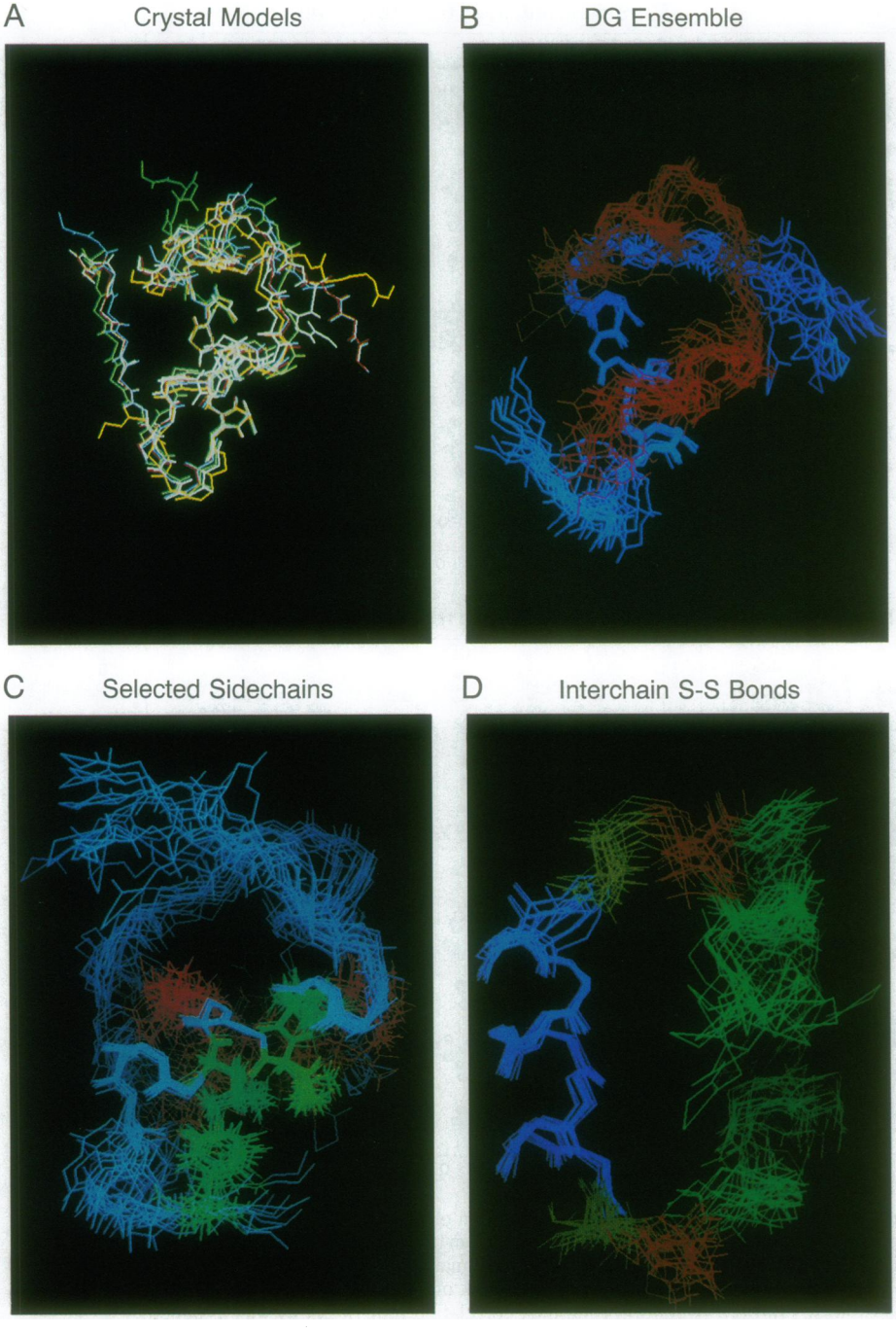

FIG. 2. (*A*) Set of crystallographic protomers: DPI (yellow), 2-Zn DPI molecule 1 (red), 2-Zn DPI molecule 2 (blue), cubic crystal form (white), and 4-Zn molecule 2 (green, with α-helical transition for residues B1–B9). (*B*) Ensemble of 14 DG/RMD structures (A chain in red, B chain in blue) with residues B9–B19 of the main chain (i.e., the B-chain α-helix) aligned shows relative displacement of A-chain α-helices and excursions of N-terminal arm of the B chain. The molecular orientation is the same as in *A*. (*C*) B-chain microdomain (blue), aligned on the B9–B19 main chain, contains well-defined side chains Phe-B24, Leu-B15, Val-B12, and Leu-B11 (bottom to top in green); the relative positions of internal A-chain residues (A2, A3, A16, and A19; shown in red) are not well correlated. (*D*) Interchain disulfide bonds (top, A7—B7; bottom, A20—B19) and adjoining α-helices, as aligned relative to B-chain α-helix (blue). The A-chain main chain is shown in green, Cys-B7 and Cys-B19 in yellow, and Cys-A7 and Cys-A20 in red.

^2H_2O, the inferred millisecond motions do not involve the breakage of peptide hydrogen bonds (i.e., local unraveling of an α-helix). Broadening is also observed among H^α and H^β resonances of cystines, suggesting slow interconversion of different disulfide configurations. Such a range of configurations is required to accommodate the variation in the A-chain/B-chain relationship observed in the NMR ensemble, as shown in Fig. 2*D*. Although correlated displacements of α-helices could account for conformational broadening of ^1H NMR resonances, no dynamic information is directly provided by DG/RMD calculations.

Predictions and Biological Implications

Thermodynamics of Protein Unfolding. A defining feature of the native state of a protein is its highly cooperative ("all-or-none") folding transition. The fluctuating tertiary structure of DPI, however, appears to partition into distinct α-helix-associated microdomains. The largest such microdomain [the B-chain α-helix (B9–B19), adjoining β-turn (B20–B23), and Phe-B24] is locally well ordered but of variable relationship to the A chain. By analogy to equilibrium molten globules (16), we predict that the thermal unfolding of DPI will not be highly cooperative. We further predict that the latent enthalpy of unfolding (as measured by differential scanning calorimetry) will be smaller than that of a conventional native state (<3 kcal·g^{-1}·deg^{-1}).

Extension to Physiological Conditions. DPI in 20% acetic acid or in aqueous solution (pH 1.8) may be analogous to acid-denatured forms of other proteins, such as the A state of α-lactalbumin and related systems (16). Extension of the molten-globule hypothesis to physiological conditions is suggested, however, by a recent NMR study of an engineered insulin monomer: its spectrum exhibits an analogous NOESY paradox (25). Corresponding partition of its hydrophobic core into microdomains would predict analogous thermodynamic evidence of a partially folded state.

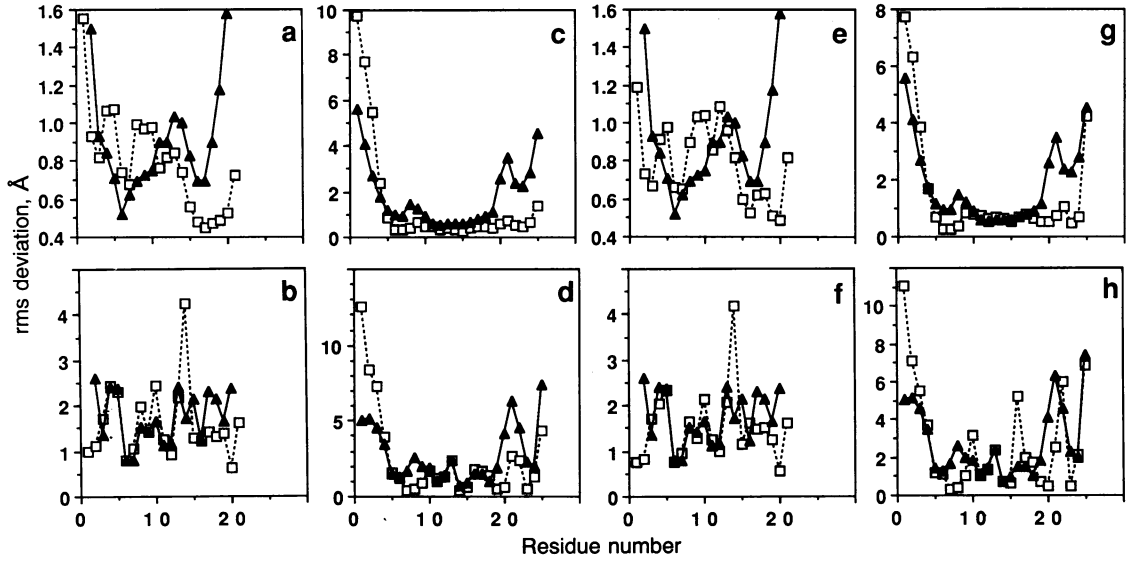

FIG. 3. (*a–d*) rms deviations within NMR ensemble (▲——▲) and collection of crystal structures (□----□) are shown as a function of residue number. (*a*) A-chain main chain. (*b*) A-chain side chains. (*c*) B-chain main chain. (*d*) B-chain side chain. rms deviations were calculated on the basis of the average of pairwise alignment relative to main-chain atoms of α-helixes. (*e–h*) rms deviations between the DPI crystal structure (7) and set of NMR structures (▲——▲) or set of crystal structures of native insulin (□----□). (*e*) A-chain main chain. (*f*) A-chain side chains. (*g*) B-chain main chain. (*h*) B-chain side chain.

Biological Implications. A recent x-ray study of an isomorphic but completely inactive insulin analogue (miniproinsulin) suggests that crystal structures of insulin depict inactive conformations (26). A complementary NMR study has revealed partial unfolding of the B chain in a mutant insulin with near-native potency, suggesting that a structural switch is required for receptor recognition (24). Partial unfolding and refolding of insulin to form a productive hormone–receptor complex would provide a biological rationale for a molten structure (27).

Conclusions

Analysis of similarities and differences among crystal structures of insulin has provided general insights into mechanisms of conformational change in proteins. To extend this analysis under solution conditions, we have determined the NMR structure of a monomeric insulin analogue (DPI). The structure of DPI in solution is similar to its structure in the crystal state in secondary structure (three α-helices and one β-turn) and overall tertiary fold. Although the NOESY spectrum does not contain sufficient information to define a precise tertiary structure (the NOESY paradox), an ensemble-correspondence is observed between experimental and simulated NOESY spectra that suggests a novel state of order and disorder (the molten-globule hypothesis). Design of biologically active species as partially folded states would extend the scope of the protein–folding problem.

We thank M. Karplus, G. G. Dodson, S. E. Shoelson, and L. J. Neuringer for advice; R. E. Chance (Eli Lilly) for generously providing DPI; D. Hare for the FELIX program; T. Havel for the DG-II program; and A. Brunger for the XPLOR program. This work was supported in part by grants from the National Institutes of Health, American Diabetes Association, and Juvenile Diabetes Foundation International to M.A.W.

1. Boelens, R., Ganadu, M. L., Verheyden, P. & Kaptein, R. (1990) *Eur. J. Biochem.* **191,** 147–153.
2. Hua, Q. X. & Weiss, M. A. (1990) *Biochemistry* **29,** 10545–10555.
3. Hua, Q. X. & Weiss, M. A. (1991) *Biochemistry* **30,** 5505–5515.
4. Pullen, R. A., Lindsay, D. G., Wood, S. P., Tickle, I. J., Blundell, T. L., Wollmer, A., Krail, A., Brandenburg, D., Zahn, H., Gliemann, J. & Gammeltoft, S. (1976) *Nature (London)* **259,** 369–373.
5. Fischer, W. H., Saunders, D., Brandenburg, D., Wollmer, A. & Zahn, H. (1985) *Hoppe-Seyler's Z. Biol. Chem.* **366,** 521–525.
6. Bi, R. C., Dauter, Z., Dodson, E., Dodson, G., Giordano, F. & Reynolds, C. (1984) *Biopolymers* **23,** 391–395.
7. Dai, J.-B., Lou, M.-Z., You, J.-M. & Liang, D.-C. (1987) *Sci. Sin.* **30,** 55–65.
8. Baker, E. N., Blundell, T. E., Cutfield, G. S., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M. C., Hubbard, R. E., Iassac, M. W., Reynolds, D. C., Sakabe, K. S., Sakabe, N. & Vjayan, N. M. (1988) *Philos. Trans. R. Soc. London Ser. B* **319,** 389–456.
9. Bentley, G., Dodson, G. & Lewitova, A. (1978) *J. Mol. Biol.* **126,** 871–875.
10. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Smith, G. D., Sparks, C. & Swenson, D. (1989) *Nature (London)* **338,** 594–596.
11. Badger, J., Harris, M. R., Reynolds, C. D., Evans, A. C., Dodson, E. J., Dodson, G. G. & North, A. C. T. (1991) *Acta Crystallogr. Sec. B Struct. Crystallogr. Cryst. Chem.* **46,** 127–136.
12. Wang, C.-C. & Tsou, C.-L. (1986) *Biochemistry* **25,** 5336–5340.
13. Schwartz, G. P., Burke, G. T. & Katsoyannis, P. G. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 458–461.
14. Chothia, C., Lesk, A. M., Dodson, G. G. & Hodgkin, D. C. (1983) *Nature (London)* **302,** 500–505.
15. Keepers, J. W. & James, T. L. (1984) *J. Magn. Reson.* **57,** 404.
16. Pfeil, W., Bychkova, V. E. & Ptitsyn, O. B. (1986) *FEBS Lett.* **198,** 287–291.
17. Kochoyan, M., Havel, T. F., Nguyen, D., Dahl, C. E., Keutmann, H. T. & Weiss, M. A. (1991) *Biochemistry* **30,** 3371–3386.
18. Havel, T. F. (1991) *Prog. Biophys. Mol. Biol.* **56,** 43–78.
19. Clore, G. M., Gronenborn, A. M., Brunger, A. & Karplus, M. (1985) *J. Mol. Biol.* **186,** 435–455.
20. Nilges, M., Habazettl, J., Brunger, A. T. & Holak, T. A. (1991) *J. Mol. Biol.* **219,** 499–510.
21. Kruger, P., Strassburger, W., Wollmer, A., van Gunsteren, W. F. & Dodson, G. G. (1987) *Eur. Biophys. J.* **14,** 449–459.
22. Caves, L. S. D., Nguyen, D. T. & Hubbard, R. E. (1991) in *Molecular Dynamics: An Overview of Applications in Molecular Biology*, ed. Goodfellow, J. (Macmillan, New York), pp. 27–68.
23. Kline, A. D. & Justice, R. M., Jr. (1990) *Biochemistry* **29,** 2906–2913.
24. Weiss, M. A., Hua, Q. X., Frank, B. H., Lynch, C. S. & Shoelson, S. E. (1991) *Biochemistry* **30,** 7373–7389.
25. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Bing, X. & Markussen, J. (1991) *J. Mol. Biol.* **256,** 1406–1412.
26. Hua, Q. X., Shoelson, S. E., Kochoyan, M. & Weiss, M. A. (1991) *Nature (London)* **354,** 238–241.
27. Mirmira, R. & Tager, H. S. (1989) *J. Biol. Chem.* **264,** 6349–6354.