Crystal Structure of a “Nonfoldable” Insulin

IMPAIRED FOLDING EFFICIENCY DESPITE NATIVE ACTIVITY*

Received for publication, July 20, 2009, and in revised form, October 19, 2009 Published, JBC Papers in Press, October 22, 2009, DOI 10.1074/jbc.M109.046888

Ming Liu1, Zhu-li Wan2, Ying-Chi Chu3, Hassan Aladdin4, Birgit Klaproth5, Meredith Choquette5, Qing-xin Hua5, Robert B. Mackin6, J. Sunil Rao7, Pierre De Meyts8, Panayotis G. Katsoyannis9, Peter Arvan10, and Michael A. Weiss11

From the 6Division of Metabolism, Endocrinology and Diabetes, University of Michigan Medical Center, Ann Arbor, Michigan 48109, the 8Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106, the 5Department of Pharmacology and Biological Chemistry, Mt. Sinai School of Medicine of New York University, New York, New York 10029, Receptor Systems Biology Laboratory, Hagedorn Research Institute, Niels Steensens Vej 6, Gentofte 2820, Denmark, the **Department of Biomedical Sciences, Creighton University School of Medicine, Omaha, Nebraska 68178, and the 11Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio 44106

Protein evolution is constrained by folding efficiency (“foldability”) and the implicit threat of toxic misfolding. A model is provided by proinsulin, whose misfolding is associated with β-cell dysfunction and diabetes mellitus. An insulin analogue containing a subtle core substitution (LeuA16 → Val) is biologically active, and its crystal structure recapitulates that of the wild-type protein. As a seeming paradox, however, ValA16 blocks both insulin chain combination and the in vitro refolding of proinsulin. Disulfide pairing in mammalian cell culture is likewise inefficient, leading to misfolding, endoplasmic reticulum stress, and proteosome-mediated degradation. ValA16 destabilizes the native state and so presumably perturbs a partial fold that directs initial disulfide pairing. Substitutions elsewhere in the core similarly destabilize the native state but, unlike ValA16, preserve folding efficiency. We propose that LeuA16 stabilizes nonlocal interactions between nascent α-helices in the A- and B-domains to facilitate initial pairing of CysA20 and CysB19, thus surmounting their wide separation in sequence. Although ValA16 is likely to destabilize this proto-core, its structural effects are mitigated by kinetic or thermodynamic partitioning among on- and off-pathway disulfide intermediates. The properties of [ValA16]proinsulin and [ValA16]proinsulin demonstrate that essential contributions of conserved residues to folding may be inapparent once the native state is achieved.

A fundamental problem in biophysics is posed by the efficiency of protein folding (1, 2). The capability of folding is an evolved property of biological sequences, i.e. not broadly representative of polypeptides as a class of heteropolymers (3). A funnel-shaped free-energy landscape enables an efficient conformational search leading by multiple trajectories to the native state (4–6). What distinguishes foldable from nonfoldable sequences (7), and how do productive trajectories avoid bottlenecks (8-10)? Despite the centrality of these questions, experimental approaches are limited. Here, we combine biophysical and cellular studies to identify a residue critical to the efficiency of oxidative folding. A model is provided by insulin, a globular protein central to the regulation of vertebrate metabolism (11).

The insulin gene encodes a single-chain precursor, preproinsulin (Fig. 1A, top) (12). A signal peptide (Fig. 1A, gray bar) is cleaved on translocation into the endoplasmic reticulum (ER) to yield proinsulin. The precursor contains successive sequence motifs, defining a B-domain (Fig. 1A, blue bar), connecting domain (black bar), and A-domain (red bar) (13). The translocated polypeptide is reduced and unfolded. Biophysical studies suggest that folding in the ER yields a well-organized A-B (insulin-like) core and disordered C-domain (Fig. 1B, dashed black segment) (14–19). Folding of proinsulin is coupled to disulfide pairing (Fig. 1, A and B, gold A6–A11, A7–B7, and A20–B19). Two cystines provide inter-struts (A20–B19 and A6–A11), whereas the third forms an external staple (A7–B7). These structural elements are required for stability and activity (20–29). Insulin disulfide isomers exhibit molten structures of marginal stability and low activity (30–32). Proinsulin isomers have been observed at low concentration in β-cells (33), and their accumulation in vivo may be inapparent once the native state is achieved.

5 The abbreviations used are: ER, endoplasmic reticulum; DKP-insulin, insulin analogue containing three substitutions in B-chain (AspB10, LysB28 and Proβ29); DM, diabetes mellitus; HPLC, high performance liquid chromatography; UPR, unfolded protein response; Tricine, N,N′-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; PDB, Protein Data Bank. Amino acids are designated by three-letter code.

8 This work was supported, in whole or in part, by National Institutes of Health Grants DK052085 (to R. B. M.), DK48280 (to P. A.), DK56673 (to P. G. K.), and DK667674 (to M. A. W.). This work was also supported by the American Diabetes Association (to M. A. W.). This article is a contribution from the Cleveland Center for Membrane and Structural Biology.

9 The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S11, Tables S1–S9, and additional references.

1 The atomic coordinates and structure factors (code 3GKY) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

2 Both authors contributed equally to this work.

3 Supported by the Hagedorn Institute of Novo-Nordisk, Ltd.

4 To whom correspondence may be addressed. E-mail: parvan@umichigan.edu.

5 To whom correspondence may be addressed. E-mail: michael.weiss@case.edu.
This study investigates a mutant proinsulin defective in folding and yet active and well organized once the native state is achieved. Substitution of Leu$^{A16}$ by Val (45) reduces the efficiency of disulfide pairing in vitro and leads in mammalian cells to accelerated degradation associated with ER stress (46, 47). Despite these perturbations, the crystal structure of the variant (as a mutant insulin) is essentially identical to wild type (48, 49). We propose that Val$^{A16}$ allows formation of off-pathway oxidative intermediates and/or destabilizes on-pathway intermediates (50). Such perturbations are similar to those proposed to underlie a newly recognized syndrome of toxic protein misfolding, permanent neonatal-onset diabetes mellitus (DM) due to mutations in proinsulin (36–39). Although to date clinical mutations (excluding insertion or removal of cysteines) have clustered in the B-domain, our results demonstrate that the A-domain is also required for efficient folding. The aberrant properties of [Val$^{A16}$]proinsulin thus provide a model for the molecular genetics of a disease of protein misfolding.

MATERIALS AND METHODS

Chemical Synthesis—Insulin was provided by Eli Lilly and Co.; S-sulfonate B-chain derivatives were obtained by oxidative sulfitolysis (51). A- and B-chain analogues were otherwise prepared by solid-phase synthesis (51). Insulin analogues were prepared by chain combination as described previously (45). Predicted molecular masses were confirmed by mass spectrometry.

Recombinant Expression of Proteins—Wild-type and variant proinsulins were expressed in Escherichia coli (52); wild-type and variant mini-proinsulins (53-residue construct) were expressed in folded secreted form in Pichia pastoris (53).

Mammalian Cell Culture—HEK293T cells were cultured at 37 °C in high glucose Dulbecco’s modified Eagle’s medium containing 10% fetal bovine serum and 0.1% penicillin/streptomycin with 5% CO$_2$. CHO-CLA14 cells were maintained in low glucose Dulbecco’s modified Eagle’s medium plus 400 μg/ml G418 and 200 μg/ml hygromycin at 37 °C with 5% CO$_2$ (54). For metabolic labeling, cells were plated into 6-well plates 1 day before transfection. Plasmid DNA (2 μg) was transfected into each well using Lipofectamine (Invitrogen). For studies of the unfolded protein response (UPR), cells were seeded into...
Bottleneck in a Folding Landscape

24-well plates 1 day before transfection. Plasmid DNAs containing proinsulin (wild type or variant) were co-transfected with plasmids (provided by R. Kaufman) encoding luciferase (driven by BiP promoter) and β-galactosidase (driven by CMV promoter) at 5:2:1 ratio using Lipofectamine; assays were performed in triplicate.

Metabolic Labeling and PAGE—At 40 h post-transfection, cells were preincubated in methionine/cysteine-deficient medium for 30 min, metabolically labeled in the same medium containing 35S-labeled Met and Cys for 1 h, washed once with complete medium, and chased for the indicated times. To examine degradation, labeled cells were chased for 4 h with or without 20 μM MG132 or lactacystin. After chase media were collected, cells were lysed in 100 mM NaCl, 1% Triton X-100, 0.2% sodium deoxycholate, 0.1% SDS, 10 mM EDTA, and 25 mM Tris-HCl (pH 7.4). Lysates and chase media were immunoprecipitated with guinea pig anti-insulin antiserum (LINCO Diagnostics, Inc.) and analyzed by Tris-Tricine/urea-SDS-PAGE under reducing and nonreducing conditions (33, 55).

UPR Assay and Statistical Analysis—The UPR assay (see supplemental Fig. S1 and supplemental Table S1) was performed as described previously (56). At 40 h post-transfection cells were washed three times with phosphate-buffered saline, lysed in luciferase assay buffer, and transferred to a 96-well plate; luminescence was measured in a Turner Biosystems plate luminometer using the Dual-Light assay system (Applied Biosystems). Statistical analysis employed a randomized block design analysis of variance model after checking for appropriateness of a normality assumption on errors using a Q-Q plot with Wilk-Shapiro test. Testing (two-tailed) was done at a significance level of α = 0.05. When significant findings were discovered, a nonparametric version of analysis of variance was run to ensure consistency.

Spectroscopy—CD spectra, obtained using an Aviv spectropolarimeter equipped with a titration unit for denaturation studies, were measured at a protein concentration of 5 μM at 4 °C (57). 1H NMR spectra were obtained at 700 MHz in D2O solution at pH 7.0 and 25 °C (21).

Thermodynamic Modeling—CD-detected guanidine denaturation data (222 nm) were fitted by nonlinear least squares to a two-state model (supplemental Table S2) as described previously (58). In brief, CD data, θ(x), were fitted by a nonlinear least-squares program according to Equation 1,

\[
\theta(x) = \frac{\theta_\alpha + \theta_\beta e^{-\Delta G^\circ_{\alpha\to\beta} - mx/RT}}{1 + e^{-\Delta G^\circ_{\alpha\to\beta} - mx/RT}}
\]

where x is the concentration of guanidine hydrochloride, and where θα and θβ are base-line values in the native and unfolded states. These base lines were approximated by pre- and post-transition lines θα(x) = θα^∞ + mαx and θβ(x) = θβ^∞ + mβx. Fitting the original CD data and base lines simultaneously circumvents artifacts associated with linear plots of DG as a function of denaturant according to ΔG^°(x) = ΔG^°αβ + mx (for reviews see Refs. 58, 59).

Crystallography—Crystals were grown by hanging-drop vapor diffusion with a 1:2.5 ratio of Zn2+ to protein monomer and 3.7:1 ratio of phenol to protein monomer (49). Drops consisted of 1 μl of protein solution (10 mg/ml in 0.02 M HCl) mixed with 1 μl of reservoir solution (0.02 M Tris-HCl, 0.05 M sodium citrate, 5% acetone, 0.03% phenol, and 0.01% zinc acetate (pH 8.0)). Crystals (space group R3) were obtained at room temperature after 2 weeks. Data were collected from single crystals flash-frozen to 100 K. Reflections from 22.58 Å to 1.8 Å were measured using synchrotron radiation (Chess, Cornell University). Data were processed with programs DENZO (version 1.9.6) and SCALEPACK (version 1.9.6). The structure was determined by molecular replacement using CNS; initial model was the wild-type TR dimer (PDB identifier 1RWE) following removal of water, zinc, and chloride ions. A translation-function search was performed following rotation-function analysis of data between 15.0 and 4.0 Å. Rigid-body refinement using CNS, employing overall anisotropic temperature factors and bulk-solvent correction, yielded values of 0.31 and 0.30 for R and Rfree, respectively, for data between 19.2 and 3.0 Å resolution. Between refinement cycles, 2Fo − Fc and Fc − Fo maps were calculated using data to 3.0 Å resolution; zinc, chloride ions, and phenol molecules were built using O (60). Water molecules were calculated and checked using DDQ (61). The geometry was monitored with PROCHECK (62); zinc ions and water molecules were built as refinement proceeded. Further refinement employed CNS (63), which implements maximum-likelihood torsion-angle dynamics and conjugate-gradient refinement. Statistical data are provided in supplemental Table S3. Potential packing defects were calculated with Surflex (64).

RESULTS

The hydrophobic core of insulin is conserved among vertebrates (12); the neighborhood of cystine A20–B19 is structurally invariant (48, 65, 66). Studies of insulin-related polypeptides have demonstrated the importance of cystine A20–B19 as a first step in disulfide pairing (67–69). Leu16, inaccessible in the native state (Fig. 2A, red), adjoins Tyr19 and Cys20 (Fig. 2B, blue and gold, respectively) along the inner surface of the A12–A19 α-helix and projects between Cys11 and Leu15 (gold and gray). Hydrophobic collapse of these side chains in the nascent polypeptide is proposed to orient Cys20 and Cys19 for initial pairing (20, 29, 57).

Val16 Impairs Insulin Chain Combination and Refolding of Proinsulin—Insulin chain combination provides a peptide model of proinsulin refolding (70). The reaction leads to native disulfide pairing, demonstrating that information required for proinsulin folding is contained within the A and B sequences (71). Yield is limited by off-pathway products under kinetic control (disulfide-linked cyclic A-chains, cyclic B-chains, B-chain dimers, and B-chain polymers). The robustness of chain combination to amino acid substitutions has enabled synthesis of hundreds of analogues in past decades. As an initial test of the importance of Leu16 in disulfide pairing, we prepared a variant A-chain containing Val16. Reactions employed 80 mg of wild-type A-chain and 40 mg of B-chain (each as S-sulfonate derivatives). Whereas this protocol ordinarily yields 8–9 mg of wild-type insulin (following purification by reverse-phase high performance liquid chromatography (HPLC) (51, 57, 72)), no product was detectable in three attempted reactions. Because the threshold of HPLC detection

DECEMBER 11, 2009•VOLUME 284•NUMBER 50

JOURNAL OF BIOLOGICAL CHEMISTRY
is <40 µg, the efficiency of chain combination was reduced by >200-fold.

Refolding of proinsulin is more efficient than chain combination (73). Accordingly, we sought to prepare ValA16-human proinsulin (position 81 of the polypeptide) by recombinant expression in E. coli. Bacterial expression gave rise to inclusion bodies containing reduced and unfolded proinsulin; following purification, oxidative refolding ordinarily yields native disulfide pairing (52, 74). Expression of wild-type and variant polypeptides within inclusion bodies achieved similar levels; the reduced polypeptides were purified in similar yield. Although refolding of wild-type proinsulin was robust (52), however, the yield of [ValA16]proinsulin was negligible. The refolding mixture gave rise to a broad and inhomogeneous HPLC elution profile, suggesting the presence of multiple products.

ValA16 Blocks the Cellular Folding of Proinsulin and Induces ER Stress—Cellular folding efficiencies of variant proinsulins were evaluated in CLA14 cells (a subclone of the Chinese hamster ovary cell line) (54) and in human cell line HEK293T. Following transfection, we examined expression, disulfide isomer formation, and secretion of newly synthesized proinsulin as radiolabeled with 35S-labeled amino acids for 1 h and chased for 4 h (Fig. 3, A–C). Denaturing PAGE (Tris-Tricine/urea-SDS-PAGE) in the absence of reduction (Fig. 3A) permitted examination of distinct proinsulin disulfide isomers as formed in the ER (33, 35); with reduction, this gel system provides a probe of total protein expression (Fig. 3B). The wild-type construct gave rise to robust expression, primarily of a fast migrating species; previous studies established that the latter is the native species (33). This species is efficiently secreted from transfected cells (Fig. 3, C, lane C to medium (lane M). Two minor species are also present as slower migrating isomers with mispaired disulfide bonds; these exhibit lower secretion efficiency (Fig. 3, brackets). In CLA14 cells, substitution of LeuA16 by Val markedly impairs expression (relative to wild-type; Fig. 3, A and B, lanes 3 and 4) and blocks its secretion (lanes 8 and 8). In HEK293T cells, folding of [ValA16]proinsulin in the ER preferentially yields non-native isomers (Fig. 3C, lane 15) that are not secreted (lane 16). Pulse-chase experiments (4 h with or without proteosome inhibitors; lactacystin in Fig. 3D and MG132 in Fig. 3F) indicate that the misfolded variant retained in the ER undergoes proteolytic degradation (lanes 22 versus 24 and 35 versus 36). A clinical mutation (CysA7 → Tyr) provided positive control for misfolding, ER retention, and degradation (Fig. 3D, lanes 19, 21, and 23); this mutation causes DM in the Akita mouse (75) and humans (36).
A probe for UPR activation was provided by up-regulation of ER resident chaperone immunoglobulin-binding protein (76). An assay for induction of the UPR transcriptional program was provided by co-transfection of a proinsulin expression plasmid with a plasmid employing the BiP promoter to drive expression of luciferase (56). Transfection efficiency was normalized by co-transfection of a proinsulin expression plasmid providing by co-transfection of a proinsulin expression plasmid (

Abraham et al., 2009). For ThrA8→His mutant (73), designated AK, (lanes 19, 21 and 23), or ValA16 variant (lanes 20, 22 and 24); an empty-vector control is provided in lane 17. Cells were chased with or without 20 μM lactacystin for 4 h. Akita and ValA16 variants exhibit enhanced degradation relative to wild type, in each case partially blocked by the proteasome inhibitor. E, control studies of unstable protein variants in transfected HEK293T cells suggest that cellular folding efficiency is uncorrelated with native-state thermodynamic stability. Folding and secretion of a molten two-disulfide analogue (containing paired substitutions CysA6 → Tyr) (Akita mutant (73); designated AK, (lanes 19, 21 and 23), or ValA16 variant (lanes 20, 22 and 24); an empty-vector control is provided in lane 17. Cells were chased with or without 20 μM lactacystin for 4 h. A probe for UPR activation was provided by up-regulation of ER resident chaperone immunoglobulin-binding protein (76). An assay for induction of the UPR transcriptional program was provided by co-transfection of a proinsulin expression plasmid with a plasmid employing the BiP promoter to drive expression of luciferase (56). Transfection efficiency was normalized by co-transfection of a proinsulin expression plasmid providing by co-transfection of a proinsulin expression plasmid.
Bottleneck in a Folding Landscape

in CLA14 cells (lanes 5 and 6 in Fig. 3A and counterparts in Fig. 3B). This surface substitution (found in avian insulins) in principle enhances the α-helical propensity of the A1–A8 segment (77, 78). We therefore tested whether HisA8 might rescue the folding defect caused by ValA16 (Fig. 3A and counterparts in Fig. 3B, lanes 9 and 10; lanes 9′ and 10′). Expression and secretion are partially restored relative to ValA16 alone but still markedly decreased relative to wild type. Refolding studies of [HisA8,ValA16]proinsulin, obtained by recombinant expression in E. coli, yielded a distribution of misfolded products similar to those observed in studies of [ValA16]proinsulin,6 likewise precluding isolation of the folded variant.

Despite the slight effects of HisA8 in partial rescue of the ValA16-associated folding defect, we undertook synthetic studies in an effort to obtain a sufficient quantity of a ValA16-containing insulin analogue for analysis of structure and function. The efficiency of wild-type chain combination was enhanced by ~20% by the stabilizing HisA8 substitution. More striking effects were observed in chemical synthesis of [HisA8,ValA16]insulin; efficiency of chain combination rose from essentially unobservable (above) to “only” 8-fold impaired. Although this yield remained anomalously low, it represented an augmentation of >25-fold and, through brute force of mass action, enabled isolation of the analogue in milligram quantities.

Folding Efficiency and Stability Are Uncorrelated—Cellular folding may be impaired by one or more of several molecular mechanisms as follows: a kinetic block en route to an on-pathway intermediate, aberrant stabilization of an off-pathway intermediate, or instability and degradation of the folded state, once reached. The above PAGE assay, by distinguishing intermediate, or instability and degradation of the folded state, the above PAGE assay, by distinguishing between these possibilities. This issue was further pursued by correlating cellular and biophysical studies. Because proinsulin consists of a folded insulin moiety and disordered connecting domain, effects of amino acid substitutions on native-state stability were investigated in corresponding insulin analogues. CD provided a convenient probe for fractional unfolding as a function of chemical denaturation (Fig. 4). Fits to a two-state model (58) in each case were characterized by R-values greater than 0.99.

The stability of [HisA8,ValA16]insulin was observed to be reduced relative to insulin and [HisA8]insulin (Fig. 4A). Its free energy of unfolding (∆Gf), as inferred from a two-state model (58), is 2.1 ± 0.1 kcal/mol (supplemental Table S2). This represents a perturbation (∆∆Gf) of 2.8 ± 0.2 kcal/mol relative to [HisA8] (∆Gf of 4.9 kcal/mol), presumably due to the following: (i) substitution of a β-branched residue in an α-helix (79), and (ii) a packing defect in the hydrophobic core (80) with possible transmitted structural changes (81). Native disulfide pairing in the three analogues was verified in each case by x-ray crystallography (below).

To test whether decreased native-state stability might in itself be responsible for the impaired folding efficiency of [ValA16]insulin and [HisA8,ValA16]proinsulin, we investigated the cellular folding and secretion of two partially folded mutant proinsulins, a two-disulfide analogue containing pairwise substitution of CysA6 and CysA11 by Ser (and so lacking cystine A6–A11; Fig. 3, lanes 29 and 30), and an analogue containing core substitution IleA2 → Gly (lanes 31 and 32). Biophysical studies of these analogues have previously been described in the context of an engineered monomer (DKP-insulin) (82). Native disulfide pairing was verified by two-dimensional NMR spectroscopy (21, 83). The two-disulfide analogue [SerA6,SerA11]DKP-insulin adopts a flexible conformation of marginal stability lacking the A1–A8 α-helix; [GlyA2]insulin also exhibits local unfolding of this α-helix (21, 24, 57, 83). Denaturation studies indicated that the stabilities of these analogues are reduced to an extent similar to or greater than that of [HisA8,ValA16]insulin (Fig. 3B and supplemental Table S2), and yet the corresponding mutant proinsulins were efficiently folded and secreted by HEK293T cells (Fig. 3E). Chain combination yields in their chemical synthesis were likewise robust. The contrast between the folding properties of ValA16-containing variants, on the one hand, and the unstable control analogues, on the other, suggests that LeuA16 contributes to efficiency of disulfide pairing in a way that is disproportionate to its role in the native state. Efficient folding, trafficking, and secretion of [SerA6,SerA11]proinsulin and [GlyA2]proinsulin further suggest that native structural organi-

6 Secretion of a corresponding variant single-chain insulin analogue in yeast P. pastoris, ordinarily robust to diverse mutations, was impaired by 40-fold by the paired substitutions HisA8,ValA16 (from 40 to 1 mg/liter).
incorporation of substitution HisB10 by dimerization. Comprising the central H9252 of LeuA16 by Val does not perturb receptor binding. A probe for these results indicate that, once folding is achieved, substitution A, spectra of [AspB10]insulin in the protein concentration range 0.05–0.30 mM at neutral pH and 25 °C (spectra a–c), b, corresponding dimerization of [ValA16,AspB10]insulin (spectra d and e), c, reference spectrum of ValA16 analogue of engineered monomer DKP-insulin (spectrum f).

Figure 5. 1H NMR studies of insulin analogues. Aromatic resonances provide probes of dimerization. A, spectra of [AspB10]insulin in the protein concentration range 0.05–0.30 mM at neutral pH and 25 °C (spectra a–c). B, corresponding dimerization of [ValA16,AspB10]insulin (spectra d and e). C, reference spectrum of ValA16 analogue of engineered monomer DKP-insulin (spectrum f).

Bottleneck in a Folding Landscape

zation (i.e. folding of the A1–A8 α-helix) is not required to pass quality control checkpoints in the ER and secretory pathway.

Structure-Function Relationships—Despite its impaired folding efficiency, the affinity of [HisA8,ValA16]insulin for the insulin receptor (Kd 0.15 ± 0.03 nM) is almost 3-fold higher than that of wild-type insulin (Kd 0.41 ± 0.06 nM). Because [HisA8]insulin exhibits similarly enhanced activity (77, 78), these results indicate that, once folding is achieved, substitution of LeuA16 by Val does not perturb receptor binding. A probe for maintenance of the receptor-binding surface was also provided by dimerization. Comprising the central α-helix and C-terminal β-strand of the B-chain, the dimer interface of insulin contains multiple side chains also involved in receptor binding (48). This surface is buttressed by LeuA16 (45), and so its maintenance or perturbation in A16 analogues provides a read-out of possible transmitted conformational changes. Because the interface contains aromatic side chains (TyrB16, PheB24, PheB25, and TyrB26), dimerization may conveniently be monitored by 1H NMR spectroscopy (82, 84). Our studies were simplified by incorporation of substitution HisB10 → Asp to block the independent trimerization surface (82, 85) (which would otherwise lead to further self-assembly) (48). Aromatic 1H NMR spectra of [AspB10]insulin and [AspB10,ValA16]insulin exhibit similar concentration-dependent resonance broadening between 50 and 300 μM (Fig. 5, A and B), indicating analogous dimerization properties. The spectrum of a related monomeric analogue retained a native-like pattern of chemical shifts ([ValA16]DKP-insulin; Fig. 5C). Although dispersion is partially attenuated, its two-dimensional NOESY spectrum contains long range inter-residue cross-peaks characteristic of native-like tertiary structure (supplemental Fig. S3).

To investigate the structure of [HisA8,ValA16]insulin in detail, single crystals were grown in the presence of zinc ions under conditions well characterized in wild-type insulin (48, 86). The crystals diffracted to a resolution of 1.8 Å with unit-cell dimensions consistent with hexamer form T3R6. The variant structure was determined by molecular replacement; statistical information is provided in the supplemental Table S3. Wild-type and variant hexamers are essentially identical (Fig. 6, A and B); root mean square deviations are provided in supplemental Tables S4–S7. The side chains of LeuA16 and ValA16 (Fig. 6, A and B, red) occupy similar core positions without perturbation of the A12–A18 α-helix. The ValA16 side chain is well defined in both T- and R-protomers without evidence of multiple conformations (Fig. 6, C and D). Surrounding density is likewise unambiguous without packing defects. (In the T-state density is weak for one branch of LeuB15 (Fig. 6C), but similar features have been observed in wild-type hexamers.) Inferred solvent accessibility of the variant side chains are given in supplemental Table S8 relative to wild type. A comparison between respective inter-residue distances involving ValA16 (variant T- and R-protomers) or LeuA16 (in multiple wild-type structures) is provided in supplemental Table S9.

Global alignment of respective T- and R-protomers relative to a collection of wild-type protomers indicates overlapping B-chain conformations and similar but not identical A-chain positions (Fig. 7). B-chain-specific alignment reveals a small rigid-body displacement of each A-chain α-helix that is systematic relative to conformational variability among wild-type structures. The structure of the variant A-chain nonetheless closely adjoins wild-type structures. Comparison of the structures of [HisA8,ValA16]insulin and [HisA8]insulin (49) in the same crystal form indicates that the small adjustment of the A-chain is likely to be due to the A16 substitution, apparently in compensation for loss of side-chain volume (Fig. 8). These subtle changes seem unlikely to provide a structural basis for the marked effects of the A16 substitution on folding. The absence of significant perturbations in structure or function suggests that LeuA16 functions transiently in an unsolved folding intermediate and, despite its contribution to global stability, is otherwise dispensable once folding is achieved.
DISCUSSION

This study investigated the folding of a globular domain. An invariant residue in proinsulin (Leu\(^{A16}\); position 81 in intact proinsulin) is required for its folding efficiency but is open to substitution in the mature hormone. Whereas expression of [Val\(^{A16}\)]proinsulin in mammalian cells leads to disulfide mispairing and degradation in association with ER stress, comparison of related insulin analogues ([Val\(^{A16}\),His\(^{A8}\)]insulin and [His\(^{A8}\)]insulin) demonstrated retention of native-like structure and function. We discuss these results in relation to the structural origins of disulfide pairing. Our findings have implications for neonatal DM as a prototypical disease of misfolding.

Substitution of Leu\(^{A16}\) by Val reduces the stability of the native state, presumably due in part to decreased \(\alpha\)-helical propensity (79, 87) and to loss of side-chain volume (80). Yet control studies of destabilizing substitutions elsewhere in the hydrophobic core (positions A2 and A6–A11) revealed unperturbed fold efficiencies, both in a transfected human cell line and as probed by insulin chain combination. These contrasting findings indicate that effects on folding can be distinguished from extent of native-state perturbation. Multiple molecular models may account for these findings. The conformational search leading to native disulfide pairing may be rendered inefficient due to destabilization of on-pathway intermediates or stabilization of off-pathway intermediates. Apparent bottlenecks may reflect thermodynamic or kinetic traps. A mutation may even create barriers not pertinent to the folding mechanism of the wild-type protein. The coarseness of our experimental probes thus prevents unambiguous interpretation of our results on the molecular level.

We suspect that more than one molecular mechanism may underlie the profound inefficiency of folding of [Val\(^{A16}\)]proinsulin. A kinetic barrier to formation of an on-pathway intermediate, for example, may lead to increased occupancy of a non-native disulfide isomer as a thermodynamic trap susceptible to degradation. Which is cause and which is effect may in turn be difficult to distinguish. We nonetheless propose, as a general framework for discussion, that Val\(^{A16}\) disproportionately impairs the stability of a species containing the single cystine A20–B19. Prior studies have suggested that this species defines an obligatory
on-pathway disulfide intermediate. Nascent native-like structure in such an intermediate could exert kinetic control of disulfide pairing, whereas folding efficiency would be unperturbed by substitutions extrinsic to this proto-core. Even as we recognize the limitations of the present data, it may illuminate future studies to consider how they relate to established biochemical features of proinsulin and the mechanism of insulin chain combination.

Disulfide Intermediates in Protein Folding—Oxidative folding may be probed by chemical trapping of populated disulfide intermediates (88). Application to proinsulin and related polypeptides is notable for the transient accumulation of one- and two-disulfide intermediates (67–69). Their partial folding may be represented by a series of trajectories on successive free-energy landscapes (Fig. 9A). We envisage that each landscape governs folding trajectories in the presence of a specific subset of disulfide bridges. Because the polypeptide acquires structure stepwise on successive pairing, the landscapes proceed from shallow to steep. A preferred sequence of disulfide intermediates, as defined by chemical trapping, hence provides a framework for visualizing a progression of multiple folding trajectories on funnel-shaped landscapes. This perspective integrates the classical disulfide-centered paradigm (89) with general physical models of folding (1, 5). Despite its elegance, the landscape perspective is difficult to establish experimentally and so in the present context should be regarded as an evocative metaphor. The utility of this metaphor lies in its potential for exploration by molecular dynamics simulations as a foundation for the design and interpretation of experiments (10).

The notion of successive landscapes is based on the observation of nonrandom disulfide intermediates. Although refolding studies of proinsulin are limited by aggregation (90), disulfide pathways of related polypeptides are well characterized (20, 23, 67–69, 91, 92). A structural pathway has been proposed based on equilibrium models (Fig. 9B) (20–22, 24–29, 93). A key role is played by initial formation of cystine A20–B19, which in the native state connects the C-terminal α-helix of the A-domain to the central α-helix of the B-domain. This bridge packs within a cluster of conserved side chains in the hydrophobic core, including LeuA16 (31, 32, 57). Neighbor- ing side chains include LeuB11, LeuB15, PheB24, and TyrA19. In one- and two-disulfide models, LeuA16 plays a prominent structural role in maintenance of a specific partial fold, recapitulating salient features of the native state (29). We envisage that general hydrophobic clustering facilitates alignment of

---

7 The fractional contribution of a given interaction to the stability of a native-like partial fold can be larger or smaller than its fractional contribution to the native state. Whereas both IleA2 and LeuA16 each contribute to native-state stability through core packing, for example, folding and chain combination are robust to A2 substitutions (45). Kinetic control via a subset of native-like contacts in a transition state differs from the concept of “pathway mutations” as originally envisioned in studies of a β-helical phage tail spike (46).
CysB19 and CysA20; once pairing has occurred, the bridge stabilizes a more specific organization of neighboring aliphatic and aromatic side chains.

This study has demonstrated that substitution of LeuA16 by Val, although compatible with native structure and function, imposes a block to the folding of proinsulin in mammalian cells. This block (recapitulated in chain combination) may arise from perturbations analogous to those that impair native-state stability but in the context of an ensemble of partial folds. Key perturbations might thus arise from the following: (a) a packing defect in the nascent core that misaligns CysA20 and CysB19, and (b) decreased H9251-helical propensity of the A-domain segment of the unfolded polypeptide due to a H9252-branched substituent.

Because AlaA16 (a helicogenic substitution of reduced volume) likewise blocks folding and secretion of a mini-proinsulin in Saccharomyces cerevisiae (94, 95), perturbed hydrophobic collapse is likely to be the predominant mechanism. Mutations elsewhere in the A20–B19-associated subdomain similarly impair chain combination and yeast biosynthesis (24, 94, 96–98). This perspective suggests that abortive cellular folding of [ValA16]proinsulin reflects a primary defect in on-pathway donor and hence strengthens the A21–B23 hydrogen bond. This study did not distinguish between kinetic and thermodynamic effects of the modifications.

We envisage as a working hypothesis that the side chain of LeuA16 and the A21–B23 main-chain hydrogen each stabilize a native-like partial fold directing pairing of cystine A20–B19. Whereas LeuA16 presumably participates in nonpolar interactions between nascent α-helices in the A- and B-domains (Fig. 9B), formation of the A21–B23 hydrogen bond apparently requires a specific orientation of the carbonyl group of GlyH9252. In the native state this conserved glycine exhibits a positive φ dihedral angle, ordinarily forbidden to l-amino acids (48). Whereas d-Ala is readily accommodated at B23, the efficiency of chain combination is markedly impaired by L-AlaB23.8 Similar stereospecific effects of L- and D-substitutions at GlyB20 imply that the productive intermediate exhibits a native-like B20–B23 β-turn. Anchored at one end by the B9–B19 α-helix and at the other end by PheH9254, this turn

8 S.-Q. Hu and M. A. W., unpublished results.

FIGURE 8. Structural environment of ValA16 in relation to LeuA16 in control structure of [HisA8]insulin. A, stereo comparison of corresponding side chains in T-state protomers. Position of ValA16 (gray) in T-state protomer in relation to selected residues in A-chain (IleA2 and TyrA19 (red); CysA6 and CysA11 (gold)) and B-chain (PheB1, LeuB11, AlaB14, LeuB15, and ValB18 (green)). [HisA8]Insulin (black; PDB entry 1RWE, molecule 1). B, stereo comparison of corresponding side chains in R-state protomers of [ValA16,HisA8]Insulin (color scheme as in A with addition of R-state-specific ligand (phenol; blue and asterisk) and [HisA8]Insulin (black; PDB entry 1RWE, molecule 2).
is retained in a peptide model of a one-disulfide intermediate (29).

Why do insulin-related polypeptides exhibit initial pairing of cystine A20–B19 when these residues are so distant in sequence? We propose that folding begins with nascent $\alpha$-helix formation within segments B9–B19 and A16–A20; these segments then interact via nonpolar clustering of Leu B11, Leu B15, Leu B18, Leu A16, Tyr A19, and the two cysteines. Diffusion-collision of $\alpha$-helices aligns the thiol moieties of A20 and B19 for disulfide pairing. The stability (or meta-stability) of this intermediate would be augmented by the A21–B23 hydrogen bond and native-like packing of Phe B24 against one edge of the disulfide bridge.

Implications for Human Genetics—Dominant mutations in the insulin gene have recently been identified as a cause of permanent neonatal-onset DM (36–39). Like Val A16, these mutations are predicted to block folding in the ER and so impair the function of pancreatic $\beta$-cells. Although expression of the wild-type insulin allele would in other circumstances be sufficient to maintain homeostasis, studies of a corresponding mouse model (75, 101, 102) have demonstrated that misfolding of the variant perturbs wild-type biosynthesis (34, 103). Impaired $\beta$-cell secretion of insulin is associated with ER stress, distorted organelle architecture, and eventual cell death (35, 104).

Clinical mutations fall in two classes as follows: (a) substitutions that introduce or remove a Cys and so unbalance disulfide pairing, and (b) substitutions at other sites apparently critical to the structural mechanism of disulfide pairing. The latter class includes Gly B23 → Val at the site of the key B23–A21 hydrogen bond (above). Additional mutational hot spots have been observed at B5, B6, and B8, conserved sites proposed to direct pairing of Cys A7 and Cys B7 (50, 55, 105, 106).
cysteine-related mutations have been observed only in the B-domain. The present results demonstrate that conserved sites in the A-domain may likewise contribute to folding efficiency and so are likely to be observed in neonatal DM. The growing catalogue of clinical mutations in the insulin gene promises to provide important insight into sequence determinants of folding in vivo. It is important to note that the distribution and frequency of clinical mutations may reflect the relative susceptibility of DNA codons to mutagenesis (due to genomic sequence and structure) as well as the consequences of the predicted amino acid substitutions in the variant polypeptides.

The genetics of neonatal DM and other diseases of misfolding suggest that the evolution of proteins has been constrained not only by its structure and function, but also by an underlying risk of toxic misfolding. The native-like crystal structure of a “nonfoldable” insulin analogue demonstrates that the transient function of a conserved side chain in the mechanism of folding may be structurally inapparent once the ground state is reached. Although neonatal DM is uncommon, the diversity of associated mutations in the insulin gene highlights the possible role of wild-type bottlenecks and resultant chronic ER stress in type 2 DM (34). The pathogenesis of progressive β-cell dysfunction in type 2 DM and the metabolic syndrome poses a medical problem of overarching societal importance.

Concluding Remarks—The evolution of proteins faces independent structural constraints at multiple levels of biosynthesis and function. A classical model is provided by insulin; its conserved residues may contribute to one or more essential processes, including folding efficiency in the β-cell, prevention of toxic misfolding, self-assembly within secretory granules, disulfide stability in the bloodstream, and receptor binding. Interlocking sets of constraints may account for the limited sequence diversity among vertebrate insulins (48). LeuA16 provides an example of an invariant residue that enables effective folding, but once the native state has been reached, ValA16 is equally compatible with canonical structure, assembly, and receptor binding. This subtle substitution impairs ground-state stability and may disproportionately perturb the relative stabilities and kinetic accessibilities of obligatory on- or off-pathway intermediates. The marked compression of structural information within the sequence of insulin suggests that other conserved sites may facilitate folding but weaken receptor binding (105); alternatively, structural features favoring receptor binding may impede disulfide pairing (107, 108). We thus envision that specific residues in insulin play distinct roles at each stage of its conformational “life cycle.” The growing clinical data base of DM-associated mutations suggests that the sequence of insulin lies at the edge of nonfoldability.

Supplemental Material—Eleven figures provide a histogram of ER stress assay, NMR spectra, additional depictions of the crystal structure, and summary of prior chemical studies of insulin chain combination; and nine tables give the original ER stress assay, NMR spectra, additional depictions of the crystal structure, and summary of prior chemical studies of insulin chain combination; and nine tables give the original data. Coordinates have been deposited in the RCSB Protein Data Bank (accession number 3GKY).

Acknowledgments—We thank S. Nakagawa for receptor binding assays; W. Jia for assistance with spectroscopy; and G. G. Dodson, D. F. Steiner, J. Whittaker, and J. Whittingham for discussions. We are grateful to a member of the Editorial Board for insights into folding mechanisms.

REFERENCES

1. Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997) Annu. Rev. Phys. Chem. 48, 545–600
2. Dobson, C. M., and Karplus, M. (1999) Curr. Opin. Struct. Biol. 9, 92–101
3. Shakhnovich, E., Farztdinov, G., Gutin, A. M., and Karplus, M. (1991) Phys. Rev. Lett. 67, 1665–1668
4. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., and Wolynes, P. G. (1996) Fold. Des. 1, 441–450
5. Dill, K. A., and Chan, H. S. (1997) Nat. Struct. Biol. 4, 10–19
6. Lazardi, T., and Karplus, M. (1997) Science 278, 1928–1931
7. Mirny, L. A., Abkevich, V. I., and Shakhnovich, E. I. (1998) Proc. Natl. Acad. Sci. U.S.A. 95, 4976–4981
8. Dillin, A. R., Abkevich, V., Shakhnovich, E., and Karplus, M. (1999) Proteins 35, 34–40
9. Tiana, G., Broglia, R. A., and Shakhnovich, E. I. (2000) Proteins 39, 244–251
10. Oliveberg, M., and Wolynes, P. G. (2005) Q. Rev. Biophys. 38, 245–288
11. Dodson, G., and Steiner, D. (1998) Curr. Opin. Struct. Biol. 8, 189–194
12. Steiner, D. F., and Chan, S. I. (1988) Horm. Metab. Res. 20, 443–444
13. Steiner, D. F. (1967) Trans. N. Y. Acad. Sci. 30, 60–68
14. Frank, B. H., and Veros, A. J. (1968) Biochem. Biophys. Res. Commun. 32, 155–160
15. Pekar, A. H., and Frank, B. H. (1972) Biochemistry 11, 4013–4016
16. Frank, B. H., Pekar, A. H., and Veros, A. J. (1972) Diabetes 21, 486–491
17. Snell, C. R., and Smyth, D. G. (1975) J. Biol. Chem. 250, 6291–6295
18. Brems, D. N., Brown, P. L., Heckenlaible, L. A., and Frank, B. H. (1990) Biochemistry 29, 9289–9293
19. Weiss, M. A., Frank, B. H., Khati, L., Pekar, A., Heiney, R., Shoelson, S. E., and Neuringer, L. I. (1990) Biochemistry 29, 8389–8401
20. Narhi, L. O., Hua, Q. X., Arakawa, T., Fox, G. M., Tsai, L., Rosenfeld, B., Holst, P., Miller, J. A., and Weiss, M. A. (1993) Biochemistry 32, 5214–5221
21. Hua, Q. X., Hu, S. Q., Frank, B. H., Jia, W., Chu, Y. C., Wang, S. H., Burke, G. T., Katsoyannis, P. G., and Weiss, M. A. (1996) J. Mol. Biol. 264, 390–403
22. Dai, Y., and Tang, J. G. (1996) Biochim. Biophys. Acta 1296, 63–68
23. Hofer, S., Uhlén, M., and Nilsson, B. (1997) Biochemistry 36, 4616–4622
24. Weiss, M. A., Hua, Q. X., Jia, W., Chu, Y. C., Wang, R. Y., and Katsoyanis, P. G. (2000) Biochemistry 39, 15429–15440
25. Guo, Z. Y., and Feng, Y. M. (2001) Biochem. Chem. 382, 443–448
26. Feng, Y., Liu, D., and Wang, J. (2003) J. Mol. Biol. 330, 821–837
27. Jia, X. Y., Guo, Z. Y., Wang, Y., Xu, Y., Duan, S. S., and Feng, Y. M. (2003) Protein Sci. 12, 2412–2419
28. Yan, H., Guo, Z. Y., Gong, X. W., Xi, D., and Feng, Y. M. (2003) Protein Sci. 12, 768–775
29. Hua, Q. X., Mayer, J. P., Jia, W., Zhang, J., and Weiss, M. A. (2006) J. Biol. Chem. 281, 28131–28142
30. Sieber, P., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Märki, F., and de Gasparo, M. (1978) Hoppe-Seylers Z. Physiol. Chem. 359, 113–123
31. Hua, Q. X., Gozani, S. N., Chance, R. E., Hoffmann, J. A., Frank, B. H., and Weiss, M. A. (1995) Nat. Struct. Biol. 2, 129–138
32. Hua, Q. X., Jia, W., Frank, B. H., Phillips, N. F., and Weiss, M. A. (2006) Biochemistry 45, 14700–14709
33. Liu, M., Hodish, I., Rhodes, C. I., and Arvan, P. (2005) Proc. Natl. Acad. Sci. U.S.A. 104, 5841–5846
34. Stay, J., Edgill, E. L., Flanagan, S. E., Ye, H., Paz, V. P., Pluzhnikov, A.,Below, J. E., Hayes, M. G., Cox, N. J., Lipkind, G. M., Lipton, R. B., Greeley,
Bottleneck in a Folding Landscape

P. G. (1987) Biochemistry 26, 6975–6979
101. Yoshioka, M., Kayo, T., Ikeda, T., and Koizumi, A. (1997) Diabetes 46, 887–894
102. Oyadomari, S., Koizumi, A., Takeda, K., Gotoh, T., Akira, S., Araki, E., and Mori, M. (2002) J. Clin. Invest. 109, 525–532
103. Izumi, T., Yokota-Hashimoto, H., Zhao, S., Wang, J., Halban, P. A., and Takeuchi, T. (2003) Diabetes 52, 409–416
104. Zuber, C., Fan, J. Y., Guhl, B., and Roth, J. (2004) FASEB J. 18, 917–919
105. Nakagawa, S. H., Zhao, M., Hua, Q. X., Hu, S. Q., Wan, Z. L., Jia, W., and Weiss, M. A. (2005) Biochemistry 44, 4984–4999
106. Hua, Q. X., Nakagawa, S., Hu, S. Q., Jia, W., Wang, S., and Weiss, M. A. (2006) J. Biol. Chem. 281, 24900–24909
107. Hua, Q. X., Xu, B., Huang, K., Hu, S. Q., Nakagawa, S., Jia, W., Wang, S., Whittaker, J., Katsoyannis, P. G., and Weiss, M. A. (2009) J. Biol. Chem. 284, 14586–14596
108. Xu, B., Huang, K., Chu, Y. C., Hu, S. Q., Nakagawa, S., Wang, S., Wang, R. Y., Whittaker, J., Katsoyannis, P. G., and Weiss, M. A. (2009) J. Biol. Chem. 284, 14597–14608